

УНИВЕРЗИТЕТ У БЕОГРАДУ
ФИЛОЛОШКИ ФАКУЛТЕТ

Тања Д. Ивановић

**РАЗВОЈ ТЕРМИНОЛОГИЈЕ ИЗ ОБЛАСТИ
ЕЛЕКТРОЕНЕРГЕТИКЕ КОРИШЋЕЊЕМ МЕТОДА
ОБРАДЕ ПРИРОДНИХ ЈЕЗИКА**

докторска дисертација

Београд, 2021.

UNIVERSITY OF BELGRADE
FACULTY OF PHILOLOGY

Tanja Ivanović

**TERMINOLOGY DEVELOPMENT IN POWER
ENGINEERING BASED ON NATURAL LANGUAGE
PROCESSING METHODS**

Doctoral dissertation

Belgrade, 2021.

УНИВЕРСИТЕТ В БЕЛГРАДЕ
ФИЛОЛОГИЧЕСКИЙ ФАКУЛЬТЕТ

Таня Иванович

**РАЗВИТИЕ ТЕРМИНОЛОГИИ В ЭНЕРГЕТИКЕ НА
ОСНОВЕ МЕТОДОВ ОБРАБОТКИ ЕСТЕСТВЕННОГО
ЯЗЫКА**

Докторская диссертация

Белград, 2021.

Ментор:

др Цветана Крстев, редовни професор, Универзитет у Београду, Филолошки факултет

Чланови комисије:

Датум одбране: _____

Наслов дисертације: Развој терминологије из области електроенергетике коришћењем метода обраде природних језика

Сажетак: У овом раду анализира се развој терминологије из области електроенергетике применом метода обраде природних језика. Рад је подељен на осам поглавља и обрађује општу теорију терминологије као научног домена, међународне и домаће институције које учествују у њеном креирању, развој специјализоване терминологије на српском језику, примену корпусне лингвистике у терминолошким истраживањима, као и корпусне алате и језичке ресурсе који се примењују приликом обраде текстова корпуса.

Паралелни корпуси представљају двојезичне односно вишејезичне корпусе текстова који су веома значајни у лингвистичким истраживањима. Развој паралелног корпуса текстова из домена електроенергетике (*ElEner*) започет је упоредо са израдом ове докторске дисертације. У поступку њене израде, анализирано је 76 докумената насталих у периоду од 2005. до 2016. године, који и чине овај корпус. Реч је текстовима законодавне, техничке и научне природе на српском и енглеском језику. У дисертацији је темељно анализиран процес одабира и прикупљања материјала за корпус, обрада текстова применом одговарајућих језичких ресурса и алата за српски и енглески језик, паралелизација текстова, екстракција терминологије на српском и енглеском језику, поравнавање и упаривање комада и термина, као и евалуација резултата добијених термина и терминолошких парова. Након завршеног процеса евалуације, сви исправно евалуирани парови су укључени у терминолошку базу података *Termi*, која подржава развој терминолошких речника у различитим областима (математика, рачунарство, рударство, библиотекарство, рачунарска лингвистика, електроенергетика, итд.), као и обраду и презентацију термина на српском, енглеском, немачком и француском језику, и извоз у различите излазне формате. Ова база је тако допуњена новим лексичким јединицама из домена електроенергетике на српском и енглеском језику, као и њиховим синонимима. Добијена листа преводних парова послужила је за генерисање двојезичног речника из домена електроенергетике.

Произведени паралелни корпус *ElEner* смештен је у дигиталну библиотеку Библиша која омогућава вишејезичко претраживање великих колекција поравнатих текстова. Претраживање ове дигиталне библиотеке обавља се помоћу лексичких ресурса који омогућавају морфолошко и семантичко проширење постављених упита.

Добијени терминолошки парови представљају основу за развој новог модерног речника из области електроенергетике, чиме се уједно отвара могућност и за унапређење и проширивање терминолошке базе Електропедија. Поступак обраде текстова предложен овом дисертацијом показао се применљивим и корисним и у другим доменима. У истраживањима која ће уследити, циљ је да се побољша предложена техника укључивањем аутоматске валидације добијених двојезичних термина кандидата у постојећу процедуру, на основу најсавременијих техника машинског учења.

Кључне речи: терминологија, корпусна лингвистика, паралелни корпуси, електроенергетика, обрада природних језика, лексички ресурси, екстракција терминологије

Научна област: Терминологија

Ужа научна област: Корпусна лингвистика, обрада природних језика

УДК број:

Dissertation title: Terminology development in power engineering based on natural language processing methods

Abstract: This paper analyzes terminology development in power engineering domain using natural language processing methods. The paper is divided into eight chapters and deals with the theory of terminology as an academic field in general, with international and domestic institutions involved in terminology development, development of specialized terminology within power engineering domain in Serbian language, the application of corpus linguistics in terminological research, as well as corpus processing tools and language resources.

Parallel corpora are bilingual or multilingual corpora of texts that are very important in linguistic research. The development of a parallel corpus composed of texts in power engineering domain (*ElEner*) started with the preparation of this doctoral dissertation. The corpus is composed of technical, scientific and legislative texts both in Serbian and English published from 2006 until 2015.

The dissertation thoroughly analyzes the process of text selection and collection, text processing techniques using appropriate language resources and tools for Serbian and English, parallelization of texts, extraction of terminology in Serbian and English, alignment and matching of chunks and terms, and evaluation of obtained results. After the evaluation process is completed, all correctly evaluated pairs are included in the *Terminology* database, which supports the development of terminological dictionaries in various fields (mathematics, computing, mining, librarianship, computational linguistics, power engineering, etc.), as well as processing and presentation of terms in Serbian, English, German and French and their export to various output formats. This database is thus upgraded with new lexical units and synonyms from the power engineering domain in Serbian and English. The obtained list of translation pairs was used for power engineering bilingual dictionary development.

The new aligned *ElEner* corpus is stored in digital library Bibliša, which enables multilingual search of large collections of aligned texts. The search of this digital library is performed using lexical resources that enable morphological and semantic expansion of the queries.

The obtained terminological pairs represent the basis for the development of a new modern dictionary in the field of power engineering, and provide an opportunity for the improvement and expansion of the terminology base of Electropedia. The text processing procedure proposed by this dissertation has proven to be applicable and useful for application in other domains as well. In the future research, the goal is to improve the proposed technique by including automatic validation of the obtained bilingual terms of the candidates in this routine, based on the state-of-the-art machine learning techniques.

Keywords: Terminology, Corpus linguistics, Parallel corpora, Power engineering, Natural Language Processing, Lexical resources, Terminology extraction

Scientific field: Terminology

Scientific subfield: Corpus linguistics, Natural Language Processing

UDC Number:

Садржај

1.	УВОДНА РАЗМАТРАЊА	1
2.	ТЕРМИНОЛОГИЈА	3
2.1.	Развој термилошких школа.....	4
2.1.1.	Бечка термилошка школа	5
2.1.2.	Руска термилошка школа	6
2.1.3.	Прашка термилошка школа.....	8
2.2.	Критика традиционалних школа терминологије	9
2.3.	Терминологија данас	12
2.4.	Термин	14
2.5.	Терминологија и стандардизација.....	15
2.5.1.	Међународна организација за стандардизацију – ISO	17
2.5.2.	Међународни информациони центар за терминологију – INFOTERM	19
2.5.3.	Европски комитет за стандардизацију – CEN	19
2.5.4.	Европски комитет за стандардизацију у електротехници – CENELEC	20
2.5.5.	Међународна електротехничка комисија – IEC	20
2.5.6.	Институт за стандардизацију Републике Србије.....	25
3.	ЈЕЗИЧКЕ И ТЕРМИНОЛОШКЕ ПРИЛИКЕ У СРПСКОМ ЈЕЗИКУ.....	27
3.1.	Историјски осврт	27
3.2.	Рани развој терминологије у српском језику	30
3.3.	Развој терминологије српског језика у 20. веку.....	32
3.4.	Термилошке активности у српском језику данас	34
3.4.1.	Утицај енглеског језика на српски данас	34
3.4.2.	Критеријуми изграђивања терминологије	36
3.4.3.	Институције задужене за језичку и термилошку политику	37
3.4.3.1.	Институт за српски језик САНУ	38
3.4.3.2.	Одбор за стандардизацију српског језика	38
3.4.3.3.	Министарство за европске интеграције РС.....	39
3.4.3.4.	Институт за стандардизацију РС.....	40
3.4.3.5.	Друштво за језичке ресурсе и технологије.....	47
4.	ФОРМИРАЊЕ И РАЗВОЈ ТЕРМИНОЛОГИЈЕ ИЗ ОБЛАСТИ ЕЛЕКТРОЕНЕРГЕТИКЕ У СРПСКОМ ЈЕЗИКУ	49
4.1.	Кратак приказ развоја електроенергетике на нашим просторима кроз науку и праксу	53
4.2.	Речници електротехнике	55
4.2.1.	Мали српско-француско-немачки електротехнички речник са скраћеницама и практичним забелешкама	55
4.2.2.	Њемачко-хрватски и хрватско-њемачки електротехнички рјечник	56

4.2.3.	Словеначки електротехнички речник.....	56
4.2.4.	Електротехнички термилошки речник, група 07, електроника	57
4.2.5.	Електротехнички речник: руско-српскохрватски, српскохрватско-руски.....	58
4.2.6.	Електротехнички речник: немачко-српскохрватски	59
4.2.7.	Речник индустријске електротехнике: немачко-српскохрватски	59
4.2.8.	Енглеско–српскохрватски електротехнички речник	60
4.2.9.	Дефиниције појмова из електротехнике на основу ИЕС (СЕД): Појмовник – речник електротехнике	60
4.2.10.	Енглеско-српскохрватски технички речник са изговором	61
4.2.11.	Међународни електротехнички ИЕС речник са терминима на српском језику: електроенергетика, електроника и телекомуникације	62
4.2.12.	Статистичка терминологија коришћена у електропривреди.....	65
4.2.13.	Речник електронике и електротехнике: енглеско-српски.....	65
4.2.14.	Речници из области електроенергетике Института за стандардизацију Србије.....	67
4.3.	Истакнуте организације у сектору електроенергетике данас	77
4.3.1.	Српски национални комитет Међународног савета за велике електричне мреже CIGRE Србија	78
4.3.2.	Српски национални комитет CIREД Србија	79
5.	ПРИМЕНА КОРПУСНЕ ЛИНГВИСТИКЕ И УЛОГА КОРПУСА У ТЕРМИНОЛОШКОМ ИСТРАЖИВАЊУ	81
5.1.	Развој корпуса	82
5.1.1.	Пределектронски корпус	82
5.1.2.	Електронски корпус	85
5.2.	Класификација корпуса.....	91
5.3.	Репрезентативност и балансираност корпуса	93
5.4.	Вишејезични корпуси.....	94
5.5.	Корпусна лингвистика у Србији	95
5.5.1.	Корпус савременог српског језика.....	95
5.5.2.	Вишејезични корпуси српског језика	97
5.5.3.	Корпус правних тековина Европске уније (Евротекa) и вишејезична база термина (Евроним)	98
5.5.3.1.	Евротекa, енглеско-српски паралелни корпус	99
5.5.3.2.	Евроним – Вишејезична термилошка база.....	100
5.5.4.	Библиша.....	103
6.	КОРПУСНИ АЛАТИ И ЈЕЗИЧКИ РЕСУРСИ У ОБРАДИ ПРИРОДНИХ ЈЕЗИКА	109
6.1.	IMS Open Corpus Workbench	110
6.2.	Unitex/GramLab	110
6.3.	Електронски морфолошки речници.....	111
6.4.	Локалне граматике.....	113

6.5.	Паралелизација текста.....	114
6.6.	Семантичка мрежа Ворднет.....	115
7.	ОБРАДА ПАРАЛЕЛИЗОВАНОГ КОРПУСА ЕЛЕКТРОЕНЕРГЕТСКОГ СИСТЕМА	117
7.1.	Процес компилације двојезичног термилошког материјала	119
7.2.	Двојезични паралелни корпус из домена електроенергетике.....	121
7.2.1.	Садржај корпуса из домена електроенергетике	121
7.2.2.	Претходна обрада (претпроцесирање) корпуса.....	122
7.3.	Екстракција терминологије на српском језику	124
7.4.	Екстракција терминологије на енглеском језику.....	125
7.5.	Поравнавање комада	126
7.6.	Упаривање термина и комада.....	127
7.7.	Евалуација резултата.....	129
7.7.1.	Статистичке мере за процену сагласности анотатора ручно анотираног корпуса.....	131
7.7.2.	Поступак евалуације	133
7.8.	Карактеристике екстрактора за српски и енглески језик.....	137
7.9.	Објављивање резултата.....	141
8.	ЗАКЉУЧНА РАЗМАТРАЊА.....	147
8.1.	Закључак.....	147
8.2.	Будући рад	149
9.	ДОДАТАК.....	150
9.1.	Списак скраћеница на ћирилици	150
9.2.	Списак скраћеница на латиници	150
БИБЛИОГРАФИЈА		155
Библиографија публикована на латиници.....		155
Библиографија публикована на ћирилици		162
Речници		164
Стандарди.....		165

Листа слика

Слика 1.	Приказ схеме обраде текста.....	13
Слика 2.	Термини на српском језику	24
Слика 3.	Примери превода термина „power system“ у стандардима SRPS IEC 60050-436:2016 и SRPS IEC 60050-603:2015	75
Слика 4.	Примери превода термина „converter“ у стандардима SRPS IEC 60050-314:2007 и SRPS IEC 60050-605:2015	75
Слика 5.	Примери превода термина „load flow“ у стандардима SRPS IEC 60050-617: 2015 и SRPS IEC 60050-603:2015	75
Слика 6.	Пример превода термина „transformer substation“	76

Слика 7. Пример превода термина „substation“	76
Слика 8. Пример превода термина „power station“	77
Слика 9. Пример превода термина „hydroelectric power station“	77
Слика 10. Пример превода термина „power electronic“	77
Слика 11. Пример одреднице из Круденових конкорданци.....	83
Слика 12. Расподела текстова Корпуса према функционалним стиловима	97
Слика 13. Евротека.....	100
Слика 14. Евроним	102
Слика 15. Терминолошки анализатор – текст који се анализира.....	102
Слика 16. Терминолошки анализатор након завршене обраде текста	103
Слика 17. Анализа хиперлинкованих термина	103
Слика 18. Почетна страница дигиталне библиотеке Библиша и одабир колекције текстова из области електроенергетике.....	106
Слика 19. Резултат претраге према кључној речи.....	107
Слика 20. Паралелизоване конкорданце сегмената као резултат претраге кључном речи „електроенергетски систем“	108
Слика 21. Дијаграм статистичког машинског превођења	118
Слика 22. Дијаграм предложеног система за терминолошку екстракцију	121
Слика 23. Пример графа којим је приказана локална граматика у форми коначног трансдукора	125
Слика 24. Евалуација двојезичне екстракције за домен електроенергетике за енглески језик	133
Слика 25. Евалуација двојезичне екстракције за домен електроенергетике за српски језик.....	134
Слика 26. Евалуација усаглашености преводних еквивалената за домен електроенергетике.....	134
Слика 27. Успешност система <i>FlexiTerm</i> и <i>TermSuite</i> у екстракцији терминологије	138
Слика 28. Одредница терминолошке базе <i>Termini</i>	141
Слика 29. Употреба базе података <i>Termini</i> у дигиталној библиотеци Библиша	142
Слика 30. Облици упарених термина у дигиталној библиотеци Библиша	143
Слика 31. Веза апликације <i>Termini</i> и корпуса за опцију јавног приступа.....	144
Слика 32. Примери конкорданци за изабране образаца	145

Листа табела

Табела 1. Терминолошка база података према предметним областима.....	22
Табела 2. Пример матрице за систематско вредновање термина	37
Табела 3. Хијерархијска схема класе 500: <i>Природне науке</i>	49
Табела 4. Хијерархијска схема класе 6: <i>Примењене науке. Медицина. Технологија</i>	51
Табела 5. Хијерархијска схема класе 5: <i>Математика и природне науке</i>	51
Табела 6. Листа речника релевантних за домене електроенергетике.....	66
Табела 7. Садржај Корпуса Универзитета Браун према типу текста, уз приказ њиховог броја и удела у корпусу	86

Табела 8. Корпус енглеског језика.....	89
Табела 9. Листа преговарачких поглавља ЕУ.....	90
Табела 10. Текстуалне колекције у дигиталној библиотеци Библиша према броју докумената и броју сегмената.....	104
Табела 11. Илустрација различитих функција упаривања.....	128
Табела 12. Резултат термилошке екстракције уз коришћење различитих функција упаривања.....	128
Табела 13. Примери евалуације.....	130
Табела 14. Израчунавање процента слагања међу анотаторима.....	132
Табела 15. Неусаглашеност на српском језику.....	134
Табела 16. Неусаглашеност на енглеском језику.....	134
Табела 17. Неусаглашеност између преводних еквивалената.....	134
Табела 18. Резултати процеса евалуације (једнојезична евалуација).....	135
Табела 19. Резултати процеса евалуације (двојезична евалуација).....	135
Табела 20. Резултати добијени применом три функције упаривања за екстракцију термина из енглеског и српског језика; прецизност екстракције дата је у заградама.....	135
Табела 21. Резултати добијени применом три функције упаривања за енглеске и српске термилошке парове.....	136
Табела 22. Пример екстрахованих термина који су пронађени и у термилошкој бази Евроним.....	137
Табела 23. Рангирање и фреквенција за 8 најпродуктивнијих синтаксичких образаца система <i>Srp-TE</i> на корпусу домена електроенергетике након његове примене на три термилошке листе из различитих домена.....	139
Табела 24. Најуспешнији односно најмање успешни обрасци примењивани у поступку термилошке екстракције.....	140

ТИПОГРАФСKE КОНВЕНЦИЈЕ

<i>курзив</i>	оригинални назив страног аутора, дела, издавачке куће или организације које се наводе у тексту, увођење термина на страном језику, наглашавање српске речи у оквиру теоријског разматрања односно увођење дефиниције на српском језику, назив дела на српском језику, скраћенице и акроними на страном језику
масна слова	истицање српског стручног термина у оквиру теоријског разматрања, истицање приликом поделе на делове, групе итд. неког термина, увођење класе УДК (први пут)
„наводници“	навођење / истицање одређеног примера из речника било на српском било на енглеском језику, превод назива дела са страног на српски језик, превод назива дела са страног језика на српски, називи удружења
умањена велика слова (енг. <i>small caps</i>)	називи корпуса, стручна терминологија коришћења у поступцима обраде текстова (токенизација, сегментација, итд.)

1. УВОДНА РАЗМАТРАЊА

Иако је корпусна лингвистика у Србији почела да се развија још средином 20. века, тек је осамдесетих година 20. века забележен нагли развој овог лингвистичког метода у Србији. У том периоду направљена је прва колекција текстова у дигиталном облику, спроведени први експерименти у морфолошком генерисању српскохрватског језика и прва истраживања у домену корпусне лингвистике. У том периоду направљен је и први значајнији искорак у развоју корпуса на нашим просторима, када је конструисан систем за генерисање конкорданци и различитих врста индекса за задати текст помоћу којег су обрађени и први паралелни корпуси. Од тада је настао изванредан број корпуса на српском језику. Тај број се, уз константан рад на развоју метода лингвистичке анализе заснованим на употреби аотираних корпуса, и данас непрекидно увећава.

Корпуси, по дефиницији, представљају колекције текстова који не морају бити предмет само лингвистичких истраживања, већ могу служити и за састављање речника односно граматичких приручника. Из наведеног се намеће закључак да се корпусна истраживања не темеље искључиво на лингвистици, већ да своје упориште проналазе и у обради природних језика (енг. *natural language processing*) и рачунарској лингвистици (енг. *computational linguistics*). Неизоставан део сваког корпусног истраживања је коришћење база података ради организовања текстова у корпус и њиховог претраживања.

Питање специјализоване терминологије, на глобалном као и на националном плану, једно је од веома заступљених како у домену лингвистике, превођења, термилошке политике, тако и у домену информатичких наука и многих других области. Напредак научне, стручне и практичне активности једне дисциплине неизоставно са собом носи и развој терминологије дате области. Таква терминологија омогућава недвосмислену комуникацију међу стручњацима као и прецизну термилошку корелацију између различитих језика. Стога, терминологија међу наведеним дисциплинама има нарочит значај.

Област обрађена овим радом припада домену електроенергетике, једне од научних дисциплина која у савременој науци на глобалном нивоу заузима веома важно место, нарочито у погледу инвестиција у овом сектору и напора за очување животне средине.

Предмет истраживања ове докторске дисертације обухвата поступак израде паралелног српско-енглеског корпуса из домена електроенергетике у циљу допуне лексичких ресурса савременог српског језика у систему двојезичне претраге паралелних колекција.

У самом раду биће образложени методи коришћени приликом прикупљања и припреме грађе из овог домена, као и методе коришћене приликом обраде корпуса одабраних текстова, од екстракције терминологије на српском и енглеском језику, преко поравнавања текстова и упаривања термина и термилошких парова на ова два језика, до евалуације резултата процеса обраде коришћењем статистичких мера за процену сагласности аотатора код ручно аотираних корпуса и укључивања исправно евалуираних парова у термилошку базу података *Termi*.

Рад се састоји од укупно осам поглавља, а тематски је подељен на шест целина. Након уводног поглавља, следе теоријски осврти на терминологију као науку, термилошку праксу у српском језику, настанак и развој стручне терминологије на нашим просторима који је текао упоредо са развојем електроенергетике као дисциплине, развој корпуса из ове области, како на глобалном тако и на домаћем тлу, као и детаљан поступак обраде термина добијених екстраховањем из специјализованог корпуса текстова за потребе справљања специјализованог електронског речника из домена електроенергетике.

Након уводног дела, следи други део односно целина која описује терминологију у општем контексту и састоји се од пет поглавља. У прва четири поглавља направљен је осврт на појаву специфичне лексике у оквиру општег језика, развој теорије терминологије, развој термилошких школа, критику традиционалних школа терминологије до терминологије данас и начина дефинисања термина као репрезента одређеног појма. Пето поглавље ове

целине описује стандардизацију терминологије кроз најзначајније међународне и домаће организације у чијој је надлежности бављење тим питањем.

Трећи део описује термилошке и језичке прилике у српском језику, кроз историјски осврт на језик, а потом и рани развој терминологије као и термилошке активности у српском језику данас. Чине је четири поглавља.

У четвртм делу описано је формирање и развој терминологије из области електроенергетике у српском језику поштујући правила универзалне класификације наука. Кроз осврт на развој ове дисциплине кроз науку и праксу, дат је преглед речника који су настајали на српскохрватском говорном подручју упоредо са развојем ове области. У последња два поглавља представљене су и истакнуте организације у сектору електроенергетике које уз велики број ангажованих стручних лица која су запослена у електроенергетском сектору (у електропривредној индустрији, при институтима или факултетима), поред стручних и научних активности у овој области, узимају учешћа и у разматрању и решавању термилошких питања. Подељена је на три поглавља и већи број потпоглавља.

Пети део подељен је на пет поглавља и већи број потпоглавља. Описује примену корпусне лингвистике и улогу корпуса у термилошком истраживању уопште, његов развој у дугачком периоду пре појаве рачунара, класификацију корпуса, као и развој корпусне лингвистике у Србији кроз изградњу различитих корпуса српског језика.

Шести део описује корпусне алате и језичке ресурсе који се користе у обради природних језика. Подељен је на шест поглавља.

У седмом делу детаљно је образложен начин обраде паралелизованог корпуса електроенергетског система. Преко процеса компилације двојезичног термилошког материјала, садржине корпуса, екстракције терминологије на српском и енглеском језику и упаривања термина, па све до евалуације резултата и статистичких мера које су коришћене за процену сагласности анотатора ручно анотираног корпуса и базе података *Termi*, која подржава развој термилошких речника у различитим областима.

У завршном делу рада, обједињени су закључци спроведених активности, изложени су резултати постигнути радом на овој грађи и представљен је значај који добијени резултати имају на развој доменске терминологије на српском језику.

Примарни циљ овог истраживања био је унапређење постојећих лексичких ресурса односно допуна електронског морфолошког речника из домена електроенергетике на српском језику и уједно проширење вишејезичне дигиталне библиотеке Библиша. Да би се овај задатак испунио спроведено је најпре прикупљање и припрема материјала из предметног домена на српском и енглеском језику, анализа садржине корпуса и његова класификација, паралелизација текстова, упаривање сегмената, поравнавање комада (енг. *chunk*), екстракција терминологије и, на крају, евалуација издвојених термина и термилошких парова.

Резултати постигнути радом на овом корпусу и њихов значај приказани су у поглављу које описује евалуацију резултата и методе које су коришћене за процену сагласности анотатора ручно анотираног корпуса. Унапређени су постојећи лексички ресурси односно електронски морфолошки речник српског језика је проширен терминима, док је база *Termi* проширена термилошким паровима из домена електроенергетике.

У дисертацији је посебна пажња поклоњена електроенергетици као веома сложеној дисциплини у којој дуги низ година, поред практичног, постоји и искуство у креирању термилошких решења али и тенденција стручњака да терминологија ове струке и убудуће буде формирана на систематичан, конзистентан, недвосмислен и прецизан начин, поштујући утврђене принципе и правила приликом њеног формирања.

2. ТЕРМИНОЛОГИЈА

Тешко је прецизно одредити када је настала специфична лексика у оквиру општег језика. Природно је да су са појавом првих заната одређене речи из опште лексике почеле добијати посебна значења која су само у одређеним областима делатности имала смисао, а да је каснији развој науке и технике условио и развој специјализованих речи и ограничио поље њихове примене (Petrović и Milenković 1988, 170).

Иако је потреба за именовањем предмета постојала још од давнина, Платоново дело *Кратил* се, у западном свету, сматра првим битним текстом о терминологији (Rey 1995, 11). Платон теорију језика не износи системски само у овом делу, његова теорија језика је расута у разним дијалозима, али полемиком у *Кратилу* је најобухватније описана проблематика језика (Damjanović 2017, 195). У свим његовим дијалозима проналазимо једну реч (термин) и мноштво разумевања једног термина, међутим, настојање да се тај термин преобликује у једно значење и један појмовни идејни поредак подразумева отклон од свих неправилних значења и тежњу за обликовањем термина као једнозначног уз поштовање принципа јединственог идентитета (Damjanović 2017, 229). Стога је Платонов допринос темељима ове науке неоспоран.

Након Платона, бројни филозофи, граматичари и лексикографи су писали о овој теми, али се значајнији помак види тек почетком 16. века када се у француском и енглеском језику појавила реч „номенклатура“ (лат. *nomen calare*), која је означавала „листу имена“, „глосар“ (Rey 1995, 11).

Појам *терминологија* први пут се појављује у другој половини 18. века, у списима немачког универзитетског професора Кристијана Готфрида Шица¹ (Rey 1995, 15), који је кованицу *Terminologie* сковао користећи латинску реч *terminus* (срп. крај, граница, циљ), и грчку реч *λογος* (срп. реч, говор, мишљење). Ова кованица заменила је *номенклатуру* – реч која се до тада пуна два века користила у истом значењу (Marzá 2009, 83).

У *Речнику српскога језика* Матице српске², појам *терминологија* дефинисан је као *систем назива, термина у некој научној области, номенклатура: граматичка ~, зоолошка ~, спортска ~.*

У основи, сама реч *терминологија* је вишезначна, па је зато и одређени број аутора тумачи на више различитих начина. Тако Бугарски (Bugarski 2007, 75) наводи најмање пет значења у којима се ова реч употребљава:

1. скуп термина који репрезентују систем појмова неке области,
2. систематски опис образовања и употребе овог скупа термина,
3. публикација у којој је систем појмова неке области репрезентован терминима,
4. специјална теорија терминологије за поједине области или језике,
5. општа теорија терминологије.

Према овом аутору „прва три значења претежно покривају практично-дескриптивне аспекте термина; ако се у тим доменима мисли само на спискове назива објеката, а не и на њихову научну систематизацију, често се употребљавају и термини *номенклатура* и *називље*. Друга два значења односе се на изразитије теоријске видове терминолошких истраживања“ (Bugarski 2007, 75).

Један од најистакнутијих представника бечке терминолошке школе, Хелмут Фелбер (Felber 1984, 1), о коме ће бити више речи касније у раду, терминологију описује као:

1. интердисциплинарну и трансдисциплинарну научну дисциплину која се бави појмовима и њиховим представљањем (помоћу термина, симбола итд.),
2. скуп термина који представљају систем појмова одређеног предметног поља,

¹ *Christian Gottfried Schütz* (1747–1832), био је немачки хуманиста познат по својим доприносима на пољу филологије и филозофије, као и по свом академском раду и књижевном уредништву и издаваштву.

² У овом раду коришћено је издање једнотомног *Речника српскога језика* Матице српске из 2011. године.

3. публикацију у којој се систем појмова одређеног предметног поља представља терминима.

Неки аутори, попут Сагера³ (Sager 1990, 3) сматрају да је у савременој употреби појма терминологије неопходно направити разлику између њена три основна значења, па тако према његовим наводима, терминологија представља:

1. скуп практичних активности и метода који се користе за сакупљање, описивање и презентацију термина,
2. скуп претпоставки, тврдњи и закључака који су потребни да би се описао однос између појмова и термина који су од кључног значаја за претходно поменуте активности,
3. речник специјализованог предметног поља.

Марија Тереса Кабре Кастелви⁴ сматра да је терминологија интердисциплинарно поље које се бави именовањем појмова одређеног предметног поља, као и њиховом реализацијом у лингвистичком или неком другом облику. Иако се аутономија терминологије као независне научне дисциплине оспорава, како даље наводи ова ауторка, терминологија има сопствена теоријска начела (терминолошку теорију) као и практичну активност (развој речника, вокабулара и глосара као и стандардизацију приликом означавања). Основни појмови и методологија од којих се састоји теорија терминологије нису оригинални, већ као и код осталих интердисциплинарних предмета, позајмљени су из суседних, сродних дисциплина као што су лингвистика, логика, онтологија и информационе науке (Cabré 1999, 32).

Према Ожеу⁵, како наводи Кабре (Cabré 1999, 12–14), постоје три основна приступа у практичној терминолошкој обради: лингвистички приступ, преводилачки приступ и терминолошко планирање. Основни циљ лингвистичког приступа, према речима овог аутора, јесте стандардизација терминологије ради прецизности, ефикасности и једноставности комуникације као и трансфера знања међу стручњацима (лингвистичку перспективу која терминологију посматра као средство изражавања делиле су руска, бечка и прашка терминолошка школа о којима ће бити говора у наставку). Преводилачки приступ је, како истиче, углавном заступљен у двојезичним и вишејезичним државама с циљем очувања и неговања језичког диверзитета, а спроводе га вишејезичне међународне организације попут Уједињених нација, Унеска, Европске уније итд., док терминолошко планирање заправо подразумева језичко планирање, а спроводи се у оним државама које имају угрожене мањинске језике и које својом институционализованом политиком кроз спровођење различитих законодавних мера, подстичу употребу и негују терминологију на мањинском језику (тако су нпр. у Квебеку спроведене посебне мере како би се обезбедио равноправан статус француског језика и његов пун развој у свим сферама употребе) (Cabré 1999, 12–14).

2.1. Развој терминолошких школа

У првој половини прошлога века, терминологија се развила као посебно поље научног испитивања. Брз напредак и развој технологије захтевао је тачно именовање одређених појмова, па се тако развој теоријске и примењене терминологије у 20. веку догодио управо захваљујући научницима и стручњацима из различитих области (Cabré 1999, 1–2).

³ *Juan Carlos Sager* – предавач и шеф катедре за језик и лингвистику на Универзитету у Манчестеру (енг. *The University of Manchester Institute of Science and Technology – UMIST*). Ова катедра је посебно специјализована за наставу у области рачунарске лингвистике, истраживање терминолошких база и машинског превођења.

⁴ *Maria Teresa Cabré Castellví* – наставник терминологије и лингвистике на Универзитету Помпеу Фабра, основала је универзитетски институт за примењену лингвистику, којим је председавала од 1994. до 2004. године. Године 2014. изабрана је за председника филолошког одељења Међународне електротехничке комисије (енг. *International Electrotechnical Commission*). Има статус професора емеритуса на истом универзитету.

⁵ *Pierre Auger* – професор лингвистичких наука на универзитету Лавал, у Квебеку. Током своје плодноне каријере бавио се проучавањем терминологије, терминографије, аутоматске обраде језика за посебну употребу, социотерминологије, лексичке неологије као и рачунарских алата који представљају помоћ приликом превођења с једног језика на други.

Три најпознатије традиционалне термилошке школе су: бечка, руска и прашка. И поред тога што су имале различите приступе термилошкој теорији, ти приступи нису били међусобно искључиви, већ су се више третирали као допунски (Šipka 2006, 149):

- Први приступ је **приступ предметног поља (стручни приступ)**, који се односи на специјалистичке аспекте оне стручне или научне области којој терминологија припада;
- Други приступ је **филозофски приступ**, који се првенствено бави општим онтолошким и логичким законитостима конструисања предметног поља и његовог означавања;
- Трећи приступ је **лингвистички приступ** који терминологију третира као поткомпоненту језичког лексикона, а стручни језик као подсистем општег језика. Овај приступ првенствено интересују термини као лексичке јединице.

2.1.1. Бечка термилошка школа

Као најугледнија и најутицајнија међу традиционалним школама терминологије, издваја се бечка термилошка школа. Ова школа је била оријентисана према посебним областима или предметима унутар људског знања, па је тако њен оснивач и најистакнутији представник, Еуген Вистер⁶, своју општу теорију о терминологији сместио између лингвистике, логике, онтологије, информатике и конкретних области чија се терминологија обрађује, а у центру његове пажње налазио се појам, односи међу појмовима и њихово подударење с терминима (Karadža-Garić 1988, 26).

Да би се боље разумео Вистеров допринос термилошкој теорији и пракси, важно је осврнути се на његову плодносну каријеру. Дипломирао је 1927. године на Електротехничком факултету Универзитета у Берлину, а докторску тезу под називом „Међународна стандардизација језика технике“⁷ одбранио је на Универзитету у Штутгарту. Управо тај рад је представљао основ за Општу теорију терминологије чији је оснивач био сам Вистер. Докторска теза је преведена на руски језик и убрзо постала обавезна литература за студенте терминологије у Совјетском савезу, а имајући у виду да је у њој дат детаљан преглед терминологије као средства комуницирања, која се бави природом, карактеристикама, описом појмова (дефиниција), односима међу појмовима, формирањем термина, стандардизацијом и интернационализацијом појмова и термина итд., чак и данас представља капитално и незаобилазно дело у области теорије терминологије (Jovković 1986, 204).

Важно је, између осталог, истаћи и да је Вистер основао први међународни истраживачки термилошки центар у Визелбургу, формирањем приватног института. Овај центар је „после Другог светског рата инспирисао на рад већи број научника и инжењера, најпре у Бечу, а затим и у СР Немачкој. Тако је настала Бечка термилошка школа. Вистерове идеје су се прошириле и на остале земље и међународне организације као што је *UNESCO (United Nations Educational, Scientific and Cultural Organization)*, *ISO (International Organization for Standardization)*, *FEANI (Fédération Européenne d'Associations Nationales d'Ingénieurs)*, итд.“ (Jovković 1986, 204–205), о чему ће бити више речи у наредним поглављима.

„Вистеров допринос Општој теорији терминологије огледао се у изради и припремању термилошких принципа који су имали за циљ да усагласе међународна термилошка истраживања и разраде друге аспекте и детаље Опште терминологије“ (Jovković 1986, 205). Од 1951. године Вистер преузима Секретаријат *ISO/TC 37* „Терминологија (начела и координација)“ при Аустријском институту за стандардизацију. Од тада па све до 1963. године, на пословима библиотекара, лексикографа, термилога и организатора састанака и конференција у Вистеровом термилошком центру, радили су Фридрих и Анелис Ланг, по

⁶ *Eugen Wüster* (1898–1977), по образовању електроинжењер, захваљујући свом доприносу на пољу терминологије, сматра се једним од пионира области термилошке стандардизације. На основу његових истраживања на пољу техничке комуникације, 1936. године основан је Технички комитет за стандардизацију терминологије Међународне организације за стандардизацију (*ISO/TC 37*). Био је један од главних сарадника на састављању и објављивању речника Међународне електротехничке комисије. Уз то, радио је и на реформи немачког правописа и универзалном децималном систему класификације. Године 1971. иницирао је оснивање Инфотерма (више о томе у поглављу 2.5.2) (Kingscott 1998).

⁷ *Internationale Sprachnormung in der Technik, besonders in der Elektrotechnik*

струци лингвистици (Infoterm 2004, 3). Године 1964. Хелмут Фелбер⁸ постаје Вистеров асистент у истраживачком раду и на тој функцији се задржава све до 1970. године. Он се уједно сматра и настављачем рада Еугена Вистера након чије је смрти 1977. године, преузео његова предавања из опште теорије терминологије, на Катедри за лингвистику Универзитета у Бечу (Infoterm 2004, 3).

Литература коју је Вистер за живота сакупио, представља јединствену библиотеку коју је после његове смрти најпре наследио Аустријски завод за стандардизацију, да би библиотека касније била пребачена у Инфотерм у Бечу (Jovković 1986, 204). Ова богата колекција литературе из области терминологије, данас је део Научно-теоријског архива при Бечком универзитету. Поред ове библиотеке, Вистер је припремио око 500 публикација из области терминологије, стандардизације терминологије, међународног кључа за терминологију, документацију, превођења, теорије симбола, класификације, теорије тезауруса, лингвистике, лексикографије, ортографије и многих других области. Многи од ових радова су, како закључује ауторка Јовковић, остали необјављени и налазе се у његовој истраживачкој архиви (Jovković 1986, 205).

2.1.2. Руска термилошка школа

Историја развоја руске термилошке науке прилично је дуга, а рад на пољу терминологије почео је још у 18. веку.

Према ауторима Лајчик и Шелов, издвојена су четири периода односно етапе кроз које су прошле термилошке активности на тлу Русије (Leitchik и Shelov 2003, 82–83):

- *Припремни период* у коме је вршен избор и примарна обрада термина и дефиниција које су у вези са одређеним појмовима, а овај период почиње од настанка редовног превођења термина и састављања првог руског термилошког речника 1780. па све до краја 1920. године;
- *Први период* (од тридесетих до краја шездесетих година 20. века) током којег се појављује теорија терминологије уз велико залагање два изузетна стручњака техничког образовања. Ти стручњаци су били Дмитриј С. Лоте⁹ и Ернест К. Дрезен¹⁰. Њихов инжењерски приступ у великој је мери одредио будућу праксу на пољу стандардизације и интернационализације терминологије. У истом периоду, значајан допринос развоју

⁸ *Helmut Felber* (1925–2005), аустријски терминолог и дугогодишњи асистент Еугена Вистера, који је значајно допринео успостављању и развоју термилошких истраживања. Студирао је грађевинско инжењерство и филозофију на Универзитету у Грацу. Служио се енглеским, француским, шпанским, италијанским и руским језиком. Радио је као библиотекар, предавач, преводилац и тумач до 1964. године када га је Вистер ангажовао као научног сарадника у Истраживачком институту у Визелбургу. Од 1970. године управља међународним секретаријатом *ISO/TC 37* на аустријском институту за стандардизацију у Бечу. У периоду од 1971–1985. године руководи Међународним информационом центром за терминологију, енг. *Infoterm* (више речи о томе у поглављу 2.5.2).

⁹ *Дмитрий Семёнович Лотте* (1898–1950), један је од оснивача Комитета научно-техничке терминологије и зачетника совјетске термилошке школе. Године 1931. објавио је чланак под називом *Очередные задачи технической терминологии* („Проблеми на пољу техничке терминологије“), исте године када је и Вистер објавио своју докторску дисертацију. Нека од његових најзначајнијих дела су: *Вопросы заимствования и упорядочения иноязычных терминов и терминологических элементов* („Проблеми позајмљивања и преузимања страних термина и њихових елемената), *Краткие формы научно-технических терминов* („Кратки облици научних и техничких термина“), *Некоторые принципиальные вопросы отбора и построения научно-технических терминов* („Основна питања избора и конструкције научне и техничке терминологије“), и постхумно објављено 1961. године *Основы построения научно-технической терминологии* („Основи структуре научне и техничке терминологије“) (Brenes 2015).

¹⁰ *Эрнест Карлович Дрезен* (1892–1937), био је летонијски и совјетски лингвиста и есперантиста, и стручњак у области научно-техничке терминологије; председавао је Савезом есперантиста совјетских република. Године 1936. помогао је оснивање међународне групе стручњака за терминологију (данашњи Технички одбор 37). Превео је на руски (заједно са још 4 преводиоца) Вистерову дисертацију под називом „Лингвистичка стандардизација у техници“ из 1931. године. Вистер је на своју листу „интелектуалних отаца термилошке теорије“, у којој је описао динамичне и далековиде људе који су током прве половине века подстицали развој терминологије, уврстио и Дрезена (Brenes 2015a).

термиолошке теорије пружили су А. А. Реформатски и Г. О. Винокур, двојица истакнутих руских лингвиста, чији су радови на овом пољу такође имали одлучујући утицај на развој руске термиолошке науке.

- *Други период* (од 1970. до 1990.) током којег терминологија постаје самостална дисциплина. Током овог периода теоретичари и практичари из области терминологије настојали су да испуне захтеве лингвиста, логичара, стручњака из области информатичких наука у циљу дефинисања предмета и задатака термиолошке науке, као и у циљу побољшања метода и дефинисања основних проблема.

Уз то, овај период карактерише и развој и интеракција термиолошких одбора у оквиру академија наука у републикама Совјетског Савеза. Уз индустрију, интензивирани су и термиолошка активност у држави, а примарна пажња поклањала се техникама које се тичу нормирања терминологије. Током овог периода, на тлу Совјетског савеза одржале су се бројне научне конференције, састанци и симпозијуми, написано око десет монографија, објављено двадесетак збирки радова и кандидовано више од 1000 докторских дисертација из области терминологије. Осим тога, развијене су хиљаде термиолошких и енциклопедијских речника – од политехничких и општих научних речника до врло ускостручних.

- *Трећи период* (од 1990. до данас) који на почетку одликује назадовање на пољу термиолошких студија и њихов велики пад у сфери научних истраживања, а што је било узроковано променама у друштвеном животу Совјетског Савеза, након чега је уследило постепено обнављање (Leitchik и Shelov 2003, 82–83).

Совјетска термиолошка школа настала је почетком тридесетих година 20. века, као резултат научно-техничког развоја тога времена. Неки од задатака који су утицали на институционално дефинисање рада на научној терминологији у свим предметним областима били су постизање неопходног разумевања међу стручњацима унутар једне предметне области односно сродних дисциплина, развој стандардизације којом се потпомажу и олакшавају међународне културне, друштвене и економске везе, спровођење различитих истраживања у области терминологије, обука научних кадрова, припрема научно-техничке литературе као и увођење рачунара у ову област. Још 1933. године основана је Комисија техничке терминологије (КТТ), при Академији наука СССР, која је касније претворена у Комитет научно-техничке терминологије (КНТТ). Његови оснивачи су Чаплигин¹¹ и Лоте који су задатак Комитета видели у:

- разрађивању принципа и метода стварања терминологије,
- сређивању терминологије фундаменталних општетехничких дисциплина – математике, физике и хемије – као и неких нових области науке и технике, које су се брзо развијале (електроника, рачунарска техника и др). Радови оснивача КНТТ развијају поставку о терминологији као систему термина који одговара систему појмова дате области науке и технике.

¹¹ *Сергѝй Алексѣевич Чаплѝгин* (1869–1942), био је стручњак за аеродинамику, аутор радова из области теоријске механике флуида и апроксимативних метода за решавање диференцијалних једначина. Сматрао је да научници треба активније да се укључе у решавање термиолошких проблема како би се олакшало научно истраживање. Године 1933. као члан Совјетске академије наука, Чаплигин је основао Совјетску школу терминологије заједно са Лотеом. Заједно су припремили документ са задацима и методама за управљање техничком терминологијом и основали Комисију техничке терминологије Академије наука, чија је улога била стварање техничке и научне теорије терминологије и установљавање начела за формирање термина и концептуалних система. Такође је помогао у развоју термиолошких система главних технолошких дисциплина као и метода којима се уређује руска техничка терминологија коју користе научници и инжењери (Brenes 2015b).

Овај Комитет пратио је и друге аспекте развоја термилошке делатности, попут искуства државне стандардизације терминологије појединих грана технике на тлу Совјетског савеза, удео совјетских лингвиста у истраживању термилошке лексике, развој радова на терминологији у иностранству, радове у области међународне координације научно-техничке терминологије, координирао рад на терминологији у Совјетском савезу и окупљао стручњаке одговарајућих области науке и технике, лингвистике и логике (Alaburić 1986, 208).

Имајући у виду тадашње прилике и развој економских, научних и техничких релација Совјетског савеза са иностранством, било је више него потребно ојачати капацитете и остварити међународну координацију терминологије због чега се и формирала совјетска научна термилошка школа, која је сарадњу углавном остваривала са Међународном организацијом за стандардизацију (*ISO*), Међународном електротехничком комисијом (*IEC*) и другим институцијама. Уз све то, још један одлучујући фактор за њено формирање била је вишејезичност на тлу бившег Совјетског савеза.

2.1.3. Прашка термилошка школа

Лингвистичко друштво под називом „Прашки лингвистички круг“ (које се неретко назива тзв. *Прашким школом*), основано је у Прагу 1926. године. Било је интернационалног карактера, а међу његовим централним фигурама истицала су се следећа имена: Р. Јакобсон¹², С. Карчевски¹³, Н. Трубецкој¹⁴, В. Матесијус¹⁵, Б. Трнка¹⁶, Б. Хавранек¹⁷, Ј. Мукаровски¹⁸, Ј. Вачек¹⁹ и др. Друштво је своју активност развијало све до немачке окупације Чехословачке, а смрћу Трубецког и одласком Јакобсона из Чехословачке, рад овог друштва замире. Тек је крајем педесетих и почетком шездесетих година прошлога века Праг поново добио репутацију снажног лингвистичког центра; око некадашњих чланова друштва окупили су се бројни

¹² *Роман Осипович Јакобсон* (1896–1982), један од најугледнијих лингвиста 20. века, пореклом из Русије. Изучавао је словенске језике и био је један од оснивача Прашког лингвистичког круга. Још као студент на Универзитету у Москви, основао је Московски лингвистички круг. Развио је револуционарне технике за анализу лингвистичких фонолошких система (Кедрова, Омелјанова и Егоров 2002).

¹³ *Сергеј Иосифович Карцевски* (1884–1955), руски лингвиста, професор руског језика и књижевности на Универзитету у Женеви. Био је један од Сосиријевих ученика, а такође и један од иницијатора оснивања женевског лингвистичког друштва, 1940. године (Amacker 1995, 241).

¹⁴ *Николај Сергеевич Трубецкој* (1890–1938), истакнути руски лингвиста, историчар језика, слависта, теоретичар структурализма, активиста Московског лингвистичког круга, један од идеолога и оснивача Прашке школе. Имао је водећу улогу у развоју лингвистике као нове научне дисциплине. Први је научник који је одвојио фонетику и фонологију и дефинисао њихове задатке. Сматра се и оснивачем морфологије (Кедрова, Омелјанова и Егоров 2002).

¹⁵ *Vilém Mathesius* (1882–1945), англиста по струци, у великој мери је допринео лингвистичком развоју своје средине. Његов рад „О potenciálnosti jevů jazykových“ који је објављен још 1911. године (у публикацији „Věstník Královské české společnosti nauk“ II на стр. 1–24) сматра се данас значајним за утирање нових путева у теоријском осветљавању језика (Ivić 2001, 201). Матесијус је имао кључну улогу у стварању Прашке школе, њене основне принципе формулисао је чак 15 година пре њеног установљења. До своје смрти 1945. године био је њен председник (Urbanová 2012, 6).

¹⁶ *Bohumil Trnka* (1895–1984), чешки лингвиста је током студија био под утицајем еминентних професора времена у ком је живео, између осталих и професора Матесијуса. Био је члан бројних удружења (једно од таквих удружења био је и Круг модерне филологије), и учествовао у раду бројних међународних научних организација. Након што је прашки лингвистички круг престао са радом 1951. године, проф. Трнка је основао радну групу за функционалну лингвистику у Кругу модерне филологије, научном удружењу повезаном са Чехословачком академијом наука у Прагу (Urbanová 2012, 29).

¹⁷ *Bohuslav Havránek* (1893–1978), чешки лингвиста који је помогао оснивање Прашког лингвистичког круга 1926. године. Најпознатији је по свом раду на теорији стандардног језика, језичкој култури и упоредној анализи славенских језика (Chovanec 2014, 27).

¹⁸ *Jan Mukařovský* (1891–1975), чешки лингвиста и професор на Универзитету у Прагу. Један од чланова прашког лингвистичког круга и значајан поборник чешког структурализма. У свом раду истицао је улогу естетске функције сматрајући је кључном карактеристиком било ког уметничког дела (Chovanec 2014, 41).

¹⁹ *Josef Vachek* (1909–1996), чешки лингвиста и англиста, један од истакнутих представника Прашког лингвистичког круга. Бавио се фонологијом, фонетиком и граматиком енглеског и чешког језика, општом лингвистиком и теоријом превођења и писаног језика (Chovanec 2014, 54).

ученици и сарадници који су својом активношћу продубљивали лингвистичке теорије (Ivić 2001, 201).

Појединци су, нарочито у почетку, називали ову школу и школом функционалне лингвистике (она се првенствено интересује за то како једна гласовна јединица функционише као знак за споразумевање). Касније се често појављује термин *школа фонолога*, јер су представници ове школе од почетка били првенствено заинтересовани за фонолошку проблематику, а данас је најраширенији назив *Прашка школа* (Ivić 2001, 156).

Ова школа била је филозофски усмерена, а ослањала се на теорије о језику као функционалном систему; део тог језика је и терминологија као један од функционалних слојева (Karadža-Garić 1988, 26).

Ауторка Ивић (Ivić 2001, 202–203) наводи више теоријских ставова које је заступала Прашка школа међу којима је и тај да је језик систем изражајних средстава који служи споразумевању, због чега треба проучавати конкретну функцију конкретних исказа: шта се саопштава, како, коме, у којој прилици. Такође, према теоријском ставу Прашке школе, језик је реалност условљен спољашњим (нелингвистичким) моментима: социјалном средином, аудиторијумом којем се саопштава, тематиком која се износи због чега је неопходно разликовати, и теоријски и практично језик културе уопште од језика лепе књижевности, језик науке од језика новина, језик улице од језика канцеларије, итд. Теорија да се језиком износе и интелектуалне и емоционалне манифестације човекове личности, да говорни и писани језик нису апсолутно идентичне појаве, да је за лингвистику подједнако важно синхроно као и дијахроно испитивање, да се језичке појаве не могу изоловано посматрати без целовитијег погледа на корелацију између свих постојећих језичких факата као и да морфолошке појаве не треба изоловати од фонолошких јер фонемске опозиције често имају значај на морфолошком нивоу (нпр. алтернација слова *к* у *ч* у српском језику, итд.), још неки су од ставова које према овом аутору (Ivić 2001, 202–203) заступа Прашка школа лингвистике.

Прашка термилошка школа развила се из прашке школе функционалне лингвистике чије теорије, засноване на раду Сосира,²⁰ истичу функционалне аспекте језика и представљају солидну основу за теорију терминологије (Jovković 1986, 206).²¹ Циљ ове школе је, према наводима ауторке Боздехове (Bozděchová 2017, 109), био структуралистички и функционалан опис специјализованог (стручног, техничког) језика. Како ова ауторка даље истиче, Прашка термилошка школа заступала је став да се термини увек морају посматрати као јединице речника, односно, делови речника народног језика, дакле, као специфични предмети лингвистичког испитивања. Узимајући у обзир вишејезичност на тлу Чехословачке, стандардизација језика и терминологије била је једна од централних тачака њеног интересовања, због чега је и рад ове школе био тесно повезан са Институтом за чешки језик, који је део Чешке Академије наука (Cabré 1999, 13).

2.2. Критика традиционалних школа терминологије

Имајући у виду да је са развојем науке и технике на глобалној сцени дошло и до одређених промена у свим научним областима, па тако и у терминологији, одређени број савремених теоретичара терминологије почео је да заступа модерније ставове у овој области, услед чега су се појавиле и бројне критике традиционалних теорија терминологије.

²⁰ *Ferdinand de Saussure* (1857–1913), швајцарски лингвиста и семиотичар који је значајно допринео развоју лингвистике и семиотике. Тачније, сматра се једним од утемељивача модерне лингвистичке теорије, а његови текстови били су манифест интелектуалног покрета 20. века, који је имао утицај на различите дисциплине попут психологије, социјалне антропологије и књижевне критике. Тај покрет данас је познат под називом *структурализам* (Harris и Taylor 1997, 209).

²¹ Прашка школа директно се надовезује на завршне теоријске спознаје руског формализма, као и на структуралистичке поставке швајцарског лингвисте Сосира. Међутим, развој школе је неодвојив и од снажног интелектуалног замаха међуратне Европе, што значи да јој се концепција обликовала на плодном научном укрштању идеја феноменологије, психоанализе, логичког позитивизма, гешталтистичке психологије (Hrvatska enciklopedija 2019).

Алтернативне, социокогнитивне термилошке парадигме, огласиле су се у серији монографија издавачке куће *John Benjamins* под називом *Terminology and Lexicography Research and Practice* („Истраживање и пракса у области терминологије и лексикографије“) (Šipka 2006, 153).

Аутор једне од таквих монографија је Рита Темерман, професорка преводилачких и термилошких студија на Катедри за примењену лингвистику Универзитета Еразмус у Бриселу²². Својом монографијом под именом *Towards New Ways of Terminology Description: The Sociocognitive Approach* („Нови начини описивања терминологије: социокогнитивни приступ“), она доводи у питање принципе и методе традиционалних школа сматрајући да је њихова једина мотивација језичка политика и планирање, а прагматичка оријентација стандардизација терминологије, што имплицира финансирање од стране владиних тела, а не језичке циљеве (Анђелковић 2017, 42). Циљ овог приручника је да објасни нову теорију, методе и примену термилошких истраживања којима се тежи да се превазиђу очигледна ограничења традиционалне терминологије коју је представљала Бечка термилошка школа (Gläser 2004, 434). Према наводима аутора Шипка, традиционалну терминологију, првенствено ону бечке школе, ауторка види као одраз сосировског структуралистичког приступа (Šipka 2006, 154).

Према мишљењу ауторке Темерман, традиционална терминологија је заправо ометала стварање, дефинисање и постављање појмова у концептуални систем, и тај аргумент се понавља и постаје водећи мотив кроз целу књигу (Gläser 2004, 434).

У првом поглављу своје књиге, које носи наслов „Од начела стандардизације до научне студије терминологије“, ова ауторка објашњава како главна начела тако и главне недостатке традиционалне термилошке школе, па тако наводи:

I начело: ономасилошки приступ – традиционална термилошка школа упућује на начин на који су појмови преузимани из спољног света и означавањем терминима, без разматрања какву улогу има језик у формирању појмова.

II начело: тврдња традиционалне термилошке школе да су појмови јасно дефинисани, па стога нејасноће односно преклапања треба да буду искључени из термилошког система који има добро осмишљену структуру. Ауторка опомиње да су многе категорије нејасне и стога не могу безусловно бити класификоване према логичким или онтолошким значењима.

III начело: додела термилошких дефиниција појмовима (појам се у идеалном случају дефинише формализованим дефиницијама према начелу традиционалне терминологије). Ауторка, пак, тврди да формализоване дефиниције нису могуће нити пожељне.

IV начело: једнозначност – једном појму припада само један термин и један термин означава само један појам чиме се, заправо, искључују случајеви синонимије и полисемије.

V начело: једном додељен тј. означен појам/термин остаје трајан, према начелу традиционалне термилошке школе. Ова ауторка ипак тврди да долази до еволуције појмова, да се категорије и термини мењају у значењу, и да се њихово разумевање шири. Временом долази до обликовања нових појмова у техничким текстовима као и специјалистичком дискурсу али то је изван предмета интересовања традиционалне школе терминологије.

Ова ауторка сумира и ставове различитих аутора који традиционалне приступе терминологији сматрају проблематичним, на првом месту, Хуана Карлоса Сагера који доводи у питање „дисциплинарни статус“ терминологије. Он терминологију сматра низом практичних активности и методологија, али не и научном дисциплином. Међутим, уколико се ослободи ограничења која јој намећу стандардизација и редукционистички приступ, терминологија може значајно допринети развоју когнитивних наука и социолингвистици, закључује Сагер (Temmerman 2000, 23).

²² На основу студија случаја о категоризацији и именовању у оквиру природних наука, ауторка Темерман (енг. *Rita Temmerman*) је развила социокогнитивну теорију терминологије. Поље њеног интересовања представљају: теорија терминологије и термилошки менаџмент, когнитивне науке и семантика, теорија превођења, студије метафоре, динамички системи у језику, интеркултурна и вишејезичка комуникација (Terminology Coordination Unit of the European Parliament 2019).

С друге стране, Темерман наводи и став Питера Вајзенхофера, који критикује основну поставку Бечке термиолошке школе према којој је једнозначност термина основни услов за недвосмислену комуникацију. Овај аутор сматра да стручњаци појединих друштвених наука могу имати различита схватања централних појмова из њихових предметних области. Он даље закључује да је разумевање појмова засновано на индивидуалном тумачењу карактеристика које неки појам има (Temmerman 2000, 26).

Аутори Завада и Свонепул сматрају да иако су природне науке прецизније од хуманистичких и друштвених, класични приступ појму није адекватан да објасни појмовну структуру природних наука (Temmerman 2000, 27). Ови аутори истичу да је потребна градација одређених категорија, односно скаларна врста дефиниција због постојања прототипски организованих појмова (Šipka 2006, 154).

Тереса Кабре Кастелви, у свом чланку *Theories of terminology: Their description, prescription and explanation* („Теорије терминологије: опис, препоруке и објашњења“) сматра да је Вистер развио термиолошку теорију на основу сопственог искуства као инжењера који је био укључен у националну и међународну стандардизацију терминологије. Према њеном мишљењу, Вистерови теоријски закључци засновани су на његовом личном искуству са ограниченим сектором терминологије, због чега је развио теорију о томе каква терминологија треба да буде како би се обезбедила недвосмислена комуникација, а не теорију о терминологији каква она заправо јесте, у свој својој разноликости и плуралности (Castellví 2003, 167). Ова ауторка, у наставку, истиче да традиционална теорија посматра термиолошке податке као садржај стандардизованог речника у ком је терминологија недвосмислена и систематична, термини из различитих језика потпуно еквивалентни а појмови увек семантички прецизни. Међутим, како даље наводи, ако посматрамо термине у њиховом природном дискурсу, приметимо варијације својствене различитим регистрима у специјализованој комуникацији, двосмисленост, редувантност, несистематичност, појаву појмовне и синонимијске варијације као и непостојање савршене међујезичне еквиваленције, због чега је неопходно шире поимање терминологије од традиционалног погледа који се своди на стандардизацију (Castellví 2003, 178).

Још једну критику традиционалне термиолошке школе упућује проф. Ингрид Мејер са Универзитета у Отави. Према речима Темерман, ова ауторка истиче важност енциклопедијске и лексикографске дефиниције, верујући да је корисницима терминологије, који нису предметни стручњаци, неопходна енциклопедијска информација како би разумели одређени термин. Поред тога, она истиче могућност вишеструке класификације појмова, с обзиром на то да се појмовне категорије могу преклопити на много начина, али и то да се терминологијом може унапредити недвосмислена комуникација (Temmerman 2000, 31), као и да супротстављеност може креирати једнозначност или се усредсредити на различитост (Šipka 2006, 154).

На француском говорном подручју појавио се нови правац под именом социотерминологија (Temmerman 2000, 31). У оквиру њега, аутори се залажу за изучавање полисемије и синонимије у оквиру термиолошке науке, што је у начелној супротности с идеалом моносемије који је заговарала традиционална школа. Друго, ови аутори доводе у питање постојање односно јасно разграничење домена и предметних поља, и залажу се за дијахронијски приступ (на супрот синхронијском) изучавању концептуализације и именовања. Аутори који заговарају овај правац, сматрају да треба изучавати стварну употребу језика (Temmerman 2000, 32).

Још један аутор чије мишљење Темерман издваја је Кагеура, предавач са Универзитета у Токију. Он, на име, појам замењује значењем и изучава употребу, а не систем. Овај аутор сматра да су јасна разграничења корисна само за нека поља, а да терминологију треба сместити у шири оквир лингвистике, уместо што је посебна, одвојена дисциплина (Temmerman 2000, 33).

Из свега наведеног, можемо закључити да за разлику од традиционалне терминологије која подразумева јасне и формализоване дефиниције, модерни правац узима у обзир развојну страну језика, па тако одбацује једнозначност и сматра да треба изучавати синонимију и

полисемију. Уз то, модерни приступ истиче важност енциклопедијске информације за кориснике који нису предметни стручњаци (Šipka 2006, 155).

Иако савремени аутори терминологије настоје да своје деловање представе као алтернативу традиционалним приступима, њихове дефиниције нису ништа друго до допуна постојећој традиционалној парадигми, која ће наставити да функционише у својој сфери употребе (Šipka 2006, 155).

2.3. Терминологија данас

У свим високоразвијеним, а све више и у недовољно развијеним друштвима и културама, терминологија постаје предмет или компонента научног развоја, језичког планирања и језичке политике, објект нормирања, усаглашавања и широко заснованих или сталних лексикографских, социолингвистичких и семантичких испитивања (Radovanović 1988, 9–10). Како овај аутор даље наводи, све већи значај добија стварање стручних речника на националним језицима, односно специјализованих гранских и монодисциплинарних речника. Терминологија је све тешње повезана с проблемима обичног и машинског превођења, с разрадом рачунарских и информационо-претраживачких система, документационо-библиотечких система и служби и система аутоматизованих истраживања и аутоматског управљања у појединим областима друштвеног рада и стваралаштва (Radovanović 1988, 9–10).

Оваква позиција терминологије покренула је идеју о настанку нове термилошке области којом се обједињује употреба рачунара и терминологије. У свом приручнику под називом *Terminology: Theory, methods and applications* („Терминологија: теорија, методе и примене“), ауторка Кабре уводи израз *Computer-aided terminology*²³ и истиче колико важну улогу рачунари имају у обављању одређених термилошких активности (Cabré 1999, 165). Ту улогу види најпре као подршку у избору грађе, састављању корпуса и екстраховању података, приликом уношења одредница као и у контексту провере информације о унесеној одредници.

И док се у раној фази употребе рачунара у терминологији, њихово коришћење сводило на складиштење и проналажење терминологије (у сврху справљања речника односно испомоћи приликом превођења) (Felber 1984, 11), дотле је општи напредак науке и технике, а све захваљујући присуству рачунара у свакодневном животу људи, учинио да савремене рачунарске методе преузму главну улогу и у терминологији као науци (Foo 2012, 13).

Данашњи рад на терминологији тешко је замислити без савремених рачунарских алата и ресурса за терминологију, који тај рад чине далеко бржим, ефикаснијим и поузданијим. Ове језичке и термилошке системе развијају рачунарски лингвисти и стручњаци за обраду природних језика, о чему ће бити више речи у поглављу 6 овог рада.

Аутори Буригол, Жакмен и Ом²⁴ (Bourigault, Jacquemin и L’Homme 2001, IX) издвајају три ставке из области рачунарске терминологије²⁵:

- аутоматска идентификација термилошких јединица и филтрирање листе кандидата (термина),
- груписање варијација и синонима,
- проналажење односа између термина.

²³ Термин *Computer-aided terminology* спомиње први пут ауторка Кабре, у поменутом приручнику у ком се реферише на Ожера и његов рад под називом *La terminotique et les industries de la langue* из 1989. године. Овај рад је написан на француском језику, а аутор није користио енглески термин *Computer-aided terminology*. Ствар још компликованијом чини и то што је оригинална књига ауторке Кабре на каталонском језику (преводиоца на енглески језик била је Ценет Децесарис а уредник Хуан Карлос Сагер). Стога се и може поставити питање да ли је поменути термин заправо сковала сама ауторка Кабре (Foo 2012, 13). Такође, ова ауторка изједначава термине *Computerized terminology* и *Computer-aided terminology*.

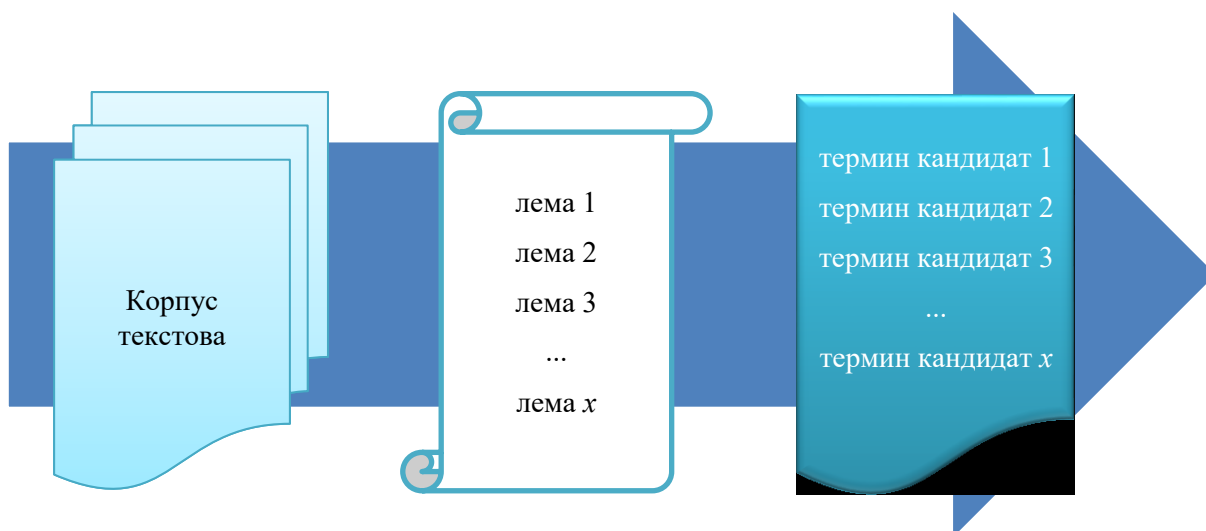
²⁴ *Didie Bourigault, Christian Jacquemin, Marie-Claude L’Homme.*

²⁵ Израз који користе ови аутори на енглеском гласи *Computational Terminology*.

У издању исте књиге из 2003. године, ови аутори говоре о *Обради природног језика у области терминологије*²⁶, тврдећи да су дефиниције и претпоставке класичне терминологије „мање прилагођене рачунарском приступу термилошкој анализи“. Као главне активности које спајају терминологију и *Обраду природног језика* ови аутори наводе: аутоматско проналажење нових термина и аутоматску идентификацију постојећих, познатих термина у корпусу текстова (Bourigault и Jacquemin 2003, 604).

Проналажење тј. откривање нових термина врши се аутоматском екстракцијом из текста што подразумева: екстракцију речи тј. јединица, било да се ради о монолексемским (енг. *Simple-word term / unit / expression* скр. *SWT, SWU, SWE*) или полилексемским јединицама (енг. *Multi-Word Term / Unit / Expression* скр. *MWT, MWU, MWE*), њихову категоризацију односно рангирање и издвајање потенцијалних термина тј. термина кандидата, док се препознавање познатих термина врши аутоматским индексирањем, којим се проналазе релевантне речи односно јединице и евидентира њихова локација у документу (Foo 2012, 14–19). Аутоматска екстракција терминологије детаљно је представљена у поглављима 7.3 и 7.4 овог рада.

На слици 1 приказан је процес аутоматске екстракције терминологије подељен на три дела/задатка, у ком се најпре из корпуса текстова врши екстракција речи које се, потом, категоризују и на крају издвајају као потенцијални термини односно термини кандидати.



Слика 1. Приказ схеме обраде текста

Екстракција термина може бити једнојезична или вишејезична. Код једнојезичне екстракције, анализира се текст односно корпус и идентификују термини кандидати, док се код вишејезичне екстракције анализира текст заједно са његовим преводом како би се идентификовали потенцијални термини и њихови еквиваленти на другом језику (Dunaevsky 2019).

Па ипак, упркос чињеници да алати за екстракцију термина олакшавају термилошки рад, помажући у попуњавању термилошке базе односно успостављању терминологије одређене области, добијену листу термина мора верификовати терминолог односно преводилац. Из тог разлога, за процес екстракције термина не можемо рећи да је у потпуности аутоматизован (Dunaevsky 2019).

Детаљније о поступцима аутоматске обраде текстова и издвајању терминологије уз помоћ рачунарских алата биће речи у поглављима 6 и 7 овог рада.

²⁶ *Term-oriented NLP*

2.4. Термин

Као што је већ истакнуто, терминологија је грана лексикологије која, према наводима ауторке Драгићевић, проучава термине, речи које се користе у некој струци или науци (Драгићевић 2010, 20).

Према Кабре, термилошка јединица, односно термин, јесте конвенционалан симбол који представља појмове унутар одређеног предметног поља (Cabré 1999, 81). Међутим, дефинисање термина није увек једноставан задатак. Зато је према аутору Шипка, једно од централних питања лингвистичке терминологије прављење разлике између *термина* и *нетермина* и то не извођењем дефиниције већ набрајањем карактеристика које се сматрају искључивим или претежним атрибутима термина (Šipka 2006, 150). Према овом аутору, карактеристике термина су следеће: термин мора да буде дефинисан и уклопљен у предметно поље за разлику од нетермина, термин је резултат свесне интервенције што није случај с већином других лексема, код термина је битно да се успоставе прецизни односи термилошке корелације између различитих језика, чиме се, онда, омогућује међујезичка и међукултурна комуникација и, најзад, у случају термина првенствено је битна његова синхрона димензија, односно његово актуелно функционисање, док је код осталих лексичких јединица подједнако важна и дијахрона. Како даље наводи овај аутор, одређење термина не треба схватати као обавезно подложно свим наведеним карактеристикама јер се не ради о потребним и довољним условима, него о списку најчешће присутних атрибута. Међутим, према његовим речима (Šipka 2006, 151), у идеалном случају, један термин, односно терминосистем треба да има следеће карактеристике:

- а. транспарентност (особина термина да се већ у самом називу виде и одређене карактеристике појма који је њиме изражен),
- б. интернационалност (особина да је термин интернационално препознатљив),
- в. устаљеност (особина да је термин дуже време присутан, општепознат и прихваћен),
- г. краткоћа (особина да термин није гломазан),
- д. системност (уклопљеност у термилошки систем),
- ђ. недвосмисленост (особина да се остварује веза термина само са једним појмом унутар једног предметног поља),
- е. прецизност (особина да термин без преливања и без остатка покрива свој појам),
- ж. несинонимност (особина да се остварује веза појма само са једним термином унутар предметног поља).

Да би се све ове карактеристике задовољиле, неопходна је сарадња како стручњака за језик – лексиколога тако и стручњака за одређену предметну област. Како наводи Драгићевић (Драгићевић 2010, 20), лексиколози су нарочито заинтересовани за питање у којој је мери термин различит у односу на лексеме општег лексичког фонда: теже ли термини да се прилагоде уобичајеним законитостима и односима својственим у општем лексикону, као што су полисемија, деривација, синонимија или антонимија, и да ли те односе успостављају мешајући се са општом лексиком или остају затворени у своје термилошке системе?

Данас постоји велики број радова који се баве питањем терминологије и одређењем термина, како у домаћој тако и у страниој стручној литератури. Међутим, у савременој теорији терминологије, како истиче Дудок, термин се углавном дефинише статички: посматра се као елемент лексичког фонда којим се именује појам одређен дефиницијом и местом у систему појмова конкретне струке, технике, привреде и сл., односно, термин се дефинише као продукт односа између појма и назива, при чему се појам одређује помоћу односа према скупу појмова који спадају у исту област (Дудок 1996, 17). Међутим, природа термина није само статичка већ је његова функција ипак много шира. Поред тога што своје место проналазе у термилошким речницима и сл., термини представљају изражено средство стручне и научне комуникације, а реализују се у стручном тексту који представља њихову природну средину, где се термини комбинују са речима из општег језика (Дудок 1996, 17). Како даље истиче овај аутор, главно питање је, дакле, да ли је термин специјална лексикална јединица или је то лексикална јединица у специјалној функцији. У лингвистички оријентисаној терминологији

превагнуо је овај други став, и како овај аутор даље закључује, иако приликом конституисања и структурације термина лингвистика има веома широк круг деловања, јер је термин пре свега језичка јединица, она овде није сама, већ је у тесној спрези са филозофијом²⁷ и конкретном струком у коју дати назив или деноминирани појам спада (Дудок 1996, 17).

Термини, према Гортан-Премк (Гортан-Премк 2004, 119), имају ограничену способност развијања вишезначности а та особина, лежи у двојакој природи термина као знака: термин припада и општем лексичком фонду, општем језичком систему, и самим тим подлеже свим лексичким законитостима, па и тежњи ка полисемантичности, али у исто време термин припада и посебном терминолошком систему па тако подлеже и његовим законитостима, па и тежњи ка моносемантичности, ка директном, симетричном односу знака и појма који му је у семантичком садржају. Термини не могу ширити своју семантичку структуру метафоричким путем, како даље наводи ова ауторка, управо зато што је семантички садржај термина појмовни, у њему нема релевантних елемената реализације, а управо су ови елементи изворишта различитих метафоричких асоцијација на основу којих се развија много значења. Они се могу полисемантички развијати метонимијом и синегдохом али ограничено потребом моносемантичности у систему коме припадају, закључује Гортан-Премк.

До проблема вишезначности, како даље наводи (Гортан-Премк 2004, 120), долази и због различите експликације коју поједини аутори дају једноме појму, а до различитог појмовног садржаја може доћи и због неједнаког развитка науке којој термин припада. Међутим, ова ауторка сматра да је у науци спорнија ситуација када се један исти термин јавља у различитим терминолошким системима и са различитим појмовним вредностима, па је недоумица да ли овакве случајеве треба сматрати терминолошким хомонимима или различитим појмовним вредностима истог термина какав је нпр. случај са термином **деривација**, који има потпуно различито значење у лингвистици и у војним наукама (Гортан-Премк 2004, 120). Како даље закључује ова ауторка, термини су, реално или потенцијално, и јединице општег лексичког фонда, па тако када изађу из својих затворених терминолошких система, из језика науке, и када уђу у општи језик, почињу да га богате и развијају полисемију (Гортан-Премк 2004, 123).

Говорећи о узроцима настанка полисемије, (Драгићевић 2010, 130) наводи да је основни разлог због којег долази до полисемије постојање знатно више предмета и појава него лексема за њих па тако, када би свака ствар око нас имала засебан назив, не бисмо могли да се споразумевамо, јер не бисмо могли да запамтимо све лексеме. Међутим, како даље наводи ова ауторка, недостатак лексике не треба тумачити сиромаштвом једног језика већ ограниченим когнитивним способностима корисника језика. Деривацијом значења од основног настаје скуп значења једне лексеме који представља њену полисемантичку структуру а састоји се из примарног/основног/номинационог значења и секундарних значења (Драгићевић 2010, 131).

Због природе терминолошких система, термини су, према (Гортан-Премк 2004, 122) неспособни за развијање синонимије, међутим, уколико до ње ипак дође, то је резултат паралелне употребе речи различитог порекла, најчешће домаћег и страног, као нпр. ЗАРАЗА – ИНФЕКЦИЈА. Међутим, термини имају већу могућност успостављања антонимских односа него лексеме општег лексичког фонда, нпр. МАТЕРИЈА – АНТИМАТЕРИЈА, ДЕМИНУТИВ – АУГМЕНТАТИВ, СИНОНИМ – АНТОНИМ (Драгићевић 2010, 21).

2.5. Терминологија и стандардизација

Опште је позната тврдња да стручни језик захтева виши степен прецизности у односу на општеупотребни језик. Сталан и брз напредак науке и технологије узрокује и брзо именовање нових појмова, па се терминологија стручних области може развити неплански, без одређеног система и прецизних правила. Насупрот томе, организована терминолошка стандардизација, коју спроводи надлежно тело за стандардизацију, обезбеђује систематичан и плански рад на терминологији с циљем да се уклоне евентуалне препреке у комуникацији међу стручњацима и успостави ред у датој области. Управо је то разлог што су научници и техничари

²⁷ Зато што је термин у тесној вези са логиком и онтологијом и најважније основно питање терминологије је, у ствари, однос појма и термина (Дудок 1996, 17).

почетком 20. века сматрали да је неопходно регулисати и уредити терминологију стручних области, а тиме и директно учествовати у процесу стандардизације (Cabré 1999, 194).

Међународна организација за стандардизацију,²⁸ основана у циљу развијања универзалних стандарда како би се унапредила комуникација и кооперација међу нацијама и уклониле баријере у области трговинске размене на међународном нивоу, дефинише термин стандардизација као процес формулисања и примене правила за систематски приступ одређеној активности ради добробити и уз сарадњу свих заинтересованих страна, а нарочито ради унапређења економије, водећи посебно рачуна о функционалним условима и безбедносним захтевима (Cabré 1999, 195).

Како Кабре даље наводи, крајњи резултат процеса стандардизације је одлука комисије донета консензусом, која се објављује у документу под именом *стандард*.

Стандардизација, која је пре свега усмерена на елиминисање непотребних разлика у циљу олакшања робне размене, највише се примењује на производима, производним методама, индустријским процесима и јавним службама (Cabré 1999, 199–200), а како даље наводи ова ауторка, имајући у виду да су термини главно средство комуникације у процесу стандардизације ових производа, сама стандардизација терминологије овде има кључну улогу.

Стандардизација терминологије је свесно обликовање језика (Felber 1977, 90). Али сам израз „стандардизација терминологије“ користи се и у општем смислу да означи да су хомонимија и синонимија елиминисани односно да је један термин недвосмислено додељен једном појму, закључује овај аутор.

Према поставкама једне од најизраженијих представница совјетске термилошке школе, М. П. Даниленко, свака уређена терминологија треба, поред осталих захтева, да задовољи и следеће лингвистичке критеријуме процене (Karadža-Garić 1988, 27):

I Општи захтеви

1. Граматичка средства изражавања стандардизоване терминологије;
2. Однос према страном, дијалекатским и простионародним речима које се јављају у улози термина;
3. Лексичко-семантичке карактеристике термина (синонимија, полисемија);
4. Кратка форма термина;

II Нормативни захтеви стандардизације терминологије:

5. Подударање термина са основним начинима и моделима грађења речи у општем језику или са специјалним моделима у терминоградњи;
6. Однос према суфиксалним варијантама термина, с обзиром на основне тенденције развоја система творбе речи (међу њима и терминоградње): тенденције ка законитости функционисања творбених модела, ка специјализацији значења творбених афикса и модела и према стварању терминолошких гнезда која се структурно и појмовно подударају;
7. Специфичности функционисања граматичких категорија у терминологији;
8. Стилистички захтеви у терминолошким конструкцијама (у сложеном термину) и у њиховим дефиницијама.

Према Фелберу (Felber 1984, 12–13), терминолошки рад се састоји из следећих активности:

- сакупљање и бележење термина који су приписани појмовима одређеног предметног поља,
- проналажење, стварање или стандардизовање система појмова одређеног предметног поља,
- проналажење или стандардизовање термина за појмове, приписивање термина одређеном појму и обрнуто,
- описивање појмова помоћу објашњења или дефиниција или стандардизовање објашњења или дефиниција,

²⁸ *International Organization for Standardization (ISO)*. Опширније о овој организацији у наредном поглављу.

- бележење термилошких и с њима повезаних података. Термилошки подаци су: термини, дефиниције, објашњења, контекст, концептуалне везе, еквиваленти у другим језицима, итд. Остали повезани подаци су: извори који се односе на појединачне податке, датум бележења итд.
- поређење појмова у традицији различитих језика и утврђивање степена еквивалентности, односно проналажење еквивалентних термина.

Област терминологије која се бави описивањем (дескрипцијом) употребе термина зове се **описна (дескриптивна)** терминологија; област која се бави прописивањем (прескрипцијом) термина за званичну употребу, што у типичном случају подразумева њихову стандардизацију зове се **прописивачка (прескриптивна)**, и коначно, она област која се бави бележењем и презентовањем термилошких података, обично у виду термилошких речника или термилошких база података зове се **терминографија** (Šipka 2006, 152).

Стандардизација терминологије која се спроводи тимским радом стручних лица у строгом смислу речи подразумева (Felber 1977, 90):

- одређивање значења термина дефиницијама;
- одређивање места појмова у систему појмова (на основу логичких или онтолошких односа међу појмовима);
- означавање сваког концепта недвосмисленим термином (одабраним међу постојећим синонимима);
- стварање термина за појам, када се међу постојећим синонимима не може наћи одговарајући термин.

Свака од ових активности одвија се унутар институција и организација задужених за стандардизацију и спровођење оваквих активности (Felber 1977, 90).

Најистакнутије организације за стандардизацију на међународном нивоу су Међународна организација за стандардизацију – *ISO (International Organization for Standardization)*, Међународни информациони центар за терминологију – *INFOTERM (International Information Centre for Terminology)*, Европски комитет за стандардизацију – *CEN (Comité européen de normalisation)*. Међу истакнутим међународним организацијама за стандардизацију у области електротехнике, која је од посебног значаја за ову дисертацију, издвајају се: Европски комитет за стандардизацију у електротехници – *CENELEC (European Committee for Electrotechnical Standardization)* и Међународна електротехничка комисија – *IEC (International Electrotechnical Commission)*. Ове европске организације за стандардизацију формирала је Европска унија, у циљу остваривања јединственог европског тржишта.

Национално тело за стандардизацију Републике Србије је Институт за стандардизацију, чији је оснивач Влада Републике Србије. Стандарди које дефинише и усваја Република Србија имају ознаку *SRPS* (некадашњи *JYS* стандард). Стандарди усвојени на међународном нивоу имају ознаку *ISO*, док стандарди усвојени у Европској унији имају ознаку *EN*.

2.5.1. Међународна организација за стандардизацију – ISO

Међународна организација за стандардизацију је независно тело, невладина међународна организација, задужена за доношење стандарда, а чине је представници завода за стандардизацију из различитих земаља. Тренутан број земаља чланица је 165, а једна земља може имати само једног представника.²⁹

Званично је основана 1947. године у Женеви, и до данас је објавила преко 23573 међународних стандарда из готово свих сфера технологије и производње. Ова организација основана је још 1926. године као Међународна федерација националних удружења за стандардизацију (*International Federation of the National Standardizing Associations – ISA*), међутим, престала је са радом 1942. године, за време Другог светског рата. Након завршетка рата, новоосновани Комитет за стандарде Уједињених нација (*United Nations Standards*

²⁹ Сви подаци у вези са Међународном организацијом за стандардизацију преузети су децембра 2020. године са веб-сајта ове организације: <https://www.iso.org>.

Coordinating Committee – UNSCC), предлаже формирање новог тела за стандардизацију које ће функционисати на глобалном нивоу.

Октобра 1946. године, представници Међународне федерације националних удружења за стандардизацију (енг. *ISA*) и Комитета за стандарде Уједињених нација (енг. *UNSCC*) оформили су нову међународну организацију за стандардизацију (енг. *International Organization for Standardization*, скр. *ISO*) (ISO 1997, 17).

Технички комитет ISO/TC 37 – језик и терминологија

У оквиру Међународне организације за стандардизацију, функционише и технички комитет *ISO/TC 37*³⁰ који је задужен за питања језика и терминологије. Овај комитет се састоји од 34 учесника и 29 посматрача, а до сада је укупно објавио 65 стандарда док је 27 у процесу израде.

У делокруг рада овог комитета спада стандардизација дефиниција, ресурса, технологија и услуга које су у вези са терминологијом, превођењем, тумачењем и другим језичким активностима у вишејезичном информационом друштву.

Један од основних стандарда овог комитета *ISO 704:2009* „Терминолошки рад – принципи и метода“ (ISO/TC 37 2009), утврђује основна начела и методе за припрему и прикупљање терминологије унутар и ван оквира стандардизације, и описује везе међу објектима, појмовима и њиховим терминолошким репрезентацијама. Уз то, утврђује општа начела по питању формирања ознака стандарда и формулације дефиниција. Одређено предзнање у области терминолошког рада је неопходно да би се у потпуности разумела ова начела. Начела су опште природе и применљива су на терминолошки рад у научним, технолошким, индустријским, административним и другим предметним областима.

Остали стандарди, објављени као и они који су још увек у процесу израде, односе се на хармонизацију појмова и термина, израду речника, представљање одредница у речницима, терминолошких одредница у стандардима, алфабетски редослед вишејезичних терминолошких и лексикографских података, библиографске референце и идентификаторе који су у вези са терминолошким радом, управљање терминолошким ресурсима, управљање смерницама за терминолошку стандардизацију, рачунарске апликације у терминологији, преводилачке услуге, терминолошке ресурсе, морфосинтаксичке, лингвистичке и семантичке анотације, лексичко означавање, вишејезичне информације, терминолошке базе података, системе за управљање терминологијом, управљање терминолошким ресурсима и сл.

Употребом стандарда које доноси овај комитет побољшава се општи квалитет терминологије и других језичких ресурса и садржаја у свим предметним областима, побољшава се управљање информацијама у различитим индустријским, техничким и научним окружењима, смањују се трошкови и повећава се ефикасност у техничкој стандардизацији и професионалној комуникацији.

Овај технички комитет је подељен у радне групе и техничке поткомитете који се баве терминолошким принципима и методама, терминолошким вокабуларом, припремом, употребом и управљањем језичким ресурсима, припремом и израдом терминолошких стандарда, рачунарски подржаном терминографијом и лексикографијом, терминолошком документацијом, кодирањем у области терминологије, применом терминологије, превођењем, тумачењем и с тим повезаним технологијама.³¹

Комитет је структурисан на следећи начин:

³⁰ Све информације у вези са Комитетом *ISO/TC 37*, преузете су јануара 2020. године, са веб-сајта Међународне организације за стандардизацију: <https://www.iso.org/committee/48104.html>

³¹ Технички комитети могу формирати један или више поткомитета. Делокруг рада поткомитета мора бити унутар делокруга матичног техничког комитета. Технички комитети као и поткомитети могу формирати Радне групе које су задужене за специфичне задатке као што је нпр. израда првог нацрта одређеног стандарда. Чланови радних група су стручњаци који се бирају на основу своје експертизе и стеченог искуства.

<i>ISO/TC 37/SC 1</i>	Принципи и методе	Поткомитет
<i>ISO/TC 37/SC 2</i>	Рад на терминологији и кодирање језика	Поткомитет
<i>ISO/TC 37/SC 3</i>	Системи за управљање терминологијом	Поткомитет
<i>ISO/TC 37/SC 4</i>	Управљање језичким ресурсима	Поткомитет
<i>ISO/TC 37/SC 5</i>	Превођење, тумачење и сродне технологије	Поткомитет
<i>ISO/TC 37/CAG</i>	Управни одбор	Радна група
<i>ISO/TC 37/JAC 0</i>	<i>ISO 639 RA</i> Заједнички саветодавни комитет	Радна група
<i>ISO/TC 37/TCG</i>	Координациона термилошка група за TC 37	Радна група
<i>ISO/TC 37/WG 10</i>	Техничка комуникација	Радна група
<i>ISO/TC 37/WG 11</i>	Општи језик	Радна група

2.5.2. Међународни информациони центар за терминологију – INFOTERM

Међународни информациони центар за терминологију – Инфотерм (*International Information Centre for Terminology*),³² основан је 1971. године на иницијативу Еугена Вистера, под покровитељством организације УНЕСКО (*United Nations Educational, Scientific and Cultural Organization – UNESCO*), с циљем да даје подршку и координира међународне активности на пољу терминологије.

Чланице овог информационог центра за терминологију су међународне, регионалне и националне термилошке институције, организације и мреже, као и остале организације које се баве термилошким активностима.

У оквиру Инфотерма је, почетком осамдесетих година прошлог века почела да се обликује Међународна мрежа за терминологију – ТермНет (енг. *International Network for Terminology – TermNet*), као једна од активности овог центра за терминологију, која је званично основана неколико година касније, 1988. године, као независна непрофитна организација (Infoterm 2004, 5–6). Ова мрежа координира међународну сарадњу на пољу терминологије.

2.5.3. Европски комитет за стандардизацију – CEN

Европски комитет за стандардизацију (*European Committee for Standardization – CEN*)³³ представља удружење које окупља национална тела из различитих европских земаља, укупно 34, које се баве пословима стандардизације. Ова организација је једна од три европске организације за стандардизацију³⁴ које званично признаје Европска унија и Европска асоцијација за слободну трговину (*EFTA*), а која је одговорна за развој и дефинисање стандарда на европском нивоу. Овај комитет обезбеђује платформу за развој европских стандарда и друге техничке документације која је у вези са различитим врстама производа, материјала, услуга или процеса. До сада, према званичним подацима са веб-сајта ове организације, Европски комитет за стандардизацију је објавио само један стандард који укључује и терминологију а његов назив је Моделовање материјала – терминологија,

³² Све информације у вези са Инфотермом преузете су августа 2020. године са веб-сајта ове организације: <http://www.infoterm.info/>.

³³ Сви подаци у вези са Европским комитетом за стандардизацију преузети су марта 2020. године са веб-сајта: <https://www.cen.eu>.

³⁴ Заједно са Европским комитетом за стандардизацију у електротехници (енг. *European Committee for Electrotechnical Standardization* скр. *CENELEC*) и Европским институтом за стандарде из области телекомуникација (енг. *European Telecommunications Standards Institute*, скр. *ETSI*).

класификација и метаподаци, објављен априла месеца 2018. године. Овај стандард обухвата дефиниције основних појмова за област рачунарског моделовања и симулације материјала.

Технички одбор (енг. *Technical Board*) је задужен за активности на пословима стандардизације, док стандарде припремају технички комитети (енг. *Technical Committees*). Сваки технички комитет има свој делокруг рада унутар којег се доносе и спроводе стандарди, а одлуке се доносе по принципу консензуса. У случају великог обима посла, унутар техничког комитета формирају се поткомитети.

2.5.4. Европски комитет за стандардизацију у електротехници – CENELEC

Ова организација (*European Committee for Electrotechnical Standardization – CENELEC*)³⁵ бави се стандардизацијом у области електротехнике на тлу Европе. Стандарди које доноси помажу трговинској размени између земаља, стварању нових тржишта, смањењу трошкова и пружању помоћи у развоју јединственог европског тржишта.

Кроз заједнички ангажман чланова ове организације и стручњака, као и индустрије и потрошача, овим стандардима се постиже технолошки развој, осигурава интероперабилност, гарантује безбедност и здравље потрошача и обезбеђује заштита животне средине.

Европски комитет за стандардизацију у електротехници основала је Европска комисија 1973. године. Данас представља непрофитну организацију која послује према законима државе Белгије.

CENELEC блиско сарађује са Међународном електротехничком комисијом (*IEC*). Како би се олакшало доношење консензуса између европских и међународних активности по питању развоја стандарда у електротехничком сектору, *CENELEC* и *CEN* су 1996. године потписали споразум о заједничком планирању новог рада и паралелном гласању, познатијим под именом Дрезденски споразум. Уколико је исход гласања позитиван, *CENELEC* ће ратификовати европски стандард, а Међународна електротехничка комисија (енг. скр. *IEC*) ће објавити међународни стандард. Сврха ове сарадње између две организације је уштеда времена приликом припреме стандарда као и елиминисање дуплог посла, а резултат су нови пројекти електротехничких стандарда од којих се већина спроводи на међународном нивоу.

2.5.5. Међународна електротехничка комисија – ИЕС

Међународна електротехничка комисија (*International Electrotechnical Commission – IEC*)³⁶ је непрофитна организација основана 1906. године, која објављује међународне стандарде из области електротехнике. Чланови ове организације су националне комисије које именују стручњаке и делегате из индустрије, државних органа, удружења и академске заједнице.

Стандарди које објављује Међународна електротехничка комисија покривају широк опсег технологија почевши од производње, преноса и дистрибуције електричне енергије све до малих кућних апарата и пословне опреме, полупроводника, оптичких влакана, батерија, соларне енергије, нанотехнологије, енергије мора и сл. Уз то, Међународна електротехничка комисија на глобалном нивоу води четири система за оцењивање усаглашености од којих сваки покрива различита подручја електротехнике (систем оцењивања усаглашености за електротехничку опрему и компоненте, систем за сертификацију стандарда који се односи на опрему за употребу у експлозивним атмосферама, систем за процену квалитета електронских компонената, систем за сертификацију стандарда који се односи на опрему за употребу у применама обновљивих извора енергије), а којима се сертификаује да су опрема, компоненте и системи усклађени са међународним стандардима.

³⁵ Сви подаци у вези са организацијом *CENELEC* преузети су марта 2020. године са веб-сајта ове организације: <https://www.cenelec.eu>.

³⁶ Сви подаци у вези са Међународном електротехничком комисијом преузети су децембра 2020. године, са веб-сајта ове организације: <http://www.iec.ch/>.

Публикације које објављује Међународна електротехничка комисија служе као основ за националну стандардизацију и као референца приликом спровођења међународних тендера и уговора.

IEC броји 62 пуноправне државе чланице. У оквиру ове организације ангажован је велики број техничких комитета и поткомитета на пословима стандардизације (укупно 210), радних група (705), пројектних тимова (194) и подршке (641), а укључен је и велики број стручњака из националних комитета (чак 22447).

На листи техничких комитета (којих има укупно 109), на првом месту се налази технички комитет за терминологију, чија је ознака *TC 1*. Поред тога што је задужен за термине и дефиниције из различитих области електротехнике као и проналажење еквивалената термина који се користе у различитим језицима, задатак овог комитета је да надгледа развој и пружа константну подршку Међународном електротехничком речнику (енг. *International Electrotechnical Vocabulary – IEV*, познатијем под називом „Електропедија“). Циљ овог речника је стандардизација и координација термина који су у вези са електротехничким наукама у сврху њихове употребе у обуци, техничким спецификацијама и робној размени, као и у проналажењу одговарајућих еквивалената у различитим језицима.

Електропедија је најисцрпнија он–лајн база термина из области електротехнике, која садржи више од 22.000 термина и дефиниција. Организована је према предметном пољу и даје информацију о еквивалентним терминима у различитим језицима међу којима је и српски. Електропедија садржи све термине и дефиниције које је Међународна електротехничка комисија објавила у Међународном електротехничком речнику односно серији стандарда 60050.³⁷

Циљ и сврха Електропедије је стандардизација међународно прихваћених термина и дефиниција у свим областима које покрива Међународна електротехничка комисија, на енглеском и француском језику. Како се области деловања ове комисије током времена развијају и проширују, тако се и садржај речника континуирано ревидира и проширује додавањем нпр. новог подручја активности.

Примарни циљ овог речника је испомоћ приликом припреме стандарда као и испомоћ у њиховој примени, како би их корисници што боље разумели и спровели. Уз то, служи и као помоћ преводиоцима техничких текстова, корисницима техничке литературе, у подучавању, справљању техничких спецификација као и у робној размени. Овај речник не покрива све појмове који се користе у различитим *IEC* стандардима, већ је пре свега речник доменских појмова. Бројни национални комитети достављају термине на свом језику за све односно неке дефиниције у овом речнику, што се обавља кроз централну канцеларију ове институције.

У фебруару 2013. године у Електропедији су се први пут појавили термини обухваћени стандардом *SRPS IEC 60050-411*, *Међународни електротехнички речник – Поглавље 411 – Обртне машине*, на српском језику. Појављивање термина на српском језику у Електропедији подразумева да ће свако будуће ревидирано или ново издање међународних стандарда серије *IEC 60050* садржати и термине на српском језику уколико постоји успостављена стандардизована терминологија из одређених области електротехнике, на чему се непрестано ради у Институту за стандардизацију Републике Србије.³⁸

До данас је терминима на српском језику Електропедија попуњена у 45 предметних области (од укупно 93). Табелом 1³⁹ приказана је терминолошка база података према предметним областима електротехнологије⁴⁰, преузета са веб-сајта Електропедије. Масним словима истакнуте су области које садрже и термине на српском језику.

³⁷ Подаци преузети децембра 2020. године, са веб-сајта: <http://www.electropedia.org/>.

³⁸ Преузето марта 2020. године са веб-сајта Института за стандардизацију Србије: http://www.iss.rs/news/news_84.html.

³⁹ Табела преузета децембра 2020. године са веб-сајта: <http://www.electropedia.org/>.

⁴⁰ Под електротехнологијом подразумевају се електричне, електронске и сродне технологије, како је и дефинисано на сајту Електропедије.

Табела 1. Терминолошка база података према предметним областима

102	Математика – Општи појмови и линеарна алгебра	561	Пијезоелектрични уређаји за регулацију и избор фреквенције
103	Математика – Функције	581	Електромеханичке компоненте за електронску опрему
112	Количине и јединице	601	Производња, пренос и дистрибуција електричне енергије – Опште
113	Физика за електротехнологију	602	Производња, пренос и дистрибуција електричне енергије – Производња
114	Електрохемија	603	Производња, пренос и дистрибуција електричне енергије – Планирање и управљање електроенергетским системима
121	Електромагнетизам	605	Производња, пренос и дистрибуција електричне енергије – Трафостанице
131	Теорија електричних кола	614	Производња, пренос и дистрибуција електричне енергије – Рад
141	Полифазни системи и електрична кола	617	Организација / тржиште електричне енергије
151	Електрични и магнетни уређаји	651	Рад под напоном
161	Електромагнетна компатибилност	691	Тарифе за струју
171	Дигитална технологија – Основни појмови	692	Производња, пренос и дистрибуција електричне енергије – Поузданост и квалитет услуге електроенергетских система
192	Поузданост	701	Телекомуникације, канали и мреже
195	Уземљење и заштита од струјног удара	702	Осцилације, сигнали и повезани уређаји
212	Електрична изолација у чврстом, течном и гасовитом стању	704	Пренос
221	Магнетни материјали и компоненте	705	Пропагација радио таласа
311	Електрична и електронска мерења – Општи услови који се односе на мерења	712	Антене
312	Електрична и електронска мерења – Општи услови који се односе на електрична мерења	713	Радиокомуникације: предајници, пријемници, мреже и рад
313	Електрична и електронска мерења – Врсте електричних мерних инструмената	714	Пребацивање и сигнализација у телекомуникацијама
314	Електрична и електронска мерења – Специфични појмови према врсти инструмента	715	Телекомуникационе мреже, телекомуникације и рад

321	Мерни трансформатори	716	Дигитална мрежа интегрисаних услуга (енг. <i>ISDN</i>) – Део 1: Општи аспекти
351	Контролна технологија	721	Телеграфија, факсимил и пренос података
371	Телеконтрола	722	Телефонија
395	Нуклеарни инструменти: Физички феномени, основни појмови, инструменти, системи, опрема и детектори	723	Емитирање: звук, телевизија, подаци
411	Обртне машине	725	Радиокомуникација у свемиру
415	Системи генератора ветротурбина	726	Преносни водови и таласоводи
421	Енергетски трансформатори и реактори	731	Комуникација оптичким влакнима
426	Опрема за експлозивне атмосфере	732	Рачунарска мрежна технологија
431	Трансдуктори	741	Интернет ствари (IoT)
436	Енергетски кондензатори	801	Акустика и електроакустика
441	Расклопни уређаји, управљачки уређаји и осигурачи	802	Ултразвук
442	Електрични прибор	806	Снимање и репродукција аудио и видео-записа
444	Елементарни релеји	807	Дигитално снимање аудио и видео- сигнала
445	Временски релеји	808	Видео-камере за неемитовање
447	Мерни релеји	811	Електрична вуча
448	Заштита електроенергетског система	815	Суперпроводност
461	Електрични каблови	821	Сигнализација и сигурносни апарати за железнице
466	Надземни водови	826	Електричне инсталације
471	Изолатори	841	Индустријско електрозагревање
482	Примарне и секундарне ћелије и батерије	845	Расвета
485	Технологије горивних ћелија	851	Електролучно заваривање
511	Нанотехнолошки електротехнички производи и системи	871	Активно потпомогнуто живљење (енг. <i>AAL</i>)
521	Полупроводнички уређаји и интегрисана кола	881	Радиологија и радиолошка физика
523	Микроелектромеханички системи (МЕМС)	891	Електробиологија
531	Електронске цеви	901	Стандардизација
541	Штампана кола	902	Оцењивање усаглашености
551	Енергетска електроника	903	Процена ризика
		904	Стандардизација животне средине за електричне и електронске производе и системе

За сваки термин и за сваки језик у Електропедији, придружена је информација о роду и броју одреднице (уколико одредница припада врсти речи именице), односно додатно

објашњење термина у тупоуглим заградама тамо где је то потребно, као што је приказано на слици 2. Предметна подручја обележена су ознакама од 102 до 904, а свако предметно подручје садржи више различитих области због чега је основној ознаци, која у примеру на слици 2 носи ознаку 605 (*Area 605: Generation, transmission and distribution of electricity – Substations*), придружен број (01, 02, итд.), којим је та специфична област предметног подручја означена (*Section 605-01: Types of substations – Substation layouts*).⁴¹ Бројеви су међусобно раздвојени цртицама, а последњи односно трећи број у низу представља редни број термина у оквиру специфичне области одређеног предметног подручја (нпр. 605-01-01).

Area: 605: Generation, transmission and distribution of electricity - Substations	Index language:
Section 605-01:	Serbian ▼
605-01-01 постројење , <електроенергетског система> с јд	Arabic
605-01-02 разводно постројење, с јд	English
605-01-03 трансформаторска станица, ж јд	Finnish
605-01-04 станица за повишење напона, ж јд	French
605-01-05 станица за снижење напона, ж јд	German
605-01-06 електровучна подстанца, ж јд	Italian
605-01-07 претварачка подстанца, ж јд	Japanese
605-01-08 подстанца за промену фреквенције, ж јд	Korean
605-01-09 подстанца са посадом, ж јд	Polish
605-01-10 подстанца без посаде, ж јд	Portuguese
605-01-11 подстанца са сталном посадом, ж јд	Serbian
605-01-12 подстанца са повременим посадом, ж јд	Slovenian
605-01-13 подстанца са даљинским управљањем, ж јд	Spanish
605-01-14 главна станица, ж јд	Swedish
605-01-15 "сателитска" подстанца, ж јд	
605-01-16 постројење са једноструким сабирницама, с јд	
605-01-17 постројење са двоструким сабирницама, с јд	
605-01-18 постројење са троструким сабирницама, с јд	
605-01-19 постројење са прстенастом сабирницом, с јд	
605-01-20 постројење са прстенастом сабирницом и са прекидачима, с јд	
605-01-21 постројење са четири склопке, с јд	
605-01-22 постројење са прстенастом сабирницом са три склопке са ремошћењем, с јд	
605-01-23 постројење са правоуганом сабирницом са четири склопке са растављачима, с јд	
605-01-24 шема са два прекидача, ж јд	
605-01-25 шема са једним и по прекидачем, ж јд	
605-01-26 распоред у којем су фазе придружене, м јд	
605-01-27 распоред у којем су фазе одвојене, м јд	
605-01-28 распоред у којем су мешовито распоређене фазе, м јд	
Section 605-02:	
605-02-01 сабирница, ж јд	
605-02-02 сабирнице, ж мн	
605-02-03 главна сабирница, ж јд	
605-02-04 резервна сабирница, ж јд	
605-02-05 обилазна сабирница, ж јд	
605-02-06 сабирница секционисана прекидачем, ж јд	
605-02-07 сабирница секционисана растављачем, ж јд	
605-02-08 секција сабирнице, ж јд	
605-02-09 поље, <подстанце> с јд	
605-02-10 напојно поље, с јд	
605-02-11 излазно поље, с јд	

Слика 2. Термини на српском језику

Мада се велики број термина из области електроенергетике на српском језику већ налази у Електропедији, Институт континуирано настоји да тај број буде што већи. У светлу тога, и резултати добијени овим истраживањем, у сарадњи са Институтом, могу пронаћи своје место у овом речнику, имајући у виду да су терминологију, добијену детаљном анализом корпуса текстова величине милион речи, верификовала и стручна лица из области електроенергетике,

⁴¹ Називи предметних подручја као и специфичних области које се налазе у оквиру њих, приказани су само на енглеском и француском језику и нису преведени ни на један од многобројних језика који су заступљени у Електропедији, па тако ни на српски.

а да су припрему (која подразумева сакупљање материјала и обраду текста), екстракцију и евалуацију терминологије, уз примену одговарајућих ресурса и алата који се користе у аутоматској обради природних језика, а за потребе обраде одабраног корпуса, обавили стручњаци Друштва за језичке ресурсе и технологије (више о раду ове организације у поглављу 3.4.3.5).

2.5.6. Институт за стандардизацију Републике Србије

Институционализован облик националне стандардизације постојао је још од 1939. године, у тадашњој Краљевини Југославији. Улогу националног тела за стандарде имао је Југословенски национални комитет за нормализацију који је функционисао као посебна установа при Министарству трговине и индустрије.⁴²

Институт за стандардизацију Србије (у даљем тексту ИСС) правни је следбеник Завода за стандардизацију (2003–2006), Савезног завода за стандардизацију (1978–2003), Југословенског завода за стандардизацију (1962–1978) и Савезне комисије за стандардизацију (1946–1962).

Према Закону о стандардизацији („Службени гласник Републике Србије“, бр. 36/2009 и 46/2015) и Одлуци о изменама и допунама оснивачког акта Института за стандардизацију Србије („Службени гласник Републике Србије“, бр. 93/2015 и бр. 27/2016), ИСС је једино национално тело за стандардизацију Републике Србије, установа која има статус правног лица и послује у складу са прописима којима се уређује правни положај јавних служби. Оснивач Института је Влада Републике Србије.

Како је даље наведено на сајту Института, ИСС доноси, развија, допуњава и повлачи српске стандарде и сродне документе, обезбеђује усаглашеност српских стандарда и сродних докумената са европским и међународним стандардима и сродним документима, води регистар српских стандарда, учествује у изради европских и међународних стандарда и сродних докумената које доносе европске и међународне организације за стандардизацију у областима за које постоје потребе и интереси Републике Србије, сарађује са европским и међународним организацијама за стандардизацију, српске стандарде и сродна документа чини доступним јавности, даје основу за израду техничких прописа, припрема програме и годишње планове доношења српских стандарда, представља и заступа интересе Републике Србије у области стандардизације у европским и међународним организацијама за стандардизацију, одобрава употребу знака усаглашености са српским стандардима и сродним документима, промовише примену српских стандарда и сродних докумената као и остало у складу са законом, актом о оснивању и статутом.

Као национално тело за стандардизацију Републике Србије, ИСС обезбеђује Републици Србији, заинтересованим странама, члановима Института и целокупној јавности српске стандарде усаглашене са међународним и европским стандардима и могућност да равноправно учествују у међународној и европској стандардизацији доносећи националне стандарде и поштујући међународно и европски признате принципе стандардизације.

Институт представља и штити интересе наше земље у следећим међународним и европским организацијама за стандардизацију:

- Међународној организацији за стандардизацију, у којој има статус пуноправног члана од 1950. године;
- Међународној електротехничкој комисији, у којој има статус пуноправног члана од 1953. године;
- Међународном *IEC* систему за испитивање усаглашености и сертификацију, у оквиру којег функционише шема признавања резултата испитивања (*IECEE/CB – Scheme*), у којој Институт има статус пуноправног члана од 1990. године;
- Европском комитету за стандардизацију, у којем од 1. јануара 2017. године има статус пуноправног члана;

⁴² Подаци преузети марта 2020. године са веб-сајта Института за стандардизацију Србије: <http://www.iss.rs/>

- Европском комитету за стандардизацију у области електротехнике у којем има статус пуноправног члана од 1. јануара 2017. године.
- Европском институту за стандарде из области телекомуникација, у којем има статус националне организације за стандардизацију (*NSO*) од јуна 2011. године.

У оквиру Института за стандардизацију РС, поред различитих комисија за стандарде, функционише и Комисија задужена за област терминологије, која носи ознаку А037. Ова комисија ради на „стандардизацији принципа, метода и примене термина и других језичких и садржинских извора у области вишејезичке и културне разноврсности“, како је наведено на веб-сајту Института, а кроз њено деловање Република Србија учествује у раду *ISO/TC 37*.

До данас, ова Комисија је донела низ стандарда из области превођења, језичких технологија и терминологије, који ће подробније бити представљени у поглављу 3.4.3.4, а којима се уређују и дефинишу принципи, методе и примене термина на српском језику.

3. ЈЕЗИЧКЕ И ТЕРМИНОЛОШКЕ ПРИЛИКЕ У СРПСКОМ ЈЕЗИКУ

Креирање терминологије у српском језику тешко је сагледати као изоловани феномен, без исцрпне анализе историје српскога језика и културе, од њених прапочетака па све до коначног уобличења, односно друштвених прилика које су током времена биле актуелне. Иако су одређени историјски факти делимично још увек нерасветљени и противуречни, непобитна је чињеница да историја српскога народа, а самим тим и језика, сеже у далеку прошлост. О свему томе најбоље сведоче сачувани споменици културе.⁴³ Они пружају недвосмислен доказ о постојању културе и писмености народа на овим просторима, до извесне мере расветљавајући прилике у којима су живели и пружајући увид у начин живота који су водили.

Међутим, имајући у виду да се у оквиру овог истраживања бавимо искључиво терминологијом, од успостављања њених основних начела на међународном нивоу, преко њеног развоја у Србији и даље, њеног успостављања и развоја у електроенергетском сектору кроз анализу корпуса и коришћење рачунарских метода обраде природних језика, анализа историје српског језика и писмености, као и културе, овај рад би одвела у сасвим другом смеру.

Стога је у најопштијим цртама, у редовима који у овом поглављу следе, представљена историја српскога језика и писмености, односно направљен осврт на историјске прилике које су довеле до настанка терминологије на нашим просторима, њеног постепеног развоја, а затим и до успостављања терминолошког система каквог познајемо данас.

3.1. Историјски осврт

Историја српског језика и писмености толико је велика и богата да би је тешко могла представити једна студија овог обима. То је поготово случај када такав историјски преглед није ни предмет њеног истраживања. Па ипак, да би се стекао општи увид у терминолошке прилике на нашем тлу, важно је указати на одређене значајне чињенице које се односе на историју језика и писмености на овим просторима, а које су биле предмет подробног истраживања наших истакнутих аутора. У том смислу, за потребе овога рада, највише је послужио *Преглед историје српскога језика* Павла Ивића, из ког су највише црпени подаци изложени у овом поглављу, а који представљају историјски осврт на постанак и развој нашег језика. Уз његов *Преглед*, коришћени су и други радови Павла Ивића као и радови других аутора попут Ирене Грицкат, Исидоре Бјелаковић итд.

Језик којим су се служили становници ових простора, а о којем за сада нема писаних трагова, био је прасловенски. Овај језик је, како истиче Павле Ивић у свом *Прегледу историје српског језика* (Ивић 1998, 5), уз балтички (литавски и летонски), германски, романски (настао даљим развојем из латинског), келтски, албански, грчки, јерменски и индоирански припадао индоевропској породици језика. Познато је да се први писани текстови на словенским језицима на овим просторима везују за долазак Ћирила и Методија 863. године. Њихов превод најважнијих богослужбених књига са грчког на словенски био је још увек сличан прасловенском језику, тако да језик тих текстова служи као врло добар извор о стању у тадашњем језику и одлично је средство за реконструкцију фаза кроз које је прасловенски језик током свог развоја пролазио (Ивић 1998, 6).

Код најстаријих сачуваних преписа старословенских текстова, који датирају с краја 10. и почетка 11. века, уочљиве су промене које су наступиле у словенском говору на Балкану, а

⁴³ Међу њима су познати старословенски глагољски споменици из Македоније с краја 10. или из 11. века који су сачувани ван македонских области (Асеманово, Зографско, Маријинско јеванђеље, Синајски псалтир, Синајски требник, Ключев зборник), ћирилски споменици из истог времена из Бугарске (Супрасалски кодекс, Савина књига), из Македоније (Добромирово јеванђеље с почетка 12. века), уз неколико мањих глагољских и ћирилских одломака, док су из последњих година 12. века, са подручја независних или полузависних српских земаља сачувана два знаменита ћирилска рукописа српске редакције: Мирослављево јеванђеље, писано око 1185. и Вуканово јеванђеље, настало око 1197–1199. године“ (Богдановић 1981, 217).

крајем 12. века у доба Немањића, оживљава и културна делатност, што се види из већег броја писаних текстова са овог тла; у језику тих текстова истовремено су се појавила два паралелна тока: српска редакција црквенословенског језика⁴⁴, често називана и српскословенским језиком, и уз њу народни језик, штокавски говор у писменој употреби (Ивић 1998, 25–28).

Занимљиво је одређење које је представила Ирена Грицкат говорећи о уметничком и неуметничком писању у средњовековљу (Грицкат 1970, 5). Имајући у виду, наиме, специфичност епохе о којој је реч, она напомиње да је критеријум израза био постављан према садржини текста, а у потрази за узорима и духовним усавршавањем, ишло се за уметничким изразом што је значило дистанцираност од обичног говора.

„У свакој уметности, а поготову у мистичарски обојеној, неопходан је патос одстојања. Уметнички израз морао је бити не само отргнут од разговорног и од евентуалног *обичног* писменог саопштења, него по могућству оптерећен свечаним асоцијацијама: асоцијацијама с амвоном и олтаром, или с нечим временски или просторно удаљеним, делимично неразумљивим. Уметничким се и иначе, у свако доба, сматра оно што носи у себи било какву метафору и што подстиче на тражење подтекста. Таква је била и тадашња руководећа идеја уметности. Како доцније, тако и онда језик уметнички конципираног састава био је пун померања и уздицања са обичног лексичког или синтаксичког тла, пун начина да преведе испричана факта на неки други, надбуквални ниво. Али не заборавимо да у старо време, при нижем ступњу опште способности за поимање подтекста, употребљавани начини су били далеко драстичнији, језик уметничке књижевности није био *уметнички* у савременом позитивном смислу речи. Где није био канонски утврђен, он је сасипао читаоца архаизмима и агломерацијама и одскакао од практичног писменог изражавања онолико колико данас, можда, одскаче језик најекстравагантније поезије од неутралног књижевног језика.“ (Грицкат 1970, 5).

И Павле Ивић, говорећи о језику у немањићкој епохи, (Ивић 1981, 634) бележи да је „у многим појединостима средњовековни језик располагао изражајним могућностима које недостају данашњем језику. Као илустрација могу послужити тешкоће на које наилазимо ако желимо да преведемо једном речју изразе као *твар* или *безначелије*. Догађа се и да преводи наоко постоје, али да стара реч има дубље или шире значење од садашње“, што превод може чинити неадекватним. „Наш модерни језик поготову заостаје за старим у погледу употребе сложеница, које су у средњем веку обилато стваране по грчком узору: *благоличје*, *богопостављен*... С друге стране, у средњовековном језику је било неуједначености штетних по функционисање језика. Тако је реч *владика* значила и владара (а онда и Христа и Бога уопште) и властелинку. У истом тексту, Душановом законнику, *потворити* значи час „нарушити, укинути“, час „обедити, лажно оптужити“. Овакве несавршености јављају се обично тамо где термилолошки систем још није у потпуности уходан и кристалисан.“ (Ивић 1981, 634).

Када говоримо о терминском фонду, врло је интересантно указати на одређена Ивићева запажања изложена у *Прегледу историје српског језика*. Познато је да је заједно с друштвеним развојем Срба у средњем веку, растао и термински фонд у различитим областима. На првом месту, како у свом тексту бележи, термински фонд писмености, био је претежно домаћег порекла (нпр. писати, писмо, књига, повеља итд.). Затим, с напретком и све већом разноврсношћу привреде код средњовековних Срба развијао се и фонд речи у тој области и то највише грађењем из домаћег језичког материјала, а мање преузимањем из других језика. Како Ивић даље запажа, има и израза за које се претпоставља да су веома старе позајмљенице из разних језика, одавно добро уклопљене у домаћи изражајни фонд, или да се у њима укрштају по једна словенска и једна туђа реч. Велики број позајмљеница је међу називима посуђа, мерама за запремину и тежину, као и новчаних јединица. Што се тиче основних привредних грана као што су земљорадња и сточарство, термински фонд је био домаћег порекла, а како даље бележи, „и у систему личних имена сачувало се код средњовековних Срба много словенске старине“. (Ивић 1998, 50–52).

⁴⁴ Црквенословенски језик се у средњем веку писао двома азбукама: глаголицом и ћирилицом.

Интересантан је и податак да се током 15. века српски језик толико раширио да је постао један од дипломатских језика у југоисточној Европи, којим су се служили и турски султани, везири као и мађарски и румунски магнати и везири (Ивић 1998, 65).

Српскословенски језик био је актуелан међу свештенством и у цркви кроз цео 16. и 17. век, како наводи Ивић, али је тим језиком била писана и књижевност, па је тако овај језик био доминантан и у српским књигама штампаним у 16. веку. У 17. веку правопис бива упрошћен и прилагођен живом српском говору, међутим црквени језик ипак у овом периоду доминира (Ивић 1998, 81–92).

Важно је истаћи да су уз турску власт у српске земље почеле улазити и турске речи, а имајући у виду дугачак период њиховог боравка на овим просторима, сасвим је природно да је њихов утицај на језик Срба био врло јак. Тако су турцизми успели да „прожму све слојеве речничког фонда, од психичког живота до јеловника, од назива разних заната до имена биљака и животиња“ (Ивић 1998, 100–101).

И у 18. веку, код Срба су били заступљени српскословенски језик, коришћен у црквене и богослужбене сврхе као и у књижевности, и народни језик. Како Ивић наводи, „језик докумената с почетка 18. века из централних и северних огњишта српске писмености био је по гласовима и облицима већ прилично близак данашњем књижевном језику екавског изговора, међутим што се тиче речника, био је врло различит од данашњег. Садржао је нешто више архаизама и огромно мноштво туђица, од којих је већи део касније ишчезао из језика“ (Ивић 1998, 110).

Услед одлуке аустријских власти да се забрани оснивање штампарије за Србе, већ почетком 18. века у Србију су почеле да пристижу књиге из Русије, а уз књиге и учитељи.⁴⁵ Заједно са њима, по природи ствари, рускословенски језик је ушао у школство као и у црквени систем, где се употребљава и данас, а одатле природно и у књижевност уместо дотадашњег српскословенског језика (Ивић 1998, 115–121).

Рускословенски језик није био много близак матерњем српском говору. Због тога је настала „мешавина рускословенског и српског језика, а врло често и руског, која се данас обично назива славеносрпским језиком“ (Ивић 1998, 129). Како Ивић даље наводи у свом *Прегледу*, овај језик је имао врло богат речнички фонд, имајући у виду да је обухватао лексику двају или чак трију језика (Ивић 1998, 130). „Писцу су стајале на располагању речи за појмове из српске грађанске и сеоске свакидашњице, али и разграната црквенословенска лексика, столећима потхрањивана разрађивањем богословских и етничких питања у ослонцу на систем значења одређен грчким и латинским вокабуларом, и уз то тадашњи руски речнички фонд, пун западноевропских позајмљеница за појмове сувремене цивилизације и обогаћен већ конституисаном терминологијом администрације, друштвеног уређења и многих наука. Славеносрпски је много боље него рускословенски задовољавао и захтев епохе просветитељства да језик књижевности буде приступачан широким слојевима народа како би књига могла допринети њиховом уздизању. Истовремено, славеносрпском се није могло пребацити да се спушта на ниво маса. Компонента црквеног језика давала му је легитимацију отмености.“ (Ивић 1998, 130).

Од аустријских власти је 1770. године добијена дозвола за објављивање уџбеника и превода дела из западноевропских литература. Оваква политика имала је за циљ да створи публику којој ће књиге на народном језику боље одговарати него оне на црквеном, као и да руски језик замени немачким језиком. Иако је 1781. године стигла наредба да се у српске школе уведе латиница, уз велики напор српских представника, дозвољено је да се књиге штампају на славеносрпском. У тим приликама појављује се и Доситеј Обрадовић који „смело и на велика врата уводи народни језик у књижевност“ (Ивић 1998, 139). Својим делом „Живот и прикљученија“ он је изричит у ставу да учени људи треба да пишу простим језиком како би знање постало доступно њиховом народу, и одбацује стари књижевни језик који је народу неразумљив. Бројни аутори су наставили да следе његов пример (Ивић 1998, 137–141).

⁴⁵ Руски учитељ Максим Суворов, синодски преводилац, отворио је 1726. године сталну „славенску школу“ у Сремским Карловцима, а повремено је држао предавања из славенског језика и у Београду. Године 1732. у Сремске Карловце долази и Емануил Козачински са неколико својих сарадника (Ивић 1998, 116).

Имајући у виду друштвено-политичке прилике крајем 17. века када се део српског народа нашао у границама Хабзбуршке Монархије, а самим тим суочио и са великим бројем нових појава и појмова, сасвим је природно што се у текстовима с почетка 18. века користио велики број немачких речи (Ивић 1998, 150). Међутим, како Ивић и сам даље бележи, и немачки језик је у истом периоду такође био подложен страним лексичким утицајима, француском, грчком, латинском и италијанском, па су тако у српски језик посредно пренети и изрази које су Немци претходно усвојили из ових језика. Крајем 18. века у српским текстовима по први пут се сусрећу и енглеске речи (Ивић 1998, 153).

Како овај аутор даље бележи, након пропасти Првог српског устанка, Вук Стефановић Караџић се састаје с Јернејем Копитаром на чији подстицај 1814. године објављује прву штампану збирку народних песама *Малу простонародњу славено-сербску пјеснарицу* и прву граматику српског језика чији је писац био Србин, *Писменицу сербскога језика по говору простога народа написану*. На Копитаров подстицај, Вук почиње прикупљати грађу и за српски речник који је објављен 1818. године. *Српски рјечник истолкован њемачким и латинским ријечима* обухвата преко 26.000 речи, а речи које су пописане у њему припадају народном говору. Овај речник је прва књига штампана новом, Вуковом ћирилицом, која се и данас употребљава. Уз свој *Рјечник*, Вук је прикључио и *Српску граматику*. Што се терминологије тиче, Вук је настојао да посрби термине, али у томе није увек имао успеха. Објављивање *Рјечника* једно је од најкрупнијих догађаја у историји српске културе имајући у виду да је Вукова језичка и правописна реформа остварена у потпуности (Ивић 1998, 176–190).

Народни језик, надаље, почиње да потискује рускословенски и славеносрпски, а три књиге Вукових *Српских народних пјесама* одушевљено су дочекане у европским књижевним круговима (Ивић 1998, 197). Српске народне песме постале су „најпривлачнија европска тековина у тој области“, а с њиховом репутацијом растао је и углед народа који их је испевао, нарочито имајући у виду велику пажњу коју су им поклонили неки од правих великана европске књижевности попут Гетеа, Јакоба Грима и Пушкина (Ивић 1998, 197).

Даљи развој српског књижевног језика обележен је појавом нових граматика међутим, и издавачка делатност из различитих струка се код Срба прилично раширила у првој половини 19. века (Ивић 1998, 205).

Управо у таквом окружењу створена је погодна клима за покретање питања стручне и научне терминологије у српском језику.

3.2. Рани развој терминологије у српском језику

Историја књижевног језика не сме да буде ограничена на историју изражавања у тзв. *лепој* књижевности, већ у обзир треба узимати и друге споменике писмености попут научних, административних и трговачких списа, статистика, наредби, расписа и осталог, имајући у виду да се и у тим различитим сферама, несумњиво појављују речи сковане посебно за такву, одговарајућу употребу, а „без којих не може да опстане ни говор представника књижевног језика, ни сама лепа књижевност“ (Грицкат 1964, 130).

Развој терминолошког система једног језика, неоспорно је у тесном садејству са друштвеним и цивилизацијским достигнућем одређеног народа. Због тога не изненађује чињеница да је друштвено–политички напредак код Срба од краја 18. века иницирао потребу за именованем многобројних недостајућих појмова.

Успостављање Кнежевине Србије која је својом политиком подстицала друштвени и културни напредак, довело је до отварања бројних школа, што је наметнуло „потребу како за образованим наставницима тако и за одговарајућим уџбеницима, приручницима и научним делима“, „на основу којих се може добити увид у процес настанка и развоја терминологије у српској средини“ (Бјелаковић 2017а, 15). Овакве прилике створиле су погодно тле за оснивање друштва чији је задатак био ширење наука на српском језику и усавршавање српског народног језика. Све то је, према наводима Бјелаковић (Бјелаковић 2017, 62), подразумевало и изразу институционализованог и систематичног, јединственог терминолошког система.

Међутим, иако је у тадашњој модерној Србији, постојала потреба за изграђеном терминологијом и термилошкоим речником, ситуација је била прилично комплексна, и то пре свега услед различитих ставова у погледу структуре будућег термилошког система: сваки покушај реализовања процеса израде термилошког апарата осујећивали су неуједначени ставови у којима се, према наводима Бјелаковић, „с једне стране, осећао изразит пуризам према страном, пре свега, интернационалној лексици (П. Соларић, Ј. Дошеновић, Ј. Милованов, С. Мркаљ, П. Болић, Ј. Стерија Поповић и др.), с друге – према славенизмима (вуковци), а с треће – постојало је и умереније мишљење према којем страну лексику није требало замењивати домаћом по сваку цену (Стратимировић, Видаковић и др.)“ (Бјелаковић 2017а, 23–24). Сам процес формирања термилошког система, дакле, није текао плански већ је, према речима Павла Ивића (Ивић 1986, 195), имао стихијски, спонтани и постепени карактер.

Године 1841, на иницијативу уважених професора Лицеја, Јована Стерије Поповића и Атанасија Николића, основано је Друштво српске словесности. Ово удружење окупљало је бројне угледне и образоване грађане тадашње Кнежевине као своје чланове.⁴⁶ Један од основних циљева овог друштва је, према наводима Бјелаковић (Бјелаковић 2017а, 24), било решавање проблема неконституисане и нестабилизане терминологије, која је углавном носила печат индивидуалног, и то кроз елаборацију српског језика и израду јединствене терминологије.

Први покушај стварања српске научне терминологије иницирало је Друштво српске словесности 1842. године (Грицкат 1964, 131). Ова идеја се затим обновила 1844, а коначно замрла 1847. године. У оба покушаја био је заступљен следећи концепт: (1) уклањање термина заснованих на грчком и латинском језику и њихово превођење или грађење одговарајућих кованица, (2) благонаклон став ка лексици рускословенске и руске провенијенције (Бјелаковић 2017а, 25).

Из *Скупљених граматичких и полемичких списа Вука Стефановића Караџића, Живота и рада Вука Ст. Караџића Љубомира Стојановића*, као и *Извештаја поводом 50-годишњице Српске академије науке* проф. Александра Белића, данас нам је познато да је Атанасије Николић 1842. године поднео Друштву *Технически речник за математическу, земљемерну, хидротехническу, хидраулическу и архитектоническу струку*, и то само речи на слово А (Грицкат 1964, 131). У Архиву САНУ постоји непотписани рукопис на једном листу хартије, у ком су на обе стране у две колоне исписане речи на а- за одговарајуће појмове из наведених струка – на латинском, немачком и српском језику (Бјелаковић 2017, 68).

Поред тога, у *Подунавци*, првом књижевном листу у Србији који је излазио у периоду од 1843–1848, објављивана су у неколико бројева термилошка решења, о којима је претходно расправљано на седницама Друштва. Радило се о преводу одређених латинских и немачких речи које су почињале на слова А, В, С, D. Овај пројекат је, међутим, био прекинут након Вукове критике да учесници ни сами не познају добро свој језик, односно (Грицкат 1964, 140) да не знају тачно шта све постоји већ у језику и шта оно значи, и да Друштво не сме да предлаже грешке, имајући у виду да представља ауторитет. Уз Вука, овом пројекту се оштро успротивио и Петар Матић, судија и књижевник из Бјеловара (а познато је да се ни сви чланови Друштва нису слагали са принципима таквога рада), што је резултирало прекидом рада на овом пројекту (Грицкат 1964, Бјелаковић 2017).

Па ипак, мада је посао који су чланови Друштва заједнички започели био прекинут, сваки аутор је и даље припремао радове мањег обима који је требало да буду објављени, што је утицало на покретање *Гласника Друштва српске словесности*. У првој свесци овог листа, Јован Стерија Поповић објављује *Називословне речи израђене Дружствомъ србске словесности* са 513 понуђених термилошких решења, као и текст у којем износи своје неслагање у вези са одлуком Друштва о прекиду рада на терминологији – *Разлози о назывословнымъ речма*.⁴⁷ (Бјелаковић 2017, 66).

⁴⁶ Неки од чланова Друштва били су: Јован Стејић, Димитрије Тирол, Вук Стефановић Караџић, Гаврило Поповић, Петар Нинковић, Јован Балугџић, Константин Коста Цукић, Ђуро Даничић итд.

⁴⁷ Гласник Друштва српске словесности из 1847. године.

Иако је исправна замисао Друштва да се рад на изради терминологије институционализује, постоји неколико кључних разлога за кратак век и неуспех у овом послу, које наводи ауторка Бјелаковић (Бјелаковић 2017, 72–74):

1. Основна идеја чланова Друштва била је да се српским ученицима приближе савремени европски научни токови, а не само конституисање термилошког система. Управо из тог разлога одбацивани су интернационализми који су углавном били на грчком и латинском језику, а који се нису предавали у тадашњим гимназијама.

2. Због малог броја образованог кадра, чланови Друштва су се расплињавали на различите области знања, не посвећујући се довољно једном одређеном пољу, што је било погубно за развој научне мисли у Србији 19. века.⁴⁸

3. Иако је идеја о српском народном језику као о будућем књижевном била блиска обема странама, чланови Друштва били су наклоњени славенизмима, што је за Вукове следбенике било неприхватљиво.

3.3. Развој терминологије српског језика у 20. веку

Говорећи о покушајима стварања српске научне терминологије средином 19. века, Ирена Грицкат бележи да „наш књижевни језик није могао понићи из неког – једног – дијалекта, којим је говорио територијално ограничени део живља, са ограниченим интересима и потребама, ни из пресађеног руског, који се растапао у славеносрпски а зачињавао народним елементима. Он је морао да се формира – и формирао се – као коине, узимајући за себе све оно што му је било потребно, из свих постојећих извора, из говора народа, из жаргона појединих слојева и сталежа, из уметничке литературе, па и из лексичког стварања обдаренијих појединаца лексикографа или сличних лингвистичких прегалаца. Специфичност наших књижевнојезичких прилика лежи и у томе што су у њима учествовали, својим књижевним и културним деловањем, и грчки и латински и руски језик (разумљиво, и остали велики европски језици у области лексике), не рачунајући миксоглотију која је настајала из веза са супстратима и суседима“ (Грицкат 1964, 130–131).

Крај 19. и почетак 20. века представљају време духовног и културног напретка у Србији. Након тешких политичких прилика и коначног ослобођења од турске владавине, „Србија се готово удвостручила, ослободила целу Стару Србију, већи део Македоније, један део Херцеговине, дошла до врата Солуна, повратила оно велико место које је некада имала на Балканском полуострву“, што је све подигло националну веру и самопоуздање, а мрачно и клонуло расположење заменио је национални оптимизам (Скерлић 1914, 349). У таквим приликама и књижевни језик се почео неометано развијати. Било је све више оних који пишу и читају, изражајне могућности језика су се шириле а у школству је настајало све више предмета и све различитијих уџбеника (Ивић 1998, 256). Овакве околности утрле су пут и научном стварању, па је тако српско друштво добијало нове појмове и нове речи за те појмове. Те речи нису „улазиле у језик без потреса и спорова, као што нису увек налажена најпогоднија решења“, али је српски књижевни језик био отворен према међународним терминима и „брзо и успешно се прилагођавао модерној урбаној цивилизацији“ (Ивић 1998, 256–257).

Српска наука је, уз велика српска имена попут Јована Цвијића, Александра Белића, Михаила Петровића и других, први пут достигла ниво светских научних токова уз коју ни језичка истраживања нису изостала. Како Ивић даље бележи (Ивић 1998, 268–269), Белић је тада стекао углед у славистици и то студијама о српскохрватским дијалектима које су откривале дубље односе међу словенским језицима у прошлости, а истовремено, Стојан Новаковић и Љубомир Стојановић објавили су велики број средњовековних текстова значајних за проучавање старог језика, као и граматичке уџбенике, док Новаковићу припада и заслуга да је иницирао израду

⁴⁸Ј. Шафарик, професор физике, предаје и историју; А. Николић, математичар, пише драме, али и расправе о језику старих српских споменика, постаје шеф полиције у Београду, итд. (Бјелаковић 2017, 73).

великог академијског речника – Речника српскохрватског књижевног и народног језика САНУ, чија израда још увек траје⁴⁹ (Ивић 1998, 268–269).

Када говоримо о развоју терминологије српског језика у 20. веку, важан је податак да је Министарство просвете Краљевине Југославије, у периоду од 1932. до 1934. године, на челу са Александром Белићем и Стјепаном Ившићем, оформило Терминолошку комисију која је била задужена за унификацију школске терминологије (Karadža-Garić 1979, XIII). Према наводима ове ауторке, у раду Терминолошке комисије учествовали су и стручњаци за поједине наставне области, а након њеног двогодишњег рада издато је пет речника школске терминологије и номенклатуре. Била је то граматичка, књижевна, ботаничка, зоолошка и хигијенска терминологија настала као резултат пописа термина и назива из тадашњих средњошколских уџбеника и унификације терминологије која је рађена на лингвистичким принципима. Основни задатак аутора била је, према њеним речима, унификација терминологије, а не њена реформа која би, иначе, подразумевала сасвим другачију врсту посла. Збивања уочи Другог светског рата омела су даље излагање оваквих речника унификоване терминологије и номенклатуре, да би се по његовом завршетку ово питање поново актуализовало, међутим, овај пут, оно је постављено знатно шире као питање терминолошке стандардизације у свим областима економског, научног и уопште културног живота (Karadža-Garić 1979, XIII–XIV). Тако је на иницијативу Летописа Матице српске 1953. године, како даље наводи ова ауторка, била расписана анкета о питањима српскохрватског језика, а 1954. године донесени и закључци Новосадског договора⁵⁰ којима је препоручено да установе од највишег научног и културног значаја озбиљно размотре питање решавања хитних и неодложних проблема по питању израде јединствене терминологије и заједничког правописа. Ови закључци обухватили су комплетну обраду система појмова датих области науке, технике и уметности као и терминолошку стандардизацију у смислу лексикографско-лингвистичког посла. Заједнички правопис српскохрватског језика је реализован као део Закључака Новосадског договора, за разлику од јединствене правописне терминологије, која је била у његовом склопу, а која је остала нереализована, па су се тако и терминолошка питања у осталим областима науке, технике и уметности спорије решавала (Karadža-Garić 1979, XIV).

За потребе школства, настајали су речници школске терминологије у издању „Школске књиге“, Загреб као и слични лексикони које је у Београду издавао „Мозаик знања“, а који су ученицима обезбеђивали разумевање појединих наставних области као и служење одговарајућим уџбеницима, наводи ауторка (Karadža-Garić 1979, XV).

Крајем седамдесетих година 20. века интензивирају се лексичка истраживања у србистици, па се тако појављују тематска истраживања и монографске студије посвећени лексички везаној за предмете материјалне културе у одређеној области (у домену рибарства, коларске, пастирске терминологије, терминологије куће и покућства, јела и пића, пчеларске терминологије, воћарске терминологије и др.) (Бјелаковић 2017а, 31–32). Међутим, иако називи ових монографија често

⁴⁹ Израда речника започета је још крајем 19. века, на иницијативу овог академика, који је и руководио радом на скупљању грађе за Речник од 1894. до своје смрти 1916. године. Први талас право сакупљачког рада на терену, који је потрајао до Првог светског рата, био је најинтензивнији и управо је у њему сабран највећи број најпотпунијих и најбољих лексичких збирки. Овај речник обухвата грађу књижевног и народног језика са целовитог простора штокавског наречја, у временском распону од Доситеја и Вука до данас. Због обима посла и вишедеценијског прикупљања грађе, прва књига *Речника* изашла је тек 1959. године. До данас је објављена 21 књига, а када буде завршен, *Речник* ће имати преко 35 књига са око 500.000 одредница. Овај пројекат чији је пун назив *Лингвистичка истраживања савременог српског књижевног језика и израда Речника српскохрватског књижевног и народног језика САНУ* један је од најважнијих и најзахтевнијих дугорочних пројеката српске науке и културе. Основни задатак сарадника на овом пројекту је израда највећег једнојезичног дескриптивног речника нашег језика – *Речника српскохрватског књижевног и народног језика САНУ*, познатог и као *Речник САНУ*. (Извор: <http://www.isj.sanu.ac.rs/projekti/rsanu/>).

⁵⁰ Новосадски договор склопљен је 1954. године, с циљем да се унапреди југословенска култура и наука, а самим тим и језик као битан чинилац културног развоја, уз језичку стандардизацију која би омогућила функционисање српскохрватског стандардног језика за све потребе које језик као инструмент модерне цивилизације треба да задовољава (Караџа-Гарић 1986, 158–159).

у себи садрже реч *терминологија*⁵¹, „лексичка грађа која је у њима регистрована не може се сматрати специјализованом струковном терминологијом јер представља израз материјалне или духовне културе једне регије“ (Бјелаковић 2017а, 31–32).

Међутим, иако постоји низ речника у којима је обрађена грађа из савремених термилошких система појединих научних области, дијахронијском испитивању овог система није посвећено много пажње, па тако ови термилошки системи представљају „радове (1) засноване на ограниченом корпусу који не нуди могућност анализе развоја термилошког система једне области знања у дужем временском периоду или оне (2) настале као резултат истраживања из перспективе дате научне области, а не лингвистике, те превасходно представљају врсту исцрпне студије научне мисли у прошлости, док је терминографском делу посвећено релативно мало пажње“ (Бјелаковић 2017а, 34).

3.4. Термилошке активности у српском језику данас

Данас је итекако јасно да се, услед научно–технолошког прогреса, стручна и научна терминологија брзо развијају, и да терминологија као дисциплина захтева систематски и стални рад стручњака свих профила, а не само лингвиста. Стога је у последњих неколико деценија на нашим просторима доста пажње посвећено управо научним истраживањима из области лингвистике, односно теми научне и стручне терминологије.

Научне и стручне области које су у последњих десетак година најчешћа тема термилошких истраживања јесу економија, маркетинг, саобраћај, спорт, библиотекарство, електротехника и друге (Анђелковић 2017, 78). Термини се, како наводи ова ауторка, углавном прате на лексичко–семантичком нивоу, а најчешћа методологија у оваквим истраживањима је социолингвистички упитник и анализа корпуса. Анализа корпуса је посебна врста истраживања која се данас обавља уз помоћ различитих рачунарских алата; више речи о томе у поглављима 5, 6 и 7 овог рада.

Термилошка активност у Србији део је свеукупне језичке и термилошке политике и планирања у Републици Србији. У највећем броју случајева, овакве активности усмерене су на поступак стандардизације терминологије, око чега се и највећи број језичких стручњака слаже, имајући у виду да се на тај начин постиже равнотежа између два паралелна процеса: процеса глобализације (односно интернационализације) стручне терминологије зарад успостављања једноставније и ефикасније комуникације, и процеса локализације, који подразумева превођење и прилагођавање терминологије у српском језику, а који омогућава очување језичког и културног идентитета (Анђелковић 2017, 77). Терминологији се, последњих година нарочито, све више приступа из угла рачунарске као и когнитивне лингвистике.

3.4.1. Утицај енглеског језика на српски данас

О доминацији енглеског језика и његовом утицају на остале језике рађен је велики број истраживања. Разлог његове супериорности несумњиво лежи у утицају британског колонијализма који је у 19. веку био у пуној снази, а затим у доминантној улози САД на светској сцени од средине 20. века.

Енглески језик данас је један од најраспрострањенијих светских језика, због чега се за њега често користи израз *lingua franca*⁵². *Lingua franca* је, дакле, језик којим се служе говорници различитих матерњих језика, у циљу успостављања међусобне комуникације. Као такав, овај варијетет енглеског језика одступа од стандардне варијанте британског или америчког, и има

⁵¹ *Терминологија куће и покућства у Војводини* ауторке Гордане Вуковић из 1988, *Војвођанска коларска терминологија* аутора Гордане Вуковић, Жарка Бошњаковића и Љиљане Недељков из 1984, *Пастирска терминологија* Срема аутора Жарка Бошњаковића из 1985 и сл. (Бјелаковић 2017а, 32).

⁵² Аутор Остлер наводи да се порекло термина *lingua franca* везује за германске Франке који су се у 5. веку преселили у Галију и прихватили локални језик који је постао познат као језик Франака или *lingua franca*. а који је временом постао незванични језик шире комуникације (Ошњански 2014).

улогу неутралног језика којим се служе говорници различитог порекла, па у том смислу има своје специфичности на нивоу лексике, граматике и прагматике у односу на стандардне варијанте енглеског језика (Ošmjanski 2014, V).

Огроман културни, технолошки, политички, војни и економски утицај англоамеричког друштва није прошао непримећено ни на нашем тлу, већ је силином пенетрирао у све поре друштва рефлектујући се, свакако, и на језик. Оно што би, како бележи ауторка Ошмјански (Ošmjanski 2014, 19), а према наводима аутора Модиаана, „могло представљати изазов у свим заједницама широм света које се суочавају са оваквом улогом енглеског језика, а свесне су својих локалних вредности, идентитета и интереса јесте *да се сачува наша изворна култура и језик, а да се користе предности глобалне интеграције путем језика шире комуникације*“.

Цитирајући Филипсона, Силашки (Silaški 2012), наводи да се велики број европских језика суочава са „маргинализацијом, губитком домена (као последицом лингвистичког империјализма), општом ерозијом важности, те губитком културног значаја – губљењем способности језика да ствара нове речи и изразе, што је резултат константног и неодбранљивог процеса продирања прилагођених или неприлагођених позајмљеница из енглеског језика које нападају све већи број семантичких домена, истискујући притом постојеће домаће или одомаћене речи и отежавајући новим домаћим речима да се укорене“.

Англицизми су у српском језику данас прилично заступљени. Свакодневно их срећемо, како у говору истакнутих појединаца тако и у говору младих, њима обилују рекламни садржаји, интернет, телекомуникационе и рачунарске технологије, често се сусрећу и на пољу банкарства, моде као и у многим другим областима. Једнотомни *речник Матице српске* описује англицизам као „особеност енглеског језика, реч или израз преузет из енглеског језика“, док *Лексикон страних речи и израза* Милана Вујаклије, уз слично објашњење, овај термин дефинише и као „енглеску реч која је ушла у неки туђ језик и одомаћила се у њему“.

Како то примећују ауторке Стијовић и Лекић, однос између домаћих и страних речи једно је од важнијих питања језичког стандарда. „Српска језичка средина није затворена према речима страног порекла. Не настоји се по сваку цену наћи домаћа реч, нарочито ако за неку страну нема добре, одговарајуће домаће или ако страна покрива оно што би се у српском језику морало препричавати. Међутим, велики, често неконтролисани прилив енглеских речи створио је и неке термилошке проблеме.“ (Стијовић и Лекић 2016, 258).

Неки од проблема које наводе ове ауторке, јесте пуко преузимање термина из енглеског језика који пак, у српском језику, имају другачије значење. Такав је случај са енглеском речи *control* чији је еквивалент на српском језику у домену електроенергетике реч *регулација*. У српском језику, именица *контрола* према дефиницији РМС (једнотомника) има значење „надзора, надгледања, провере“. Пуко преузимање ове речи довело је до тога да се она на српски преводи речју *контрола*. Тако у здравству користимо реч *контрола*, уместо надзирање болести, док се у електроенергетици често (погрешно) користи *контрола* уместо (техничко) управљање електросистемом односно регулација електросистема.

Неки од истакнутих српских лингвиста, попут Ивана Клајна и Ранка Бугарског, оправдавају потребу за увођењем англицизама, имајући у виду да је то језик глобализације и да га је у том смислу немогуће заобићи. Клајн наводи да је српски језик некада обиловао германизмима и галицизмима (речима немачког и француског порекла), а да је данас на тој позицији енглески језик. Овај аутор подсећа да „наш главни посао није да прогонимо англицизме, већ да пазимо да се српски језик не заборави“. Да се у данашњем свету не може живети без енглеских речи, али да оне треба да се преузимају и прилагођавају са мером и знањем, сматра и Ранко Бугарски. Говорећи о неприлагођеном енглеском језику у српском а не, дакле, о англицизмима, Рајна Драгићевећ нарочито истиче штетност медија односно часописа које млађа популација прати. У тим текстовима се енглеске речи намерно не избацују, услед вероватне потребе да се буде „у тренду“.⁵³

⁵³ Преузето маја 2019. према тексту са веб-сајта <http://www.politika.rs/scc/clanak/179591/Mozemo-li-da-zivimo-bez-engleskih-reci>.

Сагледавајући утицаје које енглески језик има на српски, аутор Прћић уводи појам *англосрпски* као нови варијетет језика. У свом прилогу „И ми речи за трку имамо“ (Прћић 2005, 18–24), овај аутор скреће пажњу на то да речи у српском језику ни на који начин не заостају за истозначним англицизмима, и указује на све упадљивије повођење српског језика за нормама и обичајима енглеског језика које се редовно испољава у неоправданом и објективно непотребном коришћењу речи под утицајем енглеског језика — у већини медија као и у говору и писању свих оних који се на њих угледају, а којих је све више. Такав српски језик, закључује Прћић, хибридна је врста која одступа од својих норми и прилагођава се нормама енглеског језика па га у складу са тиме и можемо назвати *англосрпским*. Треба, пак, имати на уму и да оправдани и објективно потребни англицизми не спадају у домен англосрпског већ српског језика. Под „објективном потребом“, овај аутор подразумева својеврстан допринос српском језику увођењем новог појма који се датим англицизмом именује (нпр. модем, хардвер, софтвер, интернет итд.), и који тако српски богати не само новом речју него и новим и до сада непознатим садржајем, тј. значењем. Насупрот томе, у српском језику постоји велики број англицизама који у језик не уносе нов садржај, већ служе за то да би се изразио већ познати садржај заогрнут у нов или само привидно нов облик. Таквим англицизмима врши се потискивање и истискивање постојећих речи њиховим англосрпским синонимима, тврди овај аутор.

Интересантно је његово разграничење **струковне** и **статусне** употребе англицизама, у истом прилогу, па тако овај аутор наводи да је струковна употреба најчешће мотивисана све раширенијим веровањем стручњака из готово свих области да су у данашње време једино термини из енглеског језика они који на прави, аутентичан и убедљив начин изражавају одређено стручно значење, док се постојеће домаће речи или игноришу или се, што је чешће (али и поразније), за њих уопште и не зна. Статусна употреба англицизама, с друге стране, мотивисана је веровањем да је њихова употреба у говору и писању модернија, отменија и елегантнија јер оставља утисак припадности надмоћнијем англофоном културном и језичком обрасцу (Прћић 2005, 20).

3.4.2. Критеријуми изграђивања терминологије

Процес израђивања терминологије једне научне области сложен је и врло захтеван задатак. Да би се он ваљано обавио, потребно је, како то предлаже аутор Бугарски, унапред утврдити критеријуме које термини одређене области треба да испуне. Ти критеријуми могу бити различити, зависно од научне и стручне области односно посебне сврхе за коју се испитују (нпр. за машинску обраду термина *краткоћа* је врло значајан параметар) (Bugariski 2007, 110).

Велики број критеријума који се јављају у процесу изграђивања, систематизовања, стандардизовања и прихватања лингвистичких и научних термина, Бугарски разврстава на четири велике групе и то (Bugariski 2007, 92): **формалне** (дужина, лакоћа изговора и писања, деривациони потенцијал, уклапање у терминолошки систем), **семантичке** (појмовна адекватност, мотивисаност и позорност, могућност дефиниције, прецизност, изражајност, једнозначност/вишезначност, синоними и хомоними, могућност разликовања терминолошке од општејезичке употребе), **прагматичке** (прихватљивост језичком осећању, економичност, стабилност/варијабилност, круг корисника, степен одомаћености, еквиваленти у другим језицима односно интернационалност...) и **социolingвистичке** (језици и варијанте са становишта системских могућности и преференција али и вредности које колективи за њих везују, традиције социокултурних средина, степен заступљености дате струке у науци и настави, образовни ниво и потребе корисника).

Један од начина вредновања термина приказан је табелом 2, кроз пример издвојених критеријума према избору аутора Бугарског (Bugariski 2007, 109).

Табела 2. Пример матрице за систематско вредновање термина⁵⁴

Критеријуми	Термини			
	компетенција; перформанса	језичка способност; говорна делатност	трансформација	преоблика
1. системност	–	+	+	–
2. продуктивност	0	0	+	–
3. једнозначност	–	+	0	0
4. интернационалност	+	–	+	–
5. мотивисаност	–	+	+	–
6. раширеност	0	0	+	–
7. стабилност	–	+	+	–
8. конотације	–	+	+	–
9. краткоћа	+	–	–	0
10. лакоћа	–	+	–	+
	2	6	3	1

Карактеристике које би, у идеалном случају, требало да задовољи један термин односно терминосистем, биле би: транспарентност, интернационалност, устаљеност, краткоћа, системност, недвосмисленост, прецизност и несинонимност (о чему је већ било речи у поглављу 2.4 овог рада) (Šipka 2006, 151). Према овом аутору, одређење термина не треба схватати као обавезно подложно свим наведеним карактеристикама јер се не ради о потребним односно довољним условима, већ о списку најчешће присутних атрибута. Како даље, у истом издању, овај аутор наводи, набројане карактеристике су у терминолошкој пракси често некомпатибилне и контрадикторне, па је успостављени термин често резултат компромиса између различитих захтева (Šipka 2006, 151).

Неки од критеријума које би нарочито требало узети у обзир у процесу обраде термина у области електроенергетике јесу *значање* унутар конкретног домена, *недвосмисленост* (важно је нпр. разликовати типове расклопне опреме), *несинонимност* (синонимија се не јавља често у области електроенергетике, а када се јави обично је у питању интернационална реч која се подједнако употребљава као и њен домаћи еквивалент), *краткоћа* и *прецизност* (како би се пружио краћи и најпрецизнији опис предмета или појаве који представља), *постојаност* и *трајност* током историје развоја области која је предмет анализе ове дисертације, као и *интернационалност* (домаћи термин има предност над интернационалним термином уколико постоји, међутим, уколико не постоји односно уколико је његово значење непотпуно, у пракси се прибегава коришћењу интернационализма, нпр. *generator* „генератор“).

3.4.3. Институције задужене за језичку и терминолошку политику

Језичка и терминолошка политика једне државе, у највећој мери, зависе од односа државе односно државних институција према том питању. У Републици Србији, највећи значај имају два документа којима је регулисана употреба српског језика и ћириличног писма у држави. То је, пре свега, Устав РС („Службени гласник РС“, бр. 98/2006), а потом и Закон о службеној употреби језика и писма („Службени гласник РС“, бр. 45/91, 53/93, 67/93, 48/94, 101/2005 – др. закон, 30/2010, 47/2018 и 48/2018 – испр.).

За креирање и спровођење језичке и терминолошке политике у Републици Србији задужено је неколико институција. То су Институт за српски језик САНУ и у оквиру њега Одбор за

⁵⁴ Ознаке „+“ и „–“ уписују се уколико термини одређену особину поседују у већој односно мањој мери, док „0“ означава да су термини у том погледу неутрални (Bugarski 1988, 18).

стандардизацију српског језика и Национално веће за српски језик и писмо. Уз ове институције, Министарство за европске интеграције (наследник некадашње Канцеларије за европске интеграције), у процесу превођења европског законодавства на српски језик, годинама води доследну и систематску језичку и терминолошку политику, а не смео заборавити ни Институт за стандардизацију Републике Србије, о којем је већ било речи. Уз њих, активно је и Друштво за језичке ресурсе и технологије, удружење које промовише и унапређује све гране језичких технологија на научном, стручном и практичном нивоу и доприноси изградњи језичких ресурса у дигиталном облику, с тежњом да, уз сарадњу са истраживачима и корисницима језичких технологија, очува српски језик у дигиталној ери.

3.4.3.1. Институт за српски језик САНУ

Институт за српски језик САНУ централна је научна установа у Србији задужена за систематско проучавање српског језика и његове историје, као и израду капиталних лексикографских и лингвогеографских дела – речника и атласа⁵⁵.

Пројекти који се тренутно реализују у Институту, а представљају исцрпна проучавања српског језика у његовој прошлости и садашњости су:

- Лингвистичка истраживања савременог српског књижевног језика и израда Речника српскохрватског књижевног и народног језика (од 1959. године до данас објављен је 21 том *Речника САНУ*, а планираним обимом од преко 30 томова *Речник* ће обухватити најмање 500.000 одредница).
- Обрада старог српског писаног наслеђа и израда Речника црквенословенског језика српске редакције (пројекат је израстао из Старословенистичког одсека формираног 1969. године у којем је ексцерпирана грађа из одабраних средњовековних рукописа за потребе *Речника*).
- Етимолошка истраживања српског језика и израда Етимолошког речника српског језика (од 2003. године до сада објављене су три свеске тезаурусног речника (*a–бј*), а у току је израда *Приручног етимолошког речника српског језика*).
- Дијалектолошка истраживања српског језичког простора (у оквиру овог пројекта основаног 2002. године истражују се народни говори на целом српском етничком и језичком простору. У току је и израда *Српског дијалектолошког атласа*).
- Опис и стандардизација савременог српског језика (у оквиру овог пројекта основаног 2006. године ради се на континуираном и систематском проучавању и описивању савременог српског језика).
- Примена лингвистичких истраживања у изради дигиталног појмовника (овај потпројекат започет је 2011. године као део пројекта *Интердисциплинарно истраживање културног и језичког наслеђа Србије и израда мултимедијалног интернет портала Појмовник српске културе*, чији је носилац Етнографски институт САНУ. Основни циљеви потпројекта су формирање дигиталне лексикографске базе, припрема за лексикографски и лингвогеографски рад у дигиталном окружењу и израда одредница за *Појмовник*).

3.4.3.2. Одбор за стандардизацију српског језика

Одбор за стандардизацију српског језика основан је 1997. године, а функционише при Институту за српски језик САНУ. Представља стручно тело које обједињује највећи број институција и појединаца од струке, који се научно баве изучавањем српског језика, и који су постигли готово једногласан став о најважнијим питањима у вези са статусом српског језика у Србији и на целокупном српском говорном подручју. Задатак Одбора је и да предлаже мере које би унапредиле положај и значај српског језика. Одбор има за циљ језичко планирање које би

⁵⁵ Подаци о Институту и пројектима преузети су јуна 2019. године са веб-сајта: <https://www.sanu.ac.rs/instituti/institut-za-srpski-jezik/>.

водило адекватној језичкој политици и пракси, обезбеђивању статуса српског језика као националног, како би се планирали и истицали приоритети његовог проучавања ради осавремењивања стандарднојезичке норме. Оснивачи Одбора су три академије наука и уметности (САНУ, ЦАНУ и АНУРС), Институт за српски језик САНУ и Матица српска, факултети из Србије, Републике Српске и Црне Горе на којима се студира и изучава српски језик, те Српска књижевна задруга.⁵⁶

3.4.3.3. Министарство за европске интеграције РС

Сектор за координацију превођења и припрему верзије правних тековина Европске уније на српском језику некадашње Канцеларије за европске интеграције, данас функционише као ужа унутрашња јединица Министарства за европске интеграције,⁵⁷ али је његов задатак и даље исти. Ово Одељење је, као што је и наведено на сајту Министарства,⁵⁸ задужено за координацију превођења и припрему превода прописа Европске уније на српски језик; координацију правне, стручне и језичке редактуре преведених правних аката; координацију израде предлога за успостављање јединствене терминологије Европске уније и израду појмовника терминологије Европске уније на српском језику и сарадњу са институцијама у Републици Србији, надлежним органима ЕУ и земљама у региону ради унапређења процеса превођења и редактуре, а обавља и друге послове из ове области.

У процесу стицања кандидатуре и приступања Европској унији, један од кључних услова за чланство наше државе јесте превођење правних тековина Европске уније (*acquis communautaire*) на српски језик, односно израда њихове националне верзије. Правне тековине Европске уније обухватају корпус заједничких права и обавеза који важе за све државе чланице у оквиру Европске уније. Због обима прописа који се објављују у Службеном листу Европске уније (који засад броји око 200.000 страница Службеног листа, што је око 240.000 страница стандардног формата, уз просечан годишњи пораст од 3.000–5.000 страница), припрема правних тековина ЕУ на службеном језику показала се као највећи преводилачки подухват већине нових држава чланица Европске уније. Из тог разлога, Министарству за европске интеграције додељена је улога централног координативног механизма за све фазе процеса (превођење, стручна, правнотехничка и језичка редактура), креирање терминолошке базе и преводилачких меморија, извештавање и одређивање приоритета за превођење у сарадњи са осталим државним органима. Републички секретаријат за законодавство, ресорна министарства и остали органи државне управе, у оквиру процеса припреме правних тековина ЕУ на српском језику, обављају послове правнотехничке, односно стручне редактуре превода. За коначну верификацију преведених текстова задужена је посебна Радна група за верификацију састављена од представника свих институција које учествују у процесу.

Већ је наведено да у процесу припреме верзије правних тековина Европске уније разликујемо четири фазе: превођење, стручна редактура, правнотехничка редактура и језичка редактура:

1. У фази *превођења* задатак преводилаца је да преведу текст правног акта ЕУ на српски језик, поштујући правила прописана Приручником за превођење правних аката Европске уније, и користећи Терминолошка упутства и Језичке смернице (документа која се могу пронаћи и на веб-сајту Министарства за европске интеграције: <http://www.mei.gov.rs/srp/prevodjenje/publikacije/>). Уз ова документа, у овој фази, врло је важно

⁵⁶ Подаци о Одбору за стандардизацију српског језика преузети су јуна 2019. године са веб-сајта: <http://www.isj.sanu.ac.rs/odbor-za-standardizaciju/o-odboru/>.

⁵⁷ Овај сектор је преименован у *Одељење за припрему српске верзије правних тековина Европске уније*, а функционише у саставу Сектора за комуникације, обуку и припрему српске верзије правних тековина Европске уније.

⁵⁸ Сви подаци у вези са Министарством за европске интеграције и процесом припреме верзије правних тековина Европске уније преузети су јуна 2019. године са: <http://www.mei.gov.rs/>.

користити термине и лексику који се налазе у термилошкој бази Евроним, о којој ће бити више речи у поглављу 5.5.3.2, као и у преводилачким меморијама.

2. У фази *стручне редактуре* главну улогу имају стручни редактори који су задужени за исправљање термина и превода, односно за проверу да ли смисао превода одговара ономе што у изворном тексту заиста и пише, са становишта струке. Њихов задатак је и да предлажу термилошка решења за нове појмове, избор једног од могућих термилошких решења уколико их има више, као и да предлажу боља термилошка решења од оних која се тренутно користе у домаћим прописима, уколико за тиме има потребе. У средишту њихове пажње је, дакле, стручна терминологија као и смисао одредбе, а резултат ове фазе је стручно поуздан текст и стандардизован корпус стручне терминологије, као што је наведено у Језичким смерницама Министарства.

3. У фази *правнотехничке редактуре*, стручно редигован текст обликује се у складу са номотехничким правилима, која су утврђена у Приручнику за превођење правних аката Европске уније као и оним која су својствена правном систему Републике Србије, и са општом правном терминологијом која је својствена правном систему Републике Србије, при чему се узимају у обзир посебности правног система ЕУ.

4. У фази *језичке редактуре* језички редактор пореди изворни и преведени текст, исправља све граматичке и правописне грешке, обележава недоследности у употреби лексики и терминологије у тексту, односно проверава укупни језички квалитет текста превода и уноси одговарајуће језичке исправке.

5. Фаза *верификације* је уједно и последња фаза процеса у којој учествују сви стручњаци који су радили на тексту из сваке фазе понаособ: преводилац, стручни редактор, правнотехнички редактор и језички редактор.

Један од резултата процеса превођења правних тековина Европске уније на српски језик који је уједно и својеврсна помоћ у процесу превођења јесте вишејезична база термина Евроним (поглавље 5.5.3.2). Уз Евроним, у процесу превођења правних тековина Европске уније на српски језик, за који је задужено Министарство за европске интеграције, настаје и Евротека, збирка европских правних текстова и њихових сегмената, тј. двојезични (енглеско-српски) корпус, о чему ће такође бити више речи у поглављу 5.5.3.1.

3.4.3.4. Институт за стандардизацију РС

Иако је општа слика о Институту за стандардизацију РС представљена у претходном поглављу, потребно је изнети још неке од чињеница које су у вези са овом институцијом.

Пре свега, одређени ЈУС стандарди који су били објављени још 1979. године, као што су *Терминологија – Појмови и називи: општи принципи*, *Терминологија – Системи појмова и њихово приказивање* и *Информације и документација – Смернице за састављање и даљи развој једнојезичких тезауруса*, као и одређени стандарди објављени 1994. године: *Информације и документација – Упутства за успостављање и развој једнојезичких тезауруса* и *Информације и документација – Упутства за успостављање и развој вишејезичких тезауруса*, били су усаглашени са међународним, *ISO* стандардом. Иако су наведени стандарди касније измењени и допуњени, тако да се у изворном облику више не користе, добар су показатељ да је у Југославији још седамдесетих година постојала потреба за развојем стандарда из области терминологије (Трговац 2016, 195).

Ознаку *ЈУС* носили су југословенски стандарди. Они су замењени српским *СРПС* стандардима, за чије је доношење задужен Институт за стандардизацију Републике Србије.

Комисија задужена за област терминологије Института за стандардизацију РС носи ознаку А037,⁵⁹ као што је већ поменуто раније у тексту, а има за циљ „стандардизацију принципа, метода

⁵⁹Овом комисијом руководи проф. др Ранка Станковић.

и примене термина и других језичких и садржинских извора у области вишејезичке и културне разноврсности⁶⁰. Посредством ове Комисије Република Србија учествује у раду *ISO/TC 37*.

До сада су објављени следећи стандарди из области терминологије који се могу пронаћи и на веб-сајту Института за стандардизацију:

- Термиолошки чланци у стандардима – Део 1: Општи захтеви и примери представљања (*SRPS ISO 10241-1:2018*) – овај стандард утврђује захтеве за планирање и структурирање термиолошких чланака у стандарду, са примерима термиолошких чланака у *ISO* и *IEC* документима (*SRPS ISO 10241-1 2018*);
- Термиолошки чланци у стандардима – Део 2: Усвајање стандардизованих термиолошких чланака (*SRPS ISO 10241-2:2018*) – односи се на увођење стандардизованих термиолошких чланака у друга културна и језичка окружења, а посебно на усвајање међународно стандардизованих термиолошких чланака од стране регионалних и националних тела за стандардизацију (*SRPS ISO 10241-2 2018*);
- Термиолошки рад – Вокабулар – Део 1: Теорија и примена (*SRPS ISO 1087-1:2003*) – стандард даје системски опис појмова у области терминологије (*SRPS ISO 1087-1 2003*);
- Термиолошки рад – Вокабулар – Део 2: Примене рачунара (*SRPS ISO 1087-2:2005*) – стандард дефинише термине за обраду језика и информација за примене у термиолошком раду и терминографији (*SRPS ISO 1087-2 2005*);
- Алфабетски редослед вишејезичких термиолошких и лексикографских података дат латиничним алфабетом (*SRPS ISO 12199:2018*) – овим се стандардом утврђује редослед карактера који се користе за алфабетски редослед вишејезичких термиолошких и лексикографских података (термина, елемената термина или речи) представљених латиничним алфабетом (*SRPS ISO 12199 2018*);
- Библиографске референце и идентификација извора за потребе термиолошког рада (*SRPS ISO 12615:2018*) – овај стандард се примењује на бележење, чување и размену информација о библиографским изворима за термиолошки рад и терминографију (*SRPS ISO 12615 2018*);
- Терминографија која се односи на превођење (*SRPS ISO 12616:2018*) – овим стандардом утврђује се начин повећања продуктивности преводилаца и унапређивања квалитета преведеног текста, а осмишљен је тако да помаже преводиоцима, локализаторима, терминолозима, лингвистима, менаџерима информација, администраторима података у предузећима, јавним установама, преводилачким агенцијама и агенцијама за локализацију (*SRPS ISO 12616 2018*);
- Смернице за управљање пројектима термиолошке стандардизације (*SRPS ISO 15188:2018*) – овим су стандардом утврђене смернице које успостављају фазе и процедуре које треба пратити током термиолошких стандардизацијских пројеката (*SRPS ISO 15188 2018*);
- Приказивање/представљање одредница у речницима – Захтеви, препоруке и информације (*SRPS ISO 1951:2009*) – стандардом се спецификује генеричка структура једнојезичких, вишејезичких, општих и специјализованих речника, независно од издавачког медија, а у циљу лакшег стварања, стапања, проналажења, поређења, екстракције, размене и дисеминације лексикографских података у речницима (*SRPS ISO 1951 2009*);
- Термиолошки производи и услуге – Преглед и упутство (*SRPS ISO 22128:2011*) – стандардом се утврђују и описују термиолошки производи и услуге, односно општеприхваћена термиолошка пракса коју треба поштовати приликом пружања термиолошких услуга и производа (*SRPS ISO 22128 2011*);
- Систем за управљање терминологијом, знањем и садржајем – Концептуални аспекти развоја и интернационализације класификационих система (*SRPS ISO 22274:2018*) – овим стандардом успостављени су основни принципи и захтеви који треба да осигурају да класификациони систем буде погодан за примену на светском нивоу, узимајући у обзир културну и језичку разноликост као и захтеве тржишта (*SRPS ISO 22274 2018*);

⁶⁰Преузето октобра 2019. године са сајта: https://www.iss.rs/la/tc/?national_committee_id=552

- Оцењивање и утврђивање референтних вредности (бенчмаркинг) термилошких ресурса – Општи појмови, принципи и захтеви (*SRPS ISO 23185:2010*) – стандардом се описују основни појмови који се односе на ефективну употребу термилошких података, узимајући у обзир циљеве корисника (*SRPS ISO 23185 2010*);

- Графичка нотација за моделовање појмова у термилошком раду и његова веза са *UML*-ом⁶¹ – Део 1: Смернице за коришћење *UML* нотације у термилошком раду (*SRPS ISO 24156-1:2018*) – овим стандардом дају се смернице за употребу подгрупе *UML*-ових симбола, независно од њиховог уобичајеног *UML* значења, ради представљања појмова у појмовним моделима насталим као резултат анализе појмова. Користи се за представљање појмова и односа међу њима у оквиру термилошког рада, а не описује основне појмове термилошког рада (*SRPS ISO 24156-1 2018*);

- Систем за управљање терминологијом, знањем и садржајем – Пројектовање, примена и одржавање система управљања терминологијом, (*SRPS ISO 26162:2018*) – утврђује критеријуме за пројектовање, припрему и одржавање система за управљања терминологијом (*TMS – Terminology Management System*), а намењен је термилозима, креаторима софтвера и осталима у процесу развоја знања о *TMS*-у (*SRPS ISO 26162 2018*);

- Термилошка политика – Развој и имплементација (*SRPS ISO 29383:2012*) – овај стандард даје смернице и методологију креаторима политике у владиним организацијама и управи, као и непрофитним и комерцијалним организацијама, за развој политике или стратегије у вези са планирањем терминологије и њеним управљањем (*SRPS ISO 29383 2012*);

- Термилошки рад – Принципи и методе (*SRPS ISO 704:2018*) – овај стандард утврђује основне принципе и методе за формулисање дефиниција термина и веза између предмета, појмова и њихових термилошких репрезентација (*SRPS ISO 704 2018*);

- Термилошки рад – Усаглашавање појмова и термина (*SRPS ISO 860:2018*) – овај стандард дефинише методолошки приступ усаглашавању појмова, система појмова, дефиниција и термина (*SRPS ISO 860 2018*).

Према *Стратегији даљег развоја информатике у Савезној Републици Југославији 1997.* године, дефинисано је да државни органи треба да подрже развој посебних софтверских алата за српски језик, а пре свега да стандардизују и уреде основу за јединствену примену ових алата за целокупан југословенски културни садржај на различитим рачунским платформама (Трговац 2016, 195–196).

Стандарди којима се утврђује примена рачунарских апликација у домену терминологије, а које је такође донела Комисија задужена за област терминологије Института су следећи:

- Управљање језичким ресурсима – Структуре својстава – Део 1: Представљање структура својстава (*SRPS ISO 24610-1:2018*) – овим стандардом се дефинише формат за представљање, чување и размену структура својстава (енг. *Feature structures*) у обради природних језика која се односи на анотацију, креирање или анализу лингвистичких података (*SRPS ISO 24610-1 2018*);

- Управљање језичким ресурсима – Структуре својстава – Део 2: Декларација система својстава (*SRPS ISO 24610-2:2018*) – и овим стандардом дефинише се формат за представљање, чување или размену структура својстава (енг. *Feature structures*) у обради природних језика која се односи на анотацију, креирање или анализу лингвистичких података. Креиран је са циљем да обезбеди формат података за дефинисање хијерархије типова и декларисање ограничења која се односе на скуп спецификација својстава и операција над структурама својстава, чиме се нуде средства за проверу усклађености сваке структуре својстава са референтним спецификацијама (*SRPS ISO 24610-2 2018*).

- Управљање језичким ресурсима – Оквир за морфосинтаксичку анотацију (*Morpho-syntactic annotation framework – MAF*) (*SRPS ISO 24611:2018*) – овим стандардом се дефинише оквир за представљање анотација облика речи у текстовима, такве анотације које се тичу токена,

⁶¹ Универзални језик за моделирање – *UML* (енг. *Universal Modelling Language*)

њихове везе са лексичким јединицама и њиховим морфосинтаксичким својствима (SRPS ISO 24611 2018);

- Управљање језичким ресурсима – Оквир за лингвистичку анотацију (*LAF*) (SRPS ISO 24612:2018) – овим стандардом дефинисан је оквир за лингвистичку анотацију (*Linguistic Annotation Framework – LAF*) који служи за представљање лингвистичких анотација језичких података, као што су корпуси писаних текстова, говора, знакова и видео-материјала (SRPS ISO 24612 2018).

- Управљање језичким ресурсима – Оквир за лексичко обележавање (*LMF*) (SRPS ISO 24613:2018) – овим стандардом дефинисан је оквир за лексичко обележавање (*Lexical Markup Framework – LMF*), метамодел за представљање података у лексичким базама које се користе у једнојезичким и вишејезичким софтверским апликацијама (SRPS ISO 24613 2018);

- Управљање језичким ресурсима – Подела писаног текста на речи – Део 1: Основни појмови и општи принципи (SRPS ISO 24614-1:2018) – поред основних појмова и општих принципа поделе на речи, овај стандард даје и језички независне смернице којима се омогућава да писани текст буде подељен на речи као јединице сегментације, а примењује се код превођења, уређивања садржаја, технологије говора, рачунарске лингвистике и лексикографије (SRPS ISO 24614-1 2018).

- Управљање језичким ресурсима – Подела писаног текста на речи – Део 2: Подела на речи у кинеском, јапанском и корејском језику (SRPS ISO 24614-2:2018) – основни појмови и општи принципи који су утврђени у првом делу овог стандарда примењују се на кинески, јапански и корејски језик, а текст се дели на мање текстуалне јединице (токене, речи, фразе) погодне за одређене софтверске апликације над језичким ресурсима, попут обраде природног језика, проналажења информација и машинског превођења (SRPS ISO 24614-2 2018).

- Управљање језичким ресурсима – Оквир за синтаксичку анотацију (*SynAF*) – Део 1: Синтаксички модел (SRPS ISO 24615-1:2018) – описује оквир за синтаксичку анотацију (*SynAF – Syntactic Annotation Framework*), а циљ овог стандарда је да подржи интероперабилност између језичких ресурса или компоненти за језичку обраду. Представља допуну стандарда који описује оквир за морфосинтаксичку анотацију ISO 24611 (SRPS ISO 24615-1 2018).

- Управљање језичким ресурсима – Оквир за вишејезичке информације (*Multilingual Information Framework – MLIF*) (SRPS ISO 24616:2018) – овај стандард даје основну платформу за моделовање и управљање вишејезичким информацијама у различитим областима: локализацији, превођењу, мултимедијалним анотацијама, управљању документима, подржавању дигиталних библиотека и применама информационог или пословног моделовања (SRPS ISO 24616 2018).

- Управљање језичким ресурсима – Оквир за семантичку анотацију (*SemAF*) – Део 1: Време и догађаји (*SemAF-Time, ISO-TimeML*) (SRPS ISO 24617-1:2018) – овим стандардом утврђује се формализовани језик за обележавање заснован на *XML*-у за систематично екстраховање и представљање временских информација, као и за омогућавање размене временских информација, како између различитих система за обраду језика тако и између различитих шема представљања времена (SRPS ISO 24617-1 2018).

- Управљање језичким ресурсима – Оквир за семантичку анотацију (*SemAF*) – Део 2: Дијалошке форме (SRPS ISO 24617-2:2018) – овај стандард дефинише концепте засноване на искуству и теорији за анотацију дијалога, формални језик за изражавање дијалошких анотација – дијалошки језик за обележавање (*DiAML*) и метод за сегментацију дијалога на семантичке јединице, а њиме се омогућава ручна или аутоматска анотација делова дијалога, са информацијама о активностима комуникације које учесници изводе својим учешћем у дијалогу (SRPS ISO 24617-2 2018).

- Управљање језичким ресурсима – Оквир за семантичку анотацију (*SemAF*) – Део 4: Семантичке улоге (*SemAF-SR*) (SRPS ISO 24617-4:2018) – овај стандард предлаже прихватљиву шему анотације семантичких улога, а укључује семантичке/реченичке односе између глагола и

његових аргумената, као и односе важне за друге предикативне елементе попут номинализације, именица, придева и предикатских модификатора (SRPS ISO 24617-4 2018).

- Управљање језичким ресурсима – Оквир за семантичку анотацију (*SemAF*) – Део 5: Структура дискурса (*SemAF-DS*) (SRPS ISO/TS 24617-5:2018) – овај стандард дефинише начин на који је дискурс структуриран, у смислу његове реализације/представљања и садржаја (SRPS ISO/TS 24617-5: 2018).

- Управљање језичким ресурсима – Контролисани природни језик (*CNL*) – Део 1: Основни појмови и принципи (SRPS ISO/TS 24620-1:2018) – овај стандард дефинише принципе контролисаних природних језика (*Controlled Natural Language – CNL*) и benefite које пружа, заједно са одговарајућом технологијом за подршку (SRPS ISO/TS 24620-1 2018).

- Управљање језичким ресурсима – Оквир за семантичку анотацију – Део 6: Принципи семантичке анотације (*SemAF* принципи) (SRPS ISO 24617-6:2018) – овај стандард у главним цртама дефинише стратегију за развој посебних шема анотације за неке класе семантичких појава (SRPS ISO 24617-6 2018).

- Управљање језичким ресурсима – Оквир за семантичку анотацију – Део 7: Просторне информације (*ISOspace*) (SRPS ISO 24617-7:2018) – овај стандард даје оквир за кодирање широког опсега како просторних тако и просторно-временских информација које се односе на кретање изражено у тексту на природном језику (SRPS ISO 24617-7 2018).

- Управљање језичким ресурсима – Трајна идентификација и одрживи приступ (*PISA*) (енг. *Persistent identification and sustainable access – PISA*) (SRPS ISO 24619:2018) – овим стандардом дефинишу се захтеви трајног идентификатора (енг. *Persistent Identifier – PID*) и за употребу *PID*-ова као референце и цитата језичких извора у документима, као и самих језичких извора (као нпр. дигиталних речника, терминолошких језичких ресурса, лексичких ресурса за машинско превођење, анотираних мултимедијалних/мултимодалних корпуса, корпуса текстова са анотацијама, нпр. морфосинтаксичких информација итд.) (SRPS ISO 24619 2018).

- Управљање језичким ресурсима – Предефинисане компоненте метаподатака (*CMDI*) (енг. *Component Metadata Infrastructure – CMDI*) – Део 1: Модел компоненти метаподатака (SRPS ISO 24622-1:2018) – овај стандард описује модел који омогућава флексибилно конструисање интероперабилних шема метаподатака за језичке ресурсе (SRPS ISO 24622- 1 2018).

- Управљање језичким ресурсима – Транскрипција говорног језика (SRPS ISO 24624:2018) – овај стандард дефинише, пре свега, правила за представљање транскрипције аудио и видео-записа говорне интеракције у *XML* документима засноване на *TEI*⁶² смерницама (енг. *Text Encoding Initiative*). Уз то, циљ овог стандарда је да доведе у однос транскрибоване податке са стандардима за анотираних корпуса (SRPS ISO 24624 2018).

- Систем за управљање терминологијом, знањем и садржајем – *TermBase eXchange (TBX)* (SRPS ISO 30042:2018) – овај стандард дефинише различите типове процеса који обухватају термилошке податке, укључујући анализу, описно представљање, дисеминацију и међуразмену у различитим рачунарским окружењима (SRPS ISO 30042 2018);

- Кодови за представљање назива језика – Део 1: Алфа 2 код (SRPS ISO 639-1:2018) – овај део *ISO 639* стандарда даје код који се састоји од језичких кодних елемената сажетих у двословни језички означаилац за представљање назива језика. Језички означаилац је, према овом стандарду, првобитно издвојен ради употребе у терминологији и лингвистици, али може бити прилагодљив за сваку примену која захтева изражавање језика у двословном кодном облику, а нарочито за потребе рачунарских система и апликација. Циљ овог стандарда је практична примена Алфа 2 кода, и то код оних језика који садрже значајан корпус специјализованих језика и терминологије. Језици створени искључиво за машинско коришћење, као што су језици за рачунарско програмирање, нису укључени у овај код (SRPS ISO 639-1 2018).

⁶² Иницијатива за кодирање текста (скр. *TEI*) представља конзорцијум који развија и одржава стандард за представљање текстова у дигиталном формату. Може се окарактерисати као слободно доступан скуп смерница за кодирање текстова у хуманистичким наукама уз употребу *XML*-а, међународни конзорцијум који постоји да би подржао развој ових смерница и заједницу пројеката и појединаца који користе *TEI* Смернице (Ерјавец 2010, 5).

- Кодови за представљање назива језика – Део 2: Алфа 3 кôд (*SRPS ISO 639-2:2018*) – овај део *ISO 639* стандарда дефинише две групе трословног алфаветског кода за представљање назива језика, једну за примену у терминологији а другу за примену у библиографији. Ови кодови се широко користе у библиотечкој заједници, а терминолози и лексикографи их могу прилагодити за сваку примену за коју се захтева изражавање језика у форми кода. Језици створени искључиво за машинско коришћење, као што су језици за рачунарско програмирање, нису укључени ни у овај кôд (*SRPS ISO 639-2 2018*).

- Кодови за представљање назива језика – Део 3: Алфа 3 кôд за потпуни обухват језика (*SRPS ISO 639-3:2018*) – дефинише кодове који се састоје од језичких кодних елемената састављених од трословног језичког идентификатора за представљање језика. Овај кôд објавило је Регистрационо тело за *ISO 639-3*, а његова примена је широка, посебно код рачунарских система и то онда када постоји потреба за подршком великог броја језика. Док су у фокусу *ISO 639-1* и *ISO 639-2* језици који су најзаступљенији у светској литератури, овај део стандарда укључује живе, изумрле, старе и створене језике, главне или мањинске, писане и говорне. Језици створени искључиво за машинско коришћење, као што су језици за рачунарско програмирање, нису укључени у овај кôд (*SRPS ISO 639-3: 2018*).

- Кодови за представљање назива језика – Део 4: Општи принципи кодирања и представљање назива језика и повезаних ентитета, са смерницама за примену (*SRPS ISO 639- 4:2018*) – овај стандард дефинише принципе кодирања језика који су већ утврђени у другим деловима истог стандарда, као и њихове комбинације са другим кодовима (*SRPS ISO 639- 4 2018*).

- Кодови за представљање назива језика – Део 5: Алфа 3 кôд за породице и групе језика (*SRPS ISO 639-5:2018*) – овај стандард даје кôд који се састоји од језичких кодних елемената састављених од трословног језичког идентификатора за представљање назива живих и изумрлих језичких породица и група. Језици креирани искључиво за машинску употребу, као што су програмски језици, нису укључени у овај кôд (*SRPS ISO 639-5 2018*).

Имајући у виду да су ови стандарди настали по угледу на *ISO* стандарде, природно је што је и ознака којом се домаћи стандард означава идентична ознаци у *ISO* стандарду, уз годину у којој је стандард донет.

Уз ове стандарде, Комисија задужена за област терминологије донела је и низ стандарда из области превођења:

- Услуге превођења – Захтеви за услуге превођења (*SRPS EN ISO 17100:2017*) – овим стандардом дефинишу се неопходни захтеви за испоруку квалитетних услуга превођења (*SRPS EN ISO 17100 2017*).

- Услуге превођења – Захтеви за услуге превођења – Измена 1 (*SRPS EN ISO 17100:2017/A1:2019*) – дефинише измене претходног стандарда, које се односе на квалификације преводиоца за давање услуга у складу са међународним стандардом (*SRPS EN ISO 17100:2017/A1 2019*).

- Симултано превођење – Квалитет и пренос звука и уноса слике – Захтеви (*SRPS EN ISO 20108:2019*) – овај стандард дефинише захтеве за квалитет преноса звука и уноса слике преводиоцима, и утврђује карактеристике аудио и видео-сигнала, а примењује се за превођење на лицу места, када су учесници и тумач на истом месту, као и код превођења на даљину (*SRPS EN ISO 20108 2019*).

- Симултано превођење – Опрема – Захтеви (*SRPS EN ISO 20109:2017*) – овим стандардом дефинишу се захтеви за опрему која се користи приликом симултаног превођења, а која се тиче како квалитета и преноса звука и слике који се обезбеђују преводиоцу, тако и опреме која је потребна у кабини (*SRPS EN ISO 20109 2017*).

- Симултано превођење – Сталне кабине – Захтеви (*SRPS EN ISO 2603:2017*) – овим стандардом дефинисани су захтеви и препоруке за израду и поправку сталних кабина за симултано превођење у новим и већ постојећим објектима, као и захтеви за опремом, и

квалитетом и преносом звука и слике који су преводиоцу неопходни у кабини (SRPS EN ISO 2603 2017).

- Симултано превођење – Покретне кабине – Захтеви (SRPS EN ISO 4043:2017) – дефинише захтеве и препоруке за производњу преносивих кабина за симултане преводиоце. Њихова предност у односу на уграђене кабине је у томе што могу бити расклопљене, премештене и постављене у салу за конференцију за разлику од уграђених кабина (SRPS EN ISO 4043 2017).

- Преводилачки пројекти – Опште смернице (SRPS ISO/TS 11669:2018) – Главна сврха ове спецификације је да олакша комуникацију између страна које су укључене у пројекат. Намењен је онима који потражују преводилачке услуге, онима који те услуге пружају као и онима који користе резултате пројекта – а нарочито производ превођења (SRPS ISO/TS 11669 2018).

- Примена рачунара у терминологији – Оквир за терминолошко обележавање (ISO 16642:2017). Овим стандардом утврђује се оквир за представљање података из збирки терминолошких података. Овај оквир обухвата метамодел и методе за описивање посебних терминолошких језика за обележавање изражених у XML формату. Осмишљен је тако да подржава развој и употребу рачунарских апликација за терминолошке податке и размену тих података између различитих апликација (ISO 16642 2017).

- Терминолошки рад и наука о терминологији – Вокабулар (ISO 1087:2020). У овом документу утврђују се основни термини и дефиниције за терминолошки рад и науку о терминологији, али њиме нису обухваћени термини и дефиниције који су специфични за рачунарске апликације у терминолошком раду. (ISO 1087 2020).

- Управљање терминолошким ресурсима – Спецификација категорије података (ISO 12620:2020). Овај документ одређује смернице и захтеве који регулишу спецификације категорија података за језичке ресурсе. Намењен је истраживачима и стручњацима у областима управљања језичким ресурсима који користе категорије података и спецификације категорија података (ISO 12620 2020).

- Услуге превођења – Захтеви за накнадно уређивање резултата машинског превођења (ISO 18587:2018). Овај стандард је применљив само на садржај који обрађују системи машинског превођења (ISO 18587 2018).

- Управљање терминолошким ресурсима – *TermBase eXchange (TBX)* (ISO 30042:2020). Овим документом описани су основни појмови као што су метамодел, категорије података и XML стилови (ISO 30042 2020).

У оквиру Института за стандардизацију Србије, до сада су објављени следећи речници:

- 01.040.01 Опште. Терминологија. Стандардизација. Документација
- 01.040.03 Услуге. Организација и управљање предузећима. Администрација. Транспорт. Социологија
- 01.040.07 Математика. Природне науке
- 01.040.11 Технологија заштите здравља
- 01.040.13 Животна средина. Заштита здравља. Безбедност
- 01.040.17 Метрологија и мерење. Физичке појаве
- 01.040.19 Испитивања
- 01.040.21 Машине и машински елементи за општу употребу
- 01.040.23 Пнеуматски хидраулични системи и компоненте за општу употребу
- 01.040.25 Производно инжењерство
- 01.040.27 Претварање и пренос енергије и топлоте
- 01.040.29 Електроенергетика
- 01.040.31 Електроника
- 01.040.33 Телекомуникације. Аудио и видео-техника
- 01.040.35 Информациона технологија
- 01.040.37 Технологија слике
- 01.040.39 Прецизна механика. Драгуљарство

- 01.040.43 Друмска возила
- 01.040.45 Инжењерство шинског саобраћаја
- 01.040.47 Бродоградња и бродска постројења
- 01.040.49 Ваздухопловство и космонаутика
- 01.040.53 Опрема за руковање материјалима
- 01.040.55 Паковање и дистрибуција робе
- 01.040.59 Технологија текстила и технологија коже
- 01.040.61 Индустрија одеће
- 01.040.65 Пољопривреда
- 01.040.67 Прехрамбена технологија
- 01.040.71 Хемијска технологија
- 01.040.73 Рударство и минерали
- 01.040.75 Нафта и сродне технологије
- 01.040.77 Металургија
- 01.040.79 Технологија дрвета
- 01.040.81 Индустрија стакла и индустрија керамике
- 01.040.83 Индустрија гуме и индустрија пластичних маса
- 01.040.85 Технологија папира
- 01.040.87 Индустрија боја
- 01.040.91 Грађевински материјали и високоградња
- 01.040.93 Грађевинарство (грађевински и инжењерски објекти)
- 01.040.95 Војни послови. Војно инжењерство. Оружје
- 01.040.97 Опрема за домаћинство и комерцијална опрема. Одмор и разонода. Спорт

Ознака речника се, према међународној класификацији стандарда, састоји из три нивоа. Први ниво обухвата главна подручја активности у стандардизацији и има двоцифрену ознаку. Други ниво представља њихову подгрупу и означен је троцифреном одредницом, а трећи ниво представља даљу поделу подгрупа на још уже целине. Ове одреднице одвојене су тачком (Институт за стандардизацију Србије 2018, 4).

Од посебног значаја за овај рад су речници из области електроенергетике који су у овој листи класификовани ознаком 01.040.29 (видети потпоглавље 4.2.14).

3.4.3.5. Друштво за језичке ресурсе и технологије

Друштво за језичке ресурсе и технологије развило се из Групе за језичке технологије која је функционисала у оквиру Катедре за рачунарство и информатику Математичког факултета у Београду. Примарни задатак ове Групе био је креирање и одржавање ресурса и алата за обраду српског језика.

Лексички ресурси за српски језик развијани су у оквиру Групе за језичке технологије дуги низ година. Поред корпуса српског језика, као и вишејезичних паралелних корпуса (енглеско-српског и француско-српског језика), од нарочитог су значаја систем морфолошких речника српског језика, као и семантичка мрежа речи за српски језик (српски *Wordnet*) који су развијени у оквиру међународног пројекта Балканет (Stanković, Obradović и Krstev 2009).

У оквиру Групе за језичке технологије отпочео је и рад на препознавању именованих ентитета и екстракцији информација из различитих домена, анализи сентимената, тематској класификацији, претраживању информација и машинском превођењу.

Из Групе за језичке технологије, развило се Друштво за језичке ресурсе и технологије, које је основано 2014. године. Ово Друштво је невладино и непрофитно удружење основано ради остваривања циљева у области промоције, популаризације и унапређивања свих грана језичких технологија на научном, стручном и практичном нивоу. Пружајући допринос изградњи језичких ресурса у дигиталном облику, окупља истраживаче, произвођаче и кориснике језичких технологија, а све у циљу очувања српског језика у дигиталној ери.

Неки од циљева Друштва су, креирање и реализација тематских програма и пројеката, организовање научних и других скупова, семинара, конференција, трибина, круглих столова, радионица, саветовања и других јавних догађаја и облика едукације, објављивање књига, приручника, гласила и других издања у штапаном и електронском облику, давања мишљења и изношења савета из области остваривања циљева, сарадња са образованим и истраживачким установама, удружењима, јавним органима, привредним друштвима и осталом.⁶³

⁶³ Подаци о Друштву за језичке ресурсе и технологије преузети су јула 2019. године са веб-сајта ове организације: http://jerteh.rs/?page_id=84.

4. ФОРМИРАЊЕ И РАЗВОЈ ТЕРМИНОЛОГИЈЕ ИЗ ОБЛАСТИ ЕЛЕКТРОЕНЕРГЕТИКЕ У СРПСКОМ ЈЕЗИКУ

Електроенергетика као научна дисциплина заузима једно од водећих места у савременој науци. Брз развој ове науке у последњих сто педесет година, како у свету тако и код нас, иницирао је велике промене на различитим пољима (у области индустријске производње, привреде, медицине, итд.).

Чињеница да је данас тешко замислити живот без електричне енергије указује на то у којој су мери ова наука и њене примене проникли у наше животе. Уређаји које свакодневно користимо у домаћинствима и у раду, наш живот чине недвосмислено квалитетнијим, лагоднијим и бољим. Па ипак, изузимајући оне којима је рад на електроенергетском систему свакодневица, ретко кад размишљамо о самом начину функционисања система (у смислу производње, преноса и дистрибуције електричне енергије), његовој комплексности и повезаности ове науке са другим областима.

Према Дјуијевој децималној класификацији⁶⁴ (ДДК), у којој је целокупно знање подељено на 10 главних класа, а свака класа даље на 10 главних области и тако поступно, водећи до области и подобласти различитих дисциплина, предмета и концепата, електроенергетика као наука је сврстана у *Природне науке* односно главну класу означену бројем 500.

Главне класе према Дјуијевој децималној класификацији имају следећи распоред:

000	Општа
100	Филозофија
200	Религија
300	Друштвене науке
400	Језик
500	Природне науке
600	Технологија (примењене науке)
700	Уметност
800	Књижевност
900	Географија и историја

Ова наука је даље, према ДДК, распоређена у област **530: Физика**, подобласт **537: Електрична енергија и електроника**. Надаље се разврстава на још уже области па је тако подељена на подобласти **537.5: Електроника** и **537.6: Електричне струје (електродинамика)** и термо електрична енергија (Решк 2014).

У табели 3 илустрована је класа 500 разврстана по областима и подобластима према ДДК.

Табела 3. Хијерархијска схема класе 500: *Природне науке*

500	Природне науке
530	Физика
537	Електрична енергија и електроника.
537.5	Електроника.
537.6	Електричне струје (електродинамика) и термо електрична енергија

Према универзалној децималној класификацији (УДК), којом су све области људског знања класификоване у посебна подручја односно класе,⁶⁵ електроенергетика као наука припада

⁶⁴ Дјуијева децимална класификација (енг. *Dewey Decimal Classification*) је најраспрострањенији класификациони систем на свету. Користи се у око 135 земаља и преведен је на преко 30 језика.

⁶⁵ Класе представљају научне области означене бројевима од 0–9, од којих се свака дели даље на поткласе (означене двоцифреним бројевима). Поткласе се даље разврставају на још уже целине које су означене троцифреним

различitim категоријама. Као грана електротехнике, доменски се сврстава у две различите класе односно области од укупно десет којима се у најширем смислу разврставају све научне и стручне делатности (због планираног проширења распореда класа услед брзог развоја у различitim областима знања, класа 4 остаје слободна). Разликујемо следеће класе:

Класа	
0	Наука и знање. Организација. Рачунарска наука. Информациона наука. Документалистика. Библиотекарство. Институције. Публикације.
1	Филозофија и психологија.
2	Религија. Теологија.
3	Друштвене науке.
4	–
5	Математика. Природне науке.
6	Примењене науке. Медицина. Технологија.
7	Уметност. Рекреација. Забава. Спорт.
8	Језик. Лингвистика. Књижевност.
9	Географија. Биографија. Историја.

Класа којој ова наука припада према УДК означена је бројем **6**: *Примењене науке. Медицина. Технологија*. Електронергетика је даље сврстана у поткласу **62**: *Инжењерство. Технологија опште*, поткласа **621**: *Механичко инжењерство опште. Нуклеарна технологија. Електротехника. Механизација*. Овако сужено подручје примене ове науке даље је смештено у поткласу **621.3**: *Електротехника* која се даље разврстава на **621.31**: *Производња, снабдевање и регулација електричне енергије. Електричне машине и уређаји. Електрично мерење. Примењени магнетизам. Примењена електростатика*. И ова поткласа се даље сужава на подређену поткласу означену бројем **621.311**: *Производња и снабдевање електричном енергијом. Електроенергетски објекти. Електродистрибутивни систем, опрема, конфигурација. Шеме електрификације*, која се надаље разврстава на велики број подређених поткласа међу којима су **621.311.1**: *Дистрибутивне мреже за напајање. Регионалне мреже. Области снабдевања. Трансформаторске станице. Шема електрификације*, **621.311.2**: *Електрана према врсти погона*, **621.311.4**: *Трансформаторске станице. Постројење за дистрибуцију електричне енергије*, **621.311.6**: *Напајање за електричне и електронске уређаје*, **621.311.8**: *Напајање у случају ванредног стања. Резервни генератори*. Поткласа 621.31, разврстава се и на поткласу **621.313**: *Електричне машине*, а она даље на **621.313.1**: *Генератори и мотори опште*, **621.313.2**: *Машине за једносмерну струју*, **621.313.3**: *Машине за наизменичну струју*, **621.313.5**: *Генератори и мотори под утицајем ефеката електро и магнетне индукције на флуиде*, **621.313.8**: *Електромагнетне машине. Машине са трајним магнетима*, као и на поткласу **621.314**: *Трансформација. Трансформатори. Конвертори. Исправљачи*. И ова поткласа се даље своди на велики број још ужих класа, профилишући односно сужавајући тако подручје своје примене (Universal Decimal Classification 2005, 465–535).

Класи 621.3: *Електротехника* подређена је и поткласа 621.32: *Електрични извори светлости*, затим 621.33: *Електрична вуча, локомотива, погон*, 621.35: *Електрохемијска технологија*, 621.36: *Термоелектрични претварачи. Електротермија*, 621.37: *Техника електричних и електромагнетних таласа. Осцилације. Импулси*, 621.38: *Електроника. Електронски уређаји. Електронске цеви. Фотоћелије. Акцелератори честица. Рендгенске цеви*, 621.39: *Телекомуникације. Телеграфија. Телефонија. Радиокомуникација. Видео- технологија и опрема. Даљинско управљање* (Universal Decimal Classification 2005, 540–546).

Свака од ових подобласти, иако припада подручју електротехнике, представља посебну научно-стручну делатност, јасно устројену у погледу својих задатака, циљева, техничких и

бројевима, и могу се разврставати даље на подређене класе, које се означавају бројевима до највише 12 цифара (Universal Decimal Classification 2005, XVI–XVII).

научних решења. Па ипак, јасну границу међу овим областима тешко је повући, имајући у виду њихову повезаност, међусобну зависност и испреплетеност. Тако нпр. електрична енергија се у електроенергетском систему у великом броју термоелектрана производи у нуклеарним реакторима.

Још једна класа којој област електроенергетике припада према УДК означена је бројем **5: Математика и природне науке**, поткласа **53: Физика**, **537: Електрична енергија. Магнетизам. Електромагнетизам**. Област 537 се даље сужава на **537.2: Статички електрицитет. Електростатика**, **537.3: Електрицитет. Електрична струја. Електрокинетика**, **537.5: Феномени електрона и јона** и **537.6/8: Магнетизам. Електромагнетизам**.

И док електроенергетика као област доменски припада поткласи означеној бројем 537.3, поткласе 537.5 и 537.6/8 могу се назвати њеним граничним подручјима, имајући у виду да су појаве у електроенергетском систему које омогућавају проток електричне струје незамисливе без дејства електрона и јона, као и електромагнетних појава.

Примерима приказаним у табелама 4 и 5 хијерархијски је илустрован начин разврставања подређених класа, у нашим случајевима класа 6 и 5, према УДК.

Табела 4. Хијерархијска схема класе 6: *Примењене науке. Медицина. Технологија*

6	Примењене науке. Медицина. Технологија
62	Инжењерство. Технологија – опште
621	Машинство – опште. Нуклеарна технологија. Електротехника. Механизација
621.3	Електротехника
621.31	Производња, снабдевање и регулација електричне енергије. Електричне машине и уређаји. Електрично мерење. Примењени магнетизам. Примењена електростатика.
621.311	Производња и снабдевање електричном енергијом. Електроенергетски објекти. Електродистрибутивни систем, опрема, конфигурација. Шеме електрификације.
621.311.4	Трансформаторске станице. Постројење за дистрибуцију електричне енергије.

Табела 5. Хијерархијска схема класе 5: *Математика и природне науке*

5	Математика и природне науке
53	Физика
537	Електрична енергија. Магнетизам. Електромагнетизам.
537.3	Електрицитет. Електрична струја. Електрокинетика.

Институт инжењера електротехнике и електронике (енг. *Institute of Electrical and Electronics Engineers*) је највеће техничко професионално удружење на свету чији су циљеви дефинисани као образовни и технички, усмерени ка напретку електротехничког и електронског инжењерства, електроенергетике и енергетике, телекомуникација, рачунарског инжењерства, нанотехнологије и многих других сродних дисциплина.

У оквиру Института инжењера електротехнике и електронике постоји 39 техничких друштава (енг. *societies*) која се баве разним техничким областима и у оквиру којих се објављују специјализовани научни часописи, организују конференције као и пословно умрежавање њихових чланова.

У оквиру овог Института развијају се и стандарди из разних техничких дисциплина у чијем доношењу учествују експерти из целог света. Иако формално не представља тело које је

овластила било која влада, ово удружење сарађује са организацијама из целог света у циљу што веће ефикасности као и препознатљивости стандарда у оквиру глобалне заједнице.

Међу бројним удружењима која су саставни део Института инжењера електротехнике и електронике, друштва „Енергетска електроника“ (енг. *Power Electronics Society*), „Енергетика“ (енг. *Power & Energy Society*), „Индустријска електроника“ (енг. *Industrial Electronics Society*), „Примена у индустрији“ (енг. *Industry Applications Society*), изучавају област електроенергетике. Свако од ових удружења објављује неколико различитих научних часописа из предметног домена.

Данас је, више него раније, јасно да су области које електроенергетика као научна дисциплина дотиче давно прошириле њене оквире, па је стога не можемо проучавати и практиковати без разумевања основних појмова из области заштите животне средине, обновљивих извора енергије, водопривреде, рударства, машинства и машиноградње, технологије и многих других, а нарочито рачунарских наука.

Буран развој једне дисциплине, каква је и електроенергетика, сам по себи подразумева именовање новонасталих појмова, што неизоставно утиче и на раст термилошког фонда из истоимене области. Проблем може да се јави када се услед неодложне потребе за именовањем нових појмова, терминологија такве области развије неплански, без утврђеног система и прецизних правила. Да би се таква ситуација предупредила, организују се посебна тела која на систематски утврђен начин приступају раду на терминологији, с циљем да успоставе ред у одређеној области односно отклоне препреке у комуникацији међу стручњацима, о чему је већ било речи у раду.

По својој дефиницији електроенергетика представља област електротехнике која се бави производњом, преносом и дистрибуцијом електричне енергије од произвођача до крајњег корисника тј. потрошача. Ова дисциплина се бави и проучавањем уређаја који су прикључени на електроенергетски систем, какви су трансформатори, генератори, електромотори и уређаји енергетске електронике.

Према ауторима Леви и Бекут (Levi и Bekut 1997, 1–2), данас се целокупна електроенергетика може поделити на две основне области: 1. Електроенергетске системе, и 2. Електромеханичке претвараче (односно, електричне машине и енергетску електронику). Свака од ових области је, сама за себе, прилично комплексна, и захтева даљу поделу и анализу.

Узимајући у обзир комплексност коју прорачуни електроенергетског система носе за собом, улога рачунара у електроенергетици данас је неизбежна.

Анализа електроенергетских система обухвата прорачуне односно анализу стационарних и анализу динамичких режима рада. У оба случаја подразумевана је употреба рачунарских симулационих модела система, с тим што се у првом случају употребљавају модели засновани на алгебарским једначинама, док се у другом случају користе модели засновани на диференцијалним једначинама у комбинацији са моделима заснованим на алгебарским једначинама.

Рачунарски симулациони модели система у електроенергетици су датотеке у које су уписани сви параметри елемената система који су неопходни за спровођење одређених врста системских прорачуна као што су прорачуни токова снага и напонских прилика, струја кратког споја, транзијентне стабилности итд. Ове датотеке се учитавају и прорачуни раде посебним рачунарским програмима који у себи имају имплементирани алгоритме за системске прорачуне.

Анализа стационарних процеса некада се може и проширити оптимизационим методама, па се тако проширени модели користе у оквиру квазистационарног управљања, планирања погона, као и планирања развоја електроенергетског система. С друге стране, динамички модели и одговарајућа анализа представљају основу за прорачун стабилности система, као и за синтезу појединих врста управљања (регулације) њиме.

Насупрот области електроенергетских система, у области електромеханичких претварача, анализа стационарних појава је знатно мање заступљена и она се углавном среће у фази пројектовања електромоторног погона. Данас је у овој области основни интерес усмерен ка

анализи диманичких процеса, из које затим произилазе активности као што су анализа стабилности и синтеза управљања (регулације) целокупним електромоторним погоном (Levi и Bekut 1997, 1–2).

Узимајући у обзир набројане активности и обим посла који се данас спроводе у оквиру сектора електроенергетике, јасно је да је у савременим условима за пословање у овој струци неопходан велики број математичких метода. Имајући у виду да је реч о сложеним моделима тј. моделима велике димензионалности, они се решавају искључиво применом рачунара. Стога је познавање рачунарски оријентисаних математичких метода из области нумеричке анализе, управљања и стабилности, као и оптимизације, преко потребно у бављењу овом струком (Levi и Bekut 1997, 1–2). Самим тим ни терминологију ове струке не можемо прецизно омеђити појмовима који припадају само и искључиво једном домену, јер њихова повезаност са другим предметним областима у великој мери помера утврђене и дефинисане границе.

4.1. Кратак приказ развоја електроенергетике на нашим просторима кроз науку и праксу

Практична и научна заступљеност електроенергетике као дисциплине на нашим просторима има дугачку и богату традицију. Њен неизоставни део чине и појединци који су својим залагањем, оваквом развоју, у великој мери допринели.

Када се говори о почетку развоја електроенергетике на нашим просторима, као најважнија година истиче се 1893. Тада је у Србији, тачније у Београду на Дорћолу, почела да ради прва јавна електрана – термоелектрана. Ова година представља почетак електрификације у Србији. Дотадашња београдска јавна расвета, први пут уведена још давне 1865. године, када су најпре два фењера на гас обасјавала теразијске зграде, а касније и више стотина њих, била је замењена електричним осветљењем. За реализацију овог пројекта, најзаслужнији је био чувени српски научник, професор физике и ректор Београдског универзитета Ђорђе Станојевић, који је уједно био и конструктор термоелектране на Дорћолу.

Годину дана након пуштања у погон, ова мрежа је омогућила покретање и првог београдског трамваја на електрични погон (тзв. „варошке железнице“).

Након београдске термоелектране, у Србији су почели да се нижу електроенергетски објекти. Најпре је Матеја Ненадовић, унук проте Матеје Ненадовића, изградио хидроцентралу на месту своје воденице, у Ваљевоу, на реци Градац. Хидроцентрала је завршена и пуштена у погон 1899. године (Dimitrijević 2009).

Већ наредне, 1900. године, саграђена је и хидроелектрана на Ђетињи, прва електрана у Србији која је саграђена према Теслином полифазном систему наизменичних струја. Конструисао ју је сам Станојевић, иначе велики пријатељ и поштовалац Николе Тесле, а интересантно је да је пуштена у погон само четири године након изградње прве такве централе на Нијагариним водопадима. Уз ову хидроелектрану направљени су и далековод у дужини од 1 km и седам трансформаторских станица (Dimitrijević 2009). Струја коју је ова електрана произвођила покретала је машине Ткачке радионице и осветљавала ужичке улице. Електрана ради и данас.

Радови на изградњи хидроелектране у Вучју код Лесковца, завршени су 1904. године, када је и Лесковац добио електрично осветљење. Дужина далековода износила је око 17 km. Ова електрана је у погону већ више од сто година, а фебруара 2005. године, Извршни комитет највећег светског удружења инжењера из области електротехнике, електронике, телекомуникација и сродних области *IEEE* одлучио је да се ова хидроелектрана укључи у листу објеката, проналазака и достигнућа од општег значаја за развој и историју електротехнике у свету (Dimitrijević 2009).

Хидроелектрана „Света Петка“ почела је да ради 1908. године. Ова хидроелектрана се налази у Сићевачкој клисури на реци Нишави, а име носи према оближњој цркви и манастиру. Свечаном пуштању у погон присуствовале су највише државне, политичке и јавне личности. Била је то највећа и најскупља електрана тог доба у Србији, а ради и данас.

Године 1909. са радом почињу и две хидроцентрале у Гамзиградској бањи. Електрична енергија је у почетку била коришћена за потребе Фабрике зејтина и млинова, а у вечерњим сатима за осветљавање улица, да би убрзо потом почели да је користе имућнији и угледни грађани Зајечара и ова електрана је и данас у погону.

Станојевић учествује и у подухватима изградње термоцентрале у Чачку и хидроцентрале у Нишу на Нишави, Великом Градишту на Пеку, Власотинцу на Власини, Ивањици на Моравици, а разматрао је и могућности изградње хидроцентрале на Ђердапу (Dimitrijević 2009).

Први површински коп у тзв. производњи (односно откопу) угља у Србији отворен је у Колубари 1952. И дан-данас, рударски сектор у „Електропривреди Србије“ је основна карика у функционисању електроенергетског система. Око 70 одсто произведене електричне енергије у Србији произведе се у термоелектранама, из угља са површинских копова РБ „Колубара“ и „ТЕ-КО Костолац“. Уз предузете активности за повећање ефикасности, увођење савремених технологија, кроз бољу организацију и унапређење мера заштите животне средине, угаљ ће и у будућности остати основни базни енергент Србије за производњу електричне енергије.

„Хидроелектрана Бајина Башта“ у Перућцу, највећи је хидроенергетски пројекат саграђен на реци Дрини, а пуштена је у погон 1966. године.

Највећа термоелектрана на Балкану и уједно највећи појединачни произвођач електричне енергије у електроенергетском систему Србије, ТЕ „Никола Тесла А“, пуштена је у погон 1970. године.

Хидроенергетски и пловидбени систем (скр. ХЕПС) „Ђердап 1“ пуштен је у погон 1972. године. Овај систем, који је пројектован и изграђен заједнички са Румунијом, представља највећу хидротехничку грађевину на Дунаву и највећег произвођача хидроенергије у Југоисточној Европи.

Ради препумпавања воде из Лисинског у Власинско језеро, 1978. године је изграђено пумпно акумулационо постројење (скр. ПАП) „Лисина“, чиме је омогућена већа производња „Власинских хидроелектрана“.

Хидроелектрана „Увац“ пуштена је у погон 1979. године, а 1982. године подигнута је реверзибилна хидроелектрана „Бајина Башта“.

Две највеће електроенергетске јединице у Србији почеле су да раде у оквиру ТЕ „Никола Тесла Б“, 1983. године, док је друга заједничка српско-румунска хидроелектрана на Дунаву – ХЕ „Ђердап 2“, пуштена у погон 1985. године.

Површински коп „Дрмно“ у Костолцу почео је да ради 1987. године, а хидроелектрана „Пирот“ изграђена је 1990. године.

У НАТО бомбардовању 1999. године електроенергетска постројења у Србији су претрпела велика оштећења која су након тога сукцесивно обнављана. Оштећења постројења је било и у Другом светском рату, међутим, четрдесетих година 20. века електроенергетски систем још увек није био јединствена целина, већ међусобно раздвојен по различитим локалитетима па се сходно томе ниво нанете штете не може мерити са штетом насталом 1999. године.

Све ово говори да је развој електроенергетике на нашим просторима у последњих 150 година ишао узлазном путањом, па чак и у ратним годинама када су електроенергетски објекти били потпуно уништени односно онеспособљени за даљи рад. Значај електроенергетике на време је препознат и није се правила велика пауза нити штедели људски и материјални ресурси приликом изградње свих тих објеката. Њихова изградња текла је плански и са одређеним циљем.⁶⁶

⁶⁶ Све информације о историји развоја електроенергетике на нашим просторима преузете су 8. августа 2019. године, са званичног сајта ЈП ЕПС: <http://www.eps.rs/cir/Pages/Istorija.aspx> и из електронског часописа *Планета* доступно на адреси: <http://www.planeta.rs/33/6%20portreti.htm>.

4.2. Речници електротехнике

У одељку који следи изложен је хронолошки преглед и анализа речника електротехничке струке, који су до данас објављени на простору Србије и бивше Југославије.

Преглед је рађен према Кобису, јединственом библиотечко–информационом систему са узајамном каталогизацијом преко којег је могуће претражити публикације свих библиотека које су у његовом систему, односно према личном књижном фонду аутора ове дисертације. Уколико неки од речника није уврштен у изложени попис, разлог лежи у чињеници да се до податка о њему на овај начин није могло доћи. Сваки од описаних речника представљен је као посебно потпоглавље чији је наслов сам назив речника.

Примери одредница из речника представљени су онако како се појављују у наведеним издањима речника.

4.2.1. Мали српско-француско-немачки електротехнички речник са скраћеницама и практичним забелешкама

Први стручни електротехнички речник, под називом Мали српско-француско-немачки речник, аутора Мирка Добр. Мирковића, објављен је 1927. године, у Београду, у издању Графичког института „Народна мисао“. Речник је замишљен као прва етапа у развоју наше народне терминологије односно великог техничког речника којег је тек требало саставити, као што и сам аутор наводи у Предговору.

Речник је подељен на три дела и има више од 600 одредница. Први део је Српско-француско-немачки речник. Карактеришу га прилично опсежне дефиниције појмова – одредница, као што је приказано примерима који следе (Мирковић 1927, 9–55).

Ампер замотаји m. pl. је производ броја замотаја и интензитета струје која струји кроз ове замотаје.

Ampère-tours m. pl. – Amperewindungen f. pl.

Асинхрони генератор m. по конструкцији исто је што и асинхрони мотор: сам се не буди дакле, може да ради само напајан са извора одакле узима струју потребну за стварање окретног поља. *generatrice asynchrone f. – Asynchronegenerator m.*

Асинхрони мотор m. (неистодобни – индуктни) је мотор за трофазну струју код кога ротор не врши исти број обртаја као окретно поље статора. Од синхроног генератора и мотора разликује се у томе што се не буди помоћу једносмислене струје. Магнетско поље се ствара трофазном струјом. Асинхрони мотори имају брзину 0,96 до 0,98 од синхроне брзине. Вишефазни мотори (окретно поље) крећу сами, док је за покрет једнофазних мотора потребно нарочито уређење. Асинхрони мотор при празном ходу одржава скоро синхрони број обртаја. При оптерећењу опада број обртаја а бива при пуном оптерећењу за неколико процената мањи од синхроног. *moteur m. asynchrone – Asynchrone-motor m.*

Други део је Француско-српски речник. У овом делу, појам на француском језику углавном има један преводни еквивалент на српском језику. У случајевима када преводних еквивалената на српском језику има више од једног, (тј. када се јављају синоними), они се раздвајају цртом, као у примерима који следе (Мирковић 1927, 57–70):

alimenter — напајати, обскрбити струјом

alternateur m. — алтернатор — наизменичник m.

amortissement m. — амортизација f. — умртвљивање n.

Трећи део је Немачко-српски речник, а принцип је идентичан као код другог дела, што видимо у следећим примерима (Мирковић 1927, 75–88):

Anker (einer Dynamomaschine) m. — котва, индуктор. m. Ankerbewicklung f. — замотај котве, индуктора. m.
--

Скраћенице и практичне забелешке налазе се пре почетка првог дела. Што се тиче граматичких обележја, поред сваке речи обележен је род именица (m, f, n), као и pl. уколико је у питању множина именице.

4.2.2. Њемачко-хрватски и хрватско-њемачки електротехнички рјечник

Овај речник (у оригиналу: *Njemačko-hrvatski i hrvatsko-njemački elektrotehnički rječnik*, нем. *Deutsch-kroatisches und kroatisch-deutsches elektrotechnisches Wörterbuch*), аутора Влатка Дабца, објављен је у Загребу, у издању „Школске књиге“, 1952. године, на 356. страна. У његовом састављању учествовао је „Завод за високи напон“ Техничког факултета Свеучилишта у Загребу, заједно са осталим заводима тог факултета, електроиндустријом и електропривредом. Одлуком Вијећа старјешина на Техничком факултету Свеучилишта у Загребу, хрватска електротехничка терминологија из овог речника употребљавана је као јединствена службена електротехничка терминологија у настави овог факултета. Уз то, његова основна сврха била је, као што је и предочено Предговором, да помогне хрватским стручњацима приликом коришћења опширне немачке електротехничке литературе.

Речник је двојезични, двосмерни, има 7.000 одредница, а појмови су поређани абecedним редом, најпре на немачком а потом на хрватском језику (треба имати у виду да је званични језик тада био српско-хрватски). Писмо које се користи је латиница.

Преводни еквиваленти који имају исто значење (синоними), одвајани су зарезима, а на првом месту увек је израз којем аутор даје предност у односу на други. Изрази различитог значења одвајени су тачком и зарезом, као у примерима који следе: нпр. за *Anker* синоними су *арматура* и *котва*, а *ротор* и *статор* су два различита значења. Исти поступак примењен је и на први и на други део речника (Dabac 1952, 6, 65):

Abbremsung f (des Läufers) / коћенје n (rotora)
--

Anker m 10-05-175 [Elm] / <i>арматура</i> f, <i>котва</i> f; <i>ротор</i> m, <i>статор</i> m [inducirani dio]
--

Meldetafel f 35-25-030 / <i>дојавна плоча</i> f, <i>signalna ploča</i> f

Грађа у речнику је подељена на групе, а потом на поглавља, у којима су називи означени бројевима као што је приказано претходним примерима. Прва два броја означавају групу којој термин припада, друга два означавају поглавље, а трећа три сам термин.

У округле заграде смештане су речи које употпуњују израз, док су у угласте стављане речи и скраћенице које тај израз тумаче.

У речнику се јавља и ознака уже стручне области којој термин припада, у оба дела речника, а које су предефинисане у одељку *Hrvatske kratice i znakovi*, међутим, то није чест случај.

Речник је повезан с Међународним електротехничким речником, тако што се код основних појмова (одредница), налази ознака идентичног појма из Међународног електротехничком речника. Повезано је око 800 појмова из овог речника с дефиницијама у речнику *IEC*.

4.2.3. Словеначки електротехнички речник

У периоду од 1957. до 1970. године изашло је десет томова *Slovenskog elektrotehniškog slovara*, у издању Електротехничког друштва Словеније, са седиштем у Љубљани. Сваки од ових десет томова обухвата по једну стручну подобласт електротехнике. Као основ за израду овог речника послужио је Међународни електротехнички речник, који је израдила Међународна електротехничка комисија, како сазнајемо из Предговора првог тома. Одреднице су распоређене по угледу на Међународни електротехнички речник, а из истог речника су преузимани преводи на француски, енглески, немачки и италијански језик. Што се тиче термина на српскохрватском

језику, они су преузети из Електротехничког речника чији је аутор Влатко Дабац, представљен у претходном пододељку, уколико су били доступни. Уколико недостају, на њиховом месту би стајала цртица (–). Међутим, интересантно је подсетити се да Електротехнички речник аутора Влатка Дапца, нигде у својој садржини не спомиње, у то време званични, српскохрватски језик, већ искључиво хрватски.

На крају сваког тома речника дат је попис словеначких, српскохрватских, француских, енглеских, немачких и италијанских термина.

Речник је енциклопедијског карактера и има више од 3.000 одредница. Одреднице на словеначком језику прати дефиниција појма на истом језику, а потом и преводи на (српско)хрватски, француски, енглески, немачки и италијански, као у примеру који следи (Млакар 1957–1970, 5):

25-05-015	transformácija eléktrične energije:	transformacija električne energije
	Spreminjanje električne energije brez spremembe	transformation d'énergie électrique
	frekvence.	transformation of electrical energy
		Transformierung elektrischer Energie,
		Umspannung elektrischer Energie
		trasformazione di energia elettrica

На крају речника налази се абецедни регистар израза на свих шест заступљених језика.

4.2.4. Електротехнички термилошки речник, група 07, електроника

Речник под називом *Elektrotehnički terminološki rečnik, grupa 07, elektronika* објављен је 1965. године у Београду, у издању Техничке књиге, на 221 страни. У импресуму речника, назначено је да је материјал припремила Комисија за машинску и електротехничку терминологију Савеза машинских и електротехничких инжењера и техничара Југославије, а да су на изради термина учествовали Југословенски електротехнички комитет, републичке комисије за техничку терминологију при савезима електротехничких инжењера и техничара СР Хрватске и СР Словеније и при Савезу машинских и електротехничких инжењера и техничара СР Македоније. Ангажовање најистакнутијих стручњака из привреде, са универзитета и осталих научних установа са српскохрватског, словеначког и македонског језичког подручја, указује на општи југословенски карактер речника.

Као основ за израду овог речника, а у циљу стварања јединственог српскохрватског термилошког речника из области електротехнике, Комисија за машинску и електротехничку терминологију је усвојила већ поменути Међународни електротехнички речник који је израдила Међународна електротехничка комисија. Више о српском издању овог речника у поглављу 4.2.11.

Приликом израде дефинитивних српскохрватских термина, Комисија је предлоге стручњака из Србије усаглашавала са предлозима стручњака из Хрватске како би се добио јединствен термин. Тако где то није било могуће, прихватана су два потпуно равноправна термина.

Термини на словеначком језику преузети су из речника „*Slovenski elektrotehniški slovar – Elektronika*“ који је израдила Термилошка комисија Електротехничког друштва Словеније, о којем је већ било речи. Термине на македонском језику израдила је републичка комисија за техничку терминологију при Савезу машинских и електротехничких инжењера и техничара СР Македоније.

Речник је вишејезични, има више од 800 одредница, а у њему је заступљено седам језика, према следећем редоследу: српскохрватски, словеначки, македонски, руски, енглески, француски, немачки.

Подељен је у 27 група, и то: *Општи термини, Састав материје, Електрони у материји, Радиоактивна и космичка зрачења, Наелектрисане (набите) честице, Луминесценција, Електризација и побуђивање гаса (плина), Јонизација (ионизација) гаса (плина), Пражњење (избој) у гасу (плину), Луминесцентно пражњење (избој), Проводници (водичи), полупроводници (полуводичи), изолатори, Запречни (запорни) слој, Електронска емисија, Термоелектронски ефект, Секундарна емисија, Фотоелектрични ефект, Општа класификација електронских цеви (електронки), Саставни делови електронских цеви (електронки), Радни услови електронских цеви (електронки), Карактеристичне величине електронских цеви (електронки) у електричним колима (круговима), Магнетрони, Електронске цеви (електронке) са електронским млазом, Фотоелектронске цеви (фотоелектронке) и фотоелектричне ћелије, Гасне (плинске) електронске цеви (електронке), Електронска оптика, Полупроводничке (полупроводичне) направе, Разне електронске направе.*

Уз сваку одредницу на српскохрватском језику, дата је дефиниција појма на левој страни, док су са десне стране наведени преводни еквиваленти на сваком од присутних језика као у примеру који следи (Savez mašinskih i elektrotehničkih inženjera i tehničara Jugoslavije 1965, 19).

07-08-055	
fazni ugao preleta:	fazni kot preleta.
fazni kut preleta:	фазен агол на прелетот. фазовый угол пробега. transit phase angle.
Proizvod vremena preleta i pulzacije (kružne frekvencije) u slučaju kada je struja, stvorena protokom naelektrisanih čestica, sinusna.	angle (de phase) de parcours, angle de transit. Laufwinkel, Laufzeitwinkel

У речнику је за српскохрватски језик заступљено латинично писмо и екавски изговор. На крају речника наведени су индекси термина на сваком од заступљених језика.

4.2.5. Електротехнички речник: руско-српскохрватски, српскохрватско-руски

Овај речник (у оригиналу: *Elektrotehnički rečnik: rusko-srpskohrvatski, srpskohrvatsko-ruski*), аутора Евгеније Илић, изашао је 1969. године у издању Техничке књиге из Загреба, на 388 страна. Речник је двојезични и двосмерни. Писмо заступљено у речнику је латинично, а изговор је екавски.

Као што је наведено у уводу, овај речник садржи преко 10.000 појмова из области електротехнике на руском језику, и исто толико на српскохрватском, а обухвата следеће области: електричне водове, инсталације и освету, расклопна постројења, електричне машине, трансформаторе, типове заштите и релеја, мерне инструменте и бројила, основне појмове из математике, електронике, телефоније, сигнализације и бродске електричне опреме, неке појмове из машинства, типове нареза, лежаја, спојки и кочница, као и најважније јединице мере.

На самом почетку наведене су Основне скраћенице, које се често срећу у руској техничкој литератури, као нпр. (Илић 1969, VII).

ABP (автоматический включатель резервного питания) / uređaj za automatsko uključenje rezervnog paraјanja
BK (включено) / uključeno

Што се тиче граматичких обележја, поред сваке речи обележен је род именица (*м, ж, с*) само за руски језик у првом делу речника (руско-српскохрватском). У другом делу речника (српскохрватско-руском) нема граматичких обележја, већ само скраћенице (које су дате на самом почетку, након *Основних скраћеница*) у загради и које указују на то којој ужој области електротехнике одредница припада, као нпр (Илић 1969, 2).

Поред обимне удбеничке литературе, како на руском тако и на српскохрватском језику, која је аутору служила као помоћ приликом састављања овог речника, у раду на њему аутор се, између осталог, ослањао и на *Електротехнички рјечник њемачко-хрватски и хрватско-њемачки*, аутора Влатка Дабца, из 1952. године (о чему је било речи у потпоглављу 4.2.2), као и *Енглеско-српскохрватски технички речник – електроника*, у издању предузећа „Рад“, из Београда, из 1964. године. *Енглеско-српскохрватски технички речник – електроника* није обухваћен овим истраживањем, имајући у виду да се његова садржина не односи непосредно на подручје електроенергетике.

4.2.6. Електротехнички речник: немачко-српскохрватски

Овај речник, аутора Наде С. Арсенијевић, изашао је у Београду у издању Београдског издавачко-графичког завода, 1971. године, на 150 страна. Речник је двојезични, једносмерни; изворни језик је немачки, док је циљни језик српскохрватски. Садржи више од 10.000 одредница. Писмо заступљено у речнику је латиница, док је изговор екавски. Одреднице су поређане абecedним редом, а синоними који се јављају код преводних еквивалената одвајани су зарезима. Граматичка обележја (рода и множине именица) наведена су уз одредницу на немачком језику (*m, f, n, pl*).

Након кратког предговора, наилази се на попис скраћеница: *Skraćenice osnovnih jedinica — Kurzzeichen der Grundeinheiten*, након чега следи сам речник. Након речника, на три стране на крају књиге, налазе се *Основне мерне јединице по азбучном реду — Grundmaßeinheiten in alphabetischer ordnung* (Arsenijević 1971).

4.2.7. Речник индустријске електротехнике: немачко-српскохрватски

Исте године, 1971. појављује се и Речник индустријске електротехнике: немачко-српскохрватски, аутора Ивана Драговића, Милана Павићевића и Петра Вујачића, у издању Интерпреса, Београд. Од укупно десет свезака које чине целину немачко-српскохрватског индустријско-техничког речника, томовима 9 и 9/1, на 423 односно 448 страна, означен је Речник индустријске електротехнике. Овај речник (обе тома) садржи више од 70.000 термилошких јединица, и подељен је у два тома. Први том обухвата термине од *A* до *M*, други том термине од *N* до *Z*. Речник је двојезични, једносмерни. Одреднице на немачком језику су поређане абecedним редом, а када су у питању граматичка обележја, на оба језика је наведено обележје рода (*m, f, n*). Знак тилда (~) користи се уместо одреднице (штампане масним словима), како се иста реч не би понављала и тиме речник учинио мање прегледним. У речнику се користи латинично писмо и екавски изговор. Уколико одредница има више значења, та значења су одвојена знаком тачка-зарез (;), док су синоними одвајани зарезима, као у примеру који следи (Dragović, Pavićević и Vujačić 1971, 7):

Abbild *n* слика *f*; прецрт *m*, копија *f*, модел *m*, подударни приказ *m*
Abbildung *f* слика *f*; преликавање *n*, прецрт *m*, прецртavanje *n*
 ~, elektronenmikroskopische — слика на електронском микроскопу
 ~, elektronenoptische — слика добијена помоћу електронске оптике
 ~, konforme — конформно преликавање

Друго издање овог речника, истих аутора, појављује се 1987. године, у издању Привредног прегледа из Београда, на 864. стране. У Предговору овог издања, истиче се да речник обухвата око 70.000 термилошких јединица из области опште електротехнике, електричних система, мрежа, електричних машина и уређаја, електроопреме, аутоматике, релејне заштите и телемеханике, као и из области мерне технике и електротехничких материјала. Речник је израђен

уз коришћење оригиналне техничке и научне документације. Садржина овог издања иста је као и претходног, осим што су овим издањем направљене измене и исправке погрешно одштампаних термина из претходног издања (Dragović, Pavićević и Vujačić 1987).

Треће издање овог речника објављено је у Београду, 1990. године, поново у издању Привредног прегледа, на 864. стране. И овим издањем, као и претходним, исправљене су граматичке, правописне и штампарске грешке.

4.2.8. Енглеско–српскохрватски електротехнички речник

Овај речник, аутора Јелице Марковић, излази 1986. године у Београду, у издању Југословенског центра за техничку и научну документацију, на 490 страна.

Речник је двојезични, једносмерни. Обухвата мање од 25.000 одредница. Изворни језик је енглески, циљни језик је српскохрватски. Поред стручних речи и израза, овај речник садржи и речи из помоћних наука, као што су математика и физика, као и основне структуралне речи, како то аутор наводи у предговору. Сам речник, намењен је преваходно нашим инжењерима и студентима електротехнике који, према речима аутора, нису савладали енглески језик у довољној мери (Marković 1986). Оно што га издваја од већине техничких речника јесте обележени фонетски изговор речи. Синоними су одвајани зарезима, док су појашњења, уколико их има, смештана у заграда. Граматичка обележја нису навођена. У речнику је заступљено латинично писмо и екавски изговор.

4.2.9. Дефиниције појмова из електротехнике на основу ИЕС (СЕИ): Појмовник – речник електротехнике

Овај речник, аутора Миљана М. Рашовића, изашао је 1991. године, у издању издавачке куће Сфаирос из Београда. Речник је вишејезични: термини су обрађени на српском, руском, француском, енглеском и немачком језику. Изворни језик је српски (а не, у то време још увек званични: српскохрватски), остало су циљни језици. Обухвата више од 500 одредница. За израду овог речника послужио је Међународни електротехнички речник који је израдила Међународна електротехничка комисија, као што се и наводи у уводној речи. Дефиниције и терминологија које се овде појављују договорене су на ширем међународном нивоу у оквиру Међународне електротехничке комисије, а сам речник намењен је, како уском кругу корисника и стручњака, тако и онима који се са овом облашћу маргинално сусрећу (Rašović 1991, 3–6). У Речнику је заступљено латинично писмо, изговор је екавски, а граматичка обележја нису навођена.

Речник је подељен на две целине. У првој целини обрађено је тринаест група: *Neki matematički pojmovi primenjeni u elektrotehnici (iz teorije polja)*, *Neki pojmovi u vezi sa periodičnim veličinama*, *Neki pojmovi iz oblasti elektromagnetnih talasa*, *Pojmovi iz strukture materije*, *Elektrostatika*, *Elektrokinetika (o električnoj struji)*, *Električno pražnjenje kroz gasove*, *Magnetizam*, *Elektromagnetika i elektrodinamika*, *Sistem jedinica*, *Neki karakteristični izrazi*, *Izrazi koji se odnose na uslove rada mašina i uređaja*, *Osnovni uređaji i elementi*. На крају првог дела налази се индекс појмова у абecedном редоследу, на свим заступљеним језицима у речнику. Ознака која се налази уз термин указује на место термина у речнику.

Други део садржи Додатке: *Dodatak A – O nekim osnovnim električnim uređajima koji služe za pretvaranje, korišćenje i proizvodnju električne energije*, *Dodatak B – Neki specifični pojmovi*, и *Dodatak C – O magnetizmu*, а садржи и регистре на заступљеним језицима, који су у вези са овим додацима.

У другом делу налазе се и Прилози који се односе на *Константе*, *Мерне јединице*, *Величине из електрицитета и магнетизма*, *Примере релација у различитим системима једначина* и *Светлосне величине*.

Приликом склапања дефиниције, као што и сам аутор наводи у Предговору, тежило се где год је то било могуће да се избегну сложене формулације, а да се постигне суштина, јасноћа и

једноставност изражавања. На тај начин је њихова примена омогућена ширем аудиторијуму, а не само стручњацима из области електротехнике.

Када је реч о избору појмова који су обухваћени овим речником, занимљива је напомена аутора на крају уводног дела, којом се констатује да су приликом избора појмова узети или општи и фундаментални појмови који се често срећу, или пак неки специфични изрази који се не проналазе лако.

Оно што посебно карактерише овај речник јесу дефиниције појмова на српском језику уз које се, на појединим местима наводе формуле, графикони и цртежи, као у примеру која следи (Rašović 1991, 64):

05-40-025

Srpskohrvatski	– Inpedansa
Francuski	– Impédance
Engleski	– Impedance
Nemački	– Scheinwiderstand, Impedanz

KOLIČNIK IZ NAPONA NA KRAJEVIMA KOLO I JAČINE STRUJE KROZ KOLO, KOD SINUSOIDALNE STRUJE, ODNOS EFEKTIVNIH VREDNOSTI.

$$Z = \frac{U}{I}$$

kod redne veze je

$$|Z| = \sqrt{R^2 + (L\omega - \frac{1}{C\omega})^2}$$

$$\operatorname{tg} \varphi = \frac{X_L - X_C}{R}$$

4.2.10. Енглеско-српскохрватски технички речник са изговором

Овај речник аутора Јелице Марковић, објављен је 1992. године у издању Техничке књиге из Београда, на 480 страна, а садржи више од 30.000 појмова. Области које обухвата су: *електротехника, електроника, информатика, физика, акустика, радио-техника, телевизија, енергетика, аутоматика, телекомуникације и електрична мерења.*

И овај речник, као и Енглеско – српскохрватски електротехнички речник, из 1986. године, истог аутора, намењен је пре свега нашим инжењерима и студентима електротехнике којима енглески језик није близак у довољној мери. Стога се у њему и практикује обележавање фонетског изговора речи, као и у претходно поменутом речнику.

У предговору речника, ауторка такође изражава забринутост због чињенице да код нас влада велика разноврсност и неуједначеност у погледу превода стручних речи и израза, упркос напорима Савезног завода за стандардизацију да стандардизује терминологију, наводећи да се многе стране речи практично и не преводе, већ само прилагођавају нашем језику. У том смислу, овај речник представља синтезу свих претходних речника, према њеним речима, а и сугестију стручњака како одређене термине превести (Marković 1992).

Овај речник је двојезични, једносмерни. Изворни језик је енглески, циљни језик је српскохрватски. Писмо је латинично, а изговор екавски. Одреднице су поређане абecedним редом. Синоними су раздвојени зарезима, а што се тиче граматичких обележја, она у овом речнику нису заступљена. На крају књиге, пописане се скраћенице, у истоименом поглављу.

4.2.11. Међународни електротехнички IEC речник са терминима на српском језику: електроенергетика, електроника и телекомуникације

Прво издање Вишејезичког електротехничког речника IEC објавила је 1983. године Међународна електротехничка комисија (*Dictionnaire CEI multilingue de l'électricité*). Ово издање обухватало је 40 термилолошких стандарда са приближно 12.000 термина и синонима. Поред термина на енглеском и француском језику, у овом издању заступљени су и њихови еквивалентни термини на још седам језика: руски, немачки, шпански, италијански, холандски, пољски и шведски. Прво издање речника није обухватало појмове из области телекомуникација.

Друго издање речника, проширено за термине из области телекомуникација, излази 1992. године. Ово издање, под пуним називом *Electricite, Electronique Et Telecommunications Dictionnaire Multilingue* обухвата 70 термилолошких стандарда са око 18.000 термина.

У издању Савезног завода за стандардизацију, 1996. године, појављује се прво издање овог речника на српском језику, под насловом *Međunarodni elektrotehnički IEC rečnik sa terminima na srpskom jeziku: elektroenergetika, elektronika i telekomunikacije*, по узору на свог међународног претходника. Састоји се од укупно пет књига. Прве три књиге објављене су 1996. године а преостале две, наредне, 1997. године. У свакој од пет књига, одреднице на српском језику писане су латиничним писмом, екавским изговором.

Књига 1 (која је заправо издање IEC речника из 1992. године), поред термина на енглеском, француском, руском, немачком, шпанском, италијанском, холандском, пољском и шведском, допуњена је терминима на српском језику, као и Књизи 2 овог Речника. У Књизи 1 (Savezni zavod za standardizaciju 1996) кључна реч је дата на енглеском језику, након чега следе њени преводи на преостале наведене језике, док је у Књизи 2 (Savezni zavod za standardizaciju 1996) кључна реч дата на француском језику (а након тога и њени преводи на остале језике).

Као што је наведено у Предговору првог и другог издања, основни принцип приликом припремања Речника био је омогућавање једноставне претраге термина. Термини се могу састојати од једне или неколико речи од којих немају све исту важност. Реч са главним значењем, тзв. *кључна реч*, уводи се да би се олакшала термилолошка претрага. Стога је и речник конципиран према *кључним речима*, тј. речима које су носиоци главног значења. Оне су дате по алфаветском редоследу. Испод сваке кључне речи, наведени су сви термини у којима се та реч појављује. Ти термини су такође поређани по алфаветском реду, према редним бројевима, и нумерисани су. Нумерички знак се употребљава за све унутрашње референце у Речнику. Ако се кључна реч појављује у неком од граматичких облика, за њено означавање користи се најједноставнији облик у зависности од падежа (једнина, номинатив, мушки род, неодређени начин) (Savezni zavod za standardizaciju 1996, 274):

Electricity

1. **conversion of electricity:** — *conversion 9*

2. **distribution of electricity:** — *distribution 20*

3. **electricity:** 1. The manifestation of a form of energy associated with static or dynamic electric charges. 2. The special field of science and technology dealing with electric phenomena **elektricitet**

★ *électricité* — *électricité* 1. **R** электричество. **D** Elektrizität **E** electricidad. **I** elettricità. **N** elektriciteit; elektriciteitsleer; elektrotechniek. **P** elektryczność; elektryka. **S** elektricitet. (151)

4. electricity supply system (*Syn.* electrical power system): — *system* 56

5. generation of electricity: a process whereby electrical energy is obtained from some other form of energy. ★ production d'énergie électrique — *production* 8. **R** производство электрической энергии. **D** Erzeugung elektrischer Energie. **E** producción de energía eléctrica. **I** produzione d'energia elettrica. **N** opwekking van elektrische energie. **P** wytwarzanie energii elektrycznej. **S** produktion av elenergi; elproduktion. (601)

proizvodnja električne energije

6. transformation of electricity: the transfer of electricity through a power transformer. ★ transformation d'énergie électrique — *transformation* 10. **R** трансформация электрической энергии. **D** Transformierung elektrischer Energie. **E** transformación de energía eléctrica. **I** trasformazione d'energia elettrica. **N** transformatie van elektrische energie. **P** transformacja energii elektrycznej. **S** transformerling (av elkraft). (601)

transformacija električne energije

7. transmission of electricity: the transfer in bulk of electricity, from generating stations to areas of consumption. ★ transport d'énergie électrique — *transport* 9. **R** передача электрической энергии. **D** Übertragung elektrischer Energie. **E** transporte de energía eléctrica. **I** trasmissione d'energia elettrica. **N** transport van elektrische energie. **P** przesył energii elektrycznej. **S** överföring (av elkraft). (601)

prenos električne energije

Прва и друга књига Речника садрже и податке о југословенским стандардима из којих је узет посматрани термин, у одвојеном делу на дну сваке стране. Такав термин обележен је бројем поглавља у загради, нпр. (531) Електронске цеви – JUS N.A0.531:1989.

Књига 3 садржи термине на српском, француском и енглеском језику, а начини приказивања термина су као у примеру који следи (Savezni zavod za standardizaciju 1996, 1):

triak; dvosmerni triodni tiristor	thyristor triode bidirectionnel, triac (abréviation)	thyristor 8	bidirectional triode thyristor; triac (abbreviation)	thyristor 2
--	--	-------------	--	-------------

Као што је наведено у Предговору треће књиге, свака страница књиге садржи пет стубаца. У првом ступцу су поређани термини на српском језику по абecedном реду. Синоними, термини који означавају исти појам, приказани су са онолико одредница колики је и њихов број, с тим што су у свакој одредници приказани сви синоними одређеног појма, али различито поређани тако да сваки од њих буде по једном први. Сваки хомоним представљен је одвојеном одредницом. То се односи и на сличне појмове чије се дефиниције, формулисане у различитим комисијама, незнатно разликују.

У другом ступцу су наведени приоритетни еквивалентни термини на француском језику, а у трећем њихове кључне речи са бројем (такође на француском језику). Ови бројеви одговарају бројевима под једном кључном речи у другој књизи Речника. Четврти и пети стубац садрже приоритетне еквивалентне термине на енглеском језику и њихове кључне речи са бројевима

термина у првој књизи Речника. Помоћу ових бројева лако се налазе дефиниције појмова на француском и енглеском језику. Одсуство термина у једном од ова два језика значи да није на располагању или да не постоји.

Књига 4 (руски/српски, српски/руски) (Savezni zavod za standardizaciju 1997) и Књига 5 (немачко/српски, српски/немачки) (Savezni zavod za standardizaciju 1997) објављене су 1997. године. Састоје се из два дела. У првом делу Књиге 4 дат је руско/српски речник с терминима на руском језику према азбучном реду (у руској азбуци), а у другом делу је српско/руски речник у ком су одреднице поређане према абecedном реду на српском језику. Исто се односи и на Књигу 5: у првом делу књиге дат је немачко/српски речник с терминима на немачком језику по абecedном реду, а у другом делу књиге је српско/немачки речник с терминима поређаним према абecedном реду на српском језику.

Дефиниције појмова које означавају ови термини проналазе се помоћу прве и друге књиге Речника, пошто се претходно помоћу треће књиге, преко српских термина, одреде кључне речи и одговарајући бројеви појмова. Термини који почињу грчким словима или бројевима дати су на крају Речника.

Термини су приказани на следећи начин, а пример је преузет из Књиге бр. 4 Међународног електротехничког *IEC* речника са терминима на српском језику (Savezni zavod za standardizaciju 1997, 11)

быстродействующее АПВ		brzo automatsko ponovno uključenje
быстродействующий усилитель	магнитный	poluperiodni transduktor
быстродействующий (постоянного тока)	выключатель	ultrabrzzi prekidač (jednosmerne struje)

Приликом прикупљања термилошких података за израду овог речника, поред интернационалних *IEC* стандарда, користили су се и удбеници за универзитетску и средњу наставу, стручне књиге, речници и енциклопедије. Сви преднацрти стандарда пролазили су кроз стручне комисије, сачињене од наших најкомпетентнијих стручњака, који су уз посвећен и предан рад долазили до најбољег решења (Хорват 1996, 230). Рад на утврђивању термина спроводили су наши научници и стручњаци из области електротехнике током своје 50-годишње делатности у око 100 комисија за стандарде Савезног завода за стандардизацију, како је то наглашено у Предговору Књиге 4 Речника.

Приликом дефинисања одредница, настојало се да оне буду дате једним исказом, да буду концизне, јасне и да садрже све битне карактеристике појма. Приликом избора термина, предлози који су представљали готово целу дефиницију појма били су одбацивани, а сами термини су морали испунити норму српског језика односно испољити карактеристике дате у дефиницији. Велики број термина из ове области формиран је усвајањем страних термина, и то преузимањем термина са страног језика и његовим евентуалним прилагођавањем правилима нашег језика, као и дословним превођењем страног термина. То је допринело интернационалној унификацији односно термилошкој интернационализацији. Приликом увођења скраћеница, настојало се да оне буду интернационалне, по узору на симболе величина. Тако су усвојене скраћенице енглеског језика према угледу на област телекомуникација (Хорват 1996, 229–230).

Један од главних циљева овог термилошког речника био је да олакша споразумевање између електротехничких инжењера на међународном нивоу као и да све непрецизности и недостатке који се јављају у језику електротехнике сведе на најмању могућу меру. Осим тога, Речник представља незаменљив приручник за све који се баве преводилачком делатношћу односно који на било који начин користе стране текстове. Дефиниције у њему су поуздане, а састављали су их стручњаци међународних комитета *IEC*.

4.2.12. Статистичка терминологија коришћена у електропривреди

У издању Електроенергетског координационог центра, 1997. године излази из штампе публикација под именом *Statistička terminologija korišćena u elektroprivredi*, на 154 стране. Овај приручник, у ствари, представља превод приручника *UNIPEDA*-а⁶⁷ који је објављен 1975. године под истим именом. Његов основни циљ био је да обједини статистичку терминологију која се јавља у електропривреди. Аутори овог приручника нису имали намеру да од њега створе технички речник, већ да створе јединствену терминологију намењену стручњацима широм света из области статистике која се користи у електропривреди. У приручнику су заступљена три језика: српски, француски и енглески, а текст је штампан тростубачно. Садржи више од 400 одредница.

Приручник је, у основи, подељен на три поглавља од којих је свако поглавље даље распоређено на своја потпоглавља у којима су обрађени општи термини, термини у вези са термоенергијом и хидроенергијом, електроенергетски системи и остало (Duncan и Ivandekić 1997, 3):

- Прво поглавље обрађује опште појмове (нпр. „хидроенергија и други обновљиви извори енергије“),
- Друго поглавље описује опрему која се користи у овој индустрији (реч је о појмовима који дефинишу статистичке податке који се обично не мењају временом, као нпр. „мрежни трансформатор“ или „максимална електрична снага“),
- Треће поглавље описује појмове којима се дефинише рад система и који су променљивог карактера (нпр. „фактор снаге“, „резерва снаге“ итд.).

Овај приручник састављала је група међународних стручњака, а користе га запослени у електропривредној индустрији.

4.2.13. Речник електронике и електротехнике: енглеско-српски

Овај речник од 25.000 одредница, аутора Слободана Б. Танкосића, изашао је 2006. године, у издању Грађевинске књиге, Београд, на 398 страна.

Циљ овог речника је да обједини термилошке јединице, како научних тако и индустријских делатности, а као што је у Предговору предочено, намењен је студентима, инжењерима, научницима, као и професионалним преводиоцима.

Речник је двојезични, једносмерни; изворни језик је енглески, циљни језик је српски. У речнику је заступљено латинично писмо и екавски изговор. Одреднице су сложене абecedним редом, синоними су одвајани зарезима, а ближа одређења неког појма у циљном језику, смештана су у заграде. Испод главне одреднице често се појављују речи (сложене у абecedном низу), којима се прецизира њено значење односно које са главном речи (одредницом), творе посебан појам. Граматичка обележја нису заступљена, а термини су приказани на следећи начин (Tankosić 2006, 29):

<p>arrester – zaustavljač, osigurač od prenapona, prenaponska zaštita. instrument ~ – sistem za podešavanje na nulu kazaljke instrumenta</p>
--

У табели 6 дат је хронолошки преглед речника који су анализирани у претходним поглављима. Сви речници су двојезични односно вишејезични и покривају српскохрватско

⁶⁷ *UNIPEDA – UNion Internationale des Producteurs Et Distributeurs d'Énergie électrique* (срп. Међународна унија произвођача и дистрибутера електричне енергије)

говорно подручје, а објављени су на простору Србије и бивше Југославије. Изузев приручника „Статистичка терминологија коришћена у електропривреди“ који припада области електроенергетике, ни један од преосталих наведених речника није ограничен на уско специјализовани домен електроенергетике већ покрива ширу техничку област чији је део и електроенергетика, а то је електротехника. Међу наведеним речницима, налазе се и три речника која поред електротехнике покривају и област електронике, као и један општи технички речник који обухвата термине из електротехнике, електронике, информатике, физике, акустике, радиотехнике, телевизије, енергетике, аутоматике, телекомуникација и електричних мерења.

Табела 6. Листа речника релевантних за домене електроенергетике

Речник	Година издања	Број речи	Једнојезичан / двојезичан / вишејезичан	Област
Мали српско-француско-немачки електротехнички речник са скраћеницама и практичним забелешкама	1927	* > 600	вишејезичан	електротехника
Њемачко-хрватски и хрватко-њемачки електротехнички рјечник	1952	7,000	двојезичан	електротехника
Словеначки електротехнички речник (10 томова)	1957–1970	* > 3,000	вишејезичан	електротехника
Електротехнички термилошки речник, група 07, електроника	1965	* > 800	вишејезичан	електроника и електротехника
Електротехнички речник: руско-српскохрватски, српскохрватско-руски	1969	> 10000	двојезичан	електротехника
Електротехнички речник: немачко-српскохрватски	1971	* > 10,000	двојезичан	електротехника
Речник индустријске електротехнике: немачко-српскохрватски	1971	70000	двојезичан	индустријска електроника и електротехника
Енглеско-српскохрватски електротехнички речник	1986	* < 25,000	двојезичан	електротехника
Дефиниције појмова из електротехнике на основу ИЕС (СЕИ): Појмовник – речник електротехнике	1991	* > 500	вишејезичан	електротехника
Енглеско-српскохрватски технички речник	1992	> 30000	двојезичан	техника

Међународни електротехнички ИЕС речник са терминима на српском језику: електроенергетика, електроника и телекомуникације (5 томова)	1996 – 1997	18000	вишејезичан	електротехника
Статистичка терминологија коришћена у електропривреди	1997	* > 400	вишејезичан	електроенергетика
Речник електронике и електротехнике: енглеско-српски	2006	25000	двојезичан	електроника и електротехника

* Број одредница које је проценио аутор овог рада.

4.2.14. Речници из области електроенергетике Института за стандардизацију Србије

У оквиру Сектора за електротехничку стандардизацију и информациони систем Института за стандардизацију Србије налази се Одељење за електротехнику, телекомуникације и информационе технологије.

Поред послова који се односе на производњу и коришћење, пренос и дистрибуцију електричне енергије, инсталациони прибор, сијалице и припадајућу опрему, функционисање и безбедност електричних апарата за домаћинство, мале енергетске трансформаторе и пригушнице и слично, као и послове који се односе на телекомуникације, електронске системе и уређаје, и остало, ово Одељење обавља и послове који се односе на термине и дефиниције, скупове кодираних знакова, пренос података, комуникацију и размену информација између рачунарских система, као и друге послове у складу са законом, актом о оснивању и статутом Института (Institut za standardizaciju Srbije 2016, 13).

До данас је у оквиру Сектора за електротехничку стандардизацију и информационе системе објављен велики број стандарда којима се описују електротехнички термини и дефиниције као и делови речника Међународне електротехничке комисије, Међународни електротехнички речник. Овај речник носи ознаку серије *IEC 60050*, а представља вишејезични речник за општу намену који обухвата област електротехнике, електронике и телекомуникација (о чему је већ било речи у поглављу 2.5.5). Он садржи више од 22.000 термилолошких одредница, од којих свака одговара појму. Ове одреднице су распоређене у 80 делова, при чему сваки део одговара датој области. Одреднице прате хијерархијску класификациону шему део/одељак/појам⁶⁸ док су унутар одељка појмови систематски организовани (SRPS IEC 60050-151 2019).

Институт за стандардизацију Србије је до сада објавио следеће стандарде којима се уређује област електроенергетике. Њихова класификациона група је 01.40.029:⁶⁹

- Потенцијално експлозивне атмосфере – Термини и дефиниције за опрему и заштитне системе намењене за употребу у потенцијално експлозивним атмосферама (*SRPS EN 13237:2013*). Овај стандард је израђен као помоћ пројектантима, произвођачима и осталима приликом коришћења усклађених термина и дефиниција (речника) за опрему и заштитне системе намењене за употребу у потенцијално експлозивним атмосферама. Он описује речник који се користи да би

⁶⁸ Одреднице се састоје од три елемента одвојених цртама, нпр. 151-13-82. Први елемент је број дела, други елемент је број одељка, а трећи је број конкретног појма.

⁶⁹ Ниво један обухвата подручје активности у стандардизацији, ниво два групу, док ниво три означава подгрупу. Стандарди под ознаком 01.040 представљају речнике, док су речници из области електроенергетике обележени ознаком 01.040.029.

се у свим стандардима у овој области обезбедила униформност терминологије (SRPS EN 13237 2013, 4).

- Дефиниције и номенклатура угљених четкица, држача четкица, колектора и клизних колутова (SRPS EN 60276:2010). Овај стандард успоставља речник дефиниција и термина за класе четкица, флексибилних спојница, прикључака, колектора и клизних колутова, ознаке колектора и остало. Садржи и илустрације за сваки термин (SRPS EN 60276 2010).

- Изолатори за надземне водове називног напона већег од 1 kV – Део 1: Керамичке или стаклене ланчане јединице за мреже наизменичне струје – Термини и дефиниције, методе испитивања и критеријуми за пријем (SRPS EN 60383-1:2011). Овај стандард се примењује на изолаторе од керамичких материјала или стакла за употребу на надземним водовима наизменичне струје називног напона већег од 1000 V и фреквенције од највише 100 Hz. Примењује се и за изолаторске јединице које се користе на водовима једносмерне струје, за изолаторе на крутим надземним водовима, као и за оне који се користе у трансформаторским станицама (SRPS EN 60383-1 2011). Измена овог стандарда носи ознаку SRPS EN 60383-1:2011/A11:2011.

- Изолатори за надземне водове називног напона већег од 1 kV – Део 2: Керамичке или стаклене ланчане јединице за мреже наизменичне струје – Термини и дефиниције, методе испитивања и критеријуми за пријем (SRPS EN 60383-2:2011). Као и претходни, и овај стандард се примењује на изолаторе од керамичких материјала или стакла за употребу на надземним водовима наизменичне струје називног напона већег од 1000 V и фреквенције од највише 100 Hz. Овај део се примењује и за надземне водове електричне вуче као и за надземне водове у трансформаторским станицама (SRPS EN 60383-2 2011).

- Керамички и стаклени изолациони материјали – Део 1: Дефиниције и класификација (SRPS EN 60672-1:2011). Даје дефиниције и класификацију за материјале од керамике, стакла и стаклокерамике који се користе за сврхе електричне изолације (SRPS EN 60672-1 2011).

- Спецификације за пластичне фолије за електричне сврхе – Део 1: Дефиниције и општи захтеви (SRPS EN 60674-1:2011). Даје дефиниције и опште захтеве у погледу квалитета материјала, ролни, спојева, језгара, димензије, паковања и етикетирања (SRPS EN 60674-1 2011).

- Спецификација за траке ткане од стаклених влакана и стаклених полиестерских влакана – Део 1: Дефиниције, класификација и општи захтеви (SRPS EN 61067-1:2011). Дефинише захтеве за континуално пуњене траке ткане на конвенционалним разбојима или разбојима без ткачког чуна било од стаклених влакана или комбинације стакла и полиестерских влакана. Опсеги назначених величина покривених овим стандардом су: ширине: 10 mm до 50 mm, дебљине: 0,05 mm до 0,40 mm (SRPS EN 61067-1 2011).

- Величине и јединице – Део 6: Електромагнетизам (SRPS EN 80000-6:2010). У овом стандарду су дати називи, симболи и дефиниције за величине и јединице које се користе у електромагнетици. Фактори претварања су (онда када је то примерено) такође дати (SRPS EN 80000-6 2010).

- Међународни електротехнички речник – Део 191: Сигурност функционисања и квалитет услуге (идентичан са ISO IEC 50(191):1990) (SRPS IEC 60050(191):1997). Овај стандард подељен је на два дела. Први део „Сигурност функционисања: општи термини“, садржи 18 одељака, док други део „Квалитет услуге у телекомуникацијама“, садржи одељке 19 и 20. У оба дела пописани су термини на српском, енглеском, француском и руском језику и дате су њихове дефиниције. Регистар термина на енглеском, француском и руском језику, као и Национални прилог А – Регистар термина на српском језику, дати су на крају овог стандарда (SRPS IEC 60050-191 1997).

- Међународни електротехнички речник – Део 151: Електрични и магнетни уређаји (SRPS IEC 60050-151:2019). Овај део IEC 60050 даје опште термине који се користе у различитим областима електротехнике (нпр. „електрицитет“, „магнетизам“, „електроника“, „уређај“, „компонента“ итд.), опште термине који се односе на прикључке и уређаје за повезивање, термине који се односе на електричне и магнетне уређаје опште намене, као што су отпорници,

трансформатори, релеји итд. и термине који се односе на понашање, употребу, испитивања и услове рада ових уређаја. Ова терминологија је, наравно, конзистентна са терминологијом развијеном у другим специјализованим деловима Међународног електротехничког речника (SRPS IEC 60050-151 2019).

- Међународни електротехнички речник – Део 161: Електромагнетска компатибилност (SRPS IEC 60050-161:2014). Овај део стандарда IEC 60050 се односи на термине и дефиниције који се користе у оквиру предмета и подручја примене електромагнетске компатибилности. Ово издање српског стандарда представља обједињену верзију претходног издања и измена *Amd. 1* и *Amd. 2*. Стога је редослед термина дат њиховим историјским редоследом. Поновно нумерисање и додавање или брисање термина се може очекивати у следећем циклусу одржавања IEC стандарда (SRPS IEC 60050-161 2014).

- Међународни електротехнички речник – Део 195: Уземљење и заштита од електричног удара (SRPS IEC 60050-195:2008). Овај део Међународног електротехничког речника се бави и извесним аспектима безбедности. Појам „безбедност“ који се овде примењује дефинисан је у ISO/IEC Упутству 51:1990⁷⁰ (SRPS IEC 60050-195 2008).

- Међународни електротехнички речник – Део 212: Чврсти, течни и гасовити електрични изолациони материјали (SRPS IEC 60050-212:2017). У овом делу дата је општа терминологија која се користи када је реч о чврстим, течним и гасовитим електричним изолационим материјалима. Ово друго издање повлачи и замењује прво издање објављено 1990. године и представља његову техничку ревизију. У њему су ажурирани услови и дефиниције у вези са претходним издањем. Ова терминологија је конзистентна са терминологијом развијеном у другим специјализованим деловима Међународног електротехничког речника. Овај стандард има статус хоризонталног стандарда⁷¹ у складу са IEC Упутством 108⁷² (SRPS IEC 60050- 212 2017).

- Међународни електротехнички речник – Део 411: Обртне машине (SRPS IEC 60050- 411:2013). Овај део стандарда садржи 24 одељка у којима су пописани термини на српском језику и њихови еквиваленти на енглеском језику, регистар термина на енглеском језику, као и регистар термина на српском језику (SRPS IEC 60050-411 2013). Овај стандард има статус хоризонталног стандарда према IEC Упутству 108.

- Међународни електротехнички речник – Део 426: Електрични уређаји за експлозивну атмосферу (SRPS IEC 60050-426:2003). Овај део стандарда садржи 11 одељака. За главне одреднице унутар одељака, које су на српском језику, дати су еквиваленти на француском, енглеском, руском, немачком, шпанском, италијанском, холандском, пољском и шведском језику. У другом делу стандарда су дати и регистри за сваки од поменутих језика, као и информативни прилози *NA* – Регистар појмова на српском језику и *NB* – Упоредна листа термина које треба напустити и термина који су уместо њих утврђени (SRPS IEC 60050- 426 2003). Стандард SRPS EN 13237:2013 (који је претходно описан) којим су дати термини и дефиниције за опрему и заштитне системе намењене за употребу у потенцијално експлозивним атмосферама, служи као помоћ приликом коришћења овог речника.

- Међународни електротехнички речник – Део 441: Расклопне апаратуре и осигурачи (SRPS IEC 60050-441:2014). Садржи 8 одељака у којима су пописани термини на српском језику и њихови еквиваленти на енглеском, регистар термина на енглеском језику као и регистар термина на српском језику (у оквиру Националног прилога) (SRPS IEC 60050-441 2014). Овај стандард има статус хоризонталног стандарда према IEC Упутству 108.

⁷⁰3.1 безбедност (*safety*) – одсуство неочекиваних ризика од оштећења (SRPS IEC 60050-195 2008).

⁷¹Хоризонтални стандарди представљају стандарде који се односе на основне принципе, појмове, терминологију и техничке карактеристике, релевантни су за велики број техничких комитета и од пресудног значаја за обезбеђивање усклађености корпуса стандардизационих докумената (<https://www.iec.ch/dyn/www/f?p=103:94:0>).

⁷²IEC Упутство 108 даје смернице за обезбеђивање кохерентности публикација које је објавила Међународна електротехничка комисија – хоризонталне функције, хоризонталне публикације и њихову примену.

- Међународни електротехнички речник – Део 442: Електроинсталациони прибор (*SRPS IEC 60050-442:2007*). Овај део стандарда садржи 8 одељака, као и регистре термина на енглеском и српском језику (*SRPS IEC 60050-442 2007*), док су његовој измени, чија је ознака *SRPS IEC 60050-442:2007/A1:2015* (Међународни електротехнички речник – Део 442: Електроинсталациони прибор — Измена 1), додати нови термини и један одељак.

- Међународни електротехнички речник – Део 448: Заштита електроенергетског система (*SRPS IEC 60050-448:2015*). Овај стандард је припремио *IEC* Технички комитет 1: Терминологија, у сарадњи са *IEC* Техничким комитетом 95: Мерни релеји и заштитна опрема. Садржи 6 одељака у којима су пописани термини на српском и њихови еквиваленти на енглеском језику, слике а садржи и регистре на енглеском и српском језику (регистар на српском језику дат је као информативан прилог) (*SRPS IEC 60050-448 2015*).

- Међународни електротехнички речник – Део 461: Електрични каблови (*SRPS IEC 60050- 461:2014*). Овај део серије стандарда *IEC 60050* односи се на термине и дефиниције који су коришћени у оквиру предмета и подручја примене ТК 20 „Електрични каблови“. Садржи 23 одељка, регистар термина на енглеском језику као и регистар термина на српском језику који је дат као информативан прилог (*SRPS IEC 60050-461 2014*). Ово издање је обједињена верзија првог издања и његових измена и допуна 1 и 2. Стога је редослед термина дат њиховим историјским редоследом. Поновно нумерисање и додавање или брисање термина је предложено за следећи циклус одржавања.

- Међународни електротехнички речник – Део 482: Примарне и секундарне ћелије и батерије (*SRPS IEC 60050-482:2009*). Овај део *IEV* (серије *IEC 60050*) представља хармонизацију, консолидацију и ревизију термина претходно датих у стандардима *IEC 60050- 481* „Примарне ћелије и батерије“ и *IEC 60050-486* „Секундарне ћелије и батерије“. Садржи 5 одељака у којима су пописани термини на српском и енглеском језику, регистар термина на енглеском језику као и регистар термина на српском језику који је дат као информативни прилог. Термини који су већ дефинисани у другим деловима Међународног електротехничког речника нису обухваћени овим делом (*SRPS IEC 60050-482 2009*).

- Енергетска електроника – Међународни електротехнички речник – Део 551: Енергетска електроника (*SRPS IEC 60050-551:2014*). Садржи 10 одељака у којима су пописани термини на српском и енглеском језику, регистар термина на енглеском језику као и регистар термина на српском језику који је дат као информативни прилог (*SRPS IEC 60050-551 2014*). Има статус хоризонталног стандарда према *IEC* Упутству 108.

- Међународни електротехнички речник – Део 601: Производња, пренос и дистрибуција електричне енергије – Опште (*SRPS IEC 60050-601:2015*). Овај стандард представља превод међународног стандарда *IEC 60050-601:1985* и његове измене *A1:1998*, са енглеског на српски језик. Прво је од пет поглавља међународног електротехничког речника (*IEV*) која се баве производњом, преносом и дистрибуцијом електричне енергије. Стандард садржи термине и дефиниције на српском и енглеском језику, регистар термина на енглеском језику као и регистар термина на српском језику који је дат као информативни прилог (*SRPS IEC 60050- 601 2015*). Овај стандард има статус хоризонталног стандарда према *IEC* Упутству 108.

- Међународни електротехнички речник – Део 602: Производња, пренос и дистрибуција електричне енергије – Производња (*SRPS IEC 60050-602:2015*). Овај стандард представља превод међународног стандарда *IEC 60050-602:1983*, са енглеског на српски језик. Друго је од пет поглавља међународног електротехничког речника (*IEV*) која се баве производњом, преносом и дистрибуцијом електричне енергије. Стандард садржи термине и дефиниције на српском и енглеском језику; изостављени су термини и регистри термина на француском и руском језику, а дат је регистар термина на енглеском језику. Стандард садржи регистар термина на српском језику као информативни прилог. Овај стандард има статус хоризонталног стандарда према *IEC* Упутству 108 (*SRPS IEC 60050-602 2015*).

- Међународни електротехнички речник – Део 603: Производња, пренос и дистрибуција електричне енергије – Планирање и управљање електроенергетским системом (*SRPS*

IEC 60050-603:2015). Овај стандард представља превод међународног стандарда *IEC 60050-602:1983* и његове измене *A1:1998* са енглеског на српски језик. Трећи је у серији од пет поглавља која се баве производњом, преносом и дистрибуцијом електричне енергије. Садржи термине и дефиниције на српском и енглеском језику; изостављени су термини и регистри термина на француском и руском језику, а дат је регистар термина на енглеском језику. Стандард садржи регистар термина на српском језику као информативни прилог. Овај стандард има статус хоризонталног стандарда према *IEC Упутству 108 (SRPS IEC 60050-603 2015)*.

- Међународни електротехнички речник – Део 605: Производња, пренос и дистрибуција електричне енергије – Разводна постројења и подстанице (*SRPS IEC 60050-605:2015*). Овај стандард представља превод међународног стандарда *IEC 60050-605:1983* са енглеског на српски језик. Последње је од укупно пет поглавља Међународног електротехничког речника (*IEV*) који се бави производњом, преносом и дистрибуцијом електричне енергије. Стандард садржи термине и дефиниције на српском и енглеском језику; изостављени су термини и регистри термина на француском и руском језику, а дат је регистар термина на енглеском језику односно српском као информативни прилог. Овај стандард има статус хоризонталног стандарда према *IEC Упутству 108 (SRPS IEC 60050-605 2015)*.

- Међународни електротехнички речник – Део 614: Производња, пренос и дистрибуција електричне енергије – Погон (експлоатација) (*SRPS IEC 60050-614:2018*). Овај стандард представља превод међународног стандарда *IEC 60050-614:2015* са енглеског на српски језик. Термини и дефиниције дати су на српском и енглеском језику. Стандард садржи и регистре термина, и то на српском и енглеском језику, док су регистри на осталим језицима изостављени. Овај део *IEC 60050* даје општу терминологију која се користи у производњи, преносу и дистрибуцији електричне енергије, као и опште термине који се користе за специфичне примене и одговарајуће технологије. Има статус хоризонталног стандарда у складу са *IEC Упутством 108*, и првенствено је намењен техничким комитетима за употребу (када је то могуће), у припреми стандарда у складу са принципима датим у овом Упутству. Терминологија садржана у овом стандарду је конзистентна са терминологијом развијеном у другим специјализованим деловима Међународног електротехничког речника (*SRPS IEC 60050-614 2018*).

- Међународни електротехнички речник – Део 617: Организација/тржиште електричне енергије (*SRPS IEC 60050-617:2015*). Стандард *SRPS IEC 60050-617* представља превод међународног стандарда *IEC 60050-901:2009* и његове измене *A1:2011*. Термини и дефиниције дати су на српском и енглеском језику, а стандард садржи и регистар термина на српском језику као информативни прилог. Овај део *IEC 60050* успоставља терминологију којом се дефинишу учесници на тржишту, интеракција између учесника и технички и финансијски термини који имају специфично значење у контексту тржишта електричне енергије. Овај хоризонтални стандард углавном користе технички комитети приликом припремања стандарда у складу са принципима наведеним у *IEC Упутству 108 (SRPS IEC 60050-617 2015)*.

- Међународни електротехнички речник – Део 811: Електрична вуча (*SRPS IEC 60050-811:2015*). Стандард *SRPS IEC 60050-811* представља превод међународног стандарда *IEC 60050-811:1991* са енглеског на српски језик. Стандард садржи термине и дефиниције на српском и енглеском језику. Термини и регистри термина на француском и руском језику су изостављени, а дат је регистар термина на енглеском језику односно српском као информативни национални прилог. Овај стандард има статус хоризонталног стандарда према *IEC Упутству 108 (SRPS IEC 60050-811 2015)*.

- Међународни електротехнички речник – Део 826: Електричне инсталације (*SRPS IEC 60050-826:2008*). Овај стандард представља превод међународног стандарда *IEC 60050-826:2004* са енглеског на српски језик, а односи се на електричне инсталације као што су оне у стамбеним, индустријским или комерцијалним објектима. Поред термина и дефиниција на српском, садржи и термине и дефиниције на енглеском језику. Остали језици су изостављени као и њихови регистри. Стандард садржи и национални прилог у ком је дат

абecedни регистар термина на српском језику који није идентичан са регистром термина на енглеском језику (SRPS IEC 60050-826 2008).

- Међународни електротехнички речник – Део 841: Индустијска електротермија (SRPS IEC 60050-841:2016). Овај стандард представља превод међународног стандарда IEC 60050- 841:2004 са енглеског на српски језик. Термини и дефиниције у овом стандарду дати су на српском и енглеском језику, а терминологија је у складу са терминологијом развијеном у другим специјализованим деловима Међународног електротехничког речника. Овај део IEC 60050 даје општу терминологију која се користи у индустријској електротермији, а такође и опште термине који се односе на специфичне примене и сличне технологије. Овим издањем преиспитује се и допуњује претходно издање услед развоја електродног, ласерског и ултразвучног загревања. Стандард садржи и регистар термина на српском језику дат као информативни национални прилог (SRPS IEC 60050-841 2016).

- Међународни електротехнички речник – Део 903: Оцењивање ризика (SRPS IEC 60050- 903:2014). Овај стандард је превод међународног стандарда IEC 60050-903:2013 са енглеског на српски језик. У њему су термини и дефиниције дати на српском и енглеском језику. Садржи и регистар термина на српском језику као национални информативни прилог. Овај део IEC 60050 даје општу терминологију која се користи у области оцењивања ризика. Има статус хоризонталног стандарда у складу са IEC Упутством 108. Терминологија садржана у овом делу конзистентна је са терминологијом која је донета у другим специјализованим деловима Међународног електротехничког речника (SRPS IEC 60050-903 2014).

- Микроталасни инструменти – Термини и дефиниције (SRPS L.G7.101:1982). Утврђују се термини и дефиниције за микроталасне инструменте и микроталасну мерну технику на језицима народа Југославије: српскохрватском, хрватском књижевном, македонском и словеначком, као и на страним језицима: енглеском и француском. Дати су и абecedни регистри термина према наведеним језицима (SRPS L.G7.101 1982).

Неки од преведених стандарда који нису класификовани под ознаком 01.040.029 (*Електроенергетика – речници*) Института за стандардизацију, али без којих је листа стандарда/речника домена ове струке непотпуна, су следећи:

- Међународни електротехнички речник – Део 321: Мерни трансформатори (SRPS IEC 60050(321):1995). Овим стандардом утврђују се термини и дефиниције из области мерних трансформатора. Дати су општи и заједнички термини, термини из области струјних и термини из области напонских трансформатора (SRPS IEC 60050-321 1995). Дати су и абecedни регистри на српском, енглеском, француском, руском и немачком језику.

- Међународни електротехнички речник – Део 421: Енергетски трансформатори и пригушнице (SRPS IEC 60050(421):1996). Овим стандардом утврђују се термини из области енергетских трансформатора и пригушница на српском, енглеском, француском, руском, немачком и шпанском језику, као и њихове дефиниције на српском језику. Дат је и регистар термина према наведеном редоследу језика. У дванаест секција, поред општих термина дају се термини за прикључке, намотаје, назначене карактеристике, изводе, губитке и струју празног хода, напон кратког споја, импедансу кратког споја и разлике напона, пораст температуре, изолацију, спреге, регулационе преклопке и моторни погон регулационе преклопке. (SRPS IEC 60050-421 1996).

- Међународни електротехнички речник – Део 466: Надземни водови (SRPS IEC 60050(466):1995). Утврђују се термини и дефиниције из области надземних водова. За термине и дефиниције на српском језику, дати су њихови еквиваленти на енглеском, француском, руском и немачком језику, и то: општи термини, термини за механички прорачун, распоне, профиле, распоред проводника, упоришта, облике стубова и конзоле, стубове, темелје, неизоловане проводнике, прибор за проводнике и изолаторске ланце (SRPS IEC 60050-466 1995). Дати су и абecedни регистри термина на српском, енглеском, француском, руском и немачком језику.

- Међународни електротехнички речник – Део: 121: Електромагнетизам (*SRPS IEC 60050-121:2018*). Овај стандард обухвата термине и дефиниције које се односе на електромагнетске величине, електромагнетска својства материјала и електричну проводност, на српском и енглеском језику. Дат је и регистар термина на енглеском и српском језику као национални информативни прилог (*SRPS IEC 60050-121 2018*).

- Међународни електротехнички речник – Део 300: Електрична и електронска мерења и мерни инструменти. Део 311: Општи термини који се односе на мерења. Део 312: Општи термини који се односе на електрична мерења. Део 313: Врсте електричних мерних инструмената. Део 314: Специфични термини према врсти инструмената (*SRPS IEC 60050-300*). Овај стандард представља тројезично издање на српском, енглеском и француском језику. Осим тога, уз термине на српском, дати су и термини на немачком и шпанском језику. Стандард такође садржи и регистре термина на свим наведеним језицима. Регистар термина на српском језику дат је као национални информативни прилог (*SRPS IEC 60050-300 2007*).

- Међународни електротехнички речник – Део 436: Енергетски кондензатори (*SRPS IEC 60050-436:2016*). Овај стандард садржи опште термине и дефиниције као и термине и дефиниције који указују на функцију, технологију и радне карактеристике енергетских кондензатора на енглеском и српском језику. Дат је и регистар термина на ова два језика (*SRPS IEC 60050-436 2016*).

У овом поглављу описано је укупно 38 стандарда које је објавио Институт за стандардизацију Србије, од чега 32 стандарда носе ознаку 01.040.29 (Електроенергетика), док преосталих 6 припада другим областима.

Стандарди који припадају домену Електроенергетике (01.040.29) и којих има укупно 32, броје 23 документа серије *IEC 60050* који представљају делове Међународног електротехничког речника и 9 докумената који припадају осталим серијама.

Стандарди серије *IEC 60050* из домена електроенергетике (01.040.29) представљају вишејезичне речнике. У сваком од наведених делова присутни су српски и енглески језик, док су у стандарду *SRPS IEC 60050-426:2003* поред ова два језика заступљени и француски, руски, немачки, шпански, италијански, шведски и пољски језик, а у стандарду *SRPS IEC 60050(191):1997* поред српског и енглеског заступљени су француски и руски језик. Сваки од ових стандарда односно делова Међународног електротехничког речника, садржи регистар термина на енглеском и српском језику. Стандард *SRPS IEC 60050-426:2003* поред ова два језика садржи и регистре на француском, руском, немачком, шпанском, италијанском, шведском и пољском језику, док стандард *SRPS IEC 60050-191:1997* поред регистара на српском и енглеском језику садржи и регистре на француском и руском. Једино је код стандарда *SRPS IEC 60050-617:2015* заступљен регистар само на српском језику. У свим стандардима заступљено је ћирилично писмо и екавски изговор, осим у *SRPS IEC 60050-191:1997*, *SRPS IEC 60050-426:2003*, *SRPS IEC 60050-195:2008*, *SRPS IEC 60050-442:2007*, *SRPS IEC 60050-482:2009*, *SRPS IEC 60050-826:2008* у којима је заступљено латинично писмо и екавски изговор.

Преосталих 9 стандарда из групе 01.040.29 успостављају речнике термина и дефиниција, представљају различите врсте спецификација и класификација, и служе као помоћ приликом коришћења речника. Чак седам од наведених девет стандарда прихваћени су без икаквих модификација као српски стандард на енглеском језику на основу међународног стандарда, док су преостала два стандарда, *SRPS EN 13237:2013* и *SRPS L.G7.101:1982*, објављени на српском језику (први ћирилично а други латинично, оба у екавском изговору).

Наведено је укупно 6 делова Међународног електротехничког речника који не припадају групи под ознаком 01.040.029 (Електроенергетика), већ општем домену електротехнике. У четири од 6 наведених стандарда коришћено је латинично писмо и екавски изговор, док је у стандардима *SRPS EN 60050-121:2018* и *SRPS EN 60050-436:2016* коришћено ћирилично писмо

уз екавски изговор. Абецедни регистри на српском и енглеском језику заступљени су у *SRPS IEC 60050-121:2018* и *SRPS IEC 60050-436:2016*; абецедни регистри на српском, енглеском, француском, руском и немачком заступљени су у *SRPS IEC 60050-466:1995* и *SRPS IEC 60050-321:1995*; регистри на енглеском, француском, немачком, шпанском и српском у *SRPS IEC 60050-300*; регистри на српском, енглеском, француском, руском, немачком и шпанском у *SRPS IEC 60050-421:1996*. Ови стандарди представљају мали, али репрезентативни узорак и у директној су вези са доменом електроенергетике. Међутим, број стандарда који су у непосредној или посредној вези са овим доменом знатно је већи и њима се, између осталог, прописује избор високонапонских каблова, утврђују компактни блокови за трансформаторске станице, класификују изолационе течности, испитују електрични и оптички каблови, спецификују изолационе навлаке, прописују грла за осетљиве флуоресцентне сијалице и многе друге ствари које директно или индиректно утичу на домен електроенергетике. Ипак, због обима као и теме ове дисертације, они нису предмет разматрања у овом раду.

Оно што је карактеристично за све стандарде описане у овом поглављу јесте да одреднице нису дате у азбучном односно абецедном реду, док је редни број одреднице структурисан тако да буде јединствен за сваки термин. Број одреднице се састоји од три елемента, одвојених цртама: први део означава део односно поглавље Међународног електротехничког речника и састоји се од три цифре, други део представља број одељка унутар конкретног поглавља и састоји се од две цифре, а трећи представља број појма унутар одељка и садржи такође две цифре (нпр. 151-13- 82).

Термин који означава појам зове се „препоручени термин“ и може бити праћен синонимима и скраћеницама. Сваки термин који има специфичну употребу, као и синоним, може бити праћен атрибутима који дају додатне информације, али су такви атрибути представљени обичним словима одмах после тог термина, као на пример: „**transmission line** (in electric power systems)“.

Уз дефиницију појма може стајати извор као и напомене. Извор се наводи у угластим заградама на крају дефиниције. Термини на различитим језицима пишу се у засебним редовима. Сваки језик има двословни код дефинисан у ISO 639 па се и сами термини смештају према абецедном редоследу кода.

Мада је доста пажње поклоњено уређивању стандарда на српском језику, увидом у њихову садржину уочене су одређене недоследности у превођењу извесног броја термина, као у примерима који следе. Важно је, ипак, нагласити да су у питању базични термини електроенергетске струке који су у свакодневной употреби у домену електроенергетике. Реч је о следећим терминима: први термин је „power system“ који је у стандардима *SRPS IEC 60050-436:2016* и *SRPS IEC 60050-603:2015* преведен на два различита начина: као „енергетски“ и као „електроенергетски систем“ (правилно), као што видимо на слици 3. Сличан је пример недоследности приказан и на слици 4. Термин „converter“ је у стандардима *SRPS IEC 60050-300:2007* и *SRPS IEC 60050-605:2015* преведен на два различита начина: у првом случају као „конвертор“ док у другом као „претварач“ (правилно). Још један упечатљив пример недоследности превода је код термина „load flow“ у значењу „токови снага“. Овај термин је у горњем делу слике 5 преведен као „проток оптерећења“ што је дослован, али неисправан превод.

436-02-11

спрежни кондензатор

кондензатор који се користи за пренос сигнала у енергетским системима

coupling capacitor

a capacitor used for the transmission of signals in a power system

603-01-01

планирање електроенергетског система

читав низ студија које се односе на развој техничког и економичног система снабдевања електричном енергијом

power system planning

the whole range of studies involved in the development of a technically and economically sound system of electricity supply

Слика 3. Примери превода термина „power system“ у стандардима SRPS IEC 60050-436:2016 и SRPS IEC 60050-603:2015

314-02-12

kodni konvertor

Uređaj za promenu prikaza informacije prema datom kôdu u prikaz iste informacije prema drugom kôdu.

de Codeumsetzer

es convertidor de código

code converter

transcoder

device for changing the representation of information according to a given code into the representation of the same information according to another code

605-01-07

претварачка подстанца

подстанца која садржи претвараче и чија је главна функција претварање наизменичне струје у једносмерну или обрнуто

converter substation

a substation including converters and the main function of which is to convert alternating current into direct current or vice versa

Слика 4. Примери превода термина „converter“ у стандардима SRPS IEC 60050-314:2007 и SRPS IEC 60050-605:2015

617-03-04

загушеност (у електроенергетском систему)

ситуација у преносној или дистрибутивној мрежи која, у деловима електроенергетског система, захтева ограничење протока оптерећења

НАПОМЕНА Загушење се може уклонити одговарајућим повећањем производње електричне енергије.

congestion (in electric power system)

situation in a transmission or distribution network requiring, in parts of an electric power system, a limitation of load flow

603-02-08

прорачун токова снаге

прорачун стационарног стања мреже при којем су познате променљиве улазне и излазне снаге у чворовима и могући напони у одређеним чворовима

load flow calculation

a steady-state network calculation where the known variables are input and output powers at the nodes and possibly voltages at specified nodes

Слика 5. Примери превода термина „load flow“ у стандардима SRPS IEC 60050-617: 2015 и SRPS IEC 60050-603:2015

Увидом у садржину ових стандарда, стиче се утисак да су одређени термилошки еквиваленти утврђени дословним превођењем на енглески језик, као што је приказано у примерима на сликама 4 и 5. Исто можемо констатовати и за термин „трансформаторска подстанциа“ приказаном на слици 6, који је неисправан у српском језику јер као такав не постоји, а чији је еквивалент на енглеском језику у стандарду *SRPS IEC 60050-605:2015* „transformer substation“. Треба ипак нагласити да се овај термин појављује и на Електропедији у делу 605, и ту је на српски језик преведен исправно, као „трансформаторска станица“. Термин који се у англосаксонској литератури користи за трансформаторску станицу је „substation“ док „transformer“ представља уређај – трансформатор. Имајући у виду да је трансформатор као уређај обавезан део трансформаторске станице (енг. *substation*), сам термин „transformer substation“ можемо окарактерисати као плеоназам. Важно је напоменути да се термин „подстанциа“ као што је приказано на слици 7, користи нпр. у домену топлотне енергије, али не и у домену електроенергетике (Ивановић 2016).

605-01-03

трансформаторска подстанциа

подстанциа која садржи енергетске трансформаторе и која међусобно повезује две или више мрежа различитих напона

transformer substation

a substation containing power transformers interconnecting two or more networks of different voltages

605-01-04

подстанциа за повишење напона

трансформаторска подстанциа у којој је излазна снага из трансформатора вишег напона него улазна снага

step-up substation

a transformer substation in which the outgoing power from the transformers is at a higher voltage than the incoming power

Слика 6. Пример превода термина „transformer substation“

601-03-02

подстанциа (електроенергетског система)

постројење

део електричног система, смештен на одређеном месту, који обухвата углавном крајеве преносних и дистрибутивних водова, електричне расклопне апаратуре, објекте и трансформаторе. Подстанциа обично обухвата безбедносне и управљачке уређаје (нпр. заштитни уређаји)

НАПОМЕНА Подстанциа може бити описно одређена у складу са ознаком система чији је саставни део. На пример: пренос, подстанциа (преносног система), дистрибутивна подстанциа, подстанцие од 400 kV или 20 kV.

substation (of a power system)

A part of an electrical system, confined to a given area, mainly including ends of transmission or distribution lines, electrical switchgear and controlgear, buildings and transformers. A substation generally includes safety or control devices (for example protection).

NOTE The substation can be qualified according to the designation of the system of which it forms a part. Examples: transmission, substation (transmission system), distribution substation, 400 kV or 20 kV substation.

Слика 7. Пример превода термина „substation“

Сличан је пример и са терминима „power station“ односно „hydroelectric power station“ у значењу „електране“ односно „хидроелектране“, приказаним на сликама 8 и 9. Ни један од ова два термина није адекватан, већ су за „електрану“ и „хидроелектрану“ термилошки еквиваленти „Power Plant“ и „Hydro Power Plant“, термини који су званично и доста дуго у употреби у поменутиим значењима.

601-03-01

електрана

станција за производњу електричне енергије
постројење за производњу електричне енергије, које обухвата грађевинске објекте, опрему за претварање енергије и сву неопходну пратећу опрему

power station

electrical generating station

An installation whose purpose is to generate electricity and which includes civil engineering works, energy conversion equipment and all the necessary ancillary equipment.

Слика 8. Пример превода термина „power station“

602-01-04

хидроелектрана

електрана у којој се гравитациона потенцијална енергија воде претвара у електричну енергију

hydroelectric power station

A power station in which the gravitational energy of water is converted into electricity.

602-01-05

проточна хидроелектрана

хидроелектрана која користи непромењени ток реке, при чему је период пуњења њене сопствене акумулације воде практично занемарив

run-of-river power station

A hydroelectric power station which uses the river flow as it occurs, the filling period of its own reservoir by the cumulative water flows being practically negligible.

3

Слика 9. Пример превода термина „hydroelectric power station“

Међу примерима на које свакако треба скренути пажњу јесте и термин „power electronics“, у значењу „енергетске електронике“, што је једини исправан превод. Важно је напоменути да се у англосаксонској литератури овај термин никада не појављује у једнини, као што је то случај у стандарду приказаном на слици 10 (SRPS IEC 60050-436 2016). Иако сам назив „енергетска електроника“ упућује на ширу област употребе, примена овог термина заступљена је само у домену електроенергетике (а не и енергетике), те стога нема забуне о каквим је уређајима реч. Издвојили бисмо и термине „power diode“ и „power switch“, који иако се преводе као „енергетска диода“ односно „енергетски прекидач“, спадају у домен електроенергетике, међутим ти примери нису заступљени у стандардима које смо описали.

436-02-07

кондензатор за енергетску електронику

енергетски кондензатор намењен за употребу у опреми енергетске електронике, способан да ради непрекидно у условима несинусоидалног напајања

power electronic capacitor

a power capacitor intended to be used in power electronic equipment and capable of operating continuously under non-sinusoidal conditions

Слика 10. Пример превода термина „power electronic“

4.3. Истакнуте организације у сектору електроенергетике данас

Поред предузећа која обављају јавну делатност на територији Србије из области производње, преноса и дистрибуције електричне енергије, истиче се и изванредан број организација које окупљају велики број стручних лица запослених у електропривредној индустрији, при институтима или факултетима. Стручњаци који учествују у раду ових организација баве се проучавањем свих аспеката електроенергетског система и кроз своје научне радове, који су

углавном засновани на дугогодишњој пракси у овој области, дају допринос овој науци. Радове излажу једанпут годишње на посебно организованим конференцијама које имају дугачку традицију и углед, како у земљи тако и у иностранству.

Радови подлежу стручној рецензији инжењера из института, надлежног министарства, са факултета у земљи, као и компанија и факултета из иностранства. Најистакнутије организације из ове области код нас су *CIGRE Србија* и *CIREД Србија*.

Делатност ових организација искључиво је усмерена на сектор електроенергетике, тако да су термилошка питања веома ретко предмет истраживања изложен у стручним радовима⁷³ односно предмет разматрања на организованим конференцијама. Па ипак, имајући у виду да ове организације окупљају велики број стручних лица из различитих институција, питања терминологије из електроенергетског сектора о којима се расправља на састанцима нису реткост. Термилошка решења око којих се учесници састанка сложе, примењују се у будућим радовима и у пракси. Званично, међутим, не постоје посебне секције које се баве термилошким питањима у оквиру сектора испред ових организација.

4.3.1. Српски национални комитет Међународног савета за велике електричне мреже CIGRE Србија

Српски национални комитет *CIGRE*⁷⁴ је професионално, непрофитно удружење које се на домаћем и међународном плану бави стручним и научним проблемима из области производње, преноса и дистрибуције електричне енергије, производње електричне опреме као и проблематиком планирања, изградње и експлоатације електроенергетских система и мрежа.

Важно је истаћи да је комитет *CIGRE Србија* правни следбеник Југословенског комитета *JUKO CIGRE* који је формиран 1951. године, и представља један од бројних националних комитета Међународног савета за велике електричне мреже, основаног 1921. године, чије је стално седиште у Паризу.⁷⁵

Основни циљеви *CIGRE Србија* су остваривање стручних и научних активности у области савременог планирања, изградње и експлоатације електроенергетских система и мрежа, унапређивање технологије рада у области производње, преноса и дистрибуције електричне енергије, производње електричне опреме, као и заштите животне средине у складу са научним и стручним достигнућима у земљи и иностранству, афирмација струке и науке и дефинисање основе за израду законске регулативе и стандардизације у области електроенергетике као и стручно усавршавање чланова *CIGRE Србија*, на домаћем и међународном плану.

Ради остваривања својих циљева, *CIGRE Србија* развија размену техничких информација и искустава и даје иницијативу за проучавање проблематике од интереса за наш електроенергетски систем, његове елементе, као и за електроиндустрију, а помаже и израду и примену националних стандарда из области којима се бави. У ту сврху, као што је наведено у статуту ове организације, *CIGRE Србија* организује саветовања, симпозијуме, колоквијуме, образује студијске комитете, одборе, групе, комисије и тела, сталног и повременог карактера, учествује у раду *CIGRE Paris* и њених органа, успоставља сарадњу са научно-истраживачким институцијама, предузећима, привредним друштвима и другим правним лицима и појединцима који су заинтересовани за проблематику којом се бави *CIGRE Србија*, прати и унапређује развој у појединим ужим подручјима свога рада, даје мишљење и сугестије у области рада *CIGRE Paris*, издаје и пружа помоћ у издавању публикација од интереса за делатност струке, развија и унапређује систем

⁷³ Можемо издвојити свега два рада која обрађују терминологију из домена електроенергетике. Реч је о радовима изложеним на саветовању комитета *CIGRE Србија* 1999. и 2003. године („Терминологија у области поузданости и одржавања висконапонске расклопне опреме” чији је аутор Ч. Спаић и „Стратегија одржавања постројења у циљу повећања расположивости – терминологија и споразумевање” чији су аутори Ч. Спаић и Р. Булатовић).

⁷⁴ Пун назив ове организације је *CIGRE Србија*. Све информације у вези са српским националним комитетом *CIGRE* преузете су августа 2019. године са веб-сајта: <http://www.cigresrbija.rs/>.

⁷⁵ *Conseil International des Grands Reseaux Electriques (CIGRE)* – познат под називом *CIGRE Paris*.

информисања и размену информација, објављивање материјала и препорука са својих стручних манифестација и обезбеђује финансијска средства за свој рад.

Активност *CIGRE Србија* се одвија у оквиру 16 студијских комитета. Улога ових комитета је да подстичу и координирају студије које доприносе техничком унапређењу у свакој посебној области. Студијски комитети су носиоци стручних активности и имају водећу улогу у организовању скупова *CIGRE Србија*, у избору преференцијалних тема, као и организовању и вођењу дискусија на скуповима.

Чланови ових комитета су афирмисани и врло компетентни стручњаци. Студијски комитети формирају посебне радне групе чији је циљ да обраде одређене задатке у ускоспецијализованим областима.

У саставу *CIGRE Србија* постоје следећи студијски комитети (словне ознаке студијских комитета су у складу са ознакама *CIGRE Paris*):

A1 Обртне електричне машине

A2 Трансформатори

A3 Високонапонска опрема

B1 Каблови

B2 Надземни водови

B3 Постројења

B4 Високонапонске једносмерне струје (скр. енг. *HVDC*)⁷⁶ и енергетска електроника

B5 Заштита и аутоматизација

C1 Економија и развој електроенергетског система

C2 Управљање и експлоатација у електроенергетском систему

C3 Перформансе система заштите животне средине

C4 Техничке перформансе електроенергетског система

C5 Тржиште електричне енергије и регулација

C6 Дистрибутивни системи и дистрибуирана производња

D1 Материјали и савремене технологије

D2 Информациони системи и телекомуникације

4.3.2. Српски национални комитет *CIREД* Србија

Као што је дефинисано статутом ове организације, Српски национални комитет *CIREД*⁷⁷ је професионална, непрофитна и неполитичка друштвена организација, која се на домаћем и међународном плану бави стручним и научним проблемима из области технике дистрибуције електричне енергије, производње електричне опреме и проблематиком електродистрибутивних система.

Ова организација (*CIREД Србија*) једна је од бројних националних комитета под окриљем Међународне конференције за електродистрибутивне мреже *CIREД*⁷⁸.

⁷⁶ енгл. *High-voltage direct current*

⁷⁷ Све информације у вези са организацијом преузете су августа 2019. године из Статута националног комитета и Правилника о раду стручних комисија, доступних на следећим веб-сајтовима:

<http://www.ciredserbia.org.rs/dokumenta/statut.pdf> и

http://www.ciredserbia.org.rs/dokumenta/pravilnik_o_radu_strucnih_komisija.pdf.

⁷⁸ *Congres International des Réseaux Electriques de Distribution*

Као што је дефинисано Чланом 4. Статута, овај национални комитет се састоји од признатих стручњака из области технике дистрибуције електричне енергије из електропривредних предузећа, произвођача опреме, истраживачких института и факултета, пројектних и извођачких организација, као и великих корисника (потрошача) електричне енергије.

Основни циљеви ове организације су остваривање стручних и научних активности у области планирања, изградње и експлоатације електродистрибутивних система и у области производње опреме за те системе, унапређивање делатности технике дистрибуције електричне енергије и производње електричне опреме, на стручном, научном и организационом подручју као и унапређивање положаја чланова *CIREД Србија* на стручном подручју, на домаћем и међународном плану.

У остваривању својих циљева, *CIREД Србија* развија размену техничких информација и искустава и даје иницијативе за проучавање проблематике од интереса за српске електродистрибутивне системе, његове елементе, српску електроиндустрију и помаже израду и примену српских стандарда из области којом се бави *CIREД Србија*.

У остваривању својих циљева, *CIREД Србија* активно делује на стручном изучавању, разматрању и унапређењу технике дистрибуције електричне енергије, пропагира делатност и рад конференција *CIREД*, одржава конференције о дистрибутивним мрежама Србије сваке друге године и учествује у раду међународног комитета *CIREД*, делегира представника за управни комитет *CIREД* у циљу управљања његовим радом, предлаже преференцијалне теме за конференције *CIREД* и предлаже техничком комитету *CIREД* области рада за које сматра да су од посебног интереса, обавља кореспонденцију са комитетом *CIREД* и дистрибуира информације свим заинтересованим стручњацима у земљи (позиви за писање реферата, програми конференција, часопис *CIREД* и слично), пропагира учешће на конференцијама *CIREД*, разматра предлоге реферата аутора из своје земље, бира и предлаже реферате техничком комитету *CIREД* на усвајање, пружа евентуалне додатне информације и коментаре, прикупља коначне верзије реферата са комплетном пратећом документацијом и брине о достављању и поштовању услова од стране техничког и организационог комитета конференција *CIREД*, предлаже секретаријату *CIREД* усмерења за унапређење рада *CIREД*, предлаже чланове за радне групе *CIREД*, организује стручне скупове, саветовања, семинаре и друго у Србији, прати и унапређује развој у појединим ужим подручјима свог рада и у том циљу формира стручне комисије и радне групе *CIREД Србија*, успоставља сарадњу са асоцијацијама и појединцима који су заинтересовани за проблематику којом се бави *CIREД Србија*, развија и унапређује информисање објављивањем материјала и препорука са својих стручних скупова, издаје и помаже издавање научних и стручних публикација из домена свог рада и обезбеђује финансијска средства за свој рад.

Научно-стручна тела ове организације, као и њен основни организациони облик представљају стручне комисије. Стручне комисије се баве проблематиком одређених подручја рада *CIREД Србије*.

Стручне комисије чине чланови *CIREД Србије*, истакнути стручњаци који желе и могу активно да раде на испуњавању задатака ових комисија, али и други признати српски стручни и научни радници, који својим залагањем могу допринети раду стручних комисија. Ради ефикаснијег рада, стручне комисије могу основати привремене или сталне радне групе, које обрађују специфичну проблематику из подручја комисије која их је основала.

5. ПРИМЕНА КОРПУСНЕ ЛИНГВИСТИКЕ И УЛОГА КОРПУСА У ТЕРМИНОЛОШКОМ ИСТРАЖИВАЊУ

У покушају да се одговори на питање шта је корпусна лингвистика, јављају се бројне и различите дефиниције корпусних лингвиста. Тим дефиницијама, корпусна лингвистика је у лингвистичким истраживањима описана као алат, метод, методологија, методолошки приступ, дисциплина, теорија, теоријски приступ, парадигма (теоријска или методолошка) као и комбинација свега наведеног (Taylor 2008, 180).

Тако, према дефиницији аутора Лича, који уводи термин „рачунарска корпусна лингвистика“ реферишући се на очигледно и сталнорастуће преклапање корпусне и рачунарске лингвистике након увођења рачунара у ову област, корпусна лингвистика није само методологија за проучавање језика већ представља нови филозофски приступ теми (Leech 1992, 106). И аутор Стабс (Stubbs 1993, 23–24) сматра да корпусну лингвистику не треба дефинисати као методологију закључујући да је таква дефиниција ограничава, с обзиром на то да „корпус није само алатка лингвистичке анализе већ важан концепт лингвистичке теорије“. Тојберт (Teubert 2007, 111) такође описује корпусну лингвистику као младу дисциплину која има сопствену теоријску позадину. Тврдњом да је „у протеклих неколико деценија, корпусна лингвистика постала главна методолошка парадигма у примењеној и теоријској лингвистици“, започиње свој рад аутор Грис (Gries 2006, 191). Тоњини-Бонели (Tognini-Bonelli 2001, 1) описала је корпусну лингвистику као „предупотребну методологију“ која се не држи стриктно утврђених правила већ за разлику од других дисциплина сама дефинише своја правила, што доводи до тога да различити параметри могу бити предмет језичког описа, због чега ова дисциплина и поседује статус теорије. Ауторка Малберг (Mahlberg 2006, 370–371) користи термин „корпусно теоријски приступ“, заступајући став да је језик друштвени феномен и да се значење, самим тим, посматра кроз његову употребу, а да су обрасци језичке употребе видљиви у корпусу. Ова ауторка сматра да се треба користити новим описним алатима како би се сагледала реална ситуација у тексту. Аутори Томпсон и Ханстон (Thompson и Hunston 2006, 8) тврде да у својој основи, корпусна лингвистика представља методологију усклађену са другим теоријским приступима језику. Описујући теорије које произилазе из корпусне лингвистике, ови аутори напомињу да се значење не налази у једној речи већ у „јединицама значења“ чиме се говорни дискурс одмотава кроз низ полуфиксних фраза (Thompson и Hunston 2006, 11–12).

Корпусну лингвистику као методологију дефинишу и Мекенери и Вилсон (McEnery и Wilson 1996, 1) као и Мејер (Meyer 2002, XI–XVI). На сличан начин је дефинишу и аутори Боукер и Пирсон (Bowker и Pearson 2002, 9), као „приступ односно методологију за изучавање употребе језика“. Аутори Макенери и Габриелатос (McEnery и Gabrielatos 2006, 35) сматрају да корпусна лингвистика представља и методологију и теорију наводећи да на методу истраживања одабраног корпуса утиче теоријска оријентација истраживача. Ови аутори илуструју методолошке разлике које су проистекле из теоријских разматрања. Износећи двадесет пет основних поставки корпусне лингвистике, аутор Тојберт (Teubert 2005, 4) бележи да корпусна лингвистика није метод сама по себи, већ да користи много различитих метода приликом обраде и анализе корпусних података. Уз то, овај аутор је описује и као дисциплину (Teubert 2005, 2). Као дисциплину, корпусну лингвистику описују и аутори Артс (Aarts 2002) и Вилијамс (Williams 2006) у својим радовима. Негујући традицију британске школе лингвистике, Стабс описује лингвистику као примењену друштвену науку (Stubbs 1993, 3–4), док је Тојберт сврстава у подручје хуманистике (Teubert 2005, 7–8).

У својим истраживањима, корпусна лингвистика користи велике узорке природног говора – корпусе. Иако постоји велики број дефиниција којима се описује значење речи *корпус*, многи аутори се слажу да, у лингвистичком контексту, корпус представља колекцију текстова односно његових делова, који се могу лингвистички анализирати (Meyer 2002, XI).

Сама лексема *корпус* долази из латинског језика, од именице *Corpus* (ген. *Corporis*), а има више значења (тело, супстанца, материјал, колекција, систем, структура, оквир, збирка писаних радова неког аутора). Према дефиницији коју даје једнотомни Речник српскога језика у издању

Матице српске, корпус (лингвистички) представља збирку текстова, података, докумената. Аутори Витас и Поповић, под корпусом, у лингвистичком смислу, подразумевају емпиријски материјал намењен истраживању језика (Витас и Поповић 2003, 222).

Пре него што се почео употребљавати у лингвистичкој терминологији, термин корпус се користио у значењу колекције писаног текста одређене тематике. Како наводе аутори Макарти и О'Киф (McCarthy и O'Keefe 2010, 5), Оксфордски речник енглеског језика потврђује да се лексема корпус користила у овом контексту у 18. веку, тако да су се научници могли реферисати на „корпус латинских песама“ или „корпус закона“. Први цитат Оксфордског речника лексеме корпус у лингвистичкој литератури датира из 1956. године, у раду В. С. Алена под називом *Transactions of the Philological Society* („Трансакције филолошког друштва“), где се користи у значењу „материјала писане односно говорне грађе на којима се заснива лингвистичка анализа“ (Оксфордски речник, друго издање, 2009. година) (McCarthy и O'Keefe 2010, 5). Иако се сматра да су аутори Артс и Меијс (енг. *Aarts & Meijs*) заслужни за ковање овог термина радом објављеним 1984. године под називом *Corpus Linguistics: Recent Developments in the Use of Computer Corpora in English Language Research*, термин корпусна лингвистика био је коришћен и раније, и то у раду Артса и ван ден Хеувела (енг. *Aarts & van den Heuvel*), под називом *Grammars and intuitions in corpus linguistics*, објављеном 1982. године (Taylor 2008, 179).

Како наводе аутори Макарти и О'Киф, корпусна лингвистика се данас највише повезује са електронским претраживањем линија конкорданци и речи, у циљу истраживања одређених језичких појава које се јављају у великим збиркама текстова. Овај метод, којим се детаљно претражују речи и фразе кроз огроман број текстова, практикован је још у 13. веку, када су библијски учењаци са својим сарадницима прегледали страницу по страницу Библије и ручно индексирали речи. Индексирање је настало из практичне потребе да се за друге библијске учењаке, у алфаветском реду напишу речи које се налазе у Библији, уз цитирање на ком месту су се појавиле (McCarthy и O'Keefe 2010, 3).

Данас је јасно да је за развој корпуса заслужан, пре свега, технолошки напредак. Савремена језичка истраживања подразумевају употребу рачунара који тај рад чине далеко бржим, ефикаснијим и прецизнијим. Иако је рад на првим рачунарима био изузетно тежак, потенцијал који је носила електронска ера, био је јасан од првих дана њихове употребе. Рад на текстовима уз употребу рачунара отпочео је свештеник Роберто Буза педесетих година 20. века, израдом електронског лематизованог индекса целокупног стваралаштва Томе Аквинског, под називом *Index Thomisticus*.⁷⁹ Био је то први пројекат пребацавања текста у електронску форму, који је завршен седамдесетих година 20. века (Tognini Bonelli 2010, 15).

5.1. Развој корпуса

Када говоримо о развоју корпуса, важно је пре свега направити основну разлику у погледу његове врсте. Тако разликујемо пределектронски и електронски корпус.

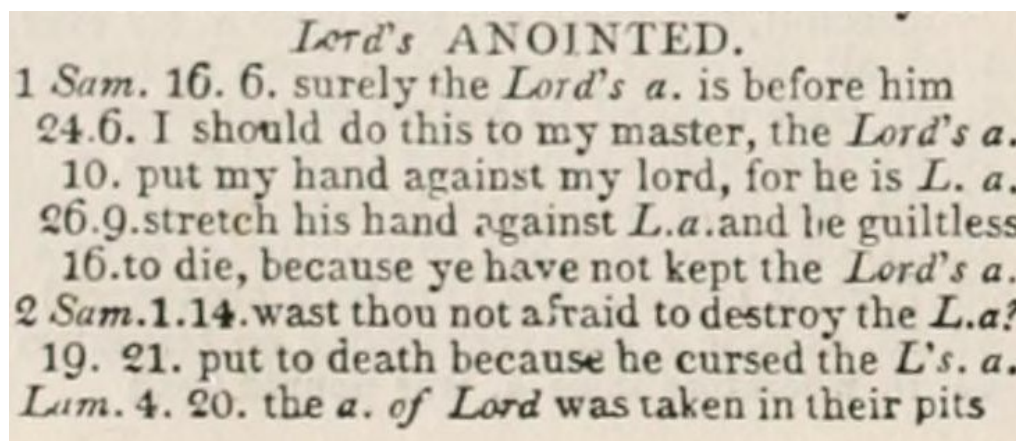
5.1.1. Пределектронски корпус

Настао је пре појаве рачунара, а састојао се од текстова на којима се вршило одређено језичко истраживање, најчешће спровођено кроз дуготрајну и веома захтевну ручну анализу (Meuer 2008, 1). Овакви пројекти ретко су били спровођени зарад општих лингвистичких истраживања већ су углавном имали одређени, специфични циљ (Utvić 2013, 3).

БИБЛИЈСКЕ КОНКОРДАНЦЕ представљају прво значајно дело засновано на корпусном истраживању језика. Састављане су на многим језицима, између осталог на латинском, грчком, хебрејском и енглеском, а неке од значајнијих су Конкорданце кардинала Ига на латинском језику, Латинске конкорданце Библије састављене у 13. веку, Јеврејске конкорданце из 15. века

⁷⁹ Овај корпус доступан је на интернет страници: <http://www.corpusthomaticum.org/it/index.age> (септембар 2019. године).

чији је аутор Ајсак Натан бен Калонимус (познат и под именом Раби Мордекаи Натан), као и две збирке конкорданци на енглеском језику – конкорданце Џона Марбека из 15. века и Александра Крудена из 18. века (Meuser 2008, 1–2). Круденове конкорданце представљају најамбициознији и најобухватнији подухват, спроведен над ауторизованом верзијом Библије краља Џејмса. Објављене су први пут 1736. године, а до 1879. године биле су објављиване у чак четрдесет и два издања (Kennedy 1998, 14). По свом обиму, ове конкорданце су веће и од саме Библије – садрже око 2.370.000 речи, а све стога што су поред појединачних речи обухватале и одређене колокације (Meuser 2008, 2). Уз то, Круден није лематизовао речи, већ је засебно уносио облике исте речи: *anoint, anointed, anointed ones, anointing, his anointed, Lord's anointed, mine anointed, anointed with oil, thine anointed, anointedst, anointing oil* (на слици 11 приказана је једна од поменутих одредница из Круденових конкорданци⁸⁰). Завршене су за изузетно кратко време, а како бележи аутор Мејер, док је на справљању конкорданци кардинала Ига учествовало пет стотина монаха, Крудену је за комплетирање његових конкорданци, на којима је свакодневно радио осамнаест сати, требало само две године. Да би их направио, сваку одредницу је ручно алфабетизовао уз помоћ оловке и папира, што је био огроман посао. Циљ оваквог прегалачког рада није било само Круденово интересовање за језик, већ лакши приступ Библији који би се омогућио свим људима (Meuser 2008, 2).



Слика 11. Пример одреднице из Круденових конкорданци

Уз то, верује се да су неке од најстаријих познатих ГРАМАТИКА класичних језика засноване управо на корпусу. Таква је, на пример, *Панинијева граматика* санскритског језика из 4. века пре нове ере, која описује језик Веда којим се већ тада увелико није говорило, а који је био доступан једино преко збирке односно корпуса сачуваних текстова. Такво је и дело *О неграматичким речима* чији је аутор Аристоник из Александрије, грчки научник из 1. века нове ере. Дело је написано у шест књига а описује неправилне граматичке конструкције у корпусу Хомерових текстова (Meuser 2008, 3).

Употреба корпуса у лексикографији има веома дугачку традицију, пре свега због тога што је корпус био главни извор илустративних цитата чија је функција била да помогне лексикографима приликом одређивања значења речи унутар неког контекста, а потом и да илуструје значење речи у састављаном речнику (Meuser 2008, 6). Коришћење корпуса за лексикографске сврхе забележено је још почетком 17. века када је Самјуел Џонсон, користећи се вишедеценијским радовима својих претходника, на папирима забележио огроман корпус реченица значајнијих писаца, како би илустровао значење и употребу енглеских речи. Помагало му је шест асистената у прикупљању преко 150.000 илустративних цитата за отприлике 40.000 одредница *Речника енглеског језика* (Kennedy 1998, 14). Овај речник је методолошки утицао на

⁸⁰ Круденове конкорданце доступне су на интернету, на страници <https://archive.org/details/crudenscompletec00crud/page/n7> (септембар 2019.)

многе речнике који су се касније појављивали, а нарочито на *Оксфордски речник енглеског језика*,⁸¹ иначе највећи речник икада објављен. Овај огроман корпус једно је од највећих лингвистичких подухвата у енглеском језику, а може се рећи и у свим језицима, а направљен је без асистенције рачунара. Паралелно са овим речником, у другом делу 19. века, сакупљена је још једна велика колекција цитата како би се употпунио Вебстеров *Амерички речник енглеског језика*,⁸² објављен 1828. године. Треће издање Вебстеровог *Новог међународног речника*,⁸³ које је објављено 1961. године, садржало је преко 10 милиона цитата којима се потврђивало и илустровало значење и употреба за готово пола милиона речи колико их је садржавао. Ово издање је вероватно последњи велики енглески речник који је завршен без помоћи електронских база података (Kennedy 1998, 14–15). Ипак, Мејер (Meuer 2008, 6–7) верује да је прва употреба корпуса у лексикографске сврхе забележена још у латинским и грчким речницима у 16. веку, као и у италијанско-енглеском речнику аутора Џона Флориа, 1598. године.

Интересовање за лингвистичке варијације у регионалним дијалектима такође је довело до тога да одређени лингвисти 19. века прикупе и анализирају корпусе различитих врста (Kennedy 1998, 15–16). Велики број радова био је заснован на лексичким варијацијама у избору речи за одређене појмове, као и на могућим варијантама одређених речи, како у правопису тако и у изговору, наводи Кенеди. *Енглески дијалекатски речник*⁸⁴ као и *Постојећа фонологија енглеских дијалеката*⁸⁵ били су два монументална резултата специјализованих студија лексичких варијација у дијалектима у Уједињеном краљевству, која су базирана на корпусима (Kennedy 1998, 15–16).

Једно од утицајнијих корпусних истраживања у првој половини 20. века поставио је Торндајк,⁸⁶ који је сакупио колекцију од 4,5 милиона речи из 41 различитог извора, како би направио листу учесталости речи којом би се побољшао наставни програм за описмењавање матерњих говорника енглеског језика у САД. Три четвртине овог корпуса су речи из Библије и класичних дела енглеске фантастике објављених у 19. веку, док остатак корпуса чине писма, новине и школска литература (Kennedy 1998, 16). Током тридесетих година 20. века, Торндајков корпус је допуњен, па је тако достигао број од 18 милиона речи из ширег спектра текстуалних извора. Лексичке анализе овог корпуса и објављена дела у којима се користио Торндајков корпус били су изузетно утицајни у области подучавања енглеског језика у многим деловима света преко тридесет година. Пратећи његов пионирски подухват, и у другим државама су састављани пределектронски корпуси у сврху подучавања, и то за холандски, француски, немачки, италијански, латински, руски и шпански језик (Kennedy 1998, 16).

Пре овог корпуса, како бележи Кенеди (Kennedy 1998, 16), у Немачкој је састављен корпус чији је циљ био сакупљање статистичких информација о употреби речи и слова у немачком језику како би се побољшала обука стенографа, чија је улога у ери пре појаве диктафона и осталих помагала, била кључна у бележењу дневних дискусија и одлука које се доносе на влади и у пословању. Немачки истраживач Кединг⁸⁷ имао је дугогодишњу помоћ асистената, којих је било преко пет хиљада, како би обрадио корпус од преко 11 милиона речи које је користио у својој анализи деведесетих година 19. века (Kennedy 1998, 16).

Иако се сматра да је корпусно истраживање отпочело раних шездесетих година 20. века, када је наступила ера електронских, машински читљивих корпуса, и пре њихове појаве је, као

⁸¹ *Oxford English Dictionary* – први том објављен 1884. године, а последњи, двадесети 1928. године, након 71 године непрекидног рада на корпусу.

⁸² *An American Dictionary of the English Language*, аутора Noah Webster.

⁸³ *New International Dictionary of the English Language* аутора Noah Webster.

⁸⁴ *The English Dialect Dictionary* (аутор Joseph Wright, објављено 1898–1905).

⁸⁵ *The Existing Phonology of English Dialects* (аутор Alexander J. Ellis, објављено 1889).

⁸⁶ Edward Lee Thorndike (1874–1949), амерички психолог, који је дао значајан допринос у области компаративне психологије и изучавања процеса учења, односно развоју научно утемељеног и ефикаснијег начина образовања (Encyclopedia Britannica 2019).

⁸⁷ Friedrich Wilhelm Kaeding (1843–1928), немачки стенограф, заслужан за развој стенографије и члан различитих стенографских удружења. Аутор је бројних радова из области стенографије.

што видимо, постојала традиција у истраживањима у сфери корпусне лингвистике, и то нарочито у библијским и студијама књижевности, лексикографији, дијалекатским, језичким и граматичким студијама (Kennedy 1998). И поред тога што влада уверење да је мануелна језичка анализа, која се спроводила коришћењем пределектронских корпуса, застарела и непотребна, управо такви, пределектронски корпуси, имали су кључан утицај на развој корпусне лингвистике какву познајемо данас (Meuser 2008, 12–13). И док данас постоји велики број доступних софтверских алата за стварање конкорданци, такви програми никада не би постојали да људи попут Александра Крудена пре тога нису концептуализовали појам конкорданци, закључује Мејер. Пределектронска корпусна лингвистика има богату историју која је допринела развоју ове дисциплине (Meuser 2008, 13).

5.1.2. Електронски корпус

Електронски корпуси саставни су део модерног доба, настали као последица рачунарске револуције, а почели су се користити шездесетих година 20. века, када се појавио први рачунарски корпус (Meuser 2008, 1). Задаци који су, пре њихове употребе, захтевали огроман људски капацитет, а какво је нпр. састављање листе учесталости и конкорданци, уз употребу рачунара могли су се са лакоћом обавити (Johansson 2008, 33). У другој половини 20. века њихове перформансе су побољшаване, постајали су све доступнији у финансијском и употребном смислу, а данас је употреба рачунара у језичким (као и осталим) истраживањима стандард. Са прогресом рачунара, развијао се и корпус, што је најочљивије у погледу раста његове величине. Постепено унапређење корпуса (подразумева се, електронских) у последњих педесетак година, условило је њихово даље разврставање и то на корпусе прве, друге и треће генерације. Корпуси прве генерације који су бројали око милион речи данас се обично сматрају корпусима узорцима, корпуси друге генерације обично броје између 7 и 30 милиона речи, док се под савременим корпусима, тзв. корпусима треће генерације, подразумевају корпуси који садрже стотине милиона речи (а наравно, данас се то мери гига речима, енг. *Gigaword corpus*) и који најчешће настају као нуспроизвод савремених електронских комуникационих система (Bratanić 1998, 173).

Већ је било речи о томе да је рад на тексту уз помоћ рачунара најпре отпочео свештеник Буза још педесетих година 20. века (реч је о електронском лематизованом индексу комплетног стваралаштва Томе Аквинског, који је завршио 1978. године).

Први електронски корпус у савременом смислу (узорак текстова), међутим, настао је на Универзитету Браун у САД, шездесетих година 20. века. Име овог корпуса је Стандардни корпус савременог америчког енглеског језика (*The Standard Corpus of Present-Day American English*), познат и под називом КОРПУС УНИВЕРЗИТЕТА БРАУН⁸⁸. Његови творци били су Хенри Кучера (енг. *Henry Kucera*) и Винтроп Нелсон Франсис (енг. *Winthrop Nelson Francis*). Овај корпус садржи готово тачно милион речи (500 текстова од којих сваки текст има отприлике 2000 речи), прикупљених 1961. године из штампаних извора са америчког подручја. Како би узорак био што бољи, текстови за корпус су прикупљени из 15 различитих текстуалних категорија. Намера аутора била је да створи уравнотежени корпус, тј. корпус који даје коректну слику једног језика (Lindquist 2009, 3). Иако се данас сматра малим, овај корпус је служио као модел у контексту плана израде других корпуса. Табелом 7 приказан је детаљан садржај Корпуса Универзитета Браун (Lindquist 2009, 4).

⁸⁸ Све информације о Корпусу Универзитета Браун доступне су на следећим веб-сајтовима:

<http://clu.uni.no/icame/manuals/BROWN/INDEX.HTM#bc3> и

https://www1.essex.ac.uk/linguistics/external/clmt/w3c/corpus_ling/content/corpora/list/private/brown/brown.html (септембар 2019).

Табела 7. Садржај Корпуса Универзитета Браун према типу текста, уз приказ њиховог броја и удела у корпусу

Тип текста		Број текстова	Удео корпуса у процентима
<i>A</i>	Штампа: репортаже (политика, спорт, друштво, новости, финансије, култура)	44	8,8
<i>B</i>	Штампа: уређивачки текстови (укључујући писма уреднику)	27	5,4
<i>C</i>	Штампа: прегледи (позориште, књиге, музика, плес)	17	3,4
<i>D</i>	Религија (књиге, периодика, трактати)	17	3,4
<i>E</i>	Вештине и хоби (књиге, периодика)	36	7,2
<i>F</i>	Традиционално знање	48	9,6
<i>G</i>	Књижевноуметнички текстови, биографије, мемоари и сл.	75	15,0
<i>H</i>	Разно (углавном државна документа)	30	6,0
<i>I</i>	Научни текстови	80	16,0
<i>J</i>	Општи жанр (романи и кратке приче)	29	5,8
<i>K</i>	Мистерија и детективски жанр	24	4,8
<i>L</i>	Научна фантастика	6	1,2
<i>M</i>	Авантуристички и вестерн жанр	29	5,8
<i>N</i>	Романса и љубавне приче	29	5,8
<i>O</i>	Хумор	9	1,8
	Некњижевни тип текста (<i>A–J</i>)	374	75
	Књижевни тип текста (<i>K–R</i>)	126	25
	Укупно	500	500

Избор текстова вршен је на основу њихове категорије и величине, случајним узорковањем. Прецизне информације о избору узорака и кодирању текстова дате су у Приручнику корпуса, чији су аутори Франсис и Кучера.⁸⁹

КОРПУС ЛАНКАСТЕР–ОСЛО/БЕРГЕН (*The Lancaster-Oslo/Bergen Corpus*, скр. *LOB*), представља британску верзију Корпуса Универзитета Браун. И овај корпус садржи 500 текстова и за сваки од текстова по отприлике 2000 речи распоређених у 15 различитих категорија, а постоји у две верзије: оригиналној верзији и аотираној верзији. Текстови који датирају из 1961. године, бирани су случајним узорковањем.⁹⁰

И КОРПУС КОЛХАПУР (*Kolhapur Corpus*) представља још једну верзију Корпуса Универзитета Браун, и то верзију на индијско-енглеском језику, с тим што су текстови прикупљени за овај корпус штампани 1978. године, а не 1961. као за претходна два (Leitner 1991, 220). Постоји још три верзије Корпуса Универзитета Браун, и то аустралијско-енглеска (енг. *Australian Corpus of English*, скр. *ACE*), која се састоји од текстова штампаних 1986. године, новозеландско-енглеска (енг. *Wellington Corpus of Written New Zealand English* скр. *WWC*), која садржи текстове штампане у периоду 1986–1990. године и мандаринско-кинеска (енг. *Lancaster Corpus of Mandarin Chinese*, скр. *LCMC*), са текстовима штампаним пар година пре и после 1991. године, и који у комбинацији са Фрајбуршким корпусима (скр.

⁸⁹ Приручник Корпуса Универзитета Браун доступан је на: <http://clu.uni.no/icame/manuals/BROWN/INDEX.HTM> (октобар 2019. године).

⁹⁰ Подаци у вези са овим корпусом преузети су септембра 2019. године са: <http://www.helsinki.fi/varieng/CoRD/corpora/LOB/>

FLOB/Frown корпусима, о којима ће бити више речи у овом поглављу) представља драгоцен извор у истраживањима контрастних студија кинеског и два главна варијетета енглеског језика.⁹¹

Доста дуго, Корпус Универзитета Браун и Корпус Ланкастер–Осло/Берген били су једини лако доступни рачунарски читљиви корпуси. Стога је и велики број истраживања из области корпусне лингвистике и заснован управо на њиховим подацима.

Технолошки напредак седамдесетих и осамдесетих година прошлог века учинио је да рачунари постану снажнији, јефтинији и једноставнији за коришћење, што је допринело и развоју интересовања за језичка истраживања односно пробоју корпусне лингвистике у овом периоду (Johansson 2008, 33). У овом периоду порастао је број машински читљивих текстова, појавили су се и први скенери, па је било могуће скенирати штампани текст. До краја седамдесетих година прошлог века, Јохансон бележи (Johansson 2008, 40), према наводима аутора Бергенхолц и Шедер (Bergenholtz и Schaefer 1979, 325–329), да су се појавила три енглеска и 14 немачких корпуса чија је величина варијала између 200.000 до 5 милиона речи. У истраживању спроведеном над енглеским машински читљивим корпусима почетком деведесетих година прошлог века (Taylor, Leech и Fligelstone 1991, 320–321), показало се да су корпуси разнолики, па тако међу њима има оних који садрже стручни тип текстова, различите варијетете енглеског језика, језик деце, историјске текстове, а затим се јављају вишејезични корпуси, корпуси са граматичким апликацијама итд. (Johansson 2008, 40). Међу њима је и Корпус Бирмингем од 20 милиона речи који представља само део укупне колекције текстова (Бирмингамске колекције енглеских текстова), која садржи преко 40 милиона речи (Taylor, Leech и Fligelstone 1991, 325–326). Рад на овом пројекту под називом Међународна језичка база података Колинса и Универзитета у Бирмингему (енг. *Collins Birmingham University International Language Database*, скр. *COBUILD*), започет је још осамдесетих година прошлог века уз сарадњу издавачке куће Колинс и Универзитета у Бирмингему. На челу пројекта био је Џон Синклер, професор савременог енглеског језика на истоименом универзитету. Пројекат израде речника *Collins Cobuild English Language Dictionary* који је објављен 1987. године, заснован је на овом корпусу (Sinclair 1990, 184).

ФРАЈБУРШКИ КОРПУСИ британског односно америчког енглеског језика (*The Freiburg-LOB Corpus of British English*, скр. *F-LOB* и *The Freiburg-Brown corpus of American English*, скр. *Frown*)⁹², представљају новију верзију Корпуса Ланкастер–Осло/Берген односно Корпуса Универзитета Браун. Рад на њиховом састављању почео је 1991. године. Циљ је био да се оба корпуса подударе по величини и композицији са својим претходницима, с једином разликом у погледу времена настанка текстова – у овим корпусима, наиме, представљен је језик раних деведесетих година. Као и оригинални Браун и *LOB* корпуси, и Фрајбуршки корпуси садрже 500 текстова од отприлике 2000 речи дистрибуираних у 15 текстуалних категорија. Вршен је пажљив одабир текстова како би се што боље подударало са ова два корпуса, а све у циљу да се лингвистима омогући емпиријска основа за проучавање језичких промена насталих током времена. Постоје две верзије овог корпуса, оригинална и верзија аотирана врстом речи чији су аутори Џефри Лич (енг. *Geoffrey Leech*)⁹³ и Ник Смит (енг. *Nicholas Smith*)⁹⁴.

Уз ове корпусе, не треба изоставити ни Корпус Лондон–Лунд (*The London-Lund Corpus of Spoken English* скр. *LLC*) који је произишао из два пројекта: Истраживања употребе енглеског језика (*Survey of English Usage*, скр. *SEU*) и Истраживања говорног енглеског језика (*Survey of Spoken English*, скр. *SSE*). Корпус се састоји од 500.000 речи транскрибованог говора британског енглеског језика. Израда овог корпуса започета је 1975. године под руководством Јана Свартвика

⁹¹ Подаци у вези са овим корпусом преузети су октобра 2019. године са:

https://www.lancaster.ac.uk/fass/projects/corpus/cbils/corpora.asp#_Toc92298887

⁹² Подаци преузети септембра 2019. године са: <http://www.helsinki.fi/varieng/CoRD/corpora/FLOB/index.html> и <http://www.helsinki.fi/varieng/CoRD/corpora/FROWN/index.html>.

⁹³ *Geoffrey Leech* (1936–2014), био је професор на Универзитету у Ланкастеру, на Одсеку за лингвистику и енглески језик. Истакао се на пољу стилистике, семантике, прагматике и корпусне лингвистике.

⁹⁴ *Nicholas Smith*, ванредни је професор примењене лингвистике на Универзитету Лестер, а главна област његовог интересовања је проучавање корпуса (писаних и говорних текстова) на енглеском језику.

(швед. *Jan Svartvik*).⁹⁵ Примарни циљ овог корпуса била је дигитализација говорног корпусног материјала како би постао доступан у машински читљивом облику (Johansson 2008, 39).

Последња деценија 20. века бележила је, уз све већи број корисника, све једноставнији приступ великом броју машински читљивих текстова и развој нових алатки за анализу текста (попут програма за конструисање конкорданци, софтвера за анализу текста итд.), а употреба се толико раширила да је обухватала широк спектар комплексних лингвистичких истраживања у области синтаксе, лексике, дискурса, језичких варијација итд. (Johansson 2008, 49).

Уз ове корпуре, ваља споменути и мрежу корпуса ЛОНГМАН⁹⁶, огромну базу података која садржи 330 милиона речи и која је драгоцен извор информација о речима, њиховој употреби, језичким трендовима и граматичким обрасцима. Корпус се састоји од пет језичких база:

- Корпус енглеског језика Лонгман/Ланкастер (*Longman/Lancaster Corpus*) садржи више од 30 милиона речи прикупљених из различитих писаних извора: књижевности, часописа, радова, па чак и са летака, амбалаже и аутобуског реда возње.
- Лонгманов амерички говорни корпус (енг. *Longman Spoken American Corpus*), садржи 5 милиона речи из свакодневног говора, прикупљених на узорку говорника различитог узраста, нивоа образовања, различитог порекла и из различитих држава унутар САД.
- Лонгманов амерички корпус писаног текста (енг. *Longman Written American Corpus*), садржи 100 милиона речи из новинских и књижних извора.
- Говорни британски корпус, који представља део Британског националног корпуса, даје информације о томе у којој мери се писани британски енглески језик разликује од говорног британског језика.
- Лонгманов корпус енглеског језика као нематерњег (енг. *Longman Learners' Corpus*), од 10 милиона речи. Ова база садржи различите писане изворе, попут есеја и студентских скрипти, прикупљене од студената из целог света, различитих националности и различитих нивоа знања језика чиме се добија недвосмислен увид у учење енглеског језика и различите проблеме и грешке који се том приликом јављају.

Још један од веома значајних савремених корпуса је КОРПУС ЕНГЛЕСКОГ ЈЕЗИКА⁹⁷. Овај корпус обједињава бројне корпуре (тј. поткорпуре, колекције, збирке) енглеског језика са различитих говорних подручја и из разноврсних извора, као што је приказано табелом 8. Константно се унапређује, а најновија колекција придружена осталим јесте Корпус коронавируса⁹⁸ који садржи текстове од јануара 2020. године и која се свакодневно допуњава. Табелом 8 приказани су корпуси који чине саставни део Корпуса енглеског језика и то према броју речи, дијалекту, периоду настанка и жанровима.

⁹⁵ Подаци преузети октобра 2019. године са: <http://www.helsinki.fi/varieng/CoRD/corpora/LLC/>.

⁹⁶ Подаци преузети октобра 2019. године са: <http://longmandictionariesusa.com/longman/corpus> и <http://www.pearsonlongman.com/dictionaries/corpus/lancaster.html>.

⁹⁷ Сви подаци у вези са Корпусима енглеског језика преузети су са веб-сајта: <https://www.english-corpora.org/> јуна 2020. године.

⁹⁸ <https://www.english-corpora.org/corona/>

Табела 8. Корпус енглеског језика⁹⁹

Корпус	Број речи	Дијалекат	Период настанка	Жанр(ови)
<i>iWeb</i> : напредни веб-корпус ¹⁰⁰	14 млрд.	6 држава	2017	веб
Корпус вести са веба (енг. <i>News on the Web</i> , скр. <i>NOW</i>)	12 + млрд.	20 држава	2010–јуче	веб: вести
Корпус коронавируса	870 + мил.	20 држава	јануар 2020–јуче	веб: вести
Корпус енглеског језика као глобалног са веба (<i>GloWbE</i>)	1.9 млрд.	20 држава	2012–2013	веб (укључујући и блогове)
Корпус <i>Wikipedia</i>	1.9 млрд.	(разно)	2014	<i>Wikipedia</i>
Корпус савременог америчкоенглеског језика (<i>COCA</i>)	1 млрд.	амерички	1990–2019	балансиран
Корпус историјског америчкоенглеског језика (<i>COHA</i>) ¹⁰¹	400 мил.	амерички	1810–2009	балансиран
Телевизијски корпус	325 мил.	6 држава	1950–2018	телевизијске емисије
Филмски корпус	200 мил.	6 држава	1930–2018	филмови
Корпус америчких сапуница	100 мил.	амерички	2001–2012	телевизијске емисије
Корпус <i>Hansard</i> ¹⁰²	1.6 млрд.	британски	1803–2005	парламент
Корпус <i>Early English Books Online</i> ¹⁰³	755 мил.	британски	1470–1690	(разно)
Корпус правних мишљења врховног суда САД	130 мил.	амерички	1790–данас	правна мишљења
Корпус магазина <i>TIME</i>	100 мил.	амерички	1923–2006	часописи
Британски национални корпус (<i>BNC</i>)	100 мил.	британски	1980–1993	балансиран

⁹⁹ Табела 8 је преузета и прилагођена са сајта Корпуса енглеског језика, а илуструје број речи, дијалекат заступљен у корпусу, период у ком је корпус настао као и жанр који је тим корпусом покривен.

¹⁰⁰ *iWeb* корпус подржава три начина претраживања: преко листе учесталости, путем речи као и путем фраза и ниски речи.

¹⁰¹ Овај корпус садржи текстове објављене у периоду од 1810–2000. године.

¹⁰² Корпус Хансард садржи скоро сваки говор одржан у британском парламенту од 1803–2005. године, и омогућава претраживање ових говора на начине који нису могући за неке друге ресурсе.

¹⁰³ Овај корпус садржи текстове дела која су штампана у периоду 1473–1700. године на територији енглеског говорног подручја, као и дела која су штампана на енглеском језику на другој територији.

Корпус <i>Strathy</i> (Канада) ¹⁰⁴	50 мил.	канадрски	1970–2000s	балансиран
Корпус <i>CORE</i> ¹⁰⁵	50 мил.	6 држава	2014	веб
Књиге са Гугла – амерички енглески	155 млрд.	амерички	1500–2000s	(разно)
Књиге са Гугла – британски енглески	34 млрд.	британски	1500–2000	(разно)

Из ове табеле уочавамо да је свеукупан број енглеских речи које се појављују у наведеним корпусима, у кумулативном смислу, не узимајући у обзир корпус нити варијанту енглеског језика која је корпусом обухваћена, већи од 222.414 милијарде.

КОРПУС ПРАВНИХ ТЕКОВИНА ЕВРОПСКЕ УНИЈЕ (фр. *ACQUIS COMMUNAUTAIRE*)¹⁰⁶ (видети поглавље 3.4.3.3) представља још једну веома значајну збирку текстова. Реч је о једном од највећих паралелизованих вишејезичних колекција законодавних текстова, и то на чак 24 језика, колико Европска унија тренутно званично и броји. Правне тековине ЕУ представљају корпус права и обавеза који важи за све државе чланице ЕУ, а подразумевају: примарно право – оснивачке уговоре, међународне уговоре, секундарно право – које се односи на институције ЕУ, праксу Суда правде као и остале обавезе које се преузимају од држава чланица у оквиру активности ЕУ. Државе које имају статус кандидата у обавези су да преведу и усвоје правне тековине Европске уније. Стога се ова база текстова непрестано увећава. Правне тековине Европске уније подељене су на 35 преговарачких поглавља којима су покривени различити предметни домени попут финансијских услуга, политике конкуренције, слободног кретања робе, радника и капитала и осталог, као што је приказано табелом 9.

Табела 9. Листа преговарачких поглавља ЕУ

Поглавље 1. Слободно кретање робе
Поглавље 2. Слободно кретање радника
Поглавље 3 . Право оснивања предузећа и слобода пружања услуга
Поглавље 4. Слободно кретање капитала
Поглавље 5. Јавне набавке
Поглавље 6. Право привредних друштава
Поглавље 7. Право интелектуалне својине
Поглавље 8. Политика конкуренције
Поглавље 9. Финансијске услуге
Поглавље 10. Информационо друштво и медији
Поглавље 11. Пољопривреда рурални развој
Поглавље 12. Сигурност хране, ветеринарска и фитосанитарна политика
Поглавље 13. Рибарство
Поглавље 14. Транспортна политика
Поглавље 15. Енергетика
Поглавље 16. Опорезивање
Поглавље 17. Економска монетарна политика
Поглавље 18. Статистика

¹⁰⁴ Овим корпусом обухваћена је канадска варијанта енглеског језика а грађа је прикупљена из говорних текстова, белетристике, популарне штампе, новина и академских текстова.

¹⁰⁵ Корпус *CORE* садржи текстове са интернета који су категорисани према већем броју различитих жанрова.

¹⁰⁶ Информације о правним тековинама Европске уније преузете су октобра 2020. године са веб-сајтова: https://web.archive.org/web/20070127161836/http://wt.jrc.it/lt/Acquis/JRC-Acquis.2.2/doc/README_Acquis-Communautaire-corpus_JRC.html и <https://web.archive.org/web/20060619034515/http://langtech.jrc.it/JRC-Acquis.html>

Поглавље 19. Социјална политика и запошљавање
Поглавље 20. Предузетништво и индустријска политика
Поглавље 21. Трансевропске мреже
Поглавље 22. Регионална политика и координација структурних инструмената
Поглавље 23. Правосуђе и основна права
Поглавље 24. Правда, слобода и безбедност
Поглавље 25. Наука и истраживање
Поглавље 26. Образовање и култура
Поглавље 27. Животна средина
Поглавље 28. Заштита потрошача и заштита здравља
Поглавље 29. Царинска унија
Поглавље 30. Спољни односи
Поглавље 31. Спољна, безбедносна и одбрамбена политика
Поглавље 32. Финансијски надзор
Поглавље 33. Финансијске и буџетске одредбе
Поглавље 34. Институције
Поглавље 35. Остало

Документа која чине правне тековине ЕУ имају јединствену ознаку која се назива целекс кôд (енг. *celex code*). Ова ознака се састоји из четири дела: прва цифра означава број сектора којем документ припада (1–уговори, 2–међународни споразуми, 3–законодавство, итд.), други број је четвороцифрени и представља годину настанка документа, словна ознака упућује на тип документа (унутар сектора означеног бројем 3, постоји три типа докумената који су означени са R–уредба, L–директива и D–одлука), и последња, четврта ознака представља јединствени број самог документа.

Овај корпус садржи више од 8.000 докумената по језику, са просечном величином броја речи од 9 милиона по језику. Сви документи су јавно доступни на веб-сајту Еурлекса (<https://eur-lex.europa.eu/>). Тим задужен за језичке технологије Заједничког истраживачког центра (енг. *Joint Research Centre*, скр. *JRC*) Европске комисије је документа правних тековина ЕУ конвертовао у *XML* формат, а текстове је очистио и појединачно паралелизовао све могуће комбинације језичких парова. Због великог броја паралелних текстова, овај корпус је погодан за спровођење свих врста језичких истраживања, као и за тестирање и упоређивање софтвера за анализу текста на различитим језицима (нпр. за паралелизовање текстова, поделу реченица, екстракцију термина, итд.) (Steinberger, и други 2006, 2142).

5.2. Класификација корпуса

Мада се, у општем смислу, корпус описује као збирка текстова намењена проучавању језика, оваква дефиниција прилично је широка и врло непрецизна. Прецизно одређивање појма корпуса подразумева утврђивање критеријума који би раздвојили корпус као емпиријску основу језичких истраживања од других насумично изабраних колекција докумената (Витас и Поповић 2003, 222).

Да би се тачно одредио појам корпуса, аутори Витас и Поповић предлажу коришћење одређених типолошких параметара којима би се једна колекција текстова описала као корпус са одређеним скупом атрибута. Овакви параметри, којима се класификује један корпус, могу се поделити према (Витас и Поповић 2003, 223):

1. носачу: пределектронски или електронски.
2. намени: лексикографски, граматички, намењени филолошким истраживањима, препознавању говора, итд.
3. домену: општи или специјализовани.

Док општи корпуси представљају основу за произвољно лингвистичко истраживање, те стога могу имати различите намене, односно користити се за лексикографска, граматичка,

семантичка, прагматичка, социолингвистичка и друга истраживања, специјализовани корпуси настају зарад одређеног специфичног лингвистичког истраживања (лексикографски корпуси, граматички корпуси, дијалекатски корпуси, регионални корпуси, нестандартни корпуси, корпуси језика као нематерњег итд.) или као помоћно средство у рачунарској лингвистици односно обради природног језика (корпуси за тренирање алата за аутоматску морфосинтаксичку анализу текста, аутоматско препознавање и генерисање говора, итд.) (Utvić 2013, 24–25).

4. периоду: синхрони или дијахрони.

Под синхроним се подразумевају они корпуси који представљају језик у датом, одређеном тренутку, док дијахрони (који се назива и историјским корпусом) представљају дужу перспективу, па тако садрже текстове из различитих временских периода (чак и различитих векова). Дијахрони корпуси омогућавају праћење промена у еволуцији језика (Xiao 2008, 401). Корпуси *LOB/FLOB* и *Brown/Frown* (о којима је било речи у поглављу 5.1.2), а који датирају из 1961. односно 1991/1992. године, један су пример дијахроних корпуса јер представљају тридесетогодишњу перспективу (Lindquist 2009, 167–168). С временом сваки корпус се може користити за историјске студије, међутим, у строгом смислу, историјски корпус представља управо онај који је створен у сврху проучавања језичких промена односно истраживања фаза кроз које је током времена пролазио један језик (Claridge 2008, 242).

5. извору: корпус писаних, разговорних или електронских (нрп. електронска пошта и сл.) текстова (Витас и Поповић 2003, 222).

Док корпуси писаних текстова, према наводима ауторке Хунт, обухватају публиковане и непубликоване материјале, корпуси говорних текстова су транскрибовани материјали који су, исто као и писани, представљени у текстуалном облику (Hundt 2008, 168). Имајући у виду да је технички врло компликована, а самим тим и скупља, колекција говорних корпуса одувек је заостајала за писаном (Lindquist 2009, 11). Електронска пошта, форуми, блогови, друштвене мреже само су неки од извора електронских текстова (Utvić 2013, 26). Уз њих, издвајају се и тзв. *мултимодални* корпуси који, према наводима аутора Олвуд (Allwood 2008, 208) представљају дигитализовану колекцију језичког и комуникационог материјала за који се користи више чулних модалитета, у које спадају гестикалација, додир, поглед, говор. Примери мултимодалних корпуса су дигитализоване колекције текстова илустрованих сликама или дијаграмима или дигитализоване колекције филмова са припадајућим транскрипцијама говора (Allwood 2008, 208).

6. обиму: статички или динамички.

Статички корпуси представљају коначне колекције текстова, непроменљиве величине, док динамички подразумевају колекције које се стално допуњавају новим текстовима како би пратили брзе језичке промене, развој језика и животни циклус неологизама (Xiao 2008, 394). Још један израз који се користи за динамичке корпусе је *monitor corpora* (Xiao 2008, 394). Ови корпуси се на годишњем, месечном а некада чак и дневном нивоу допуњују свежим материјалом, због чега и њихов обим стално расте. Када се говори о статичком корпусу, обично се мисли на корпус који се састоји од узорака текстова, енг. *sample corpus* (Xiao 2010, 150). Један од најпознатијих динамичких корпуса је Корпус савременог америчког енглеског језика (енг. *The Corpus of Contemporary American English*, скр. *COCA*).

7. начину аотирања: с обзиром на морфолошке и синтаксичке аотације.

Аотација корпуса представља придруживање додатних информација оригиналном тексту (Витас и Поповић 2003, 225). У зависности од врсте информације, разликујемо неколико нивоа аотације (Utvić 2011, 37а):

- информације у вези са извором текста (одговарајуће библиографске референце), информације о величини текста (број токена и тип), информације о настанку и ажурирању електронске верзије текста (датум креирања, оригинална верзија

текста, процедуре које се користе приликом преношења у машински читљив облик, одговорна лица за стварање електронске верзије, корекција грешака, итд.), што се све може придружити корпусном тексту.

- подељеност текста према његовој логичкој структури (поглавља, наслови, пасуси, реченице), енг. *structural annotation/markup*.
- свакој речи унутар корпуса могу се придружити следеће информације:
 - врста речи (именица, придев, глагол, итд.), енг. *Part of Speech tagging*, скр. *POS tagging*,
 - лема (номинатив једнине именице, инфинитив глагола, итд.), енг. *lemmatization*,
 - вредности флективних категорија (род, број, падеж, време, глаголски вид, итд.), односно флективни основ и наставци, енг. *grammatical annotation*,
 - творбена основа, префикси, суфикси и инфикси,
 - изговор (акценат),
 - границе слогова,
- придруживање одговарајућег значења сваком токену односно реализација семантичке аотације (енг. *semantic annotation*).
- низу једночланих или вишечланих корпусних речи придружује се заједничка информација о функцији речи (да ли је у питању субјекат, предикат, објекат, глаголска одредба), или се такав низ може обележити као синтагма (именска, придевска, итд.), енг. *parsing* (синтаксичка аотација).
- на нивоу дискурса могу се реализовати релације кореференције (анафорички и катафорички односи) између корпусних речи (енг. *coreference annotation*).
- говорни чин може се аотирати прагматичким односно стилистичким информацијама, енг. *pragmatic annotation* и енг. *stylistic annotation*.

И поред бројних предности које нам пружа аотација корпуса, међу којима је њихова ефикаснија претрага, проналажење изостављених информација (елиминисане слике, табеле, фусноте и сл.), односно поједностављена статистичка анализа корпуса, аотацију корпуса прате и одређени недостаци, због чега се поједини лингвисти противе самој идеји аотације. Управо је то разлика између лингвиста који заступају приступ заснован на корпусу (енг. *corpus-based linguists*) и лингвиста који заступају приступ вођен корпусом (енг. *corpus-driven linguists*). Лингвисти који заступају приступ заснован на корпусу инсистирају на детаљној аотацији корпуса и сматрају да је улога корпуса да тестира, исправља и допуњава постојеће теорије. Лингвисти који заступају приступ вођен корпусом тврде да приликом анализе корпуса треба елиминисати унапред замишљене теорије како би се постулирале језичке категорије које су засноване искључиво на подацима добијених анализом. Стога закључују да је аотација корпуса непотребна, зато што представља једну сасвим одређену анализу корпуса коју обављају аотатори, па би поновљена анализа дала и поновљене резултате туђе анализе (Utvić 2011, 38a).

5.3. Репрезентативност и балансираност корпуса

Репрезентативност узорка представља једну од најважнијих карактеристика које један корпус треба да задовољи. Управо та карактеристика разликује корпус од обичне архиве (насумичне збирке текстова). За разлику од архиве, корпус је направљен тако да представља одређени језик односно језички варијетет (Хиао 2010, 149).

Иако се истраживачи углавном фокусирају на величину узорка као најважнији критеријум у постизању репрезентативности (у смислу количине текстова коју треба укључити у корпус и броја речи по узорку текста), Бајбер сматра да репрезентативност представља обим у којем узорак обухвата целокупан опсег варијабилности унутар једне популације (Biber 1993, 243).

У основи, корпус је узорак језика односно језичког варијетета (тј. популације), а узимање узорка је обавезно приликом састављања сваког корпуса (када су живи језици у питању). У том

смислу, репрезентативност већине корпуса у највећој мери одређују два фактора: опсег жанрова у корпусу односно њихова балансираност и избор текстова за сваки жанр појединачно (тј. узимање узорака) (Xiao 2010, 149).

Без особине репрезентативности, резултат који произилази из одређеног корпуса односи се само и искључиво на њега, а такви резултати нису употребљиви у ширем смислу (Leech 2007, 135).

Да би један корпус био балансиран, он мора садржати примерке варијетета текстова који изражавају што је могуће шири обим морфолошких, синтаксичких и семантичких феномена у језику (Витас и Поповић 2003, 223). Односно, како то наводи Лич (Leech 2007, 136), за корпус кажемо да је балансиран онда када је величина његових поткорпуса (који представљају одређене жанрове) сразмерна тј. пропорционална релативној учесталости појављивања тих жанрова у целокупном текстуалном универзуму тог језика. Балансираност се, дакле, према речима аутора Лича, изједначава са пропорционалношћу.

Бајбер (Biber 1993, 247–248), међутим, одбацује пропорционалност сматрајући да она подразумева узорковање говорних и писаних облика из језичке заједнице сразмерно њиховој припадности демографским класама (нпр. према старости, полу, социоекономској класи, итд.), а то би довело до искривљене слике корпуса, са становишта представљања целокупног спектра језичких варијација. Овај аутор полази од претпоставке да је 90% језичке активности конверзационог карактера, а да конверзација у целини има релативно мало одступање у поређењу са другим језичким варијететима. То би значило да преостали варијетети језика имају малу заступљеност (нпр. статути, телевизијске вести), јер само мали део језичке заједнице учествује у настајању таквих текстова (Biber 1993, 247–248).

Један од принципа са којима се већина аутора слаже је, према наводима Лича, и тај да што је корпус већи и разноврснији у погледу жанрова и осталих језичких варијетета, то је и балансиранији односно репрезентативнији (Leech 2007, 138).

5.4. Вишејезични корпуси

Узимајући у обзир језичка својства корпуса, разликујемо једнојезичне и вишејезичне корпусе (међу које спадају и двојезички корпуси). Једнојезични тј. монолингвални представљају корпусе у којима је заступљен један језик, двојезични тј. билингвални представљају корпусе у којима је заступљено два језика, док је вишејезичним тј. мултилингвалним корпусима заступљено три и више језика.

Вишејезички корпуси могу бити паралелизовани (енг. *parallel corpora*, *aligned corpora*) и упоредиви (компаративни, енг. *comparable corpora*). Под паралелизованим се подразумевају корпуси који паралелно садрже изворни текст и његов превод. Паралелизовани текстови су семантички еквивалентни и упарени до одређеног нивоа препознавања, као што је пасус, реченица или реч (Krstev и Vitas 2009, 495). Паралелизовани текстови чине паралелизовани корпус. Овакви корпуси могу бити једносмерни (с једног језика на други), двосмерни (када садрже и изворни и циљни текст на два језика), и вишесмерни (исти текст на више различитих језика) (Xiao 2010, 160). Такви су нпр. законодавни текстови који настају у институцијама Европске уније (скр. ЕУ), а који морају имати онолико званичних верзија колико је и званичних језика ЕУ (ЕУ тренутно броји 24 званична језика и 27 држава чланица). Паралелизовани корпуси су предмет различитих врста истраживања, нарочито у домену обраде природног језика (Krstev и Vitas 2009, 495), користе се код креирања термилолошких база и преводилачких меморија (више о термилолошкој бази правних тековина ЕУ на српском језику у поглављима 5.5.3.1 и 5.5.3.2), у језичкој настави односно практичној обуци преводилаца као и код машинског превођења. Ови корпуси не морају бити вишејезични већ могу садржати два или више независна превода истог текста на једном језику (Krstev и Vitas 2009, 496). У пракси се за паралелне корпусе користи и термин „преводни корпуси“ (енг. *Translation corpora*), међутим овим термином се често означава и колекција преведених текстова без њиховог извора, па је стога термин донекле збуњујућ (Aijmer 2008, 277).

С друге стране, упоредиви корпуси не садрже семантички еквивалентне текстове већ текстове који припадају истом домену, истом временском периоду, итд. Овакав корпус је лакше припремити и, мада је његова сфера употребе ограничена, користи се у одређеним апликацијама за обраду природног језика које су засноване на статистичким методама. Користе се и за веома блиске језике за које је тешко пронаћи преведене текстове, нпр. за српски, хрватски и бошњачки, чешки и словачки, итд. (Krstev и Vitas 2009, 496).

5.5. Корпусна лингвистика у Србији

Развој корпусне лингвистике у Србији можемо поделити на два периода, неелектронски и електронски. Крајем педесетих година прошлог века у оквиру пројекта „Квантитативни опис структуре српског језика“, посвећеног статистичком изучавању српског писаног наслеђа од 12. века до данашњих дана, на Институту за експерименталну фонетику и патологију говора, развијен је дијахронијски корпус српског језика, чији је аутор Ђорђе Костић, оснивач и дугогодишњи директор поменутог Института (Костић 2009, 11). Основу Костићевог пројекта чини КОРПУС СРПСКОГ ЈЕЗИКА чији је и сам аутор, а који се састоји од око 11 милиона речи. Реч је о неелектронском корпусу који обухвата српски језик од 12. века до савременог језика (Костић 2009, 11). Главна сврха Костићевог дијахронијског корпуса који је састављен ручним процесирањем наведених узорака текста био је квантитативни опис структуре српског језика (Krstev 2008, 11).

Средином деведесетих година прошлог века, у сарадњи Института за експерименталну фонетику и патологију говора и Лабораторије за експерименталну психологију Филозофског факултета у Београду, почиње дигитализација овог Корпуса (Костић 2009, 12).

5.5.1. Корпус савременог српског језика

Први помак у развоју корпуса на нашим просторима представља конструисање система под називом Аурора, који је генерисао конкорданце и различите врсте индекса за задати текст. Аурора је названа по програмском систему „Аурора“ Душка Витаса, у којем је први пут примењена (Vitas 1981). На Математичком институту је 1981. године настао пројекат под називом Математичка и рачунарска лингвистика, чији је циљ, између осталог, био и формирање корпуса савременог српског језика. У периоду од 1981–1985. године, под руководством Душка Витаса, забележени су и први резултати овог пројекта међу којима је прва колекција текстова у дигиталном облику, први експерименти у морфолошком генерисању српскохрватског, прва истраживања у домену корпусне лингвистике (односи се на анализе језика уџбеника, језика закона и осталог), као и успостављање контаката са истакнутим европским истраживачима у домену корпусне лингвистике. Током осамдесетих година 20. века, коришћењем система Аурора састављени су и обрађени први паралелни корпуси, а 1989. године чак и припремљен српски превод *SGML* (енг. *Standard Generalized Markup Language*) стандарда.¹⁰⁷

На иницијативу и под руководством Душка Витаса, процес сакупљања и дигитализације текстова за корпус отпочела је Група за језичке ресурсе и технологије 2002. године, а његов циљ био је изградња и унапређење постојећих језичких ресурса за српски језик. Прва фаза на изградњи Корпуса савременог српског језика, обухватала је прикупљање текстова за корпус, тестирање различитих алата за креирање и претрагу корпуса, израду базе података и веб- сучеља који би омогућили контролисан приступ корпусу (регистрацију и ауторизацију корисника), администрацију корпуса (управљање корпусима и корисничким налозима) и једноставну претрагу корпуса (Utvić 2013, 248–249).

Неетикетирани корпус српског језика (скр. НЕТК), представља прву верзију Корпуса савременог српског језика, која је настала 2003. године и која, као што и само име каже, није садржала ни лингвистичку анотацију ни библиографске податке о текстовима. Па ипак, корпус

¹⁰⁷ Више информација на <http://www.korpus.matf.bg.ac.rs/prezentacija/istorija.html>.

је био (и још увек је) доступан на вебу, а могуће га је претраживати помоћу регуларних израза. Реч је о једнојезичком корпусу чији је развој имао за циљ изградњу нових и унапређење постојећих језичких ресурса за српски језик (Utvić 2013, 248).¹⁰⁸

Пре изградње овог Корпуса, требало је превазићи одређене проблеме, за које искуство односно решења у другим језицима нису постојали. Неки од њих су: употреба два писма – ћириличног и латиничког, употреба различитих схема кодирања за електронске текстове, употреба два изговора, екавског и ијекавског као и проблеми приликом дефинисања опсега српског језика у широј заједници која је раније била позната као српско-хрватска (Krstev и Vitas 2005).

Колекција текстова који чине овај корпус величине је 22,2 милиона корпусних речи. Како би се превазишла односно неутрализовала разлика која потиче од различитих ћириличних и латиничних кодних схема, корпусни текстови су кодирани коришћењем карактерског скупа *ASCII*, односно кодном схемом *Aurora*.¹⁰⁹

У периоду од 2005–2010. године, у оквиру пројекта под називом „148021 Теоријско-методолошки оквир за модернизацију описа српског језика“, Група за језичке технологије и ресурсе наставила је са развојем нових и одржавањем постојећих језичких ресурса, а посебна пажња била је посвећена конструкцији паралелизованих корпуса и развоју одговарајућих алата, проширивању електронског морфолошког речника вишечланим лексемама, као и развоју лексичких ресурса за препознавање и класификацију именованих ентитета. У том периоду, настављено је и са прикупљањем нових текстова за Корпус савременог српског језика (скр. *SrpKor*) и унапређење веб-сучеља (претрага паралелних корпуса, екстракција и снимање изабраних конкорданци) (Utvić 2013, 252–253).

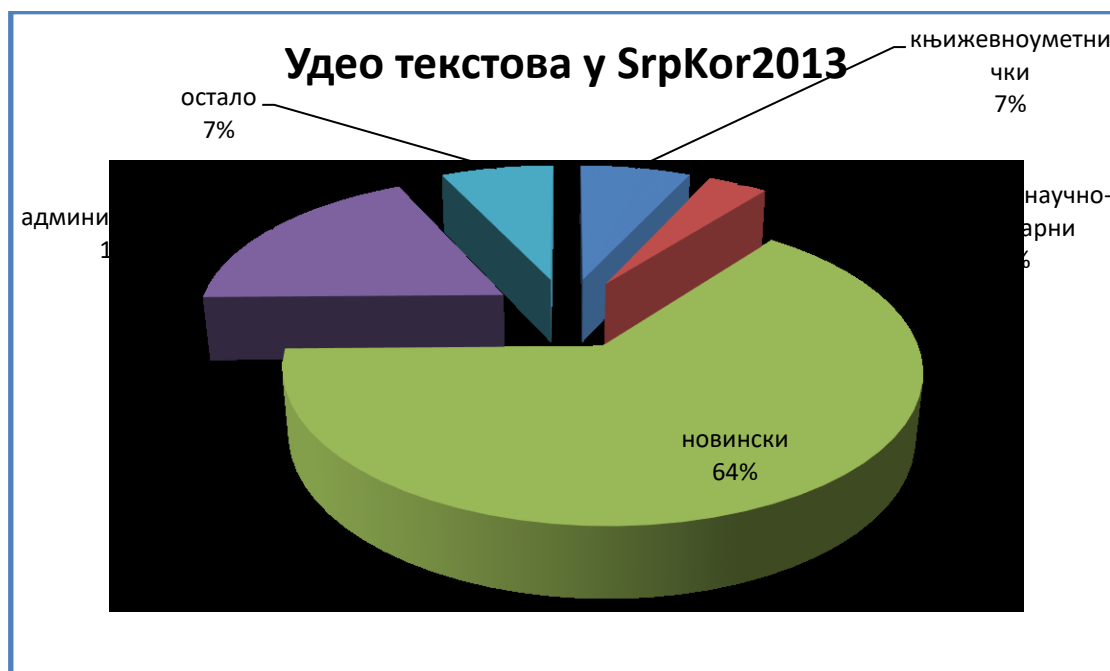
Пре генерисања последње, треће, верзије Корпуса савременог српског језика, која садржи 122 милиона корпусних речи и која носи ознаку *SrpKor2013*, у периоду од 2011. до 2013. постојале су и претходне две верзије, прва под ознаком *SrpKor2011*, од 113 милиона корпусних речи, и друга под ознаком *SrpKor2012*, од 118 милиона корпусних речи, као претече последње верзије. Све три верзије аутоматски су морфолошки анотирани, а текстовима корпуса су придружене и одговарајуће библиографске информације (Utvić 2011, 40).

Садржина *SrpKor2013*¹¹⁰ јесу књижевно–уметнички текстови српских писаца у 20. и 21. веку, научни и научно-популарни текстови из различитих домена (природне и друштвене науке), административни и општи текстови. Један део текстова чине и преводи на српски језик. На слици 12 приказана је расподела текстова овог корпуса према функционалним стиловима (Utvić 2013, 258).

¹⁰⁸ Рад на реализацији овог корпуса био је само делимично потпомогнут пројектом под називом „1743 Интеракција текста и речника“, који је финансирало Министарство за науку Републике Србије у периоду од 2002–2004. године. Па ипак, већи део самог посла су волонтерски обавили сарадници на овом пројекту.

¹⁰⁹ Кодна схема Аурора пресликава дијакритичке карактере и диграфе који се користе у српској латиници на начин као што је приказано у Корисничком приручнику за употребу *CQP*-а кроз веб-интерфејс, на веб-сајту <http://korpus.matf.bg.ac.rs/prezentacija/uputstvo.html#aurora>.

¹¹⁰ <http://www.korpus.matf.bg.ac.rs/prezentacija/korpus.html>



Слика 12. Расподела текстова Корпуса према функционалним стилевима

Корпусни текстови су кодирани коришћењем карактерског скупа *ISO-8859-1*, односно кодном схемом *Aurora*, а текстови су анотирани библиографски и морфолошки. Библиографска анотација корпусног текста, поред уобичајеног библиографског описа, садржи и информације о функционалном стилу којем припада текст, као и информације о томе да ли је текст написан на изворном српском језику или је у питању превод на српски са неког страног језика. Морфолошка анотација садржи само информације о одговарајућој лем и врсти речи. Иако је непотпуна, оваква анотација омогућава да се корпус претражује помоћу великог броја различитих упита, што пре тога није било могуће.¹¹¹

Део корпуса *SrpKor2013* који садржи текстове који нису под заштитом ауторских права, величине је 3,7 милиона корпусних речи, а могуће га је преузети и дистрибуирати у складу са лиценцом *CC_BY-NC*.¹¹²

SrpKor2013 представља електронски, синхрони и динамички корпус савременог српског језика, а као такав, овај корпус се непрекидно ажурира, у настојању да постигне што већу репрезентативност и балансираност, и то константним додавањем нових и заменом постојећих текстова (Utvić 2013, 317–318).

У току 2021. године се очекује нова верзија корпуса са знатно више текстова, коригована и допуњена анотација као и унапређена претрага.

5.5.2. Вишејезични корпуси српског језика

Први паралелни енглеско-српски и француско-српски корпуси доступни за претраживање на вебу, настали су у периоду од 2011–2013. године, а у истом периоду довршено је и вишејезично електронско издање романа „Пут око света за 80 дана“ Жила Верна (Utvić 2013, 256–257). Вернов роман одабран је, с једне стране, због тога што је обухватио верзије романа на 18 језика, међу којима је и српски, а с друге стране што, имајући у виду његов садржај, представља погодан текст за различите врсте анализа, нарочито у домену препознавања именованих ентитета (географски појмови и разне врсте мере) (Vitas, и други 2008).

¹¹¹ <http://www.korpus.matf.bg.ac.rs/prezentacija/korpusi.html>

¹¹² <http://www.korpus.matf.bg.ac.rs/prezentacija/korpusi.html>.

Вишејезично електронско издање романа Џорџа Орвела „1984“, најпре је морфосинтаксички анотирано у формату LADL/DELA, а потом у MULTEXT-EAST (енг. *Multilingual Text Tools and Corpora for Eastern and Central European Languages*) формату који подржава језике централне и источне Европе. Да би анотирање било могуће, претходно је развијена MULTEXT-EAST спецификација за српски језик (Krstev, Vitas и Erjavec 2004).

Сви преводи су паралелизовани са неком од верзија на француском, енглеском или српском језику, тако да ресурс садржи укупно 32 битекта у форматима *TMX* и *HTML* (Utvić 2013, 257).

ПАРАЛЕЛИЗОВАНИ ЕНГЛЕСКО-СРПСКИ КОРПУС представља део корпуса *SrpEngKor*, а настао је формирањем два поткорпуса: *SELFEH* и *BALKANTIMES* (Utvić 2013, 252). Овај корпус се може преузети и самостално користити.

SELFEH (од енг. *Serbian-English Law Finance Education and Health*)¹¹³ је паралелни српско-енглески поткорпус који садржи текстове из области финансија, здравства, права и образовања. Развијен током пројекта Интера¹¹⁴. Садржи око милион речи на сваком од језика, а паралелизован је коришћењем формата *TMX*. Састоји се од преко 150 докумената у *XML* формату, у којима су корпусне речи анотирани одговарајућом лемом и врстом речи.

BALKANTIMES је паралелни српско-енглески поткорпус који садржи текстове вести са сајта *Southeast European Times* (скр. *SETimes*), који више не постоји, а које обухватају информације о Југоисточној Европи (текући догађаји, економија, дипломатија, филм, туризам, спортови и наука), и то на следећим језицима: албанском, бошњачком, хрватском, енглеском, македонском, српском и турском.

ЕНГЛЕСКО-СРПСКИ ПАРАЛЕЛИЗОВАНИ КОРПУС такође је део *SrpEngKor*¹¹⁵, а чине изворни текстови на енглеском језику који су преведени на српски, као и изворни српски текстови који су поравнати са својим преводима на енглески језик. Текстови припадају различитим жанровима: књижевност, новинске вести, научни часописи, веб-журнализам, здравље, право, образовање чак и филмски титлови. Садржину корпуса чине и шест романа ауторке Џејн Остин (енг. *Jane Austen*) (Krstev и Vitas 2009, 500). Величина корпуса је око 4.5 милиона речи.

ФРАНЦУСКО-СРПСКИ ПАРАЛЕЛИЗОВАНИ КОРПУС (*SrpFranKor*)¹¹⁶ чине текстови преведени на српски језик, као и српски текстови преведени на француски језик. У корпусу су заступљена два жанра: књижевност и новинске вести. Избор текстова вођен је њиховом књижевном вредношћу као и доступношћу у електронском облику, а важно је истаћи и чињеницу да књижевни текстови, по правилу, имају прецизније преводе и дају увид у језик и граматичке феномене који се не могу сагледати у другим жанровима (Vitas и Krstev 2007, 148–149). Паралелизација односно поравнање текста вршено је аутоматски, на нивоу реченице (Vitas и Krstev 2007, 152–153). Корпус је величине око 1.7 милиона речи.

СРПСКО-НЕМАЧКИ КОРПУС (*SrpNemKor*) чине књижевна дела српских писаца превођена на немачки језик као и књижевна дела немачких писаца превођена на српски језик. У питању су савремени аутори и књижевни прозни текстови, романи. Садржину корпуса чини четрнаест паралелизованих романа, преко 1.6 корпусних речи и 48.004 преводних парова (Андоновски 2019).

5.5.3. Корпус правних тековина Европске уније (Евротекa) и вишејезична база термина (Евроним)

Према значају односно месту које на домаћем националном плану заузима, уз корпусе који су до сада овим радом изложени, истиче се и Корпус правних тековина Европске уније

¹¹³ <http://www.korpus.matf.bg.ac.rs/prezentacija/selfeh.html>.

¹¹⁴ <https://www.mpi.nl/intera/>.

¹¹⁵ Подаци у вези са корпусом *SrpEngKor* доступни су на вебу:

http://www.korpus.matf.bg.ac.rs/SrpEngKor/SrpEngKor_2013_01.pdf (преузети јануара 2020. године).

¹¹⁶ Подаци у вези са корпусом доступни су на вебу

http://www.korpus.matf.bg.ac.rs/SrpFranKor/SrpFranKor_2012_07.pdf (преузети јануара 2020. године).

(Евротекa), односно вишејезична база термина (Евроним), Министарства за европске интеграције, Владе Републике Србије.

Ове базе настале су као производ дугогодишњег рада некадашње Канцеларије за европске интеграције, односно данашњег Министарства за европске интеграције Владе Републике Србије. Имајући у виду да је превођење правних тековина ЕУ односно израда њихове националне верзије један од кључних услова за стицање кандидатуре односно приступање Европској унији, а да се обим прописа који се појављују у Службеном листу ЕУ непрестано увећава (укупно броји око 200.000 страна службеног листа закључно са мартом 2021. године), припрема правних тековина ЕУ показала се као нимало једноставан задатак, о чему је било речи у поглављу 3.4.3.3. У том комплексном процесу, настао је и Корпус правних тековина ЕУ (Евротекa) као и термилошка база (Евроним), које се током овог процеса непрестано допуњују новим терминима.

5.5.3.1. Евротекa, енглеско-српски паралелни корпус

ЕВРОТЕКА је збирка европских правних текстова и њихових сегмената, тј. двојезични (енглеско-српски) корпус. Настала је у процесу превођења правних тековина Европске уније на српски језик који је отпочео 2009. године, а рад на њој још увек траје. Тренутно садржи 753380 речи (или 21068 преводачких јединица/сегмената).¹¹⁷ Правне тековине ЕУ (фр. *acquis communautaire*), обухватају примарно и секундарно законодавство ЕУ односно корпус заједничких права и обавеза који важе за све државе чланице ЕУ (видети поглавље 3.4.3.3).

Превођење правних тековина ЕУ (а потом и њихова хармонизација са националним законодавством), један је од основних услова за чланство државе кандидата у Европској унији и претходи процесу њеног приступања односно прикључења ЕУ, о чему је већ било речи у поглављу 3.4.3.3. Имајући у виду да се превођење ових текстова показало као највећи преводачки подухват у држави, било је неопходно развити ресурсе који би послужили као помоћ у послу тако великог обима. Евротекa, као енглеско-српски корпус правних текстова и њихових сегмената, значајан је ресурс који се користи у процесу превођења и припреми верзије правних тековина ЕУ на српском језику.

Корисници овог корпуса су пре свега професионални преводиоци и то у највећем броју случајева они који су ангажовани на превођењу европског законодавства, међутим ова збирка прописа је истовремено доступна и ширем кругу корисника.

Евротекa, за сада, садржи само енглеско-српске сегменте, јер се као извор за превођење правних тековина ЕУ на српски језик користе текстови на енглеском језику. Корпус се може претраживати на оба језика (и оба писма на српском језику), као што је и резултат претраге увек двојезичан. Претрага се врши на нивоу речи и скупова речи.

За превођење правних текстова, а самим тим и за креирање овог корпуса, користи се програм *SDL Trados*,¹¹⁸ тј. његов алат *Translator's Workbench*, који служи за формирање преводачких меморија.

Процес превођења правних аката Европске уније подразумева више корака: само превођење, стручну, правнотехничку и језичку редактуру, после којих следи фаза верификације. На основу фазе након које је сегмент текста унет у корпус одређује се степен поузданости (енг. *reliability*). Поузданост се дефинише на скали од 1 до 5. Та фаза (превод, стручна редактура, правнотехничка редактура, језичка редактура, верификација) наведена је у заглављу сваког сегмента.

По задатом упиту, ЕВРОТЕКА приказује тражени сегмент текста и његове припадајуће атрибуте: поље (енг. *field*), ступањ обраде на ком се текст налази, тј. степен поузданости – сегмент који има највећи степен поузданости – 5, биће први приказан и идентификациони број (енг. *ID*)

¹¹⁷ Подаци преузети јуна 2021. године са веб-сајта <http://prevodjenje.mei.gov.rs/evroteka/index.php?jezik=srps>.

¹¹⁸ *SDL Trados (TRANslation & DOcumentation Software)* – један је од најпознатијих софтвера за превођење помоћу рачунара (енг. *CAT – Computer Assisted Translation*). Његови алати засновани су на технологији креирања базе превода (енг. *TM – Translation Memory*), базе података која се креира и ажурира током процеса превођења.

документа који је заправо целекс број из ког је одговарајући сегмент издвојен. Он је истовремено и линк ка датом документу у бази *EUR-Lex*. Такође, приказује се и директоријум у који је дати документ распоређен у *EUR-Lex* бази. На слици 13 приказан је резултат претраге према кључној речи.

Речи за претрагу се уносе у поље у горњем левом делу почетне странице. Притиском на тастер „Тражи“, у десном делу екрана појављују се резултати претраге: сегмент у којем се тражени упит налази, на два језика – енглеском и српском, као и линк ка Еврониму (терминолошкој бази), уколико је тражени низ уједно и термин који у њему постоји.

На слици 13, приказан је пример претраге термина *power*. Једноставним упитом, без помоћних знакова, добија се листа погодака у којима се тражена реч појављује у било којој комбинацији и на било којој позицији у тој комбинацији. У примеру који је издвојен на слици 13, термин *power* се приказује у значењу „овлашћења“, „пуномоћја“ и „погона“.

У претрази Евротекке могу се користити и помоћни знаци као што су доња црта (), тачка (.) или упитник (?), а који могу заменити било који карактер у одређеној речи, као и знакове проценат (%) и астерикс (*) који замењују било који број односно ниску карактера.

ЕВРОТЕКА Тирница Latinica English

Енглеско-српски паралелни корпус

Евроним **Евротекка** Терминолошки анализатор Публикације Корисни линкови Упутства

Унесите реч или фразу:

Тражи Избриши

Корпус садржи 743514 речи.
 Ажурирано: 4. 2. 2021.
 Copyright (©) 2009–2021. Влада Републике Србије. Сва права задржана. Забрањено неовлашћено умножавање и дистрибуција садржаја сајта.

Тражени израз: **power**
 Број погодака: **255**
 Превод са Евронима:

- [електрична енергија](#)
- [овлашћење](#)
- [снага](#)
- [моћ](#)

01 Finance and institutions	Верификација	ID: 32013R1021 , Приручник (2019)
The delegation of power shall be tacitly extended for periods of an identical duration, unless the European Parliament or the Council opposes such extension not later than three months before the end of each period.		Делегирање овлашћења прећутно се продужава за периоде једнаког трајања, осим ако се Европски парламент или Савет не успротиве том продужењу најкасније три месеца пре истека сваког периода.
00 Primary legislation	Верификација	ID: 12016ME
WHO, having exchanged their full powers , found in good and due form, have agreed as follows:		КОЈИ су се, разменивши своја пуномоћја, за која је утврђено да су у ваљаном и прописаном облику, споразумели о следећем:

13 Industry and internal market	Језичка редактура	ID: 32009R0079
For the purposes of vehicle type-approval, manufacturers shall equip hydrogen-powered vehicles with hydrogen components and systems that comply with the requirements of this Regulation and its implementing measures and are installed in accordance with this Regulation and its implementing measures.		За сврхе хомологације типа возила произвођачи опремају возила са водоничним погоним водоничним деловима и системима који испуњавају захтеве из ове уредбе и њених спроведбених мера и који су урађени у складу са овом уредбом и њеним спроведбеним мерама.

Слика 13. Евротекка

5.5.3.2. Евроним – Вишејезична термилошка база

ЕВРОНИМ¹¹⁹ је вишејезична база термина, која је настала као један од резултата процеса превођења правних тековина Европске уније на српски језик, а која је, истовремено, и важно помагало у том процесу. Циљ креирања ове термилошке базе јесте успостављање јединствене терминологије на српском језику, из читавог низа различитих дисциплина, области односно општег домена људског деловања. Поред опште лексике, опште термилошке лексике, правне терминологије и терминологије из осталих струка, у овој термилошкој бази налази се и изванредан број најчешћих фраза и устаљених израза. Поред унификације терминологије, овом термилошком базом се олакшава и посао преводиоцима правних аката, а уз то, доступна је и

¹¹⁹ <http://prevodjenje.mei.gov.rs/evronim/index.php?jezik=srps>

ширем кругу корисника за различите потребе. Терминолошка база Евроним тренутно садржи 46505 термина.¹²⁰

Сам процес превођења правних аката Европске уније подразумева неколико фаза кроз које сваки документ (односно термини који се у оквиру њега јављају) појединачно мора да прође. То су фазе: превођења, стручне, правнотехничке и језичке редактуре. Када термин прође кроз сваку од наведених фаза, стручно тело га као таквог верификује. То је уједно и последња фаза, којом се наведени процес завршава. У одређеном броју случајева, поједини термини који су формално исти, појављују се у две или више области. У тим случајевима, подразумева се, исти термини имају различито значење и третирају се као засебне одреднице.

Имајући у виду да се правни документи ЕУ преводе са енглеског језика (у Еврониму означени са *EN*), сваки унос садржи термин на енглеском, превод термина на српски језик (*SR*), а где год је то било могуће и оцењено као сврсисходно и корисно за поређење или боље разумевање, наведени су и еквиваленти у другим језицима. Осим енглеског, званични језици ЕУ заступљени у Еврониму јесу: бугарски (*BG*), чешки (*CS*), немачки (*DE*), грчки (*EL*), шпански (*ES*), француски (*FR*), мађарски (*HU*), италијански (*IT*), пољски (*PL*), португалски (*PT*), румунски (*RO*), хрватски (*HR*), словачки (*SK*), словеначки (*SL*). У бази се могу наћи и термини на латинском језику (*LAT*).

На слици 14 приказан је пример претраге термина *power cable*. Поред његових еквивалената на српском, француском, немачком и словеначком језику, у оквиру ове одреднице налазе се и подаци о датуму његовог уноса, датуму измене, редном броју термина у терминолошкој бази, пројекта превођења у оквиру којег је термин преведен, изворни језик са ког је термин преведен, област којој термин припада на српском и енглеском језику, термин – референца термина којом се упућује на документ из којег је термин издвојен, институција која је препоручила или потврдила превод термина, дефиниција појма означеног термином, документ који је извор дефиниције као и поузданост превода термина на скали од 1 до 4¹²¹.

Уз ове податке, одредницу могу пратити и следећи подаци: термин на српском – референца термина којом се упућује на документ у којем је термин потврђен на српском језику, контекст у ком се на изворном или српском језику јавља термин, документ из ког је издвојен контекст, напомене – додатне информације, белешке и коментари, напомена о документу који је извор додатних информација, пример адекватне употребе термина, употребни статус термина (колоквијално, професионални жаргон, препоручује се, непотврђено), граматичке напомене, врста речи којој термин припада (именица, придев, глагол, прилог), напомене о граматичком броју (лат. *singularia tantum*, *pluralia tantum*), напомене о граматичком роду, хипоним – подређени појам у односу на појам означен термином, хипероним – надређени појам у односу на појам означен термином. Ипак, у базу Евроним још увек није унет ни један термин којег прате сви ови подаци.

Имајући у виду да је ова терминолошка база тесно везана за процес превођења правних тековина ЕУ, заједно са процесом, и сама база се непрестано мења, допуњује, напредује, прилагођава и усавршава.

¹²⁰ Податак преузет јуна 2021. године.

¹²¹ 1 = предлог за превод термина, који тек треба потврдити (по правилу за нове концепте), 2 = термин који се користи, а није потврђен у фази стручне, односно правнотехничке редактуре, 3 = термин који су потврдили и препоручили стручни или правнотехнички редактори, 4 = верификован термин (из примарног законодавства).

The screenshot shows the EuroTerm database interface. At the top, there are navigation tabs: "Евроним", "Евротекста", "Термилошки анализатор", "Публикације", "Корисни линкови", and "Упутства". A search bar contains the text "power cable". Below the search bar are buttons for "Тражи" and "Избриши". There is a link for "Напредна претрага" and a note "Број погодака: 1". A list of results shows "1. power cable". To the right, detailed information is provided for the term "power cable" in EN, "енергетски кабл" in SR, "cable d'alimentation" in FR, "Stromkabel" in DE, and "električni kabel" in SL. The EN entry includes the date of entry (11. 8. 2017), date of change (27. 6. 2019), project (Secondary legislation), field (electronics and electrical engineering), region (CP), source language (EN), and entry number (17451). It also lists the term, reference (32011R0305), definition, and definition reference (IATE). The SR entry includes the source language (SR), confidence level (3), and institution (Ministry of Construction, Transport and Infrastructure). The FR entry includes the term and references (IATE, Woerterbuch.info, LEO, Wikipedia, DWDS). The DE entry includes the term and references (IATE, Woerterbuch.info, LEO, Wikipedia, DWDS). The SL entry includes the term. At the bottom left, the number "687187" is visible.

Слика 14. Евроним

Термилошки анализатор представља програм који анализира текстуалне датотеке у односу на термилошку базу Евроним тако што препознаје оне термине који у њој постоје. Након обраде текста сви термини који постоје у Еврониму постају хиперлинкови. Постављањем курсора миша на хиперлинк, приказује се превод датог термина и у загради степен његове поузданости. Кликом на хиперлинк, отвара се нови прозор са свим доступним подацима о изабраном термину из Евронима (слике 15, 16 и 17).

The screenshot shows the EuroTerm database interface for the text analysis tool. At the top, there are navigation tabs: "Евроним", "Евротекста", "Термилошки анализатор", "Публикације", "Корисни линкови", and "Упутства". Below the tabs, there is a section titled "Упишите текст за анализу:". A text area contains the following text: "Article 28 of Regulation (EU) No 608/2013 provides that a right holder is to be liable for damages towards the holder of the goods where, inter alia, the goods in question are subsequently found not to infringe an intellectual property right." Below the text area are buttons for "Обрада текста" and "Брисање текста". At the bottom, there are dropdown menus for "Изворни језик:" (English/Serbian) and "Текст је:" (Single language / Two language - horizontal display / Two language - vertical display).

Слика 15. Термилошки анализатор – текст који се анализира

Евроним	Евротекa	Терминолошки анализатор	Публикације	Корисни линкови	Упутства
---------	----------	--------------------------------	-------------	-----------------	----------

memory usage: 35/256MB

Article 28 of Regulation (EU) No 608/2013 provides that a right holder is to be liable for damages towards the holder of the goods where, inter alia, the goods in question are subsequently found not to infringe an intellectual property right.

интелектуална својина (4)

Слика 16. Терминолошки анализатор након завршене обраде текста

CreationDate	17. 12. 2009.
ChangeDate	16. 9. 2019.
Project	Примарно законодавство
Subject EN	intellectual property
Subject SR	интелектуална својина
Source Language EN	
Entry Number	5446

EN	intellectual property Term Reference 12016P Charter of Fundamental Rights of the European Union
SR	интелектуална својина Reliability 4 Note Обухвата патенте, жигове, индустријски дизајн, топографију интегрисаних кола, ауторска права, географске ознаке и трговачке тајне. Note Reference Валтер Гуд, Речник термина међународне трговинске политике, 2005.
CS	duševní vlastnictví
FR	propriété intellectuelle
DE	geistiges Eigentum
SK	duševné vlastníctvo

687268

Слика 17. Анализа хиперлинкованих термина

5.5.4. Библиша

Дигитална библиотека БИБЛИША¹²², развијена у оквиру Друштва за језичке ресурсе и технологије Универзитета у Београду, представља алат који омогућава вишејезичко претраживање великих колекција поравнатих текстова. Упити којима се врши претраживање ових колекција могу бити састављени од једне односно више кључних речи на различитим језицима. Упите је могуће семантички и морфолошки проширити што је нарочито важно у флективним језицима какав је српски (Stanković, Krstev, и други 2017). Морфолошко проширење врши се помоћу веб-сервиса *Vebran (ref)* који се заснива на *Unitex (ref)* рутинама (о чему ће бити више речи у одељку 6.2), и морфолошким е-речницима (*ref*) којима се на упит додају флективни облици.

Библиша за сада садржи 12 паралелизованих колекција, од чега пет часописа: Инфотека¹²³, Подземни радови¹²⁴, Архитектура и урбанизам¹²⁵, Стоматолошки гласник Србије¹²⁶, *Management*¹²⁷; три паралелизоване збирке настале у међународним пројектима: *BEAKTEL TEMPUS (Blending Academic and Entrepreneurial Knowledge in Technology Enhanced Learning)*¹²⁸, *CESAR (Central and South-East European Resources)*¹²⁹, *Intera (Integrated European language data*

¹²² Библиша, <http://jerteh.rs/biblisha/>

¹²³ Инфотека, <http://infoteka.bg.ac.rs/index.php/sr>

¹²⁴ Подземни радови, <http://ume.rgf.bg.ac.rs/index.php/ume>

¹²⁵ Архитектура и урбанизам, <http://www.iaus.ac.rs/code/navigate.aspx?Id=111>

¹²⁶ Стоматолошки гласник, <https://scindeks.ceon.rs/journalDetails.aspx?issn=0039-1743>

¹²⁷ *Management*, <http://management.fon.bg.ac.rs/index.php/mng>

¹²⁸ *BEAKTEL TEMPUS*, <http://www.erasmusplus.ac.me/?tempus-projekat=aeoebleshooting-academic-and-entrepreneurial-knowledge-in-technology-enhanced-learningae%C2%9D>

¹²⁹ *CESAR*, <http://www.meta-net.eu/projects/cesar/>

Repository Area)¹³⁰; паралелизовани корпус наставних текстова *Education*; паралелизовани корпус књижевних текстова на српском и немачком језику (СрпНемКор); две паралелизоване доменске збирке: *Mining* – паралелизовани корпус текстова из области рударства и *EIEner* – паралелизовани корпус текстова из области електроенергетике, подељен на два дела, део у отвореном приступу и заштићени део, који је уједно и предмет ове дисертације (више о овом корпусу у одељку 7.2).

У табели 10 дат је преглед текстуалних колекција у Библиши према броју докумената и броју сегмената.

Табела 10. Текстуалне колекције у дигиталној библиотеци Библиша према броју докумената и броју сегмената

Колекција	Број докумената	Број сегмената
Инфотека	144	20.762
Подземни радови	55	4.831
Архитектура и урбанизам	9	1.641
Стоматолошки гласник Србије	10	1.279
<i>Management</i>	181	29.326
<i>BEAKTEL</i>	10	2.745
<i>Education</i>	2	7.799
<i>CESAR</i>	5	922
<i>Intera</i>	157	46.630
СрпНемКор	14	48.004
<i>Mining</i>	34	712
<i>EIEner</i>	76 (68+8)	29.943 (19725+10218)
Укупно	697	194594

Лексички ресурси који омогућавају морфолошко и семантичко проширење постављених упита у Библиши могу се поделити у три групе:

1. електронски морфолошки речници српског језика који се користе за генерисање флективних облика кључних речи у упиту за претрагу на српском језику. Процес изградње електронских речника и граматика је прилично спор имајући у виду да се већи део њихове конструкције обавља ручно, међутим интеракција коју имају са корпусима омогућава напреднију обраду коју би без њихове подршке било тешко постићи (Vitas и Krstev 2012). Изградња оваквог речника захтева темељну и прецизну анализу флективних својстава речи, при чему се сматра да сваки низ абecedних карактера који је ограничен сепараторима представља реч (једночлану односно вишечлану лексичку јединицу) (Vitas и Krstev 2012). Ови речници су развијени у формату DELA који је организован у два одвојена подсистема: речник једночланих лексичких јединица (DELAS и DELAF) и речник вишечланих лексичких јединица (DELAC и DELACF) (Vitas и Krstev 2012). На основу ових модела, у оквиру Групе за језичке ресурсе и технологије Универзитета у Београду, конструисан је систем електронских морфолошких речника српског језика чији су аутори Цветана Крстев, професор Филолошког факултета Универзитета у Београду и др Душко Витас, професор Математичког факултета Универзитета у Београду. Ови морфолошки речници садрже коначне аутомате и трансдукторе за препознавање непознатих речи (Vitas и Krstev 2012). Свакако, одређене карактеристике српског језика попут употребе два писма – ћириличног и латиничног, правописа који је заснован на фонологији – употреби различитих варијанти српског језика као што су екавски и ијекавски изговор (и сходно томе постојању дублета), богатог морфолошког система који се рефлектује на флексију и деривацију као и слободан редослед речи у смислу субјекта, предиката, објекта и осталих делова реченице и

¹³⁰ *INTERA*, <https://www.mpi.nl/intera/>

употреба енклитика, морају бити узете у обзир приликом обраде текстова на српском језику (Vitas и Krstev 2005). Више о електронским морфолошким речницима у поглављу 6.3.

2. српски и енглески ворднет (енг. *WordNet*) којим се омогућава семантичко проширење постављених упита на српском и енглеском језику. Ворднет представља лексичко-семантичку мрежу која се састоји од појмова груписаних у скупове синонима, тзв. синсетове (енг. *synset* или *synonymous set*), који су међусобно повезани семантичким везама. Речнички појмови у ворднету нису представљени алфаветским редом као у традиционалним речницима, већ концептуално, на основу семантичке меморије и поимања (Митровић 2018, 45). Поред упоредне лексичке анализе, глобални ворднет омогућава примену у осталим областима као што је вишејезично управљање информацијама, знањем и садржајем, руковање текстуалним подацима, вишејезичну екстракцију, машинско превођење, итд. (Krstev, Koeva и Vitas 2006). Више о ворднету у поглављу 6.6.

3. термилошке базе података из различитих домена и онтологије као што су Речник библиотекарства и информационих наука¹³¹, *GeolISSTerm*¹³², *RudOnto*¹³³ и *Termi*¹³⁴.

- Речник библиотекарства и информационих наука садржи терминологију која се користи у теорији и пракси библиотекарства и информационих наука као и других сродних области на српском, енглеском и немачком језику. Верзија речника доступна на вебу тренутно садржи 40.000 одредница (приближно 14.000 на српском, 12.400 на енглеском и 14.000 на немачком језику), 900 дефиниција или анотација термина који су део библиотечких стандарда, 2.300 акронима међународних и националних организација и институција и 190 адреса релевантних локација на вебу.
- *GeolISSTerm* је тезаурус геолошких термина на српском и енглеском језику, развијен на Рударско-геолошком факултету Универзитета у Београду. Развијен је за потребе Министарства животне средине и просторног планирања Републике Србије са циљем обезбеђивања стандардног електронског геолошког речника за логички конзистентан опис, интерпретацију и класификацију геолошких јединица, геолошке структуре, минералних и хидрогеолошких ресурса, као и других геолошких својстава у домену примењених дисциплина (Станковић, и други 2011, 54). Овај тезаурус садржи више од 3000 термина и њихових еквивалената и подељен је у неколико подобласти: петрологија, минералологија, хидрогеологија, геофизика, итд. (Stanković, Obradović и Utvić 2014, 13–14).
- *RudOnto* је још један термилошки ресурс на српском и енглеском језику, развијен на Рударско-геолошком факултету Универзитета у Београду са циљем да покрије већи део стручне терминологије из области рударског инжењерства и геологије и тако постане референтни извор рударске терминологије на српском језику (Stanković, Obradović и Utvić 2014, 15).
- *Termi* је вишејезична база термина која подржава развој термилошких речника из различитих области попут математике, рачунарства, рударства, геологије, језичких технологија, пољопривреде, електроенергетике, итд. Ова апликација тренутно подржава обраду и приказ термина на енглеском и српском језику, док је развој за друге језике у плану. Више о овој бази у поглављу 7.9.

Колекције докумената чије претраживање омогућава Библиша, састоје се од поравнатих паралелних текстова конвертованих у формат *TMX* (енг. *Translation Memory eXchange*, срп. размена преводилачких меморија). *TMX* је отворени стандард заснован на *XML* формату, који служи за складиштење преводилачких меморија и омогућава једноставнију размену података из преводне меморије односно поравнатих паралелних текстова између различитих софтверских

¹³¹ Речник библиотекарства и информационих наука који је у Библиши, настао је на основу речника који је објавила Народна библиотека Србије, прилагођен и допуњен. Доступан на веб-адреси <http://rbi.nb.rs/srlat/dict.html>.

¹³² Доступан на веб-адреси <http://geoliss.mre.gov.rs/recnik/>

¹³³ Доступан на веб-адреси <http://rudonto.rgf.bg.ac.rs/>

¹³⁴ Доступан на веб-адреси <http://termi.rgf.bg.ac.rs/>

алата (Stanković, Krstev, и други 2017, 3). Преводачке меморије представљају збирке сегмената у којима је текст изворног језика повезан са еквивалентним преводом текста циљног језика односно произведени *TMX* документ састављен је од добијених јединица превођења (Андоновски 2019). *TMX* документ садржи заглавље и тело. Заглавље се састоји од метаподатака који описују поравнате текстове док тело садржи скуп преводачких јединица (енг. *Translation Units*, скр. *TU*), са две и више варијанти преводачких јединица (енг. *Translation Unit Variant* скр. *TUV*) које садрже исти сегмент текста на различитим језицима. У Библиши, *TUV* сегменти одговарају реченицама али сегмент такође може бити и део реченице односно може се састојати и од неколико реченица. Сваки *TUV* има атрибут *xml:lang* са *ISO* кодом језика текста сегмента као његове вредности. За производњу *TMX* докумената користи се програмски пакет *ACIDE* који подразумева употребу алата *Xalign* који служи за аутоматско поравнање реченице и конкорданцера за визуализацију поравнања и мануалну корекцију грешака приликом поравнавања (Stanković, Krstev, и други 2017, 3–4).

На слици 18 приказана је почетна страница дигиталне библиотеке Библиша на којој корисник задаје кључну реч за претрагу, одабира језик кључне речи као и текстуалну колекцију коју претражује. Са леве стране екрана смештени су лексички ресурси који служе за семантичко проширење упита, а које корисник може и мануално укључити у претрагу. На слици 19 илустрован је резултат проналажења кључне речи „електроенергетски систем“ у лексичким ресурсима који су инкорпорирани у ову дигиталну библиотеку и који се могу користити за проширивање упита на други језик и синонимима. Резултати претраге могу се семантички проширити хипернимима односно хипонимима.

The screenshot shows the 'BIBLIŠA: ALIGNED COLLECTION SEARCH TOOL' interface. At the top, there is a navigation bar with links: Home, Metadata browse, Metadata search, Mongo search, Manage data, Help, Tutorial, About. The main content area is titled 'WELCOME TO ALIGNED TEXT COLLECTION SEARCH TOOL!'. It features a search form with a 'Keyword' field, a language dropdown set to 'sr', and a 'Text collection' dropdown set to 'All'. Below the search form, there is a table of synonyms and text collections. The 'Synonyms' column lists various resources like WordNet, Dictionary of Librarianship, Biblimir, GeolISSTerm, RudOnto, and Termi. The 'Text collection' column lists various collections like All, INFOtheca / Infoteka, Underground Mining Engineering / Podzemni Radovi, Architecture and Urbanism / Arhitektura i urbanizam, Serbian Dental Journal / Stomatološki Glasnik Srbije, BAKTEL TEMPUS project, INTERA project Management, EIEner (highlighted), CESAR, Mining, and SrpNemKor. At the bottom, there are checkboxes for 'Include Hypernyms' and 'Include Hyponyms', a 'Preview and modify terms for query' button, and a 'New Search' button.

Слика 18. Почетна страница дигиталне библиотеке Библиша и одабир колекције текстова из области електроенергетике

WELCOME TO ALIGNED TEXT COLLECTION SEARCH TOOL!

Keyword Text collection

Synonyms	en	sr
<input checked="" type="checkbox"/> WordNet...		
<input checked="" type="checkbox"/> Dictionary of Librarianship ...		
<input checked="" type="checkbox"/> Biblimir ...		
<input checked="" type="checkbox"/> GeolISSTerm...		elektroenergetski sistem
<input checked="" type="checkbox"/> RudOnto ...		elektroenergetski sistem
<input checked="" type="checkbox"/> Termi ...	electric power system,power system,electricity system,system power	elektroenergetski sistem,elektroenergetski sistem,elektroprivredni sistem

Include Hypernyms Include Hyponyms
 Match query: both sentences , only one language sentences , '-' no filtering.

Morphological query expansion

Слика 19. Резултат претраге према кључној речи

На основу резултата постављеног упита, систем генерише паралелизоване конкорданце сегмената као што је илустровано на слици 20. Сваку линију конкорданце прати и податак из ког је документа преузета, веза до метаподатака који додатно описују документ као и веза до самог документа.

SEARCH RESULTS

	Number of concordances (en/de/fr): 120	Broj konkordansi (sr): 120
Elektroprivreda apstrakti, ID: 8.2.3.1 metadata	n1788 Key words: SCADA, PHP, MySQL, Power System , energy consumption	n1788 Ključne reči: SCADA, PHP, MySQL, elektroenergetski sistem , potrošnja električne energije
Elektroprivreda apstrakti, ID: 8.2.3.1 metadata	n513 Key words: power systems, wind power plants, intermittency, energy storage, plant's economy	n513 Ključne reči: elektroenergetski sistemi , vetroelektrane, prekidnost, akumulacija energije, ekonomičnost elektrane
Elektroprivreda apstrakti, ID: 8.2.3.1 metadata	n995 Key words: relay protection, distribution network, power plant, embedded generation, power system	n995 Ključne reči: relejna zaštita, distributivna mreža, mala elektrana, distribuirana proizvodnja, elektroenergetski sistem
Elektroprivreda apstrakti, ID: 8.2.3.1 metadata	n1710 The paper gives overview of problems faced during reconnection of Power System of Serbia and Montenegro back to The First Synchronous UCTE Zone.	n1710 U oktobru 2004. godine elektroenergetski sistem Srbije i Crne Gore rekonektovan je sa nekolicinom okolnih zemalja iz druge sinhronne zone u prvu sinhronu UCTE zonu.
Elektroprivreda apstrakti, ID: 8.2.3.1 metadata	n581 Those analyses have been established on the model of real interconnection, which consists of the EPSs of Serbia, Montenegro, Bosnia and Herzegovina, Hungary, Croatia, Macedonia, Romania, Bulgaria, Greece and Albania.	n581 Pri tome, modelovana je realna elektroenergetska interkonekcija, koju sačinjavaju pored elektroenergetskog sistema Srbije, elektroenergetski sistemi Crne Gore, Bosne i Hercegovine, Hrvatske, Mađarske, Makedonije, Rumunije, Bugarske, Grčke i Albanije.
Elektroprivreda apstrakti, ID: 8.2.3.1 metadata	n340 The system for excitation regulation, unlike the excitation system, includes a regulation structure, i.e. synchronous generator which is on the one side connected with the excitation system and on the other side with the power system .	n340 Sistem za regulaciju pobude, za razliku od sistema pobude, uključuje i objekat regulacije, tj. sinhroni generator spojen s jedne strane za sistem pobude, a s druge za elektroenergetski sistem .
Zakon o energetici, ID: 8.3.2.7 metadata	n37 21) energy system is an electricity system , a natural gas, oil or oil derivatives system, and distance heating and cooling system comprising energy facilities interlinked into a single technical and technological system;	n37 21) energetski sistem je elektroenergetski sistem , sistem prirodnog gasa, nafte ili derivata nafte i sistem daljinskog grejanja i hlađenja, koji se sastoji od energetskih objekata međusobno povezanih tako da čine jedinstven tehničkotehnoški sistem;
Elektroprivreda apstrakti, ID: 8.2.3.1 metadata	n1790 Automated Meter Reading (AMR) system is officially in use in National Dispatching Center in Podgorica since May, 1st 2005. After that date, electric power system controlled from this Center became the best equipped system for billing and monitoring of power flows in this part of Europe.	n1790 Sistem za automatsko daljinsko očitavanje brojlara (AMR sistem) u Nacionalnom dispečerskom centru u Podgorici zvanično je pušten u rad 01. 05. 2005. godine, čime je elektroenergetski sistem kojim ovaj Centar upravlja, dobio najsavremeniju infrastrukturu za kontrolu i obračun tokova električne energije u ovom dijelu Evrope.
32009L0028_SR2.txt, ID: 8.1.1.29 metadata	n490 Member States shall take the appropriate steps to develop transmission and distribution grid infrastructure, intelligent networks, storage facilities and the electricity system , in order to allow the secure operation of the electricity system as it accommodates the further development of electricity production from renewable energy sources, including interconnection between Member States and between Member States and third countries.	n490 Države članice treba da preduzmu odgovarajuće korake da razviju mrežnu infrastrukturu za prenos i distribuciju, pametne mreže, objekte za skladištenje i elektroenergetski sistem , da bi se omogućio bezbedan rad elektroenergetskog sistema i omogućio dalji razvoj proizvodnje električne energije iz obnovljivih izvora, uključujući i interkonekciju između država članica i između država članica i trećih zemalja.
32009L0072_SR2.txt, ID: 8.1.1.30 metadata	n198 'interconnector' means equipment used to link electricity systems;	n198 „interkonektivni dalekovod“ - uređaji i oprema za povezivanje elektroenergetskih sistema ;
Elektroprivreda apstrakti, ID: 8.2.3.1 metadata	n430 Two characteristic cases of PSS's parameter settings are analyzed.	n430 Razmatrana su dva karakteristična slučaja podešenja stabilizatora elektroenergetskih sistema .
Elektroprivreda apstrakti, ID: 8.2.3.1 metadata	n427 Phase compensation method for tuning of Power System Stabilizers (PSSs), which are the main resources for damping of low-frequency electromechanical oscillations, is applied.	n427 Primenjen je postupak fazne kompenzacije za podešavanje stabilizatora elektroenergetskih sistema (PSS – "Power System Stabilizer"), kao osnovnog sredstva za prigušenje sporo promenljivih elektromehaničkih oscilacija.

Слика 20. Паралелизоване конкорданце сегмената као резултат претраге кључном речи „електроенергетски систем“

Корпус електроенергетике који је предмет ове дисертације, смештен је у дигиталну библиотеку Библиша. Више о начину и поступцима обраде овог корпуса на српском и енглеском језику биће приказано у поглављу 7 овог рада.

6. КОРПУСНИ АЛАТИ И ЈЕЗИЧКИ РЕСУРСИ У ОБРАДИ ПРИРОДНИХ ЈЕЗИКА

У претходном поглављу представљени су, између осталог, и паралелни корпуси које је до сада саставила Група за језичке ресурсе и технологије Универзитета у Београду. У овом поглављу биће речи о софтверским решењима односно корпусним алатима и језичким ресурсима који су коришћени у обради и анализи наведених паралелних корпуса, међу којима је и корпус електроенергетике за који су поступци обраде детаљно образложени у наредном поглављу.

Обрада природног језика као једне од области рачунарске лингвистике подразумева спровођење различитих поступака као што су лексичка, морфосинтаксичка и семантичка анализа. То значи да се у контексту обраде природног језика између осталог врши сегментација текста (препознавање речи и реченица), токенизација, лематизација, етикетирање врсте речи (енг. *Part of Speech Tagging*, скр. *POS tagging*), граматичко парсирање, а обављају се и специфични задаци: препознавање именованих ентитета, екстракција терминологије, машинско превођење, итд.

Избор корпусних алата зависи од бројних чинилаца од којих су само неки: врста софтвера (слободан, бесплатан, отвореног кода, власнички, комерцијални), платформа на којој програм ради (односи се на архитектуру рачунара, оперативни систем, програмско окружење у ком програм може да ради), архитектура клијент–сервер (због огромне величине савремених електронских корпуса као и брзог интернета, изгубила се потреба за складиштењем корпуса на локалним рачунарима као и брзом хардверу који омогућавају њихову претрагу, већ се велики део операција извршава преко веба), постављање нових верзија (након уклањања одређених неисправности у раду програма и уградње нових могућности, као и системска подршка приликом коришћења одабраног алата.

Корпусни алати користе се приликом претпроцесирања односно претходне обраде текста, а обухватају уређивање текста, оптичко препознавање карактера (уколико је потребно), конверзију текста из једног формата у други, као и индексирање ради једноставнијег претраживања.

Под језичким ресурсима дефинишемо скуп језичких података и описа у машински читљивом облику који се користе за развијање, унапређивање или евалуацију алгоритама или система за обраду природног језика (Stanković 2009).

Корпусни алати користе разне врсте језичких ресурса (Utvić 2013):

- Корпусе које претражују и анализирају;
- Референтне корпусе које консултују за потребе статистичке анализе;
- Лексичке ресурсе попут електронских речника, семантичких мрежа, вишејезичних база података, онтологија;
- Граматичке ресурсе (банке синтаксичких стабала, итд.)

Корпусни алати користе се и приликом претраживања корпуса. Задавањем упита одговарајућег упитног језика добијају се резултати у виду конкорданци. Упитни језици углавном се заснивају на употреби регуларних израза, а неки од њих омогућавају претрагу аотираних корпуса по придруженим информацијама (лингвистичким, метаподацима који се односе на текстове корпуса односно на сам корпус). Помоћу упитних језика могу се претраживати и паралелизовани текстови, и тада се као резултат добијају паралелизоване конкорданце (Utvić 2013).

Корпусни алати користе се и за статистичку анализу корпуса која подразумева генерисање листе учесталости, њено сортирање према леми, типу, врсти речи, итд., генерисање *n*-грама, колокациону анализу, одређивање кључних речи текста, итд. (Utvić 2013).

У досадашњем раду, Група за језичке технологије Универзитета у Београду користила је, између осталог, софтверска окружења *IMS OCWB* (енг. *IMS Open Corpus Workbench*) и *Unitex* за обраду текстова.

6.1. IMS Open Corpus Workbench

*IMS OCWB*¹³⁵ је бесплатан програмски пакет, отвореног кода, развијен за рад са великим текстуалним корпусима. Настао је на Институту за аутоматску обраду природних језика (нем. *Institut für Maschinelle Sprachverarbeitung*, скр. *IMS*) у Штутгарту. Једна од његових најзначајнијих компонената је алат за постављање упита за претрагу *CQP* (енг. *Corpus Query Processor*) који се заснива на регуларним изразима, у складу са стандардом *POSIX*¹³⁶. Тренутно препоручена верзија за употребу овог софтвера је 3.4, док је наредна 3.5 још увек у фази израде. Верзија 3.5 је у складу са стандардом *PCRE*¹³⁷ (Evert и The CWB Development Team 2020, 62). *CQP* омогућава претрагу према логичким изразима, као и претрагу додатних позиционих атрибута тј. претрагу по морфосинтаксичкој анотацији. Као резултат претраге *SrpKor*-а добијају се конкорданце које корисник може да прегледа, направи избор које конкордаце жели да задржи на приказу као и да сачува избор издвојених конкорданци.

Група за језичке ресурсе и технологије је, помоћу алата *IMS OCWB* развила онлајн сумеђу за претраживање развијених корпуса, а помоћу њега је реализована и прва верзија Корпуса савременог српског језика.

6.2. Unitex/GramLab

*Unitex/GramLab*¹³⁸ представља софтверски пакет отвореног кода, развијен за потребе анализе и обраде корпусних текстова на природним језицима уз коришћење језичких ресурса и алата. Ови ресурси садрже електронске морфолошке речнике, локалне граматике и табеле лексикон-граматика, а иницијално су били развијени на француском језику у Лабораторији за аутоматску документацију и лингвистику (фр. *Laboratoire d'Automatique Documentaire et Linguistique*, скр. *LADL*), под руководством Мориса Гроса (*Maurice Gross*) (Paumier 2020, 13).

Unitex је настао као замена за софтверски алат *Intex*, чији је творац Макс Силберштајн (енг. *Max Silberztein*), настао је у истој лабораторији *LADL*, и служио је за обраду текстова уз помоћ електронских морфолошких речника у форматима *DELA* (фр. *Dictionnaire 'Electronique du LADL*, скр. *DELA*) (Silberztein 1993). Након прекида сарадње аутора *Intex*-а и лабораторије *LADL*, развијен је систем *Unitex* са подршком за *Unicode* кодну шему.

Unitex је стона апликација односно не подржава архитектуру клијент–сервер. Имплементирана је коришћењем програмских језика *C++* и *Java*, и има могућност да се компилира на платформама *Linux*, *Windows* и *Mac OS X*. Најновија верзија система *Unitex* је *3.3alpha*¹³⁹ и доступна је на његовом званичном веб-сајту. На сајту је такође доступан и детаљан приручник за употребу овог алата.¹⁴⁰ Међу 22 основна језичка ресурса, колико је тренутно покривено системом *Unitex*, налазе се и ресурси за српски језик.

У обради текстова корпуса, *Unitex* користи електронске морфолошке речнике као и локалне граматике у форми графова. Сви лексички ресурси се приликом обраде текстова компилирају у коначне аутомате и трансдукторе који се користе приликом претпроцесирања тј. претходне обраде текста као и његовог претраживања. Претпроцесирање текста (видети поглавље 7.2.2) врши се на његовој копији, а обухвата токенизацију и сегментацију текста. У овој фази могуће је применити и лексичке ресурсе попут електронских морфолошких речника и речничких трансдуктора на текст при чему се добија морфолошки речник текста као пресек примењених лексичких ресурса и самог текста.

Избор језика са којим ће се радити је корак којим се започиње рад на систему *Unitex*. Сваки језик представља један поткаталог приватног каталога који је изабран приликом првог покретања

¹³⁵ Информације доступне на веб-сајту: <http://cwb.sourceforge.net/>

¹³⁶ *Portable Operating System Interface for uniX*.

¹³⁷ Perl Compatible Regular Expressions

¹³⁸ <http://unitexgramlab.org/>

¹³⁹ Податак од 24. фебруара 2021. године.

¹⁴⁰ Приручник је доступан на: <https://unitexgramlab.org/releases/3.2/man/Unitex-GramLab-3.2-usermanual-en.pdf>

система *Unitex*. Корени каталог (енг. *root directory*) одабраног језика се копира у приватни каталог како би се избегла евентуална оштећења системских датотека приликом њихове модификације.¹⁴¹

Овакав поткаталог садржи већи број других поткаталога у које се смештају текстови које желимо да анализирамо (поткаталог *Corpus*), компилирани речници који служе за њихову обраду (поткаталог *Dela*), графови који се користе за претраживање текстова (поткаталог *Graphs*), графови који служе за отклањање вишезначних анотација у тексту (поткаталог *Elag*), поткаталози у које се смештају изворни облици речника лема и флективни графови који се користе за производњу речника облика (поткаталог *Inflection*). У истом поткаталогу налазе се и конфигурационе датотеке којима се дефинише алфавет одабраног језика, као и конфигурација програма *Unitex* – системске датотеке. Свака измена подешавања коју корисник направи чува се као конфигурациона датотека која је смештена у директоријуму актуелног језика (Krstev 2016).

Обрада текста започиње након ових почетних подешавања система. Улазне текстуалне датотеке које ће бити садржај корпуса морају бити у формату чистог, сировог текста: **.txt *.xml *.html* (енг. *raw* или *plain text*). Имајући у виду да *Unitex* користи кодну шему *Unicode*, подразумева се да су текстови већ у том формату. Текстуалне датотеке које производи сам *Unitex* у фази претпроцесирања односно претходне обраде текста добијају суфикс **.snt.* и многе функције система *Unitex* могу се применити само на ову врсту текстова (Krstev 2016).

Фаза претходне обраде састоји се од више корака: нормализација сепаратора, подела у реченице, нормализација недвосмислених облика, токенизација и примена речника.

Под сепараторима се подразумевају бланко карактери, карактери за нови ред и карактери за табулацију.

Подела у реченице подразумева коришћење локалних граматика које зависе од језика који се користи. Њиховом применом се одређује основна јединица лингвистичке обраде. Нормализација недвосмислених облика подразумева да се неки облици у тексту могу по потреби нормализовати. То значи да се нпр. скраћеница попут „ОДС“ замени њеним пуним обликом који гласи „Оператор дистрибутивног система“.

Токенизација подразумева поделу текста на токене. Токен може бити сепаратор реченица, ознака за заустављање, лексичка етикета, ниска слова, појединачни несловни карактер као што су интерпункцијски знаци, цифре, итд. Као производ ове фазе добијају се листе токена креиране у посебним датотекама и разврстане према редоследу токена по ком се јављају у тексту, по фреквенцији појављивања и алфаветском редоследу. Уз датотеке које садрже ове листе, налазе се и оне које садрже целе бројеве који одговарају индексу токена у датотеци која је задужена за њихов редослед у тексту, као и статистику у смислу укупног броја граничника реченица, укупног броја токена (и укупног броја различитих токена), укупног броја простих облика (потенцијалних) речи, укупног броја различитих речи и цифара.

У последњој фази налази се примена речника, а састоји се у томе да се направи подкуп примењених речника за оне облике који су се појавили у тексту. *Unitex* формира још два речника и то речник полилексемских речи и речник непознатих тј. непрепознатих речи.

6.3. Електронски морфолошки речници

Одмах на почетку важно је скренути пажњу на то да се електронски речници разликују од машински читљивих речника, јер су електронски речници намењени рачунарској обради текста (софтверу), за разлику од машински читљивог речника који је намењен човеку као кориснику. Човек као корисник углавном у речнику тражи значење неке речи или њен превод, док се електронски речници користе у истраживањима језика и креирању језичких алата (Стијовић, Сабо и Станковић 2017)

Електронски речници представљају неопходан лексички ресурс у различитим фазама аутоматске обраде текстова. То је речник намењен обради и анализи текста, а садржи неопходне информације које му олакшавају процес сегментације и морфолошке анотације. Ипак,

¹⁴¹ Напомена: програм не копира речнике у приватни директоријум због уштеде простора на диску.

конструкција електронских речника и локалних граматика је спора јер се највећи део процеса њихове израде обавља ручно, али би без њихове подршке напредна обрада текста била немогућа (Vitas и Krstev 2012, 279).

На изградњи електронских морфолошких речника и граматика дуги низ година радила је међународна мрежа лабораторија која се бави рачунарском лингвистиком, *RELEX* (Laporte 2003). Ову лабораторију основали су Морис Грос и његов *LADL* тим. Систем *Unitex* ради са речницима које су изградили чланови ове мреже, а који покривају француски, енглески, грчки, португалски, руски, тајландски, корејски, италијански, шпански, норвешки, арапски, немачки, пољски, српски и друге језике.

Конструкција оваквог речника подразумева прецизну и исцрпну класификацију флективних својстава речи, при чему се под речју подразумева ниска алфаветских карактера ограничена сепараторима (интерпункцијски знаци или контролни карактери). Овако дефинисана реч се обично назива проста реч (енгл. *simple word*) за разлику од сложене (полилексемске) речи (енгл. *compound word*) која се састоји од контингентног низа простих речи. (Vitas и Krstev 2009).

Ови речници омогућавају да се текст аутоматски сегментира на речи и анализира на морфолошком нивоу методом лексичког препознавања. Поступак се састоји у поређењу речи у анализираном тексту са садржајем речника. Уколико се реч налази у речнику, њој ће бити додељена информација о могућим лемама и одговарајућим граматичким категоријама, а ако се у њему не налази, биће сврстана у категорију непознатих речи (Vitas и Krstev 2009, 88).

Морфолошки речници са којима ради систем *Unitex* су у формату *DELA* и организовани су кроз речнике простих речи *DELAS* и *DELAF* и речнике полилексемских речи *DELAC* и *DELACF*. *DELAS* (фр. *Dictionnaires électroniques des mots simples*) је речник канонских облика (лема) и користи се за генерисање речника *DELAF* (фр. *Dictionnaires électroniques des formes fléchies*), који је речник подређених (флективних) облика и користи се у морфолошкој анализи и анотацији текста. Везу између канонског облика и свих његових подређених (флективних) облика омогућавају трансдуктори коначних стања под системом *Unitex* (Krstev 2016a).

Речник простих речи састоји се, дакле, из три дела (Krstev, Vitas и Pavlović-Lažetić 2008, 3):

- Речник лема *DELAS*,
- Скуп трансдуктора који описују особине флективних парадигми,
- Речник флективних облика *DELAF*

Структура *DELAS*-а је таква да је свакој одредници (леми) у речнику придружена ознака за врсту речи тј. утврђени код од једног или више великих слова; веза са подређеним облицима (ако постоје), и то је обично нумеричка (или алфанумеричка) ознака која уз врсту речи омогућава аутоматско генерисање свих подређених облика за речник *DELAF*; синтаксни, семантички, дијалекатски, употребни, доменски, итд. маркери који додатно описују лему. Флективна својства су изражена кодом флективне класе који описује флективне наставке и њима додељену граматичку вредност (Vitas и Krstev 2009, 88).

Одредницу у речнику *DELAF* прати: канонски облик (или лема); ознака врсте речи (преузета од леме); синтаксни, семантички, дијалекатски, употребни, доменски, итд. маркери преузети од леме; скуп кодова који представљају вредности граматичких категорија које се односе на облик (Krstev 2016a).

Попут речника простих, и морфолошки речник полилексемских речи се састоји од два речника а то су: *DELAC* и *DELACF*. Одредницу у речнику *DELAC* (фр. *Dictionnaire électronique des noms composés*) прате леме које су саставни део сложеница и које су као такве дефинисане према речнику *DELAS* заједно са флективним кодом класе као и синтаксичким и семантичким маркерима који се односе на дату полилексемску јединицу (Krstev и Vitas 2009a, 205).

Флективни облици у речнику *DELACF* (фр. *Dictionnaire électronique de mots composés fléchies*) генеришу се аутоматски на основу информација које стоје уз одредницу у речнику *DELAC* и на основу података у речницима *DELAS/DELAF*. Одредницу у *DELACF* речнику прате исте информације као и одреднице у речнику простих облика *DELACF*. За генерисање одредница за речник *DELACF* потребна су два типа трансдуктора: трансдуктори који производе флективне

облике монолексема и други који регулишу однос између конституената у одређеној полилексеми (Андоновски 2019, 86).

Током деведесетих година двадесетог века, Група за језичке технологије отпочела је рад на развоју морфолошког електронског речника српског језика, на основу ових модела, да бисмо данас имали развијене електронске речнике на ћириличном и латиничном писму, чији су аутори др Цветана Крстев, професор Филолошког факултета Универзитета у Београду и др Душко Витас, професор Математичког факултета Универзитета у Београду.

Према подацима из маја 2021. године, речник простих речи се састоји од 205.003 леме, са 207.312 значења, док је број полилексемских јединица обухваћених речником 22.865, са 23.014 значења. Најзаступљеније врсте речи су именице (116.192 лема), придеви (64.274 лема) и глаголи (21.159 лема).

За опис појединих језичких феномена су изграђене локалне граматике које омогућавају да се опише структура и додели граматичко значење нивовима простих речи који се не могу експлицитно навести у речнику. Неки примери локалних граматика су граматике за препознавање низова бројева записаних речима, цифрама и њиховим комбинацијама (Vitas и Krstev 2009) или граматике за препознавање датума и других временских израза у тексту (Jaćimović 2019).

6.4. Локалне граматике

У општем смислу, локалне граматике које систем *Unitex* користи могу се представити као алат којим се описује већина језичких феномена. Користе се за задавање упита, а представљене су у форми регуларних израза или у форми графова.

Регуларни изрази у систему *Unitex* су:

- Токени и лексичке маске
- Конкатенација односно спајање два регуларна израза
- Унија два регуларна израза
- Клинијево затворење два регуларна израза (са значењем неограниченог појављивања).

У регуларном изразу токен може бити сепаратор реченица {S}, ознака за заустављање {STOP}, лексичка маска као нпр. <N:fs4q> (именица женског рода у једнини у акузативу), ниска слова, појединачни несловни карактер као што су интерпункцијски знаци, цифре, и остало, о чему је већ било речи у поглављу 6.2. Токени и лексичке маске се могу користити помоћу морфолошких филтера за постављање упита кроз регуларни израз у *POSIX* формату. (Krstev 2016b).

Лексичка маска је упитни образац који може да се сравни са једним или више токена. Лексичке маске су најзанимљивије када користе информације из речника.

Постоји више врста лексичких маски (Krstev 2016b):

- *Предефинисане лексичке маске* су уграђене у систем *Unitex* и не зависе од језика, као нпр: <E> – празно слово (ε је симбол из теорије формалних језика) и указује да на одређеном месту у тексту нечега не мора бити; <TOKEN> – сравњује се са сваким токеном осим са бланко карактером; <MOT> или <WORD> – сравњује се са токеном кога чини ниска слова односно све што је потенцијално реч језика.
- *Лексичке маске које се реферишу на речник текста* зависне су од језика тј. зависе од тога којим речницима је текст обрађен. Најједноставније маске се реферишу на лему или врсту речи. Нпр. <divan.N> се реферише на све токене чији је канонски облик „диван“, а врста речи именица. Сравниће се са „дивану“, али неће са „дивнијој“.
- *Лексичке маске које користе синтаксичке и семантичке маркере*. У овим маскама се семантички и синтаксички маркери раздвајају знаком +, а из речника текста се издвајају само они улази који имају све наведене синтаксичке или семантичке маркере. Нпр. <N+Num> издваја из речника улазе који су именице (енг. *N* – *Noun*) и које означавају људско биће (+*Num*).

- *Лексичке маске са искључивањем семантичких маркера.* У овим маскама се семантички и синтаксички маркери раздвајају знаком тилда ~, док се из речника текста издвајају само они улази који немају ни један од наведених синтаксичких или семантичких маркера. Нпр. Маска $\langle N+Nprop+Hum \rangle$ издваја из речника маркере $N+NProp$ – који имају властите именице а које уз то имају и маркер Hum – људско биће, док маска $\langle N+NProp\sim Hum \rangle$ издваја из речника маркере $N+NProp$ – који имају властите именице, али само оне које нису означене маркером Hum – људско биће. Редослед навођења маркера није од значаја.
- *Лексичке маске са флективним условима.* У овим маскама ограничења се могу постављати коришћењем флективних кодова. Користе се исти кодови који се користе и у примењеним речницима. Да би се ова ограничења користила, неопходно је да им претходи бар ознака речи или један позитиван семантички или синтаксни маркер. Нпр. $\langle N:s:pI \rangle$ препознаје у тексту све облике који су према речнику именице у једнини или именице у множини у номинативу.

Лексичке маске се могу и негирати и то тако што се уметне знак узвика одмах после знака за мање, нпр. $\langle !MOT \rangle$.

Конкатенација представља спајање регуларних израза на више начина. Могуће је експлицитно користити оператор за конкатенацију, а то је тачка. Нпр. $\langle PRO+ProA+Pos \rangle$. $\langle N \rangle$ се савлађује са присвојном заменицом иза које следи именица. Уместо експлицитног оператора, може се користити и бланко карактер и празна ниска. Када се формира сложенији израз у коме се користе различити оператори, тада се могу користити заграде (Krstev 2016b).

Унија два регуларна израза добија се коришћењем оператора + и савлађује се са једним од операнда (представља дисјункцију). Нпр. $\langle A \rangle + \langle PRO+ProA \rangle$ $\langle ljubav \rangle$ савлађује у тексту секвенцију од две речи од којих је прва придев или придевска заменица, а друга је реч „љубав“ у било ком облику (Krstev 2016b).

Клинијево затворење два регуларна израза или Клинијева звезда, омогућава да се у тексту савлађује нула, једно, или више појављивања регуларног израза на који се примењује. Користи се оператор *. Нпр. $\langle A \rangle * \langle N \rangle$ у тексту препознаје секвенције од једне, две, три и више речи од којих је последња именица, а претходе јој придеви. Ако желимо да регуларни израз препозна бар једно појављивање израза на који се примењује Клинијева звезда, довољно је да се тај регуларан израз понови тј. споји са Клинијевом звездом. Нпр. $\langle A \rangle \langle A \rangle * \langle N \rangle$ препознаје секвенције које се састоје од бар једног придева иза кога следи именица (Krstev 2016b).

Локалне граматике могу се формулисати помоћу графова, који представљају врло погодну визуелизацију регуларних израза. Свака кућица односно чвор у графу садржи регуларне изразе. Међутим, графови постају брзо веома сложени, а да би се олакшала њихова израда и да би се делови графова могли уклапати у друге графове за различите примене, често се користе подграфови (видети пример на слици 19 у одељку 7.3).

6.5. Паралелизација текста

Паралелни текст се формира од два текста од којих је један оригиналан, док је други настао његовим превођењем. Овакви (паралелни) текстови се упарују, односно поравнавају, и тада од паралелног текста настаје паралелизовани. Поступак превођења паралелног текста у паралелизован има две фазе. У првој фази се паралелни текстови деле на сегменте, који могу бити веће или мање целине попут пасуса, реченица или речи. У другој фази врши се поравнавање овако сегментираних текстова помоћу неке од расположивих метода за поравнавање. Поравнавањем се врши повезивање еквивалентних сегмената у два или више паралелних текстова (Stanković 2009, 147).

Група за језичке технологије Универзитета у Београду развила је програмски пакет за паралелизоване корпусе, тзв. интегрисано развојно окружење за паралелне корпусе, *ACIDE* (енг.

Aligned Corpora Integrated Development Environment) (Obradović, Stanković и Utvić 2008).¹⁴² Ово окружење, између осталог, обезбеђује графички кориснички интерфејс (енг. *Graphical User Interface*, скр. *GUI*) за:

- паралелизацију;
- визуелизацију паралелизованог текста, контролу и корекцију;
- генерисање датотека у *TMX* формату;
- разлагање датотека у *TMX* формату на датотеке појединачних језика;
- вертикализацију текста.

ACIDE обавља паралелизацију коришћењем програмских пакета *XAlign* и *Concordancier* које је развила Лоренска лабораторија за информатичка истраживања и примене (фр. *Laboratoire Lorrain de Recherche en Informatique et ses Applications*, скр. *LORIA*), у Француској.

О паралелизованим корпусима које је развила Група за језичке ресурсе и технологије било је више речи у поглављу 5.5.2.

6.6. Семантичка мрежа Ворднет

Под руководством професора Џорџа Милера (енг. *George Armitage Miller*), група психолингвиста и лингвиста је 1985. године у Лабораторији за когнитивне науке Принстонског универзитета почела да развија семантичку мрежу ворднет. Овај тим истраживача имао је идеју да направи ресурс који би повезао традиционалне лексикографске информације са модерним рачунарским могућностима (Miller, и други 1990).

Ворднет (енг. *Wordnet*) је велика лексичка база података, подељена према врстама речи, и то према именицама, глаголима, придевима и прилозима. Врсте речи груписане према скуповима синонима, тзв. синсетовима (енг. *synset* или *synonymous set*), од којих сваки сет изражава посебан концепт. Синсетови су међусобно повезани појмовним и лексичко-семантичким релацијама.

Именички део базе организован је као хијерархијска мрежа именичких чворова која се успоставља на основу односа постојања релације надређености и подређености (хипероним/хипоним) између појмова које ти чворови представљају. За неки појам можемо рећи да је подређен другом појму ако поседује сва својства која поседује и надређени појам, али наравно, има и нека своја специфична својства (Крстев, и други 2008). Пример за релацију надређености и подређености могу бити уређаји као што су „прекидач“ (надређени) и „укључивач“ (подређени). „Прекидач“ има функцију прекидања струје квара, док „укључивач“ осим ове функције прекидања има и функцију поновног успостављања струјног кола.

Још неке од релација које се јављају у ворднету су, најпре, део–целина и члан–целина (холоним/мероним) која повезује два појма од којих је један део другог. Пример за релацију део–целина и члан–целина могу бити уређаји као што су „агрегат“ (целина) и „генератор“ (део). „Агрегат“, осим генератора, садржи турбину, турбински систем, побудни систем и, евентуално, стабилизатор система. У Ворднету се јавља и релација антонимије која синсетове повезује супротним појмовима, на пример „произвођач – потрошач (електричне енергије)“. Постоје и многе друге семантичке релације, али ово су само неке од најчешћих које се успостављају у ворднету.

Структура српског ворднета заснива се на структури Принстонског ворднета. Српски ворднет почео је да се развија у оквиру Балканет пројекта, након чијег завршетка је рад на овој мрежи настављен волонтерски као и на постдипломским студијама на Групи за библиотекарство и информатику Филолошког факултета Универзитета у Београду (Крстев, и други 2008).

¹⁴² Аутори софтверског решења су проф. др Ранка Станковић и др Милош Утвић, под руководством проф. др Цветане Крстев, проф. др Душка Витаса и проф. др Ивана Обрадовића (Група за језичке технологије 2008).

Изградња ворднета заснована је на вишејезичности и омогућава компаративну језичку анализу, управљање знањем, управљање садржајем, екстракцију и проналажење информација, машинско превођење, итд. (Krstev, Koeva и Vitas 2006).

7. ОБРАДА ПАРАЛЕЛИЗОВАНОГ КОРПУСА ЕЛЕКТРОЕНЕРГЕТСКОГ СИСТЕМА

Велики број научноистраживачких активности је претходних година био посвећен екстракцији двојезичне терминологије из паралелизованих и упоредних корпуса у оквиру различитих научних домена.

Аутоматска екстракција терминологије (енг. *Automatic terminology extraction*, скр. *ATE*) као један од задатака у оквиру обраде природног језика, подразумева идентификацију доменски релевантних термина применом рачунарских метода. Аутоматска екстракција термина корисна је за широк спектар задатака као што су машинско превођење, конструкција тезауруса, класификација, индексирање, проналажење информација, идентификација термина за израду „традиционалних“ ресурса, итд.

Двојезичну екстракцију терминологије до сада је примењивао велики број аутора за различите језичке парове. Неки од примера су:

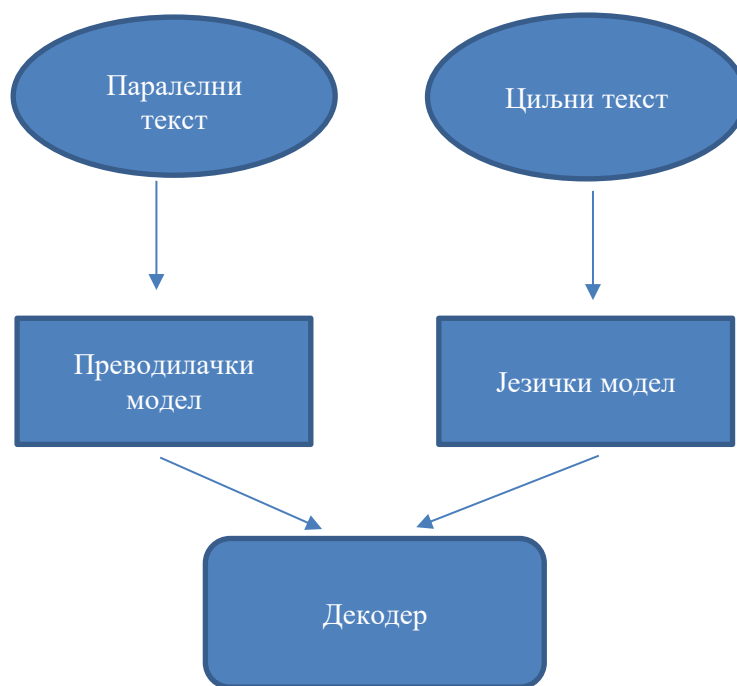
- енглески / холандски (Hoste, и други 2010),
- енглески / француски (Hazem и Morin 2016, Nakami и Bollegala 2017, Liu, Morin и Saldarriaga 2018, Semmar 2018),
- енглески / шпански (Oliver 2017, Haque, Penkale и Way 2018),
- енглески / турски (Haque, Penkale и Way 2018),
- енглески / немачки (Arcan, и други 2017),
- енглески / хинди (Haque, Penkale и Way 2018),
- енглески / вијетнамски (Tien, и други 2019),
- енглески / италијански (Arcan, и други 2017, Liu, Morin и Saldarriaga 2018),
- енглески / словеначки (Repar, и други 2019),
- енглески / српски (Šandrih, Krstev и Stanković 2020), итд.

Упоредне доменске корпусе као улазне ресурсе користили су највише аутори (Hazem и Morin 2016, Nakami и Bollegala 2017, Liu, Morin и Saldarriaga 2018). Аутори (Oliver 2017, Arcan, и други 2017, Haque, Penkale и Way 2018, Semmar 2018, Šandrih, Krstev и Stanković 2020) су предлагали технике засноване на паралелним корпусима и статистичким табелама фраза за машинско превођење, док су неки од наведених предлагали коришћење и осталих ресурса као што су двојезични речници (Hazem и Morin 2016, Nakami и Bollegala 2017, Semmar 2018, Liu, Morin и Saldarriaga 2018, Šandrih, Krstev и Stanković 2020), преводилачке меморије (Oliver 2017), морфосинтаксички обрасци (Semmar 2018, Šandrih, Krstev и Stanković 2020) и термилошке базе података (Oliver 2017, Nakami и Bollegala 2017). Неки истраживачи су користили чак и једнојезичне корпусе међутим, у таквим случајевима, екстракција двојезичне терминологије је обављана у каснијим корацима (Tien, и други 2019).

Екстракција терминологије вршена је у оквиру различитих предметних домена међу којима је:

- биомедицина (Hoste, и други 2010),
- вулканологија, ветроенергија, преглед вести (Hazem и Morin 2016, Liu, Morin и Saldarriaga 2018),
- законодавни текстови (Hazem и Morin 2016, Semmar 2018),
- политика и финансије (Oliver 2017, Haque, Penkale и Way 2018),
- медицина (Hazem и Morin 2016, Oliver 2017, Nakami и Bollegala 2017, Arcan, и други 2017, Liu, Morin и Saldarriaga 2018, Tien, и други 2019),
- хемија, техника, наука, туризам, уметност, видео-игре, фармацеутски производи (Haque, Penkale и Way 2018),
- информациона технологија (Arcan, и други 2017, Haque, Penkale и Way 2018),
- библиотекарство и информационе науке (Šandrih, Krstev и Stanković 2020),
- менаџмент (Анђелковић, Сеничић и Станковић 2018), као и
- електроенергетика, који је и предмет ове докторске дисертације (Ivanović, и други 2021).

Велики број аутора користи систем за статистичко машинско превођење *Moses*¹⁴³, који од 1999. године развија тим за статистичко машинско превођење (енг. *Statistical Machine Translation*, скр. *SMT*) у Центру за обраду језика и говора на Универзитету Џон Хопкинс. Он омогућава корисницима да на својим корпусима од почетка до краја обуче и евалуирају целовит систем за статистичко машинско превођење: обучавање језичког модела, обучавање преводачког модела, њихово обједињавање (компонента декодер), а затим и аутоматско тестирање модела превођења за било који језички пар. На слици 21 илустрован је дијаграм статистичког машинског превођења.



Слика 21. Дијаграм статистичког машинског превођења

Тестирање се спроводи на корпусу паралелних текстова у којем алгоритам за претрагу, на основу модела превођења, брзо проналази најадекватнији превод између већег броја понуђених случајева. У статистичком машинском превођењу, системи за превођење се тестирају на великим узорцима паралелних текстова (из којих се може установити како систем преводи мање сегменте текстова), као и на великим узорцима једнојезичних текстова (на основу чега се утврђује како циљни језик треба да изгледа) (Koehn 2020). Посебно битна компонента система *Moses* је алат за поравнавање сегмената реченица, *GIZA++*. Овај алат користи приступ фразно заснованих паралелних текстова што подразумева поделу реченице на текстуалне одломке који се потом упарују са одломцима циљног језика. Још један од алата који представља систем за аутоматску двојезичну термиолошку екстракцију и служи као проширење постојећег алата за превођење *TonD* (енг. *Terminology on Demand*), развио је аутор Оливер (Oliver 2017).

За екстракцију билингвалне терминологије за српски из области библиотекарства и информатике коришћењем упарених сегмената (*GIZA++* и *Moses*) коришћени су следећи ресурси и алати: листа терминологије из домена библиотекарства и информатике на изворном језику односно екстрактор за изворни језик (у овом случају енглески), доменски упарени корпус изворног и циљног језика, на основу чега је извршена екстракција терминологије за циљни језик (српски) употребом постојећег екстрактора терминологије за циљни језик (Šandrih, Krstev и Stanković 2020).

¹⁴³ <http://www.statmt.org/moses/>

Оливер и Васкез су развили класу *TBXTools* у програмском језику *Python* за рангирање термина кандидата у односу на њихову учесталост. Овај алат примењује статистичке и лингвистичке методе за екстракцију вишечланих термина и омогућава корисницима да на једноставан начин идентификују вишечлане термине из доменских корпуса као и преводне кандидате (Oliver и Vázquez 2015, 474). Након тога спроведена је аутоматска детекција преводних еквивалената коришћењем паралелног корпуса помоћу алата *GIZA++*. Преводни кандидати су затим прегледани у табели фраза, а као резултат, понуђено је 5 највероватнијих решења за сваког изворног кандидата.

У откривању термина у медицинским информацијама о пацијенту, у домену биомедицине, аутори Осте, Фанопстал, Лефевер и Делер (Hoste, и други 2010) користе приступ који се заснива на речнику и машинском учењу, при чему приступ машинском учењу не користи само лексичке информације, већ у обзир узима и морфолошке и статистичке информације, локални контекст, итд.

Једнојезично екстраховање доменских термина уз коришћење различитих алата за екстракцију терминологије први пут је приказано у раду аутора Аркан, Турчи, Тонели, Ботелар (Arcan, и други 2017). Имајући у виду изворни термин и паралелни реченични пар у којем се појављује, добијен је скуп могућих превода и то превођењем термина са постојећим системом статистичког машинског превођења који је испробан на истом корпусу помоћу алата *GIZA++*, односно применом поравнања речи.

Аутори Хак, Пенкале и Веј (Haque, Penkale и Way 2018) предложили су двојезичну екстракцију термина која интегрише спољне ресурсе за проналажење превода изворних кандидата који постоје у табели фраза алата *GIZA++*. У првом кораку, и на изворној и на циљној страни паралелног корпуса изведено је једнојезично екстраховање термина на основу теста логаритамске веродостојности (енг. *log-likelihood*). Двојезична терминологија са екстрахованом листом једночланих термина извучена је помоћу табеле фраза и процедуре која користи базе података вишејезичних веза на Википедији за избор одговарајућих циљних термина за изворни појам.

За састављање двојезичних листи вишечланих израза (енг. *MultiWord Expressions* скр. *MWE*) аутор Семар (Semmar 2018) представља тзв. хибридни приступ за екстракцију и поравнање у једном кораку и то, екстракцију улазног паралелног корпуса поравнатог помоћу алата *GIZA++* и функције рангирања, након чега се вишечлани кандидати (*MWE*) филтрирају морфосинтаксичким обрасцима.

Систем за двојезичну терминолошку екстракцију и поравнање термина који се заснива на једнојезичној терминолошкој екстракцији из енглеских и словеначких текстова и аутоматском поравнању листе уз коришћење скупа од седам метода двојезичног поравнања, који се најпре извршавају одвојено, а затим се обједињују помоћу рачунарских алгоритама развили су и приказали аутори Репар, Потпечан, Вавпетич, Лаврач, Полак (Repar, и други 2019).

Метод за екстраховање терминологије из домена електроенергетике коришћен за анализу енглеско/српских језичких парова се ослања на паралелни корпус и табеле фраза статистичког машинског превођења евалуираних помоћу алата *GIZA++*. Попут аутора који су применили једнојезично екстраховање термина (Arcan, и други 2017), за потребе ове анализе изведена је једнојезична екстракција термина из домена електроенергетике, коришћењем постојећих терминолошких екстрактора како за српски тако и за енглески језик посебно. За разлику од принципа рангирања који су у свом раду представили аутори Оливер и Васкез (Oliver и Vázquez 2015), за потребе овог истраживања анализиране су различите функције подударања које поравнавају добијене једнојезичне терминолошке листе уз коришћење претходно добијених табела фраза, што је детаљније обрађено у одељцима који следе.

7.1. Процес компилације двојезичног терминолошког материјала

Аутоматска екстракција двојезичне терминологије за терминолошке парове на српском и енглеском језику ослања се на паралелизоване двојезичне доменске текстове, терминолошке

екстракторе за оба језика као и на алатке за поравнавање комада (енг. *chunk*), где комади представљају подреченичне низове токена.

За екстракцију двојезичне терминологије коришћен је систем *BilTE* (енг. *Bilingual Terminology Extraction*), који је развила Група за језичке ресурсе и технологије Универзитета у Београду. Овај алат омогућава екстракцију двојезичне терминологије на основу паралелне двојезичне колекције текстова и терминолошких (једнојезичних) листи изворног и циљног језика (Šandrih 2020).

BilTE је веб-апликација¹⁴⁴ помоћу које је извршена екстракција терминологије из корпуса електроенергетике. Садржи три модула, а то су: унос, паралелизовање и завршна обрада као и резултат.

Приликом уноса корисник отпрема (енг. *upload*) две текстуалне датотеке упарене на нивоу реченица које имају исти назив, али различите екстензије које имплицирају језик текста који садржи датотека (нпр. *elektroenergetika.sr* и *elektroenergetika.en*). Овакве датотеке потом потом обрађује алат *GIZA++*, чији је рад укључен као део процедуре. Након тога учитава терминолошку листу српских, а потом и терминолошку листу енглеских термина (о производњи ових листа биће више речи у одељцима 7.3 и 7.4).

Упаривање алатом *GIZA++* даје тзв. *табелу фрази*, а изводи се тако што *GIZA++* паралелно учитава два унета текста. Када год се два вишејезична комада појаве заједно, њихово појављивање се записује у текстуалну датотеку (*f_phrases*). Након тога, *f_phrases* се сортира на два начина (према изворном и циљном термину), и тако се добијају две табеле. Термини кандидати који имају вероватноћу нижу од одређеног прага се одбацују. Након тога, наступају два корака у фази завршне обраде. У првом кораку се одбацују термини који не припадају домену, а у другом кораку се лематизују енглески комади користећи интерфејс за употребу тезауруса *WordNet* кроз модул *NLTK* за програмски језик *Python*¹⁴⁵ и српски комади са електронским речницима за српски језик (Krstev 2008).

Код завршне обраде основни кораци су: да се задрже само кандидати присутни у листи на енглеском језику који се потом укрштају са екстрахованим српским вишечланим терминима (Šandrih и Stanković 2019, 131–133).

Двојезични терминолошки пар се добија ако се оба комада у паралелизованом пару подударе са екстрахованим терминима из одговарајућих једнојезичних листи. Архитектура таквог система приказана је на слици 22. Једнојезична доменска терминологија на оба језика се добија применом терминолошких екстрактора. Након добијања коначних једнојезичних терминолошких листи, одговарајући парови се аутоматски поравнавају, затим редукују и потом дају двојезичну доменску терминолошку листу преводних парова кандидата који су спремни за мануелну евалуацију.

¹⁴⁴ <http://bilte.jerteh.rs/>

¹⁴⁵ <https://github.com/nltk/wordnet>



Слика 22. Дијаграм предложеног система за терминолошку екстракцију

7.2. Двојезични паралелни корпус из домена електроенергетике

У овом поглављу детаљно је описан начин израде двојезичног корпуса из домена електроенергетике који је и предмет ове докторске дисертације. Корпус је састављен од техничких, научних и законодавних текстова на српском и енглеском језику објављиваним у периоду од 2006. до 2015. године. Имајући у виду да је састављен за потребе одређеног лингвистичког истраживања и да припада специфичном домену, овај корпус се сврстава у специјализоване корпуре, за разлику од општих (референтних) чији је циљ да представе језик у целини.

Већина текстова који чине овај корпус доступна је на вебу, на сајту Јавног предузећа „Електропривреда Србије“ (скр. ЈП ЕПС) и Агенције за енергетику Републике Србије (скр. АЕРС), док мањи део корпуса чине законодавни текстови из домена електроенергетике који још увек нису јавно доступни имајући у виду процедуру кроз коју морају проћи пре свог коначног објављивања, као и нека интерна документа ЈП ЕПС.

7.2.1. Садржај корпуса из домена електроенергетике

Први корак у изради ове дисертације био је избор текстуалног материјала који ће чинити корпус из домена електроенергетике, а који је, пре свега, морао бити у довољној мери обиман да би био репрезентативан. Имајући у виду да су на вебу доступна законска и подзаконска акта, захтеви и техничке спецификације као и научни текстови који у доброј мери и са различитих аспеката описују и обухватају терминологију коришћену у овој струци, тако је и избор материјала био усмерен на ову грађу, уз додатак једног интерног извештаја ЈП ЕПС као и докумената који припадају правним тековинама ЕУ, а који на српском језику нису јавно доступни на вебу. Осим правних тековина ЕУ, од којих је сваки документ преведен са енглеског на српски језик, сва остала грађа је преведена са српског на енглески језик.

Списак грађе која чини корпус из домена електроенергетике је следећи:

1. Извештај о заштити животне средине ЈП ЕПС из 2006. године (интерни документ који није јавно доступан, преведен са српског на енглески језик),

2. Функционални захтеви и техничке спецификације *AMI/MDM*¹⁴⁶ система (документ објављен 2012. године, преведен са српског на енглески језик и доступан на веб-сајту ЈП ЕПС на оба језика),
3. Апстракти научних радова часописа „Електропривреда“ у издању ЈП ЕПС (часопис је престао да излази 2011. године тако да су и текстови обухваћени овом колекцијом настали најкасније те године. Апстракти су преведени са српског на енглески језик).
4. Законодавни текстови који обухватају законска и подзаконска акта преведена са српског на енглески језик, и то су следећа документа:
 - Закон о енергетици објављен 2014. године, доступан на веб-сајту Агенције за енергетику Републике Србије;
 - Уредба о условима испоруке и снабдевања електричном енергијом објављена 2013. године, доступна на веб-сајту Агенције за енергетику Републике Србије;
 - Уговор о оснивању Енергетске заједнице објављен 2006. године, доступан на веб-сајту Агенције за енергетику Републике Србије;
 - Методологија за критеријуме и начин утврђивања трошкова прикључења на систем за пренос и дистрибуцију електричне енергије објављен 2015. године, доступна на веб-сајту Агенције за енергетику Републике Србије;
 - Правилник о енергетским дозволама објављен 2014. године, доступан на веб- сајту Министарства рударства и енергетике.
5. Правне тековине Европске уније (*acquis communautaire*) које сачињавају препоруке, уредбе, директиве, прописе и одлуке ЕУ из домена електроенергетике. Међу свим наведеним колекцијама, само је ова збирка преведена са енглеског на српски језик.

Сви наведени текстови који су део овог електроенергетског корпуса иницијално су били двојезични и у електронској форми, тако да није постојала потреба за обрадом текста у смислу његове дигитализације односно сканирања штампаних материјала, већ је текст од самог почетка био у машински читљивом облику. То је у великој мери поједноставило иницијалну припрему грађе за овај корпус.

Још једна погодност приликом припреме грађе за корпус била је и та што је двојезична колекција докумената правних тековина ЕУ већ била поравната, а за поравнање је коришћен софтверски пакет *SDL Trados*. Остали текстови су прошли кроз поступак сегментације и поравнања сегмената (енг. *align*), о чему ће бити више речи у наредном одељку.

7.2.2. Претходна обрада (претпроцесирање) корпуса

Под претходном обрадом корпуса која се још назива и претпроцесирање корпуса (енг. *corpus preprocessing*), подразумева се спровођење неопходних радњи којима се лингвистичка грађа из свог изворног облика преноси у облик погодан за формирање електронских корпуса, какав је и овде случај.

Претходна обрада корпуса спроведена у овом истраживању обухвата:

- Прикупљање односно преузимање текстова у *docx* формату, углавном са веб- сајтова,
- Класификацију текстова,
- Конверзију текстова из *docx* формата у формат чистог електронског текста,
- Обраду текстова коришћењем платформе *Unitex*.

¹⁴⁶ Напредна мерна инфраструктура (*AMI* енг. *Advanced Metering Infrastructure*) и управљање мерним подацима (*MDM* енг. *Meter Data Management*).

Први корак претходне обраде корпуса подразумева прикупљање односно преузимање текстова са веб-сајтова ЈП ЕПС и АЕРС на локални рачунар, чиме је отпочето креирање овог корпуса.

Други корак подразумева класификацију текстова према потколекцијама којима логички припадају, а које се у овом случају разликују према врсти текста (техничка спецификација, подзаконски акт или научни текстови), и према језику са којег, односно на који се преведе.

Прикупљени и класификовани текстови се, у трећем кораку, из *docx* формата конвертују у формат чистог електронског текста. Чист електронски текст подразумева само просту секвенцу карактера, формата **.txt*, или **.xml*, без било каквог обликовања слова, форматирања текста, табела, слика и осталог.

У четвртом кораку врши се обрада текста која подразумева токенизацију текста, његову сегментацију на реченице и поравнање сегмената, као и морфосинтаксичка анотација уз примену електронских морфолошких речника једночланих и вишечланих речи.

* ТОКЕНИЗАЦИЈА ТЕКСТА подразумева уклањање вишка белина у тексту (размаци, знак за нови ред, табулатор) који се у рачунарству третирају као сепаратори, међутим поред белина треба водити рачуна и о интерпункцијским знацима имајући у виду њихову вишезначност која је, пре свега, условљена контекстом у ком се појављују. Тако је некада тешко проценити да ли интерпункцијске знаке треба третирати као делове неког токена или као појединачне токене. Нпр. цртица која се појављује у полусложеницама, али која служи и за растављање речи на крају реда; апостроф који се јавља у функцији наводника, индикатора за изостављено слово, а може бити и део презимена; запета која служи и као сепаратор делова реченица али се користи и у децималном запису бројева; тачка која је такође сепаратор у запису бројева, може бити део скраћенице, а служи и као интерпункцијска ознака за крај реченице. Уз све то, белине се појављују и као делови вишечланих лексема (као у примеру термина „одводник пренапона“), али и у оквиру мерних израза (нпр. 50 MWh) и именованих ентитета (нпр. Термоелектрана „Никола Тесла“). Свакако, у тексту се могу јавити и грешке, нарочито када су изостављени размаци између две суседне речи. У таквим случајевима велику помоћ пружају програми за проверу правописа као и електронски морфолошки речници (Utvić 2013, 94–95, 100).

* СЕГМЕНТАЦИЈА ТЕКСТА подразумева разлагање текста на мање смислене целине, реченице или пасусе. У овој фази обраде корпуса електроенергетике, сегменти су углавном биле реченице, повремено мање или веће јединице (пасуси). Сегментација је аутоматски обављена коришћењем *Unitex* система који се употребљава за креирање и претраживање корпуса. Реченице су сегментирани помоћу локалних граматика креираних у систему *Unitex* за обраду текста на српском језику. Локалне граматике (енг. *local grammars*) представљају алтернативни приступ синтаксичкој анализи реченица природног језика, са циљем да се избегне формализам лексикализованих контекстно слободних граматика огромних размера (Utvić 2013, 113–114).

Локалне граматике, о којима је било више речи у поглављу 6.4, имплементирани су као коначни аутомати и трансдуктори које корисник може мењати односно прилагодити одређеним, специфичним потребама користећи њихов графички приказ, тј. графове. Локалне граматике за препознавање краја реченице прилагођене су правопису српског језика и саставни су део лингвистичких ресурса за српски језик који се дистрибуирају са програмом *Unitex*.

За поравнање текста односно њихову паралелизацију, Група за језичке технологије је користила софтверски пакет *SDL Trados*. Текстови који су већ били у *Trados word* формату, рашчитани су и преведени у *TMX* облик и потом увезени у базу. Остали текстови који нису били доступни у *Trados word* формату поравнати су помоћу програмског пакета *ACIDE*, а произведени *TXM* је потом импортован у базу.

Колекција текстова која чини корпус електроенергетике састоји се од 76 докумената са 29.943 сегмената. Прва потколекција која садржи правне тековине ЕУ садржи 60 докумената са 17.935 сегмената и 8 докумената са 1.790 сегмената. Друга потколекција се састоји од правних текстова који садрже 7 докумената са 8.399 сегмената, док један документ из ове збирке садржи

апстрактне научноистраживачких радова из часописа „Електропривреда“ са 1.819 сегмената. Енглески део корпуса садржи 612.454, док српски део садржи 545.346 једночланих лексема.

Корпус текстова из области електроенергетике подељен је на два дела, на заштићени део и део у отвореном приступу. Јавно доступан део ове колекције броји 8 докумената односно 10218 сегмената док други, заштићени део, садржи 68 докумената односно 19725 сегмената.

* **МОРФОСИНТАКСИЧКА АНОТАЦИЈА** подразумева посебну врсту лингвистичког обележавања којом се сваком токену у тексту придружују додатне информације из електронског речника као што су врста речи, канонски облик или лема и граматичка анотација тј. вредности морфолошких категорија попут рода, броја, падежа, лица и сл., односно барем неке од наведених.

Уколико се токену додељују морфолошки дескриптори којима се дефинише само врста речи, такав процес се назива етикетирање врстом речи (енг. *Part of Speech Tagging*, скр. *POS tagging*), о чему је већ било речи у поглављу 5.2. У случају да се придружује информација о леми, такав процес се назива лематизација (енг. *lemmatization*). Док лексичка реч тј. лексема представља скуп облика који имају исту основу, припадају истој врсти речи и имају исто значење, лема (енг. *lemma*) представља парадигму лексеме редуковану на један основни облик. Лема је, дакле, основни облик речи којим је лексема представљена у стандардном речнику.

Грамматичка анотација која прецизније описује етикету, може стајати уз сваку односно само одређену врсту речи. На самом почетку, пре него што се приступи морфосинтаксичкој анотацији, неопходно је дефинисати ознаке које ће приказивати одређене морфолошке информације: морфолошке дескрипторе или етикете (енг. *tag*). Скуп етикета (енг. *tagset*) може бити различите величине, од најједноставнијих до изузетно детаљних, у зависности од особина природног језика за који се дефинишу. Међутим, за исти природни језик могуће је, такође, формирати скупове етикета који су различите величине, што свакако зависи од намене аотираног корпуса.

Подела на врсте речи може се разликовати од једног језика до другог, међутим неке врсте речи су заједничке за већину језика. У српском језику разликује се десет врста речи, и то: именице, заменице, глаголи, придеви, бројеви, прилози, предлози, везници, речце и узвици. Свака врста речи може бити одређена скупом дефинисаних морфолошких категорија које прецизније одређују одговарајућу етикету. Оне могу бити карактеристичне за више врста речи, као на пример, падеж – карактеристично за именице, придеве, заменице, или само за поједине врсте речи, као на пример, степен за поређења придева или глаголски облик за глаголе.

Код аутоматске морфосинтаксичке анотације могу се појавити и одређени проблеми као што су: вишезначност и обрада препознатих односно непознатих речи (Vitas 2007). До вишезначности (енг. *ambiguity*) долази када се једном токену може придружити више морфосинтаксичких описа који се међусобно искључују, док препознате речи представљају токене о којима програм за анотацију не поседује никакву информацију (Utvić 2013, 143–144).

Двојезични текстови који чине ову колекцију припадају домену електроенергетике. Доступни су на веб-сајту дигиталне библиотеке двојезичних паралелних текстова – Библиша (<http://jerteh.rs/biblisha/>), и регистровани корисници им слободно могу приступити (Stanković, Krstev, и други 2017). Уз то, све информације у вези са величином ове збирке као и детаљни метаподаци докумената доступни су на веб-сајту Библише.

7.3. Екстракција терминологије на српском језику

За екстракцију терминологије из области електроенергетике из колекције на српском језику, коришћен је хибридни систем *Srp-TE* који се ослања на доменске термине у постојећим терминолошким и лексичким ресурсима као и на примену синтаксичких образаца. Први део овог система припада одређеном домену јер користи само оне лексеме које се налазе у постојећим лексичким ресурсима, а које су релевантне за тај предметни домен, док други део не припада ни једном одређеном домену уз претпоставку да се синтаксичка структура термина не разликује у зависности од домена. Такође, први део овог система препознаје како монолексеме тако и полилексеме, док други део препознаје само полилексеме.

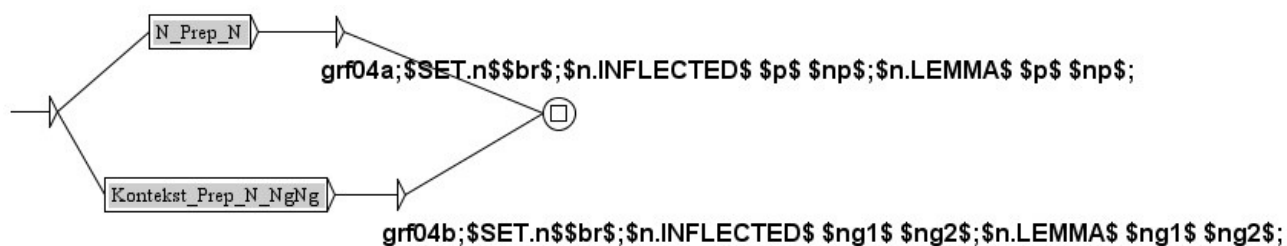
Лексички ресурс који користи систем *Srp-TE* састоји се од свеобухватних електронских морфолошких речника за српски језик (*SrpMD*) који садрже монолексеми и полилексеми, обухватају општу лексичку, властита имена, топониме, енциклопедијско знање и терминологију бројних домена (Krstev 2008). Пре допуне електронског морфолошког речника из домена електроенергетике, овај ресурс је садржао само 296 монолексеми и 43 полилексеми из домена електроенергетике. Наравно, број њихових флективних облика у морфолошком речнику био је знатно већи.

Синтаксички обрасци за препознавање потенцијалних термина су изражени у облику коначних аутомата у којима чвор може садржати било литерал (токен) било референцу за унос у *SrpMD*. Сви уноси у *SrpMD*, независно од њиховог типа, детаљно су описани морфолошким, синтаксичким, семантичким и доменским карактеристикама које синтаксички обрасци користе како би се обезбедила прецизност претраживања. Овај систем је подешен тако да препознаје 29 најчешће коришћених синтаксичких образаца: 4 за полилексеми са две компоненте, 6 за полилексеми са три компоненте, 8 за полилексеми са четири компоненте и 11 за полилексеми са пет компоненти. Најчешћа синтаксичка структура полилексеми одређена је на основу више од 10.000 полилексеми из различитих домена које се већ налазе у *SrpMD*.

Најчешћа синтаксичка структура полилексеми у српском језику, независно од домена, према укупном броју ових лексеми унетим у српски морфолошки речник је *A~Noun* (именица којој претходи придев који се с њом слаже у броју, падежу и роду, 59,5% свих термина), на пример „прикључни вод“. Остале синтаксичке структуре које се често појављују су *N Ng/i* (именица иза које долази именица у генитиву или инструменталу, укупно 9,1%), на пример „стабилизатор система“ и *N Prep~N* (именица иза које долази предлошка фраза, укупно 9,2%), на пример „потрошња по домаћинству“.

У првој итерацији систем *Srp-TE* је екстраховао 98.922 термина кандидата, да би после лематизације и уклањања дупликата, листа садржала укупно 68.579 српских термина кандидата (*Srp-TC*).

На слици 23 је приказана локална граматика у форми коначног трансдуктора која препознаје термине облика именица иза које следи предлошка фраза (нпр. „потрошња по домаћинству“) – горњи пут трансдуктора – и именица праћена фразом у генитиву или инструменталу (именица иза које следи придев, именица у генитиву/инструменталу или две именице у генитиву – нпр. „регулација напона“). У трансдуктору је назначен и излаз за сваку препознату фразу: облик, лема и граматички број (нпр. „\$SET.n\$\$\$br\$;\$n.INFLECTED\$ \$p\$ \$np\$;\$n.LEMMA\$ \$p\$ \$np\$;\$ \$ng1\$;\$n.LEMMA\$ \$ng1\$ \$ng2\$“).



Слика 23. Пример графа којим је приказана локална граматика у форми коначног трансдуктора

7.4. Екстракција терминологије на енглеском језику

Постоји неколико слободних алата за екстракцију енглеских термина. Један од њих је *FlexiTerm*¹⁴⁷, самостална апликација отвореног кода развијена за потребе решавања задатака аутоматске идентификације термина у текстуалним документима. Код ове апликације, препознавање термина се врши у две фазе. У првој се лексичко-синтаксичке информације

¹⁴⁷ Веб-сајт <http://users.cs.cf.ac.uk/I.Spasic/flexiterm/>

користе за избор термина кандидата након чега се такав кандидат бодује коришћењем формуле која процењује њихову колокациону стабилност, али и узимајући у обзир могуће синтаксичке, морфолошке, деривационе и ортографске варијације (Spasić, и други 2013).

Апликација *TermSuite*¹⁴⁸ се користи за изградњу терминологије из корпуса. Овај програмски пакет функционише као и остали алати за екстракцију терминологије: најпре се идентификују, а потом и рангирају термини кандидати. Међутим, у овај пакет су поред основних функција имплементирани и нове па је тако *TermSuite* вишејезично дизајниран, може да се надограђује и да обрађује варијанте термина (Cram и Daille 2016).

На основу детаљне анализе изложених алата, за развој терминологије из области електроенергетике одабрана су два програмска пакета *FlexiTerm* (Šandrih, Krstev и Stanković 2020, 6) и *TermSuite*, па је као излазна листа енглеских термина кандидата коришћена унија скупова израза који су екстраховани овим алатима.

Директном применом на улазне реченице, *FlexiTerm* је екстраховао 8.484 термина кандидата после лематизације помоћу алата *spaCy*¹⁴⁹, а затим и уклањања дупликата. *FlexiTerm* аутоматски искључује термине са фреквенцијом мањом од 3. На сличан начин, након елиминације термина са фреквенцијом мањом од 3 и применом исте лематизације и уклањања дупликата, *TermSuite* је произвео 7.888 термина кандидата. Заједно су *FlexiTerm* и *TermSuite* произвели листу термина кандидата на енглеском језику (*Eng-TC*) која има 13.312 различитих термина, при чему је само *FlexiTerm* екстраховао 5.424 (41%), *TermSuite* 4.829 (36%), док су преосталих 3.059 термина кандидата (23%) екстраховала оба екстрактора.

7.5. Поравнавање комада

За поравнавање (енг. *alignment*) **комада** коришћен је паралелизовани реченични корпус из домена електроенергетике који је и описан у одељку 7.2. Поравнати корпус из овог домена унапређен је листом двојезичних парова који су добијени из српског (Krstev, Pavlović-Lažetić, и други 2004) и енглеског ворднета у којем се синсетови поравнавају путем интерлингвалног индекса који омогућава повезивање појмова на више језика којима се означава исти концепт, као и енглеско-српског глосара који садржи општу лексику (Šandrih, Krstev и Stanković 2020). Одреднице из ових ресурса су моделоване за оба језика имајући у виду њихове флективне особине, а одговарајући облици су поравнати применом *ad hoc* хеуристичког приступа (на пример, енглеска реч *bigger* је поравната са различитим облицима компаратива придева „велик – већи, већег, већа“, итд.). Добијени резултати показали су да ово проширивање корпуса унапређује поравнавање термина и коначне резултате. Претходна истраживања на другом домену су показала да се допуном корпуса овим двојезичним паровима успешност упаривања термина повећава за 20%, изражено F_1 мером која балансирано одражава одзив и прецизност (Šandrih, Krstev и Stanković 2020). Стога је и корпус из домена електроенергетике унапређен проширеним скупом који броји 458.749 двојезичних парова речи.

За поравнавање речи, екстраховање и рангирање фраза коришћен је алат *GIZA++* система *Moses*¹⁵⁰ (Och и Ney 2003), о чему је већ било речи на почетку поглавља 7. Овај систем служи за машинско превођење и омогућава аутоматско тестирање модела превођења за било који одабрани језички пар (Koehn и Schroeder 2007, 177). Производ *GIZA++* алата су поравнати **комади**, наведени у *табели фраза*, уз које долази рангирање превода фразе. Једна од функција овог алата је аутоматско одбацивање поравнатих комада за које је вероватноћа превода тих фраза испод одређеног прага. Након што су из табеле фраза одбачени парови са вероватноћом превода испод 0.85, у следећем кораку примењен је корак одбацивања оних сегмената чији је пресек са

¹⁴⁸ Веб-сајт <http://termsuite.github.io/>

¹⁴⁹ *SpaCy* је модул отвореног кода. Написан је у програмском језику *Python* и намењен је за напредну обраду природних језика. Веб-сајт <https://spacy.io/>

¹⁵⁰ *Moses*, <http://www.statmt.org/moses/>

лематизованом „врећом речи“ (енг. *Bag-of-Words*, скр. *BoW*)¹⁵¹ за екстраховане термине на енглеском језику празан. Модел вреће речи за листу екстрахованих енглеских термина добијен је коришћењем модула *spaCy* програмског језика *Python* за токенизацију на енглеском и уклоњене су незначајне речи (енг. *stop words*), елиминацијом функционалних речи присутних у стоп листи модула *spaCy*. Иста лематизација примењена је и на поравнате комаде на енглеском језику. Елиминисани су они комади у којима се није нашла ниједна лематизована реч из „вреће речи“ (комад садржи само граматичке речи). Након овог поступка, табела фраза је имала 342.983 поравнатих комада.

7.6. Упаривање термина и комада

Када се комади поравнају, потребно је одабрати оне у којима оба дела садрже термин кандидат из домена електроенергетике. Да би се направила листа српских термина кандидата, у претходним истраживањима на корпусу из домена библиотекарства и информатике, коришћена су два приступа. У првом приступу екстрактор *Srp-TE* је примењен на комаде текста на српском језику, док је у другом приступу примењен на текст српског дела корпуса електроенергетике у целини (Šandrih, Krstev и Stanković 2020). Бољи резултати добијени су применом другог приступа којим се издваја већи број термина јер екстрактор не ометају границе комада. Тачније, и прецизност и одзив су се повећали за готово 1% када је екстрактор примењен на цео текст. Из тог разлога је листа српских термина кандидата екстрахована из електроенергетског корпуса (скр. ЕЕ) применом алата *Srp-TE* на цео текст, чиме је добијена листа *Srp-TC*. Као резултат, добијена је листа од 98.922 термина кандидата, која је након лематизације *реч по реч*, применом морфолошких речника и уклањањем дупликата смањена на 68.579 ставки. За енглеске термине кандидате *Eng-TC*, директно је примењен алат *Eng-TE* на улазни текст.

Када се направе листе терминолошких кандидата *Eng-TC* и *Srp-TC*, следећи корак је њихово поклапање са одговарајућим деловима поравнатих комада. Да би се уклониле терминолошке варијације као и појављивање интерпункцијских знакова и незначајних речи, у поменутом претходном истраживању коришћена је функција лабавог упаривања (енг. *loose match*) за оба језика (названа **токен**) (Šandrih, Krstev и Stanković 2020): наиме, пре упоређивања, интерпункцијски знаци и незначајне речи уклоњене су како из комада тако и из термина, и представљене као неуређени скупови лематизованих речи. Анализа добијених резултата показала је да је то резултирало бројним нетачним упаривањима што је утицало на свеопшту прецизност. Из тог разлога је извршена упоредна анализа три различите функције упаривања:

- **Token** функција која подразумева упоређивање нефункционалних речи, редослед није битан.
- **StringL** функција која подразумева упоређивање свих речи, редослед није битан.
- **StringS** функција која подразумева прецизно упоређивање облика, тј. строго поравнање ниски.

Табелом 11 приказана су два примера која илуструју извршавање наведених функција упаривања за енглески језик. У случају функције упаривања *Token*, комад „share of energy from renewable source“ и термин кандидат „share of renewable energy source“ своде се на идентичне скупове који се пореде {*share, energy, renewable, source*}.

У случају функције упаривања *StringL*, комад „demand management of energy“ и термин кандидат „management of energy demand“ се савлађују јер се пореде идентични скупови {*demand, management, of, energy*} и {*management, of, energy, demand*}.

У примеру *StringS* функције упаривања, и комад и термин кандидат „energy efficiency“ су идентичне ниске.

¹⁵¹ „Врећа речи“ представља скуп речи у једном тексту са израчунатом фреквенцијом њиховог појављивања у том тексту независно од њихових граматичких облика и реда речи у реченици.

Јасно је да су сви кандидати упарени функцијом *StringS*, такође упарени и функцијом *StringL*, док су сви кандидати упарени функцијом *StringL*, такође упарени и функцијом *Token*.

У доњем делу табеле 11 приказани су тзв. *негативни примери* функција упаривања: *Token/StringL* приказује термин који се подудара са *Token* функцијом, али не и са функцијом *StringL*, док је сличан пример и термина упареног са функцијом *StringL*, али не и са функцијом *StringS*, приказан у реду *StringL/StringS* функције.

Табела 11. Илустрација различитих функција упаривања

Функција упаривања	Српски комад	Енглески лематизовани комад	Екстраховани енглески термин
<i>Token</i>	удео енергије из обновљивих извора	share of energy from renewable source	share of renewable energy source
<i>StringL</i>	управљање енергијом	demand management of energy	management of energy demand
<i>StringS</i>	енергетска ефикасност	energy efficiency	energy efficiency
<i>Token / StringL</i>	развој дистрибутивног система	development of the distribution system	distribution system development
<i>StringL / StringS</i>	потрошња воде и енергије	consumption of water and energy	consumption of energy and water

У табели 12 приказане су вредности добијене као резултат описаних процедура, уз коришћење различитих функција упаривања.

Колона *IntS* представља број преосталих поравнатих комада у табели фраза након укрштања (знак &) енглеских комада са листом енглеских термина кандидата *Eng-TC*. Српски комади из редуковане табеле фраза су затим укрштени са листом српских термина кандидата *Srb-TC*, што је резултирало одређеним бројем поравнатих комада наведених у колони *IntT*.

Табела 12. Резултат терминолошке екстракције уз коришћење различитих функција упаривања

Функција упаривања	Корпус EE+	Табела фраза	<i>Eng-TC</i>	<i>Srb-TC</i>	Табела фраза & <i>Eng-TC</i> (<i>IntS</i>)	<i>IntS</i> & <i>Srb-TC</i> (<i>IntT</i>)
<i>Token</i>					72.944	5.146
<i>StringL</i>	488.692	342.983	13,312	68.579	8.724	3.435
<i>StringS</i>					8.247	3.267

7.7. Евалуација резултата

Развој терминологије из области електроенергетике, као и осталих дисциплина, дугачак је, комплексан и захтеван задатак. Темелјан приступ овом процесу подразумева јасно дефинисање критеријума за избор терминологије који се од једне до друге области истраживања, као и сврхе због које се анализа спроводи, могу битно разликовати.

Термини добијени током процеса овог истраживања потичу из релевантних извора као што су, у највећем броју случајева, јавна законска и подзаконска акта из области коју обрађујемо. Из таквог извора настала је листа термина која је добијена поступцима описаним у овом поглављу, и која је потом евалуирана мануалним методом.

Евалуацију резултата обавила су два независна оцењивача, од којих је један дугогодишњи експерт у домену електроенергетике, а други преводилац специјализован за исти домен (аутор ове дисертације).

Критеријуми који су нарочито узети у разматрање приликом обраде термина у области електроенергетике су:

- **ЗНАЧЕЊЕ** унутар одређеног домена, на пример: „укључивач“ (*recloser*) – уређај који у неколико узастопних покушаја затвара електрично коло и прекида га уколико квар и даље постоји;
- **НЕДВОСМИСЛЕНОСТ** – важно је направити разлику између одређених врста електроенергетске опреме. На пример, неопходно је разликовати типове расклопне опреме као што су прекидач и укључивач. Оба уређаја, и прекидач и укључивач прекидају струју кvara, али прекидач не врши поновно затварање струјног кола као што то чини укључивач;
- **НЕСИНОНОМНОСТ** – ако је има, обично се ради о страниј речи која се подједнако користи као и њен домаћи еквивалент, на пример: називни (дом.) = назначени (дом.) = номинални (лат.);
- **КРАТКОЋА И ПРЕЦИЗНОСТ** – како би се пружио најкраћи и најпрецизнији опис предмета или појаве који представља, на пример: „осигурач“ – елемент који прекида струју кvara;
- **ПОСТОЈАНОСТ И ТРАЈНОСТ** током историје развоја дате области, на пример: „електроенергетски систем“, „далековод“, „разводно постројење“, „трансформаторска станица“, „електрана“, итд. Ово су примери веома постојаних и трајних термина имајући у виду да се у овој струци користе више од једног века;
- **ИНТЕРНАЦИОНАЛНОСТ** – страна реч би се могла сматрати решењем ако не постоји домаћи еквивалент, попут неких од основних термина у области електроенергетике као што су нпр. *generator* „генератор“, или *transformer* „трансформатор“. Ипак, важно је напоменути да у стручним круговима, у последњих неколико година, влада велико интересовање за усвајањем стручне терминологије која се не заснива на пуком преузимању већ на иницијативи да кад год могућности то допуштају буде направљен и усвојен његов домаћи еквивалент; такав је пример за већ поменути термин *recloser* за кога је усвојен домаћи термин „укључивач“.

Једнојезични термини кандидати на енглеском и српском језику као и њихово поравнавање, су посебно евалуирани.

За њихову евалуацију коришћене се следеће вредности:

ЕЕ	термин који припада ЕЕ домену
Т	термин који припада неком сродном домену (нпр. електротехника, машинство, итд.)
F	општеприхваћена фраза или колокација
X	остало (није термин)

Еквивалентност термина рангирана је према следећим вредностима:

OK	Тачно
NOK	Нетачно
X	није применљиво јер су један или оба кандидата нетачни

Укупан број евалуираних парова је 3.833. Скуп је садржао термине кандидате који су екстраховани најмање једном функцијом упаривања. Једнојезични кандидати евалуирани су само ако су били укључени у преводилачки пар.

У табели 13 приказане су различите вредности евалуације додељене једнојезичним кандидатима и преводним паровима.

Табела 13. Примери евалуације

Евал. парови	Евалуација EN_SR	Енглески кандидат	Српски кандидат
OK	EE_EE	terminal	priključnica
	T_T	technical specification	tehnička specifikacija
	F_F	combined technique	kombinovana tehnika
	EE_F	connection cost	troškovi priključenja
NOK	EE_EE	power plant	male elektrane
	T_T	efficiency class	energetska efikasnost
	F_F	global cost	ukupni troškovi
X	EE_X	energy customer	energija za kupce
	T_X	regulatory authority	regulatorno telo o tim
	F_X	conventional energy	konvencionalni izvori energije sa energijom
	X_EE	final customer for electricity	krajnji potrošač električne energije
	X_F	active power maximum	aktivne snage
	X_X	energy for loss	energija za gubitke

Повремено су парови евалуирани као тачни, иако су енглеским и српским кандидатима додељиване различите вредности (видети ред 4 у табели 13). То је обично био случај у примерима када термину по значењу одговара предложени кандидат на другом језику, али се овај не може прихватити као термин.

7.7.1. Статистичке мере за процену сагласности анотатора ручно аотираног корпуса

Анотација корпуса представља поступак којим се деловима корпуса придружују додатне информације. Поступке за оцену квалитета ручне анотације веома је значајно развити и применити на мањим деловима електронског корпуса због њихове касније примене на великим електронским корпусима као и због примене у задацима обраде природних језика који користе поступке аотирања, а то су: лематизација, кореновање (енг. *stemming*), означавање врстама речи, итд. (Младеновић 2016).

Многе мере које се данас користе за процену валидности доприноса учесника у пројектима групне расподеле рада (енг. *crowdsourcing*) развијене су за потребе мерења у психолошким истраживањима. Природа мерења у тим истраживањима налагала је да дође до развоја теорија о менталним тестовима који узимају у обзир грешке у мерењу и поузданости (Schaer 2012).

Информације које се прикупљају током истраживања класификују се у зависности од нивоа мерења, од чега зависи њихова статистичка обрада. Различити нивои мерења података садрже различите количине информација без обзира на предмет мерења.

У систему класификације који је представљен у (Stevens 1946) нивои мерења су: номинални (категорички), ординални (уређени), интервални и размерни (рацио) ниво. Како ниво мерења расте од најнижег (номинални ниво) до највишег (размерни ниво), расте и количина информација коју садрже подаци, као и математичке операције које се могу изводити над тим подацима.¹⁵²

Подаци на номиналном нивоу мерења се називају номинални или категорички. Категорички подаци бележе квалитет или карактеристику неке особе, као што су: боја очију, припадност полу, нацији или политичкој партији, мишљење о неком друштвеном проблему, итд. Категорички подаци сврставају индивидуе у групе и ови подаци се обично представљају као број и/или проценат броја или особа који спадају у одређене групе. На номиналним подацима, дозвољене су статистичке технике као што су: мод, пропорција, χ^2 тест, Φ коефицијент корелације, коефицијент контингенције C .

Ординална (уређена) скала одређује шта је веће или мање, али разлике између појединих јединица скале нису једнаке. На овај начин се прикупљају следећи подаци: рангови, школске оцене и др. На уређеним скалама, дозвољене су статистичке технике као што су: мод, коефицијент корелације r и коефицијент W .

Карактеристика интервалне скале је да одређује шта је веће или мање, а разлике између појединих јединица скале су једнаке на сваком делу скале и у сагласности са мереном особином. Ове скале немају апсолутну нулу, а уколико је нула на њој постављена, сматра се да је она арбитарна. Пример интервалних скала су резултати на психолошким тестовима, иако неки теоретичари сматрају да је примереније податке добијене психолошким тестирањем третирају као ординалне (уређене). Поред статистичких техника наведених у претходне две скале, може се користити и: аритметичка средина, стандардна девијација, z -вредност, коефицијент корелације укључујући и парцијалну и мултиваријантну корелацију.

Размерни (рацио) ниво мерења има све особине интервалне скале и још има апсолутну 0. То значи да су бројчани односи исти с односима у мереној појави. Мерења у физици су на размерној скали, и то су: дужина, тежина, отпор и др. Поред свих статистичких техника наведених у претходним скалама, код размерног нивоа мерења може се користити и: геометријска средина и коефицијент варијабилности V .

Методе које се најчешће користе за оцену квалитета ручне анотације два или више анотатора из класе статистичких оцена сагласности су Коенова капа (*Cohen's kappa*), Флајсоф капа (*Fleiss' kappa*) и Крипендорфов алфа (*Krippendorff's alpha*) коефицијент (Младеновић 2016).

¹⁵² Статистика у друштвеним наукама, <http://www.e-statistika.rs/index.php?pa=56&idTeksta=35>, последњи приступ децембра 2020.

Флајсоф капа узима у разматрање већи број анотатора, а Крипендорфов алфа коефицијент приказује степен слагања учесника у једном таквом пројекту.

За анализу података у оквиру пројеката групне расподеле рада, аутори Хајес и Крипендорф (енг. *Hayes* и *Krippendorff*) (Hayes и Krippendorff 2007) предлажу коришћење Крипендорфовог алфа коефицијента јер, за разлику од других, сличних мера које се такође користе, ова мера испуњава све услове за успешно приказивање степена слагања учесника у једном таквом пројекту. Исти аутори наводе да свака ефикасна мера за процену тзв. слагања међу анотаторима (енг. *inter-annotator agreement*) треба да задовољи следеће услове:

1) Треба да процени слагање између два или више посматрача (анотатора, учесника) који описују сваку јединицу анализе независно једни од других. У случају када постоји више посматрача, ова мера не би требало да зависи од броја посматрача нити да се мења у зависности од пермутације и селективног учешћа посматрача. У случају да су та два услова испуњена, на слагање доприноса учесника не би утицали индивидуални идентитети посматрача који стварају или процењују податке у неком пројекту.

2) Треба да буде заснована на расподели категорија или скала бодова које сами посматрачи користе. Тако поузданост доприноса посматрача не зависи од разлике која би настала између тога како су аутори упутства за учешће у неком пројекту замислили да ће подаци изгледати и тога какви су подаци заиста добијени.

3) Треба да произведе нумеричку скалу између најмање две тачке која ће имати разумну интерпретацију поузданости. Савршеним слагањем се, по конвенцији, сматра 1.000 (тј. 100%), или у случају недостатка слагања, типично означеног са 0.000 (што не мора обавезно бити доња тачка скале поузданости).

4) Мера би требало да одговара нивоу мерења података. То значи да приликом примене статистичких мера на неколико врста података, мера мора задржати своју математичку структуру, осим и једино у случају одлика мерног система. То омогућава поређења између различитих мерних система, ако је то потребно за стандард поузданости.

5) Начин прикупљања узорача би требало да буде познат, или да га је бар могуће израчунати.

Процент слагања (енг. *percent agreement*) је пропорција јединица анализе са својим одговарајућим описима, за које се два посматрача слажу (табела 14). Ова мера се веома лако израчунава, тако што се број категорија у којима су се анотатори сложили подели са бројем укупних категорија. У табели која је наведена као пример, у којој су анотатори бирали између номиналних вредности (у три случаја анотатори су дали исте одговоре) тј. могућих одговора 2,3,4 или 5, резултат процента слагања је 3 подељено са 5, дакле проценат слагања је 60%.

Табела 14. Израчунавање процента слагања међу анотаторима

Јединица анализе	Анотатор 1	Анотатор 2	Слагање
1	3	4	0
2	2	2	1
3	5	4	0
4	4	4	1
5	3	3	1
			3 / 5

За потребе истраживања које је предмет ове дисертације, коришћен је метод Коенова капа, због тога што овај процентни рачун не узима у обзир вероватноћу да су оба анотатора случајним избором (а не знањем), дошли до обележеног резултата.

КОЕНОВА КАПА (енг. *Cohen's kappa coefficient*) је статистичка мера која се користи за одређивање сагласности у анонотацији између два оцењивача номиналном скалом мерења

(Cohen 1960), где је мера усаглашености удео заједничких одлука у којима постоји сагласност, након изузимања случајне сагласности.

Коенов капа коефицијент се користи само када су два посматрача независна један од другог.

Његов опсег $[-1,1]$, којим се приказује степен сагласности, подељен је на шест класа:

- $<0,00$ слабо,
- $0,00-0,20$ ниско,
- $0,21-0,40$ тачно,
- $0,41-0,60$ умерено,
- $0,61-0,80$ високо,
- $0,81-1,00$ врло високо.

Коенова капа степен сагласности ($A_m - agreement\ measure$) изражава се као однос:

$$A_m = \frac{P_o - P_e}{1 - P_e}$$

Ознаке:

P_o – релативни опсервирани број аотација у којима постоји слагање (процент опсервација у којима постоји слагање) и

P_e – очекивана вероватноћа слагања аотација.

$P_o - P_e$ представља оне случајеве који су се појавили као слагање изван очекиваног слагања.

$1 - P_e$ представља вероватноћу очекиваног неслагања.

A_m се тумачи као удео слагања изван очекиваног у делу у коме се очекује неслагање.

7.7.2. Поступак евалуације

Поступак евалуације независних евалуатора је обављен у две фазе. У првој фази, мањи скуп од 3833 термина је евалуиран независно, док су у другој фази анализирани разлике и усаглашени критеријуми којих су се евалуатори придржавали у наставку евалуације. Овакав „двофазни“ приступ резултирао је добром сагласношћу евалуатора.

Мера усаглашености екстрахованих термина на српском и енглеском језику, израчуната је за три класификациона задатка, као што је приказано на сликама 24, 25 и 26, и износи:

1) термини на енглеском: $en_A_m = 0.8614$,

Count of Teval	EN	Column Label	Yes	No	p0	pe	EE	F	T	X	(blank)	Grand Total
Row Labels	EE		3460	373	0.9027	0.2977	0.0494	0.0942	0.1463	0.0078		
EE			827	13								852
F			8	1014								1182
T			16	118								1463
X			1	26								336
(blank)												341
Grand Total			852	1171	1469	341	3833					

$$\kappa \equiv \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e}$$

$$p_e = \frac{1}{N^2} \sum_k n_{k1} n_{k2}$$

Слика 24. Евалуација двојезичне екстракције за домен електроенергетике за енглески језик

2) термини на српском: $sr_A_m = 0.8197$, и

Табеле 18 и 19 представљају укупне резултате процеса евалуације. Треба напоменути да су се сви евалуирани парови разликовали, али је било дупликата међу енглеским као и међу српским кандидатима. Пронађено је 3.034 јединствених енглеских и 3.368 јединствених српских термина.

У одређеном броју случајева дупликати су били тачно повезани са синонимним терминима, нпр.

direct measurement ↔ директно мерење, и *direct measurement* ↔ непосредно мерење;

цена електричне енергије ↔ *electrical energy price*, и цена електричне енергије ↔ *price of electricity*.

Табела 18. Резултати процеса евалуације (једнојезична евалуација)

Једнојезична евалуација	Енглески	Српски
ЕЕ	885 (23.1%)	732 (19.1%)
Т	1.466 (38.2%)	1.219 (31.8%)
F	1.177 (30.7%)	1.344 (35.1%)
X	305 (8.0%)	538 (14.0%)
Укупно	3.833	3.833

Табела 19. Резултати процеса евалуације (двојезична евалуација)

Двојезична евалуација	ЕН/СР парови
ОК	2.684 (70.2%)
НОК	546 (14.2%)
X	603 (15.7%)
Укупно	3,833

Међу 3.833 екстрахована двојезична пара, 3.234 су издвојена помоћу свих примењених функција упаривања, од чега 77 помоћу *StringL* и *Token* функција упаривања, док је 522 пара екстраховано само функцијом *Token*.

У табели 20 приказани су резултати једнојезичне термилошке екстракције са три различите функције упаривања. Приметно је да се за оба језика прецизност правилне екстракције повећава применом стриктније функције упаривања и то, мање за енглески (пораст од 0,4%), а више за српски (пораст од 1,5%) (колоне ЕЕ+Т+F). Сходно томе, апсолутни број правилно екстрахованих термина и фраза опада.

Табела 20. Резултати добијени применом три функције упаривања за екстракцију термина из енглеског и српског језика; прецизност екстракције дата је у заградама

Функција упаривања	Језик	ЕЕ	Т	F	ЕЕ+Т+F	X	Укупно
<i>Token</i>	EN	632 (20.8%)	1.157 (38.1%)	1.019 (33.6%)	2.808 (92.6%)	226 (7.4%)	3.034
	SR	590 (17.6%)	1.047 (31.1%)	1.236 (36.7%)	2.873 (85.3%)	494 (14.7%)	
<i>StringL</i>	EN	546 (20.6%)	1.030 (38.9%)	886 (33.4%)	2.462 (92.9%)	189 (7.1%)	2.651

	SR	518 (17.5%)	934 (31.5%)	1.101 (37.1%)	2.553 (86.0%)	415 (14.0%)	2.968
<i>StringS</i>	EN	531 (20.3%)	1.025 (39.1%)	879 (33.6%)	2.435 (93.0%)	184 (7.0%)	2.619
	SR	503 (17.3%)	927 (31.9%)	1.093 (37.6%)	2.523 (86.8%)	383 (13.2%)	2.906

У табели 21 приказани су резултати евалуације парова екстрахованих применом три различите функције упаривања. Из табеле видимо да се прецизност повећава за 2% када се примењује најстриктнија функција упаривања (*StringS*), у односу на најмање стриктну функцију (*Token*).

Када се упореде резултати једнојезичне евалуације термина кандидата, примећујемо да је српски екстрактор мање успешан од енглеског (У табели 18, колона X, видимо 7,4% неправилних енглеских термина кандидата наспрам 14,7% неправилних српских термина кандидата).

Табела 21. Резултати добијени применом три функције упаривања за енглеске и српске термилошке парове

	OK	NOK	X	Укупно
<i>Token</i>	2.684	546	603	3.833
	(70.0%)	(14.2%)	(15.7%)	
<i>StringL</i>	2.355	460	496	3.311
	(71.1%)	(13.9%)	(15.0%)	
<i>StringS</i>	2.330	447	457	3.234
	(72.0%)	(13.8%)	(14.1%)	

Добијени резултати упоређени су са два термилошка извора доступна на вебу, у машински читљивом облику, а то су Евроним и Електропедија. Обе термилошке базе података су релевантне због великог броја термилошких уноса у области електротехнологије. Поред тога, Евроним је релевантан због процедуре и корака који се спроводе у процесу превођења европског законодавства (*acquis communautaire*) на српски језик (видети поглавља 3.4.3.3, 5.5.3.1 и 5.5.3.2), а који укључују стручну, правнотехничку и језичку редактуру преведеног термина.

Евроним тренутно садржи 46.505 уноса¹⁵³ и већ више од једне деценије се непрестано употпуњује и ревидира. За рад на овој бази задужени су стручњаци из различитих предметних домена. Поузданост појмова дефинисана је на скали од 1 до 5 што ову базу термина чини поузданом и доследном. Због тога је током процеса евалуације сваки издвојени енглески термин кандидат додатно претраживан у Еврониму. Уколико би се нашао у тој бази (било да је енглески или српски термин), његов превод би био забележен као прилог уз издвојене српске и енглеске термине кандидате (табела 22). Они су касније коришћени за формирање синсетова, о чему ће бити више речи у следећем одељку.

¹⁵³ Подаци су од 22. јуна 2021. године.

Табела 22. Пример екстрахованих термина који су пронађени и у терминолошкој бази Евроним

Евал. парови	Евалуација EN_SR	Енглески термин кандидат	Српски термин кандидат	Евроним
OK	T_F	significant incident	značajni incidenti	značajna nezgoda
	T_T	investment plan	plan ulaganja	investicioni plan
	T_T	sidewall	bočni zid	bočna strana

Међутим, иако је база Евроним релевантна за поређење са екстрахованим терминима из области електроенергетике током процеса евалуације, укупно је било 52 слагања и ни једно од њих не припада домену који испитујемо.

Екстраховани термини и парови поређени су са још једним јавно доступним двојезичним енглеско-српским ресурсом, Електропедијом (верзија од фебруара 2018). Установљено је да 136 различитих енглеских термина већ постоје у Електропедији од чега је 61 припада домену електроенергетике (ЕЕ), 44 је евалуирано као општи термин (Т), док је 31 евалуирано као смислена фраза која није термин (F).

Најзаступљеније предметно подручје из Електропедије је под ознаком 617 (Организација / Тржиште електричне енергије) са 15 термина, од којих само један није евалуиран као ЕЕ термин, и 601 (Производња, пренос и дистрибуција електричне енергије – Опште) са 12 термина од којих су сви евалуирани као термини који припадају домену електроенергетике (ЕЕ).

У Електропедији постоји 25 различито евалуираних српских термина: 13 евалуираних као ЕЕ, 8 као Т, 4 као F. Најзаступљенија предметна област је 102 (Математика – Општи појмови и линеарна алгебра) са 4 термина, од којих су сви евалуирани као Т. Од 2.684 енглеско-српских парова који су евалуирани као правилни, има 18 енглеско-српских парова за које и енглески и српски термин постоје у Електропедији. У само једном случају се српски екстраховани и евалуирани термин разликује од термина који је предложила Електропедија:

EngTE=EngEPed *service standard* ↔ СрпTE *standard usluga*; СрпEPed *standard za usluge*.

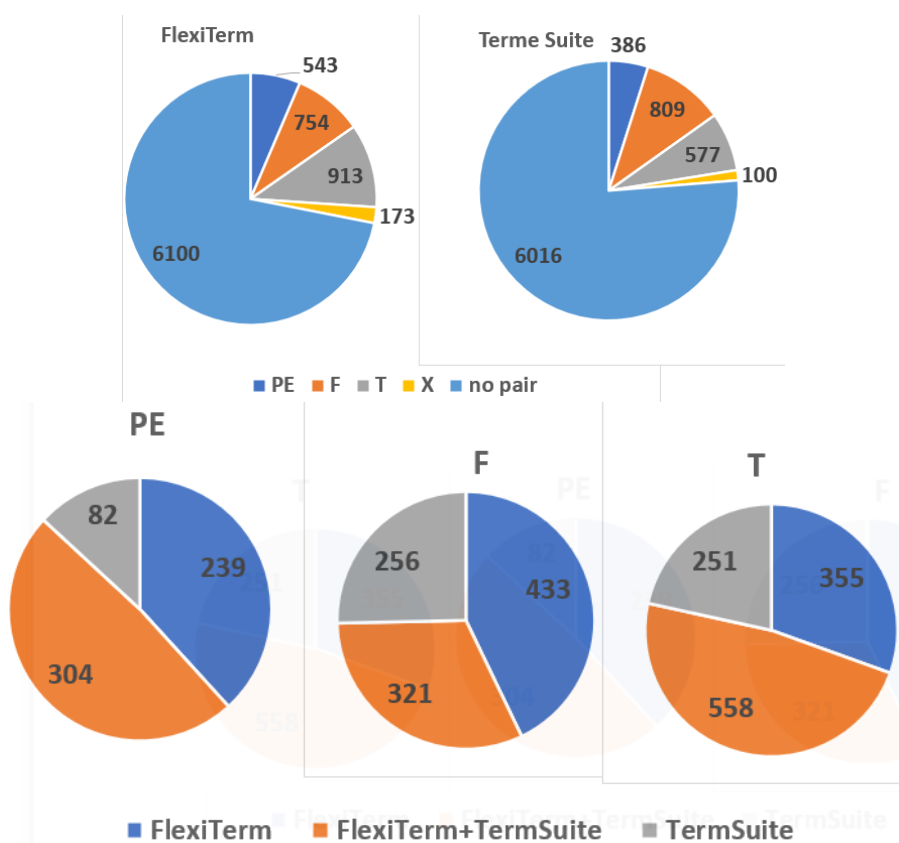
У овом случају имамо два еквивалентна термина са различитом синтаксичком структуром. Преклапање екстрахованих термина и термина који постоје у Електропедији је сувише мало да би се извукли чврсти закључци. Ипак, заједнички терминолошки парови (и они који су у Електропедији и они који су екстраховани за потребе овог рада), показују да ова процедура издваја парове који су правилни. Мало преклапање такође указује да је рад на терминолошком обогаћивању Електропедије неопходан.

7.8. Карактеристике екстрактора за српски и енглески језик

О примени различитих система за екстракцију терминологије већ је било речи у претходним поглављима. За екстракцију доменских термина на енглеском језику, коришћена су два система у слободном приступу, *FlexiTerm* и *TermSuite*. Будући да ни један од ових система није оперативан за српски језик, за издвајање појмова из доменских текстова на српском примењен је једини доступни систем који се ослања на скуп правила изражених у облику синтаксичких образаца.

У највећем броју случајева, систем *FlexiTerm* екстрахује полилексема, док су једине монолексема које је овај систем препознао били акроними (1,2% од свих издвојених термина који су прешли праг постављене вредности), нпр. *DNS (Domain Name Server)* и *EPS* (Јавно предузеће Електропривреда Србије). *TermSuite* је екстраховао већи број монолексема (33,6% од свих

екстрахованих термина који су прешли утврђени праг). Најдужи термини које је екстраховао *FlexiTerm* имају по 7 компонената (има укупно 23 таква термина), док најдужи термин који је екстраховао *TermSuite* има 8 компонената (1 термин од осам компонената и 18 термина са по 7 компонената). Ни један од ових термина није упарен са српским терминима, што не изненађује имајући у виду да систем *Srp-TE* екстрахује само термине дужине до 5 компонената. Најдужи термини који су успешно упарени са српским терминима, а које су екстраховала ова два екстрактора, имали су по 6 компонената. На слици 27 приказани су нумерички подаци о успешности ова два система. У првом реду може се видети да је мањи број термина који су екстраховани помоћу алата *TermSuite* упарено са српским терминима од оних који су екстраховани помоћу алата *FlexiTerm* – 76,3% термина је екстраховано помоћу алата *TermSuite* који немају свој пар, наспрам 71,9% таквих парова екстрахованих помоћу алата *FlexiTerm*. У другом реду се види да је међу терминима на енглеском језику који су били упарени са терминима на српском, *FlexiTerm* екстраховао више оних који су евалуирани као термини из домена електроенергетике, општих појмова и фраза.



Слика 27. Успешност система *FlexiTerm* и *TermSuite* у екстракцији терминологије

Да бисмо проценили успешност екстрактора *Srp-TE* у проналажењу корисних термина кандидата, применили смо га на три различите термилошке листе: једну из домена сродног домену електроенергетике – Електропедију и две из различитих домена: психологије и рачунарства.¹⁵⁴ Из ових термилошких листи коришћене су само полилексеми, и то само оне чије су све компоненте већ постојале у *Srp-TE* систему, што је гарантовало успешност. Редуковане листе и даље су имале 73,20% оригиналних полилексеми из Електропедије, 75,62% термина из домена психологије и 70,12% термина рачунарских наука. Осам најуспешнијих синтаксичких образаца, представљених у табели 23, препознало је више од 85% свих

¹⁵⁴ Листа термина из домена рачунарских наука добијена је из Оксфордског речника рачунарства (Oksfordski rečnik računarstva 1990), који је превод другог издања истоименог речника из 1986. године издавача *Oxford University Press*. Листа термина из домена психологије резултат је сличног пројекта, превода истоименог речника издавача *Oxford University Press* чији превод никада није објављен.

прихваћених термина (из електроенергетике и сродних домена). Ових 8 образаца препознано је још више термина на листама психологије и рачунарства, упркос чињеници да ове листе садрже и дуже термине (али не дуже од 7 компонената). Исти обрасци били су мање успешни са терминима из Електропедије. Може се приметити да је број термина који садрже више од 5 компонената на листама из домена психологије и рачунарских наука веома мали, док таквих термина има знатно више у Електропедији, више од 5% свих полилексема. Најдужи термин у Електропедији има 17 компонената *napon isključenja pomoćnog izvora za uređaj diferencijalne struje koji se isključuje automatski u slučaju otkaza pomoćnog izvora* и представља тачан превод енглеског термина *opening voltage of the auxiliary source for residual current devices which open automatically in case of failure of the auxiliary source* који има 21 компоненту. Међу терминима који нису препознати у Електропедији постоје термини који користе релативне клаузе (попут оне најдуже), или нису именичке фразе, нпр. *linearно завистан* „linearly dependent“ и *krstariti u računarskim mrežama* „surf in computer networks“. Даљи експерименти са екстракцијом терминологије у електроенергетском и другим доменима требало би да разјасне да ли су ове специфичности откривене у Електропедији својствене домену или типу ресурса (*Електропедија* је дигитални терминолошки ресурс, док су друга два традиционални терминолошки речници).

Табела 23. Рангирање и фреквенција за 8 најпродуктивнијих синтаксичких образаца система *Srp-TE* на корпусу домена електроенергетике након његове примене на три терминолошке листе из различитих домена

Обрасци	Екстраховани ЕЕ термини (ЕЕ+Т)	Електропедија	Психологија	Рачунарске науке
A~N	1 (764)	1 (1676)	1 (2395)	1 (1046)
NN _g	2 (328)	2 (711)	2 (515)	2(306)
(A N) _{gi} (A N) _{gi}	3 (132)	3 (437)	3 (134)	3(96)
N (Prep~N)	4 (87)	6 (196)	5 (109)	6–8 (39)
A~A~N	5 (79)	4 (351)	4 (110)	4 (56)
A~N(A N)gi	6 (65)	5 (223)	6 (71)	6–8 (39)
A~N(A N)gi (A N)gi	7 (27)	8 (171)	11(11)	11 (6)
N (Prep~A~N)	8 (27)	7 (191)	9 (19)	6–8 (39)
Укупан број за највише 8 образаца	1509	3956	3364	1627

Укупан број полилексема		5080	3532	1798
Које препознаје екстрактор <i>Srp-TE</i>	1736	4469	3463	1711
		87.97%	98.05%	95.16%
Који препознају највише 8 образаца	86.92%	77.87%	95.24%	90.49%
Најдуже полилексема	5	17	7	7
Термини чији је број компонената већи од 5		274	7	9
		5.39%	0.20%	0.50%

(Табела 23)

Легенда:

A–придев, N–именица, Prep–предлог, N_{gi}–именица у генитиву или инструменталу, (A|N) –придев или именица, Prep~N–предлог иза кога стоји именица у одговарајућем падежу, N~N– две именице које се слажу у падежу и броју, A~N–именица којој претходи придев који се слаже са именицом у падежу, броју, роду и аниматности.

У табели 24 приказани су највише и најмање успешни обрасци међу обрасцима који су екстраховали најмање 50 термина кандидата. Може се извести закључак да се најуспешнији обрасци ослањају на A~N фразу, док се најмање успешни ослањају на предлошку фразу. То показује да је тешко бавити се предлошким фразама у локалном контексту јер ови обрасци, или не захватају целу фразу (нпр. *ugalj za proizvodnju čega?*) или се не односе на именицу која им претходи (*dimenzija sijalice sa dimenzijama*). Ово је веома важно јер је N (Prep~N) међу високо ранжираним обрасцима – од 4. до 8. ранга у зависности од термилошке листе (табела 23). Ова запажања ће се примењивати у будућим истраживањима да би се додали неки нови потенцијално корисни обрасци у поступку екстракције термина и елиминисали они који се нису показали корисним односно како би се проширио контекст њиховог претраживања.


Табела 24. Најуспешнији односно најмање успешни обрасци примењивани у поступку термилошке екстракције

Образац	EE	T	F	EE+T+F	EE+T+F (%)	X	Укупно
Најуспешнији							
A~A~N	100	133	112	345	98.3	6	351
A~N	428	824	958	2210	94.0	140	2350
N (A N) _{gi} (A N) _{gi}	64	32	28	124	93.9	8	132

A~N (A N) _{gi} (A N) _{gi}	78	52	28	158	90.8	16	174
A~N (A N) _{gi}	98	148	61	307	88.0	42	349
Најмање успешни							
A~N (Prep ~ N)	20	145	44	209	64.3	116	325
N~N	10	20	29	59	64.1	33	92
N N _{gi} (Prep ~ N)	8	48	4	60	60.0	40	100
N (Prep ~ N)	75	375	156	606	52.3	552	1158

7.9. Објављивање резултата

Након завршеног процеса евалуације, сви исправно евалуирани парови су укључени у термилошку базу података *Termi*, која подржава развој термилошких речника у различитим областима као што су: математика, рачунарство, рударство, библиотекарство, рачунарска лингвистика, електроенергетика итд. (Radoјићић, и други 2018). Апликација *Termi* тренутно подржава обраду и презентацију термина на српском, енглеском, немачком и француском језику (Andonovski, Šandrih и Kitanović 2019) и извоз у различите излазне формате: *Excel*, *TBX*, *lemon rdf*¹⁵⁵, *LMF XML*, *CSV*. У завршном кораку у *Termi* је уврштено 2.972 терминска пара, што премашује број евалуираних парова услед чињенице да су евалуатори у извесним случајевима екстрахованим терминима додавали синониме (који су нпр. пронађени у Еврониму), формирајући тако синонимне скупове (синсетове). Број синсет парова уврштених у *Termi* је 2.359, од чега 552 припадају домену електроенергетике, 1.035 сродним доменима технике, а 772 су општи парови термина. Након завршене обраде, неколико термилошких парова је допуњено дефиницијама и библиографским референцама. Један такав пример приказан је на слици 28.

	
Termin:	 visoki napon
Definicija:	U rudarstvu uglja, naponi iznad 1.000 V.
Sinonimi:	
Skraćenice:	
Nadređeni koncept:	elektroenergetika Rečnik elektroenegetike
Bibliografija :	Dictionary of Mining, Mineral and Related Terms, eduMine,
Termin:	 high voltage
Definicija:	In coal mining, voltages above 1,000 V.
Sinonimi:	high-voltage
Pretraži korpuse:	matematika ▾ Konkordance Frekvencije oblika

© 2021 - Univerzitet u Beogradu, Rudarsko-geološki fakultet

Слика 28. Одредница термилошке базе *Termi*

¹⁵⁵ Модел лексикона за онтологије (енг. *LEXicon Model for ONtologies*, скр. *lemon*) је *RDF* (енг. *Resource Description Framework*) модел који служи за опис лингвистичких ресурса (лексикона и машински читљивих речника) на различитим језицима, дизајниран у складу с принципима семантичког веба у циљу пружања повезаних података на вебу (Gracia, и други 2014).

База података *Termi* има широк спектар примене. Нпр, у дигиталној библиотеци Библиша користи се за проширивање упита синонимима као и за проширивање упита на другим језицима.

Употреба базе *Termi* у Библиши приказана је на сликама 29 и 30. На слици 29 налази се екран за претрагу библиотеке Библиша на ком је појам за претрагу *asynchronous generator* на енглеском (*en*), а колекција која се претражује је *електроенергетика*. Претрага је проширена морфолошки као и терминима који се налазе у бази *Termi*. Добијени термилошки пар из базе *Termi* садржи синониме како за оригинални енглески термин, тако и за одговарајући српски. Сви су морфолошки проширени и користе се за поравнавање конкорданци као што је приказано на слици 30, на којој су истакнути облици термина који су упарени.

BIBLIŠA: ALIGNED COLLECTION SEARCH TOOL [Log In] [Register]

Home Metadata browse Metadata search Mongo search Manage data Help Tutorial About

WELCOME TO ALIGNED TEXT COLLECTION SEARCH TOOL!

Keyword Text collection

Synonyms	en	sr
<input checked="" type="checkbox"/> WordNet...		
<input type="checkbox"/> Dictionary of Librarianship ...		
<input type="checkbox"/> Biblimir ...		
<input type="checkbox"/> GeolISSTerm...		
<input type="checkbox"/> RudOnto ...		
<input checked="" type="checkbox"/> Termi ...	asynchronous generator, asynchronous generator, induction generator	asinhroni generator, induktivni generator

Include Hypernyms Include Hyponyms

Match query: both sentences [-], only one language sentences [-], '-' no filtering.

Morphological query expansion

Слика 29. Употреба базе података *Termi* у дигиталној библиотеци Библиша

Elektroprivreda apstrakti, ID:8.2.3.1 metadata	n992 There are several effects of embedded generator to the existing relay protection: – Short circuit currents in the network are changed, – The nature of fault currents of asynchronous generator is different compared to synchronous generator – The system is more complex, with a higher risk of a short circuit occurrence, – Fault current from the embedded generation may make protection difficult or impossible, – Nuisance tripping, – Changed condition of system earthing, – Possibility that embedded generation protection at the low-voltage side does not detect fault at the medium-voltage side, – Possible great problems when automatic resetting is applied, and – Increased risk of ferroresonance.	n992 Postoji više uticaja mini elektrana na rad postojeće relejne zaštite distributivnih mreža: – Struje kratkih spojeva u mreži sa distribuiranom proizvodnjom su promenjene, – Priroda struja kvara se menja ukoliko se radi o asinhronim generatorima ili generatorima upravljanim energetsom elektronikom, – Sistem je složeniji, sa većim rizikom da nastane kratak spoj, – Zbog struje kvara iz mini elektrane rad zaštite može biti onemogućen ili otežan, – Neselektivno isključenje, – Promena uslova uzemljenja, – Mogućnost da zaštita mini elektrane na niskom naponu ne detektuje kvar na srednjem naponu, – Mogući veliki problemi kada je primenjeno automatsko ponovno uključenje, i – Povećana opasnost od ferorezonanse.
Elektroprivreda apstrakti, ID:8.2.3.1 metadata	n1802 On the base of typical sheme of induction generator connection to distribution network, the Simulink model of this system was formed.	n1802 Na bazi tipične šeme veze asinhronog generatora sa distributivnom mrežom formiran je SIMULINK model takvog sistema.
Elektroprivreda apstrakti, ID:8.2.3.1 metadata	n378 ANALYSIS OF THREE-PHASE SELF-EXCITED INDUCTION GENERATOR OPERATION USING GENETIC ALGORITHM	n378 ANALIZA RADA TROFAZNOG SAMOPOBUĐENOG ASINHRONOG GENERATORA PRIMENOM GENETIČKOG ALGORITMA
Elektroprivreda apstrakti, ID:8.2.3.1 metadata	n1801 IMPACT OF SHORT CIRCUITS AND IRREGULAR DISCONNECTION IN DISTRIBUTION NETWORK ON ASYNCHRONOUS GENERATOR OPERATION	n1801 UTICAJ KRATKIH SPOJEVA I NEREGULARNIH ISKLJUČENJA U DISTRIBUTIVNOJ MREŽI NA RAD ASINHRONOG GENERATORA
Elektroprivreda apstrakti, ID:8.2.3.1 metadata	n380 A genetic algorithm has been employed for determination of the unknown variables by minimizing the impedance module of the equivalent circuit.	n380 Nepoznate veličine stacionarnog stanja samopobuđenog asinhronog generatora određuju se minimizacijom modula ekvivalentne impedanse sistema primenom genetičkog algoritma.
Elektroprivreda apstrakti, ID:8.2.3.1 metadata	n1803 This model was applied for existing network and dynamic behaviour of induction generator was analysed in the case of short circuit and unregular disconnection near the generator.	n1803 Primenom tog modela na praktičnom primeru praćeno je dinamičko ponašanje asinhronog generatora pri kratkim spojevima i prekidima veze sa distributivnom mrežom u njegovoj blizini.
Elektroprivreda apstrakti, ID:8.2.3.1 metadata	n379 In this paper a procedure for the steady-state analysis of a three-phase self excited induction generator supplying balanced/unbalanced loads is proposed.	n379 U ovom radu je predložen postupak za određivanje performansi trofaznog samopobuđenog asinhronog generatora u uslovima simetričnog/nesimetričnog opterećenja.
Elektroprivreda apstrakti, ID:8.2.3.1 metadata	n1805 The analyse includes symetric and unsymetric short circuits, possible self-excitation after the circuit breaker opening and induction generator operation without one phase.	n1805 Analizom su obuhvaćeni simetrični i nesimetrični kratki spojevi, operacije otvaranja prekidača koje mogu dovesti do neželjenog samopobuđivanja, kao i rad asinhronog generatora na dve faze.
Elektroprivreda apstrakti, ID:8.2.3.1 metadata	n1750 Key words: micro hydropower plant, crossflow turbine, induction generator	n1750 Ključne reči: mikro hidroelektrana, turbina sa povratnim protokom, induktivni generator

Слика 30. Облици упарених термина у дигиталној библиотеци Библиша

Апликација *Termi* је спрегнута и са корпусом, тако да се може изабрати неки од понуђених корпуса и из њега могу добити конкорданце (слика 31 доле) и фреквенције облика (слика 31 десно). Ова опција сада је доступна за једночлане термине, а у току је проширење и на вишечлане термине. Напоменимо да је коришћење ових опција јавно доступно и не захтева се никакво логовање на систем.

https://termi.rgf.bg.ac.rs/Dodaj/PojedinacniPrikaz?id=116216

Concordances – Profil 1 – Microsoft Edge
https://termi.rgf.bg.ac.rs/CQP/NoskeFreq

transformator

Termin: transformator
Definicija:
Sinonimi:
Skracenicice:
Nadređeni koncept: elektroenergetika Rečnik elektroenergetike

Termin: distribution transformer
Definicija:
Sinonimi: transformer

Pretraži korpus: BiKes_sr Koncordance Frekvencije oblika

Plain lemma

transformatora	66
transformator	29
TRANSFORMATORA	13
transformatori	9
transformatorima	7
transformatoru	4
transformatorom	1
transformatore	1
TRANSFORMATORIMA	1

Concordances – Profil 1 – Microsoft Edge
https://termi.rgf.bg.ac.rs/CQP/Noske

Plain lemma

režim 'stationary regime' energetski	transformator	'power transformer' Kao disciplina koja se
prekidač , energetski pretvarač /	transformator	, energetska analiza , kondezatorska baterija
. Ključne reči : elektrostatički filter , VNVF	transformator	, energetski pretvarač , IGBT ,
se objektima i opremom (energetski i mer-ni	transformatori	, prekidači , kondenzatori , akumulatorske
identifikacije otkaza sistema (pregrevanje	transformatora	, lom armature , gubitak zapornih svojstava
, gubitak zapornih svojstava spojeva plašta	transformatora	i cevnih vodova , destrukcija visokonaponske
scenariji udesa su isticanje ulja iz	transformatora	i visokonaponske opreme i kontaminacija
i kontaminacija životne sredine , eksplozija	transformatora	i uticaj udarnog talasa , požar na
i uticaj udarnog talasa , požar na	transformatoru	i uticaj toplotnog zračenja . U radu je izvršena
voda , zemljovodno uže , struja kroz zvezdište	transformatora	ISKUSTVA U EKSPLOATACIJI HIBRIDNOG NAPAJANJA
izdvajajuća . Zadržava se jedinica	transformator	/ ispravljač klasičnog tiristorski
, bez oštećenja ili ubrzanog starenja 50 Hz	transformatora	. Hibridno rešenje može biti značajno u slučaju
ANALIZA RADA DIFERENCIJALNE ZAŠTITE	TRANSFORMATORA	U POGONU KOJI NAPAJA ELEKTROLUČNE PEĆI
pogon se sastoji od peći i specijalnog	transformatora	koji se napaja od strane mrežnog
koji se napaja od strane mrežnog	transformatora	. Energija za topljenje posledica je
izobličenja , koja se prenose i na mrežni	transformator	. Tokom eksploatacije dolazilo je do više
prorada diferencijalne zaštite mrežnog	transformatora	, koje su posledica rada elektrolučnog pogona .
elektroenergetskih sistema (vodovi ,	transformatori) . Posebno je pitanje kako proizvedena polja
je da su oštećeni samo namotaji srednjeg stuba	transformatora	i na osnovu svih raspoloživih činjenica
koji su ranije razvili autori , utvrđeno je da je	transformator	spreman za pogon . Izmerena parcijalna

Слика 31. Веза апликације *Termi* и корпуса за опцију јавног приступа

Када је у питању претраживање које обављају овлашћени корисници, тада је могуће изабрати и опцију конкорданци према обрасцима, на пример: придев+изабрана именица (слика 32 горе) или придев+придев+изабрана именица (слика 32 доле). Систем је снабдевен обрасцима који су се показали као најпродуктивнији, али је уједно и отворен за једноставну даљу допуну, како новим обрасцима, тако и новим корпусима.

stacionarni režim 'stationary regime'	energetski transformator	'power transformer' Kao disciplina koja se
obavlja se objektima i opremom (energetski i	mer-ni transformatori	, prekidači , kondenzatori , akumulatorske
PEĆI Elektrolučni pogon se sastoji od peći i	specijalnog transformatora	koji se napaja od strane mrežnog
transformatora koji se napaja od strane	mrežnog transformatora	. Energija za topljenje posledica je
strujna izobličenja , koja se prenose i na	mrežni transformator	. Tokom eksploatacije dolazilo je do više
prorada diferencijalne zaštite	mrežnog transformatora	, koje su posledica rada elektrolučnog pogona .
praznjenja koji se javljaju kod on-site	popravljenih transformatora	koji su ranije dugo bili u pogonu , što je u radu
Srpske RAZVOJ PROTOTIPA VISOKONAPONSKOG	VISOKOFREKVENTNOG TRANSFORMATORA	ZA NAPAJSANJE ELEKTROSTATIČKIH IZDVAJAČA Sve
, koji se koriste u ovim aplikacijama , jeste	energetski transformator	podizač napona . Njegova realizacija zahteva
dobijeni u toku razvoja visokonaponskog	visokofrekventnog transformatora	0 , 4 / 60 kV / kV , radne učestanosti 10 kHz ,
" . Ključne reči : elektrostatički izdvađač ,	visokofrekventni transformator	, energetski pretvarač , visokofrekventno
. Oba pretvaračka kanala se napajaju iz	popudnog transformatora	, koji se napaja sa izvoda generatora .
, specifična toplota PROBLEMI U PRIMENI	STRUJNIH TRANSFORMATORA	MALIH NAZNAČENIH SNAGA Primena strujnih
MALIH NAZNAČENIH SNAGA Primena strujnih	mernih transformatora	malih naznačenih snaga (od 2 , 5 VA do 5 VA) može u
sekundarnog opterećenja na tačnost ovakvih	strujnih transformatora	. Takođe se osim razmatranja navedenih
prevazilaženje . Ključne reči : strujni	merni transformatori	, sekundarno opterećenje , amplitudna i fazna
PROCENE OSTARELOSTI IZOLACIJE	ENERGETSKIH TRANSFORMATORA	PREKO SAVREMENIH DIJAGNOSTIČKIH TEHNIKA I
dijagram temperature vruće tačke u toku	rada transformatora	ili na kontinualnom merenju temperature vruće
i uporedni rezultati na primerima velikih	energetskih transformatora	. Ključne reči : energetski transformator ,
energetskih transformatora . Ključne reči :	energetski transformator	, izolacija , dijagnostika , temperatura ,

koriste u ovim aplikacijama , jeste energetski	transformator podizač napona	. Njegova realizacija zahteva rešavanje niza
zaštitu trofaznih trostubnih energetskih	transformatora sprega YNyn	. Algoritam detektuje zemljospoj u
U DIJAGNOSTICI POGONSKOG STANJA ENERGETSKIH	TRANSFORMATORA Izolaciono ulje	pored funkcije izolovanja delova pod naponom i
gubicima , ili je nužna zamena sa novim	transformatorima povećane snage	. Prva varijanta zahteva mnogo manja početna
kojih je izvedena procena preostalog veka	transformatora prvog bloka	snage 380 MVA napona 420 kV . Na osnovu ukupnog

Слика 32. Примери конкорданци за изабране образаца

Као што је већ поменуто, апликација *Termin* омогућава експортовање речника у различите формате како би се омогућило коришћење развијених ресурса у различитим апликативним окружењима, а сагласно принципима Отворене науке¹⁵⁶ публиковали у машински читљивим облицима. Наводимо исечак датотеке генерисане сагласно *ISO* стандарду за размену термилошких података *TBX* о ком је било речи у поглављу 3.4.3.4.

```

<conceptEntry id="c116089">
  <langSec xml:lang="SR">
    <termSec>
      <term>probojni napon</term>
      <termNote type="termType">entryTerm</termNote>
      <descripGrp>
        <descrip type="definition">Napon pri kojem pukne izolator ili dielektrik; ili napon na kome započinje jonizacija i provodljivost u gasu ili pari.</descrip>
      </descripGrp>
    </termSec>
    <termSec>
      <term>probojni napon</term>
      <termNote type="termType">synonym</termNote>
    </termSec>
  </langSec>
  <langSec xml:lang="EN">

```

¹⁵⁶ <https://www.open.ac.rs/>

```

<termSec>
  <term>breakdown voltage</term>
  <termNote type="termType">entryTerm</termNote>
  <descripGrp>
    <descrip type="definition">The voltage at which an insulator or dielectric
ruptures; or the voltage at which ionization and conduction begin in a gas or vapor</descrip>
  </descripGrp>
</termSec>
<termSec>
  <term>breakdown voltage</term>
  <termNote type="termType">synonym</termNote>
</termSec>
</langSec>
</conceptEntry>

```

Добијени термилошки парови представљају основу за развој новог модерног речника из области електроенергетике, на основу чега ће се побољшати и проширити и термилошка база Електропедија. Уз све то, поступак се показао применљивим и у другим доменима. Овако постављена инфраструктура омогућава накнадно уређивање аутоматски унетих термилошких одредница, додавање дефиниција и сродних података.

У истраживањима која ће уследити, циљ је да се побољша предложена техника укључивањем аутоматске валидације добијених двојезичних термина кандидата у постојећу процедуру, на основу најсавременијих техника машинског учења.

У будућим истраживањима посебна пажња ће се посветити даљем усавршавању термилошког екстрактора за српски *SrpTE*, пре свега предлошким фразама са којима рад у локалном контексту није једноставан. Осим тога, *SrpTE* ће бити обogaћен новим корисним обрасцима у поступку екстракције термина док ће се искључити они који се нису показали корисним. Поред наведеног, полуаутоматска екстракција дефиниција је једна од истраживачких активности која се тренутно одвија и која ће бити саставни део даљег термилошког рада. Предвиђено је и поређење резултата за оба преводилачка смера: на текстовима преведеним са енглеског на српски, као и текстовима преведеним са српског на енглески, али за то је неопходан корпус знатно већих димензија јер се на тај начин обезбеђује корпус који је уравнотежен не само по смеру превођења него и по стилском регистру (научни радови, законска акта, и сл.).

8. ЗАКЉУЧНА РАЗМАТРАЊА

8.1. Закључак

Предмет истраживања ове докторске дисертације је развој терминологије из домена електроенергетике на српском језику уз коришћење различитих метода обраде природних језика. У свом дугогодишњем раду на паралелним корпусима, истраживачи окупљени око Друштва за језичке ресурсе и технологије (Јертех), израдили су већи број вишејезичних паралелних корпуса, који су представљени у поглављу 5.5 овог рада. Корпус из домена електроенергетике још један је у низу корпуса који је настао уз подршку ових истраживача, и производ је рада на овој докторској дисертацији.

Колекција текстова из домена електроенергетике која се састоји од укупно 76 докумената, садржи материјал на српском и енглеском језику, а састављена је од законодавних, техничких и научних докумената објављиваним у периоду од 2006. до 2015. године. Разлог за избор докумената насталим у овом периоду лежи у чињеници да је услед процеса дерегулације електроенергетског сектора у Републици Србији, као и у целој југоисточној Европи овај период био веома динамичан. Процес дерегулације електричне енергије односио се на успостављање оператора система (преносног и дистрибутивног), раздвајање активности производње и снабдевања електричном енергијом, успостављање регулаторног тела и успостављање слободног тржишта електричне енергије. Све ове интензивне активности обављане у периоду од 2005. до 2015. године, праћене су релевантним законима и подзаконским актима и усклађене са прописима ЕУ. Поред тога, у овом периоду објављен је велики број истраживачких радова из области дерегулације енергетског сектора. Све је то, последично, услед динамичних промена у сектору довело до увођења великог броја нових термина.

Питање специјализоване терминологије, како у свету тако и код нас, једно је од веома заступљених у домену лингвистике, превођења, термилошке политике, као и у домену информатичких наука и многих других области. Развој једне дисциплине упоредо са собом носи и развој терминологије исте области. Стога је и њен значај међу свим наведеним дисциплинама нарочит.

Овим радом су представљени методи коришћени приликом обраде корпуса одабраних текстова, почев од екстракције терминологије на српском и енглеском језику, преко поравнавања текстова и упаривања термина и термилошких парова на ова два језика, до евалуације резултата процеса обраде коришћењем статистичких мера за процену сагласности анотатора и укључивања исправно евалуираних парова у термилошку базу података *Termi*.

Рад се састоји од укупно осам поглавља. После уводног поглавља, следи теоријски осврт на терминологију као науку, термилошку праксу у српском језику, настанак и развој стручне терминологије на нашим просторима који је текао упоредо са развојем електроенергетике као научне дисциплине, примену корпусне лингвистике и улога корпуса у лингвистичким истраживањима, развој корпуса из области која је предмет ове дисертације као и детаљан поступак обраде термина добијених екстраховањем из специјализованог корпуса текстова за потребе справљања специјализованог електронског речника из домена електроенергетике.

Примарни циљ овог истраживања био је унапређење постојећих лексичких ресурса у систему двојезичне претраге паралелних колекција односно допуна електронског морфолошког речника из домена електроенергетике на српском језику, као и истовремено проширење вишејезичне дигиталне библиотеке Библиша.

Допуни електронског морфолошког речника претходио је поступак израде паралелног српско-енглеског корпуса из домена електроенергетике, чија је припрема, обрада и паралелизација била слична обради осталих, већ постојећих корпуса приказаних у поглављу 5.5.2.

За екстракцију терминологије из колекције на српском језику, коришћен је систем *Srp- TE* који се ослања на доменске термине у постојећим термилошким и лексичким ресурсима као и на примену синтаксичких образаца. Лексички ресурс који користи систем *Srp- TE* састоји се од

свеобухватних електронских морфолошких речника за српски језик (*SrpMD*) који садрже монолексеми и полилексеми, обухватају општу лексику, властита имена, топониме, енциклопедијско знање и терминологију бројних домена. Пре допуне електронског морфолошког речника, овај ресурс је садржао 296 монолексеми и 43 полилексеми из домена електроенергетике.

За екстракцију терминологије на енглеском језику коришћени су програмски пакети у слободном приступу: *FlexiTerm* и *TermSuite*. Ови алати су заједно произвели листу термина кандидата на енглеском језику (*Eng-TC*) која има 13.312 различитих термина. Применом алата *Srp-TE* добијена је листа *Srp-TC* од 98.922 термина кандидата у првој итерацији, која је након лематизације реч по реч, применом морфолошких речника и уклањањем дупликата смањена на 68.579 ставки.

За поравнавање **комада** коришћен је паралелизовани реченични корпус из домена електроенергетике (описан у одељку 7.2). Поравнати корпус из овог домена унапређен је листом двојезичних парова који су добијени из српског и енглеског ворднета и других двојезичних листа речи, који су представљени у поглављу 7.5. Одреднице из ових ресурса су проширене за оба језика имајући у виду њихове флективне особине, а одговарајући облици су поравнати применом *ad hoc* хеуристичког приступа као што је описано у поглављу 7.5. Добијени резултати показали су да ово проширивање корпуса унапређује како поравнавање тако и коначне резултате, те је стога и корпус из домена електроенергетике унапређен проширеним скупом који броји 458.749 двојезичних парова речи.

За поравнавање речи, екстраховање и рангирање фраза коришћен је *GIZA++* алат система *Moses* (представљени у одељку 7), који служи за машинско превођење и омогућава аутоматско тестирање модела превођења за било који одабрани језички пар. Овај алат омогућава аутоматско одбацивање поравнатих комада за које је вероватноћа превода тих фраза испод одређеног прага. Коришћењем модула *spaCy*, у наредном кораку је примењена лематизована „врећа речи“ за екстраховане термине на енглеском језику за токенизацију на енглеском и уклоњене су незначајне речи применом уграђене стоп листе. Иста лематизација примењена је и на поравнате комаде на енглеском језику па су тако елиминисани они комади у којима се није нашла најмање једна лематизована реч из „вреће речи“. Након овог поступка, *табела фраза* је имала 342.983 поравнатих комада.

Екстраховани термини на српском језику евалуирани су мануалном методом, узимајући у обзир јасно дефинисане критеријуме за избор термина као што су, пре свега, значење појма унутар домена, недвосмисленост, несинонимност, краткоћа, прецизност, постојаност, трајност и интернационалност.

Евалуацију резултата обавила су два независна оцењивача, од којих је један дугогодишњи стручњак у домену електроенергетике, а други преводилац специјализован за исти домен. Укупан број евалуираних парова је 3.833.

Као статистичка мера за одређивање сагласности у анотацији између два оцењивача номиналном скалом мерења коришћен је Коенов капа коефицијент. Усаглашеност у анотацији између два оцењивача у овом истраживању била је врло висока. Термине и парове термина око којих је дошло до неслагања између евалуатора, у завршном кораку је размотрила већа група стручњака, када су и додељене коначне вредности таквим терминима.

Након завршеног процеса евалуације, сви исправно евалуирани парови су укључени у термилошку базу података *Termi*. Ови термилошки парови представљају основу за развој новог савременог речника из домена електроенергетике, чиме су створени предуслови за побољшање и проширење термилошке базе Електропедија. Поступак обраде текстова предложен овом дисертацијом показао се применљивим и корисним и у другим доменима. У будућим истраживањима, циљ је да се побољша предложена техника укључивањем аутоматске валидације добијених двојезичних термина кандидата у постојећу процедуру, и то на основу најсавременијих техника машинског учења.

8.2. Будући рад

Развој терминологије једне области никада није једноставан процес. Обавља се у већем броју итерација, почев од прикупљања текстова за корпус (при чему се увек може поставити питање да ли су одабрани текстови репрезентативан узорак одређене струке), примене различитих рачунарских метода и алата приликом њихове обраде све до екстракције термина и коначне евалуације добијених резултата. Уз то се свакако подразумевају лингвистичке, информатичке као и стручне компетенције истраживача из области чију терминологију развијамо.

Као једна од најважнијих привредних грана сваке земље, па тако и наше, електроенергетика је област у којој веома брзо настају нови термини за предмете, појмове и појаве у оквиру ове струке. Уз то, продор страних термина из ове области захтева озбиљну посвећеност у проналажењу адекватних терминолошких решења.

Речници који су релевантни за домен електроенергетике, а који су описани у поглављу 4.2, углавном покривају општу област електротехнике и нису доступни у дигиталном облику. Стога је креирање једног целовитог, систематичног, свеобухватног и јавно доступног дигиталног двојезичног терминолошког речника из области електроенергетике преко потребно. У оваквом речнику терминологија српског језика ове струке била би упоредо дата са њеним еквивалентима, терминологијом на енглеском језику. Уз сваку терминолошку одредницу у оваквој бази, биле би придружене додатне информације о значењу, пореклу и употреби одређеног термина и у другим областима у којима се јавља. Уз све то, одреднице би биле подржане и колекцијама примера употребе специфичног термина.

Последњих година улажу се велики напори да се све појаве у свету науке и технике терминолошки адекватно прате и означавају због чега је интензиван и рад на нормирању, стандардизацији термина и стварању вишејезичких речничких база. Креирање једног оваквог, информатички подржаног речника, имао би за циљ унапређење размене научних, техничких и других информација које су важне за опстанак и даљи развој савременог друштва, али би имао и значајну улогу у савладавању језичких баријера. Стога је у будућем раду планиран развој терминолошке базе података која би преводиоцима и стручњацима из области електроенергетике омогућила брз приступ подацима и њихово ефикасно претраживање, али која би била и од помоћи у процесу унификације терминологије приликом израде стручних, техничких и других текстова, коришћењу приликом израде стандарда, итд. Из тог разлога, корпус текстова који би били похрањени у бази неопходно је стално ажурирати додавањем нових и заменом постојећих текстова.

За креирање оваквог доменског компјутеризованог речника са директним приступом неопходно је укључити експерте из области рачунарства, рачунарске лингвистике, филологе, преводиоце, као и експерте из области електроенергетике, што у основи подразумева интердисциплинарни приступ.

Овакво софтверско решење требало би да обезбеди брз и лак приступ подацима у бази, једноставну и брзу комуникацију са терминолошком базом података, приступ бази података путем веб-интерфејса као и путем андроид апликације.

9. ДОДАТАК

9.1. Списак скраћеница на ћирилици

КТТ	Комисија техничке терминологије
КНТТ	Комитет научно-техничке терминологије
ИСС	Институт за стандардизацију Србије
САНУ	Српска академија наука и уметности
ЦАНУ	Црногорска академија наука и уметности
АНУРС	Академија наука и уметности Републике Српске
ЈУС	Ознака за југословенски стандард
СРПС	СРПС је ознака за стандарде и сродне документе које доноси Институт за стандардизацију Србије
ДДК	Дјуијева децимална класификација
УДК	Универзална децимална класификација
ХЕПС	Хидроенергетски и пловидбени систем
ПАП	Пумпно акумулационо постројење
ЕУ	Европска унија
НЕТК	Неетикетирани корпус српског језика
ЈП ЕПС	Јавно предузеће „Електропривреда Србије“
АЕРС	Агенција за енергетику Републике Србије
ЕЕ	електроенергетика

9.2. Списак скраћеница на латиници

<i>ISO</i>	енг. <i>International Organization for Standardization</i> (Међународна организација за стандардизацију)
<i>ISA</i>	енг. <i>International Federation of the National Standardizing Associations</i> (Међународна федерација националних удружења за стандардизацију)
<i>TC</i>	енг. <i>Technical Committee</i> (технички комитет)

<i>UNSCC</i>	енг. <i>United Nations Standards Coordinating Committee</i> (Координациони комитет за стандарде Уједињених нација)
<i>Infoterm</i>	енг. <i>International Information Centre for Terminology</i> (Међународни информациони центар за терминологију – Инфотерм)
<i>UNESCO</i>	енг. <i>United Nations Educational, Scientific and Cultural Organization</i> (Образовна, научна и културна организација Уједињених нација)
<i>FEANI</i>	фр. <i>Fédération Européenne d'Associations Nationales d'Ingénieurs</i> (Европска федерација националних инжењерских удружења)
<i>NLP</i>	енг. <i>Natural Language Processing</i> (обрада природних језика)
<i>TermNet</i>	енг. <i>International Network for Terminology</i> (Међународна мрежа за терминологију – ТермНет)
<i>CEN</i>	енг. <i>European Committee for Standardization</i> (Европски комитет за стандардизацију)
<i>EFTA</i>	енг. <i>European Free Trade Association</i> (Европска асоцијација за слободну трговину)
<i>CENELEC</i>	енг. <i>European Committee for Electrotechnical Standardization</i> (Европски комитет за стандардизацију у електротехници)
<i>IEC</i>	енг. <i>International Electrotechnical Commission</i> (Међународна електротехничка комисија)
<i>IEV</i>	енг. <i>International Electrotechnical Vocabulary</i> (Међународни електротехнички речник познатијем под називом „Електропедија“)
<i>UML</i>	енг. <i>Unified Modeling Language</i> (графички језик за моделовање софтверског система)
<i>TMS</i>	енг. <i>Terminology Management System</i> (систем за управљање терминологијом)
<i>MAF</i>	енг. <i>Morpho-syntactic annotation framework</i> (оквир за морфосинтаксичку анотацију)
<i>LAF</i>	енг. <i>Linguistic Annotation Framework</i> (оквир за лингвистичку анотацију)
<i>LMF</i>	енг. <i>Lexical Markup Framework</i> (оквир за лексичко обележавање)
<i>SynAF</i>	енг. <i>Syntactic Annotation Framework</i> (оквир за синтаксичку анотацију)
<i>MLIF</i>	енг. <i>Multilingual Information Framework</i> (оквир за вишејезичке информације)
<i>XML</i>	енг. <i>Extensible Markup Language</i> (прошириви мета језик за означавање текстуалних докумената)

<i>PISA</i>	енг. <i>Persistent identification and sustainable access (PISA)</i> (трајна идентификација и одрживи приступ)
<i>CMDI</i>	енг. <i>Component Metadata Infrastructure</i> (предефинисане компоненте метаподатак)
<i>TEI</i>	енг. <i>Text Encoding Initiative</i> (стандард за представљање текста у дигиталном облику)
<i>TBX</i>	енг. <i>TermBase eXchange</i> (систем за управљање терминолошким ресурсима, знањем и садржајем)
<i>CNL</i>	енг. <i>Controlled Natural Language</i> (контролисани природни језик)
<i>IEEE</i>	енг. <i>Institute of Electrical and Electronics Engineers</i> (институт инжењера електротехнике и електронике)
<i>UNPEDE</i>	фр. <i>UNion Internationale des Producteurs Et Distributors d'Energie electrique</i> (Међународна унија произвођача и дистрибутера електричне енергије)
<i>CIGRE</i>	фр. <i>Conseil International des Grands Reseaux Electriques</i> (Национални комитет Међународног савета за велике електричне мреже)
<i>CIREN</i>	фр. <i>Centre International de Recherche sur l'Environnement et le Développement</i> (Међународна конференција за електродистрибутивне мреже)
<i>HVDC</i>	енг. <i>High-voltage direct current, HVDC</i> (Једносмерна струја високог напона)
ЕЕС	Електроенергетски систем
<i>LOB</i>	енг. <i>The Lancaster-Oslo/Bergen Corpus</i> (Корпус Ланкастер-Осло/Берген)
<i>F-LOB</i>	енг. <i>The Freiburg-LOB Corpus of British English</i> (Фрајбуршки корпус британског енглеског језика)
<i>Frown</i>	енг. <i>The Freiburg-Brown corpus of American English</i> (Фрајбуршки корпус америчког енглеског језика)
<i>ACE</i>	енг. <i>Australian Corpus of English</i> (Аустралијски корпус енглеског језика)
<i>WWC</i>	енг. <i>Wellington Corpus of Written New Zealand English</i> (Корпус писаног новозеландског енглеског језика Универзитета у Велингтону)
<i>LCMC</i>	енг. <i>Lancaster Corpus of Mandarin Chinese</i> (Мандаринско-кинески корпус Универзитета у Ланкастеру)
<i>LLC</i>	енг. <i>The London-Lund Corpus of Spoken English</i> (Корпус Лондон-Лунд)
<i>COBUILD</i>	енг. <i>Collins Birmingham University International Language Database</i> (Међународна језичка база података Колинс Универзитета у Бирмингаму)

<i>COCA</i>	енг. <i>The Corpus of Contemporary American English</i> (Корпус савременог америчког енглеског језика)
<i>SEU</i>	енг. <i>Survey of English Usage</i> (истраживања употребе енглеског језика)
<i>SSE</i>	енг. <i>Survey of Spoken English</i> (истраживања говорног енглеског језика)
<i>BNC</i>	енг. <i>British National Corpus</i> (Британски национални корпус)
<i>JRC</i>	енг. <i>Joint Research Centre</i> (Заједнички истраживачки центар)
<i>SGML</i>	енг. <i>Standard Generalized Markup Language</i> (стандард за дефинисање језика за обележавање)
<i>SrpKor</i>	Корпус савременог српског језика
<i>TU</i>	енг. <i>Translation Units</i> (преводиљачке јединице)
<i>SMT</i>	енг. <i>Statistical machine translation</i> (статистичко машинско превођење)
<i>PoS tagging</i>	енг. <i>Part of Speech Tagging</i> (етикетирање врсте речи)
<i>TUV</i>	енг. <i>Translation Unit Variant</i> (варијанта преводиљачке јединице)
<i>TonD</i>	енг. <i>Terminology on Demand</i> (алат за превођење)
<i>CQP</i>	енг. <i>Corpus Query Processor</i> (упит за претрагу <i>CQP</i>)
<i>POSIX</i>	енг. <i>Portable Operating System Interface for uniX</i> (стандард)
<i>PCRE</i>	енг. <i>Perl Compatible Regular Expressions</i> (стандард)
<i>BoW</i>	енг. <i>Bag-of-Words</i> (врећа речи)
<i>TMX</i>	енг. <i>Translation Memory eXchange</i> (размена преводиљачке меморије)
<i>SWT</i>	енг. <i>Simple-Word term</i> (монолексема)
<i>MWT, MWU, MWE</i>	енг. <i>Multi-Word Term / Unit / Expression</i> (полилексема)
<i>DELAS</i>	фр. <i>Dictionnaires électroniques des mots simples</i> (речник лема)
<i>DELAF</i>	фр. <i>Dictionnaires électroniques des formes fléchies</i> (речник подређених флективних облика)
<i>DELAC</i>	фр. <i>Dictionnaire électronique des noms composés</i> (речник сложених облика)
<i>DELACF</i>	фр. <i>Dictionnaire électronique de mots composés fléchies</i> (речник флективних облика)
<i>BilingTE</i>	енг. <i>Bilingual Terminology Extraction</i> (екстракција двојезичне терминологије)

<i>AMI</i>	енг. <i>Advanced Metering Infrastructure</i> (напредна мерна инфраструктура)
<i>MDM</i>	енг. <i>Meter Data Management</i> (управљање мерним подацима)

БИБЛИОГРАФИЈА

Библиографија публикована на латиници

- Aarts, Jan. 2002. „Does corpus linguistics exist? Some old and new issues.“ Уредник L.E. Breivik и A. Hasselgren. *From the COLT's mouth and others': Language corpora studies in honour of Anna-Brita Stenström* (Rodopi) 1–19.
- Aijmer, Karin. 2008. „Parallel and comparable corpora.“ У *Corpus linguistics : an international handbook*, уредник Anke Lüdeling and Merja Kytö, 275–292.
- Allwood, Jens. 2008. „Multimodal corpora.“ У *Corpus linguistics : an international handbook*, уредник Anke Lüdeling and Merja Kytö, 207–225. Berlin: Walter de Gruyter.
- Alaburić, Jelica. 1986. „Razvoj sovjetske terminološke škole.“ Уредник Nadežda Vinaver. *Terminološki pregled* (Udruženje naučnih i stručnih prevodilaca Srbije), 207–211.
- Amacker, René. 1995. „Geneva School, after Saussure.“ У *Concise History of the Language Sciences: From the Sumerians to the Cognitivists*, уредник R.E. Asher E.F.K. Koerner, 239–243. Cambridge, UK: Pergamon.
- Andonovski, Jelena, Branislava Šandrih, и Olivera Kitanović. 2019. „Bilingual Lexical Extraction Based on Word Alignment for Improving Corpus Search.“ *The Electronic Library* (37) 4: 722–739.
- Arcan, Mihael, Marco Turchi, Sara Tonelli, и Paul Buitelaar. 2017. „Leveraging Bilingual Terminology to Improve Machine Translation in a CAT Environment.“ *Natural Language Engineering* 23 (5): 763–788.
- Bergenholtz, Henning, и Burkhard Schaefer. 1979. *Empirische Textwissenschaft: Aufbau und Auswertung von Text-Corpora*. Königstein: Cornelsen Verlag.
- Biber, Douglas. 1993. *Representativeness in Corpus Design*. Т. 8, у *Literary and Linguistic Computing*, 243–257.
- Bourigault, Didie, Christian Jacquemin, и Marie-Claude L'Homme. 2001. *Recent Advances in Computational Terminology*. Уредник Didier Bourigault, Christian Jacquemin и Marie-Claude L'Homme. Amsterdam /Philadelphia: John Benjamins Publishing Company.
- Bourigault, Didie, и Christian Jacquemin. 2003. „Term Extraction and Automatic Indexing.“ Уредник R. Mitkov. *The Oxford Handbook of Computational Linguistics* (Oxford University Press) 599–615.
- Bowker, Lynne, и Jennifer Pearson. 2002. *Working with specialized language: a practical guide to using corpora*. Routledge.
- Bozděchová, Ivana. 2017. „Prague School of Terminology.“ Уредник Sambor Grucza Marta Małachowicz. *Polskie i europejskie nurty terminologiczne* (Wydawnictwo naukowe Instytutu Komunikacji Specjalistycznej i Interkulturowej, Uniwersytet Warszawski) 108–120.
- Bratanić, Maja. 1998. „Korpusna lingvistika na kraju 20. stoljeća i implikacije za suvremenu hrvatsku leksikografiju.“ *Filologija* 30–31: 171–177.
- Brenes, Patricia. 2015. *Who is Who in Terminology Dmitrij Semënovič Lotte (1898–1950)*. Последњи приступ октобар 2019. <http://inmyownterms.com/who-is-who-in-terminology-dmitrij-semenovic-lotte-1898-1950/>.
- Brenes, Patricia. 2015a. *Who is Who in Terminology: Ernest K. Dresen (1892–1937)*. Последњи приступ октобар 2019. <http://inmyownterms.com/terminology-ernest-k-dresen-1892-1937/>.

- Brenes, Patricia. 2015b. *Who is Who in Terminology: Sergej Alekseevič Čaplygin (1869–1942)*. Последњи приступ октобар 2019.
- Bugarski, Ranko. 1988. „Kriterijumi izgrađivanja terminologije.“ Уредник Nadežda Vinaver. *Terminologija: načela, normiranje, usaglašavanje rada* (Prevodilac) 17–20.
- Bugarski, Ranko. 2007. *Lingvistika u primeni*. Beograd: Čigoja štampa.
- Cabré, Maria Teresa. 1999. „Terminology: Theory, methods and applications.“ Уредник Juan C. Sager, 13. John Benjamins Publishing Co.
- Castellví, Maria Teresa Cabré. 2003. „Theories of terminology: Their description, prescription and explanation.“ *Terminology* (John Benjamins Publishing Company) 9 (2): 163–199.
- Chovanec, Jan. 2014. „Chapters from the History of Czech Functional Linguistics.“ Последњи приступ новембар 2019. <https://digilib.phil.muni.cz/data/handle/11222.digilib/131558/monography.pdf>.
- Claridge, Claudia. 2008. *With the passage of time, every corpus will eventually turn into one that can be used*. Т. 1, у *Corpus Linguistics: An International Handbook*, уредник Anke Lüdeling and Merja Kytö, 242–259. Berlin: Walter de Gruyter.
- Cohen, Jacob. 1960. „Coefficient of Agreement for Nominal Scales.“ *Educational and Psychological Measurement* 20 (1): 37–46.
- Cram, Damien, и Béatrice Daille. 2016. „Terminology Extraction with Term Variant Detection.“ *Proceedings of ACL–2016 System Demonstrations* (Association for Computational Linguistics) 13–18.
- Damjanović, Dragiša. 2017. „Poimanje jezika kod Platona.“ *Matica crnogorska* (Matica crnogorska) 191–242.
- Dimitrijević, Milan S. 2009. „Đorđe Stanojević – Astronom koji je elektrificirao Srbiju.“ *Planeta*, Mart – April. Последњи приступ 8 9, 2019. <http://www.planeta.rs/33/6%20portreti.htm>.
- Dunaevsky, Carolina. 2019. *Term extraction and Terminology tools*. European Parliament. Последњи приступ септембар 2019. https://termcoord.eu/discover__trashed/free-term-extractors/.
- Encyclopedia Britannica. 2019. *Encyclopedia Britannica*. Последњи приступ септембар 2019. <https://www.britannica.com/biography/Edward-L-Thorndike>.
- Evert, Stefan, и The CWB Development Team. 2020. *The IMS Open Corpus Workbench (CWB): CQP Query Language Tutorial – CWB Version 3.4.24*. http://cwb.sourceforge.net/files/CQP_Tutorial.pdf.
- Felber, Helmut. 1977. „International Efforts to Overcome Difficulties in Technical Communication.“ *Overcoming the Language Barrier*. München: Third European Congress on Information Systems and Networks Organised by the Commission of the European Communities.
- Felber, Helmut. 1984. *Terminology manual*. Paris: Unesco, Infoterm.
- Foo, Jody. 2012. „Computational Terminology: Exploring Bilingual and Monolingual Term Extraction.“ Linköping, Sweden: LiU Tryck.
- Gläser, Rosemarie. 2004. „Rita Temmerman. Towards New Ways of Terminology Description: The Sociocognitive Approach.“ *Lexikos*, 434–439. <http://lexikos.journals.ac.za/pub/article/view/709/308>.
- Gracia, Jorge, Elena Montiel-Ponsoda, Daniel Vila-Suero, и Guadalupe Aguado-de-Cea. 2014. „Enabling Language Resources to Expose Translations as Linked Data on the Web.“ 409–413. Последњи приступ фебруар 2021. http://www.lrec-conf.org/proceedings/lrec2014/pdf/863_Paper.pdf.

- Gries, Stefan Th. 2006. „Some proposals towards more rigorous corpus linguistics.“ *Zeitschrift für Anglistik und Amerikanistik* 54 (2): 191–202.
- Grupa za jezičke tehnologije. 2008. *ACIDE – integrisano razvojno okruženje za paralelizovane korpuse*. Последњи приступ јун 2021. <http://korpus.matf.bg.ac.rs/soft/acide.html>.
- Hakami, Huda, и Danushka Bollegala. 2017. „A Classification Approach for Detecting Cross-Lingual Biomedical Term Translations.“ *Natural Language Engineering* 23 (1): 31–51.
- Haque, Rejwanul, Sergio Penkale, и Andy Way. 2018. „TermFinder: Log-Likelihood Comparison and Phrase-Based Statistical Machine Translation Models for Bilingual Terminology Extraction.“ *Language Resources and Evaluation* 52 (2): 365–400.
- Harris, Roy, и Talbot J. Taylor. 1997. *Landmarks in Linguistic Thought I: The Western Tradition from Socrates to Saussure*. 2. New York.
- Hayes, A. F., и K Krippendorff. 2007. „Answering the Call for a Standard Reliability Measure for Coding Data.“ *Communication Methods and Measures* 1 (1): 77–89.
- Hazem, Amir, и Emmanuel Morin. 2016. „Efficient Data Selection for Bilingual Terminology Extraction from Comparable Corpora.“ *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. The COLING 2016 Organizing Committee. 3401–3411.
- Hoste, Veronique, Klaar Vanopstal, Els Lefever, и Isabelle Delaere. 2010. „Classification-based scientific term detection in patient information.“ *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 16 (1): 1–29.
- Hrvatska enciklopedija. 2019. *Praška škola*. Leksikografski zavod "Miroslav Krleža". Последњи приступ новембар 2019. <http://www.enciklopedija.hr/natuknica.aspx?ID=49963>.
- Hundt, Marianne. 2008. „Text corpora.“ У *Corpus linguistics : an international handbook*, уредник Anke Lüdeling and Merja Kytö, 168–187. Berlin: Walter de Gruyter.
- Infoterm. 2004. „30 YEARS OF INFOTERM.“ Последњи приступ октобар 2018. http://www.infoterm.info/pdf/about_us/30_years_infoterm.pdf.
- Institut za standardizaciju Srbije. 2016. „Informator o radu Instituta za standardizaciju Srbije.“ Institut za standardizaciju Srbije. Последњи приступ јул 2019. https://www.iss.rs/images/upload/Dokumenta/inf_o_radu_lat.pdf.
- Ivanović, Tanja, Ranka Stanković, Branislava Šandrih Todorović, и Svetana Krstev. 2021. „Corpus-based bilingual terminology extraction in the power engineering domain.“ Уредник Куо Kageura и Rita Temmerman. *Terminology (John Benjamins)*. (Прихваћено за штампу).
- Ivić, Milka. 2001. *Pravci u lingvistici*. Deveto izdanje, dopunjeno poglavljem "Lingvistika u devedesetim godinama". Уредник Ivan Čolović. Beograd: Biblioteka XX vek.
- Jaćimović, Jelena. 2019. „Recognition and normalization of temporal expressions in Serbian medical narratives.“ *Информека* 19 (No. 2).
- Johansson, Stig. 2008. „Some aspects of the development of corpus linguistics in the 1970s and 1980s.“ *Corpus Linguistics: An International Handbook*, 33–53.
- Jovković, Ljiljana. 1986. „Bečka terminološka škola.“ Уредник Nadežda Vinaver. *Terminološki pregled (Udruženje naučnih i stručnih prevodilaca Srbije)* 204–206.
- Karadža-Garić, Mevlida. 1988. „Principi i metode rada u terminološkoj leksikografiji.“ Уредник Nadežda Vinaver. *Terminologija: načela, normiranje, usaglašavanje rada (Udruženje naučnih i stručnih prevodilaca Srbije)* 25–32.

- Karadža-Garić, Mevlida. 1979. *Višestrukosti u nastavno-naučnoj terminologiji i njihova upotreba u školama SRBiH*. Sarajevo: Institut za jezik i književnost u Sarajevu, odjeljenje za jezik.
- Kennedy, Graeme D. 1998. *An Introduction to Corpus Linguistics*. Longman.
- Kingscott, Geoffrey. 1998. *Tribute to the founder of terminology: 1998 marks the centenary of Eugen Wüster's birth*. Последњи приступ October 2019. <http://www.mt-archive.info/jnl/LangToday-1998-Wuster.pdf>.
- Koehn, Philipp. 2020. *Moses – Statistical Machine Translation System: User Manual and Code Guide*. Edinburgh: Cambridge University Press.
- Koehn, Philipp, и Josh Schroeder. 2007. „Experiments in domain adaptation for statistical machine translation.“ *Proceedings of the second workshop on statistical machine translation*. Prague: Association for Computational Linguistics. 224–227. Последњи приступ 2020. <http://www.statmt.org/wmt07/WMT-2007.pdf#page=238>.
- Krstev, Cvetana. 2016a. „Elektronski rečnici.“ *Kurs: Leksičko prepoznavanje u obradi prirodnih jezika*. Последњи приступ новембар 2020. <http://poincare.matf.bg.ac.rs/~cvetana/Nastava/1516/nastava1516-new.html>.
- Krstev, Cvetana. 2016. „Početak rada sa Unitex-om i učitavanje teksta.“ *Kurs: Leksičko prepoznavanje u obradi prirodnih jezika*. Последњи приступ новембар 2020. <http://poincare.matf.bg.ac.rs/~cvetana/Nastava/1516/nastava1516-new.html>.
- Krstev, Cvetana. 2008. *Processing of Serbian: Automata, Texts and Electronic Dictionaries*. Belgrade: Faculty of Philology of the University of Belgrade.
- Krstev, Cvetana. 2016b. „Pronaženje obrazaca u tekstu – regularni izrazi.“ *Kurs: Leksičko prepoznavanje u obradi prirodnih jezika*. Последњи приступ новембар 2020. <http://poincare.matf.bg.ac.rs/~cvetana/Nastava/1516/nastava1516-new.html>.
- Krstev, Cvetana, Duško Vitas, и Gordana Pavlović-Lažetić. 2008. „Resources and methods in the morphosyntactic processing of Serbo-Croatian.“ *Formal Description of Slavic Languages: The Fifth Conference, Leipzig 2003*. Leipzig: Peter Lang: Frankfurt am Main. 3–17.
- Krstev, Cvetana, Duško Vitas, и Tomaž Erjavec. 2004. „MULTEXT-East Resources for Serbian.“ *Zbornik 7. mednarodne multikonference "Informacijska družba IS 2004", Jezikovne tehnologije*.
- Krstev, Cvetana, Gordana Pavlović-Lažetić, Duško Vitas, и Ivan Obradović. 2004. „Using Textual and Lexical Resources in Developing Serbian Wordnet.“ *Romanian Journal of Information Science and Technology* (Publishing House of the Romanian Academy) 7 (1–2): 147–161.
- Krstev, Cvetana, Svetla Koeva, и Duško Vitas. 2006. „Towards the Global Wordnet.“ *Conference Abstracts of the First Interantional Conference of Digital Humanities Organisations (ADHO) Digital Humanities*. Paris-Sorbonne. 114–117.
- Krstev, Cvetana, и Duško Vitas. 2009. „An Aligned English-Serbian Corpus.“ *ELLSIIR Proceedings (English Language and Literature Studies: Image, Identity, Reality)* 495–508.
- Krstev, Cvetana, и Duško Vitas. 2009a. „An Effective Method for Developing a Comprehensive Morphological E-dictionary of Compounds.“ *Arena Romanistica* 204–212.
- Krstev, Cvetana, и Duško Vitas. 2005. „Corpus and Lexicon - Mutual Incompleteness.“ *Proceedings of the Corpus Linguistics*.
- Laporte, Éric. 2003. *The RELEX Network*. Последњи приступ новембар 2020. <http://infolingua.univ-mlv.fr/english/Relex/introduction.html>.

- Leech, Geoffrey. 1992. „Corpora and theories of linguistic performance.“ Уредник Jan Svartvik. *Directions in Corpus Linguistics: Studies and Monographs, Proceedings of Nobel Symposium 82* 65: 105–122.
- Leech, Geoffrey. 2007. „New Resources, or Just Better Old Ones? The Holy Grail of Representativeness.“ *Corpus Linguistics and the Web* 133–149.
- Leitchik, Vladimir M., и Sergej D. Shelov. 2003. „Terminology: Where is Russian Science Today?“ *LSP and professional communication* 3.
- Leitner, Gerhard. 1991. „The Kolhapur Corpus of Indian English — intravarietal description and/or intervarietal comparison.“ У *English Computer Corpora: selected papers and research guide*, уредник Anna-Brita Stenström Stig Johansson, 215–232. Berlin, New York: Mouton de Gruyter.
- Levi, Viktor A., и Duško D. Bekut. 1997. *Primena računarskih metoda u elektroenergetici*. I izdanje. Novi Sad: Stylos.
- Lindquist, Hans. 2009. *Corpus Linguistics and the Description of English*. Edinburgh University Press.
- Liu, Jingshu, Emmanuel Morin, и Peña Saldarriaga. 2018. „Towards a Unified Framework for Bilingual Terminology Extraction of Single-Word and Multi-Word Terms.“ *Proceedings of the 27th International Conference on Computational Linguistics*. Association for Computational Linguistics. 2855–2866.
- Mahlberg, Michaela. 2006. „Lexical cohesion: Corpus linguistic theory and its application in English language teaching.“ *International Journal of Corpus 11 (3)* (John Benjamins Publishing Company) 363–383.
- Marzá, Nuria Edo. 2009. *The Specialised Lexicographical Approach: A Step Further in Dictionary-making*. Bern: Peter Lang AG, International Academic Publishers.
- McCarthy, Michael, и Anne O’Keeffe. 2010. „Historical perspective: What are corpora and how have they evolved?“ У *The Routledge Handbook of Corpus Linguistics*, 3–13. New York: Routledge.
- McEnery, Tony, и Andrew Wilson. 1996. *Corpus linguistics*. Edinburgh: Edinburgh University Press.
- McEnery, Tony, и Costas Gabrielatos. 2006. „English corpus linguistics.“ *The Handbook of English Linguistics* 33–71.
- Meyer, Charles F. 2002. *English Corpus Linguistics: An Introduction*. Cambridge: Cambridge University Press.
- Meyer, Charles F. 2008. „Pre-electronic corpora.“ Уредник Anke Lüdeling и Merja Kytö. *Corpus Linguistics, An International Handbook* (Walter de Gruyter · Berlin · New York) 1: 1–14.
- Miller, George A., Richard Beckwith, Christiane Fellbaum, Derek Gross, и Katherine Miller. 1990. „Introduction to WordNet: An On-line Lexical Database.“ *International Journal of Lexicography*.
- Obradović, Ivan, Ranka Stanković, и Miloš Utvić. 2008. „Integrirano okruženje za pripremu paralelizovanog korpusa.“ *Die Unterschiede zwischen dem Bosnischen/Bosniakischen, Kroatischen und Serbischen* (LitVerlag) 563–578.
- Och, Franz Josef, и Hermann Ney. 2003. „A Systematic Comparison of Various Statistical Alignment Models.“ *Computational Linguistics* 29 (1): 19–51.
- Oliver, Antoni. 2017. „A System for Terminology Extraction and Translation Equivalent Detection in Real Time.“ *Machine Translation* 31 (3): 147–161.
- Oliver, Antoni, и Mercè Vázquez. 2015. „TBXTools: A Free, Fast and Flexible Tool for Automatic Terminology Extraction.“ *Proceedings of the International Conference Recent Advances in Natural Language Processing*. Hissar: Incoma Ltd. Shoumen. 473–479.

- Ošmjanski, Vera V. 2014. *Koncept engleskog jezika kao lingua franca i njegova percepcija u akademskoj zajednici u Srbiji, dokt. disertacija*. Beograd: Univerzitet u Beogradu, Filološki fakultet.
- Paumier, Sébastien. 2020. *Unitex 3.2 User Manual*. Université Paris-Est Marne-la-Vallée. Последњи приступ 11 2020. <https://www.cis.uni-muenchen.de/people/lg3/ManuelUnitex.pdf>.
- Peck, Bob. 2014. „DDC - 500 - Natural sciences & mathematics.“ University of Illinois, Library and Information Science. Последњи приступ март 2021. http://bpeck.com/references/DDC/ddc_mine500.htm.
- Petrović, Ljubica, и Andrijana Milenković. 1988. „Specifičnosti naučnih i tehničkih termina i problemi njihovog prevođenja sa stranog na srpskohrvatski i sa srpskohrvatskog na strani jezik.“ Уредник Nadežda Vinaver. *TERMINOLOGIJA: načela, normiranje i usaglašavanje rada* (Prevodilac) 170–173.
- Radojičić, Marija, Ivan Obradović, Ranka Stanković, Miloć Utvić, и Sebastijan Kaplar. 2018. „A Mathematical Learning Environment Based on Serbian Language Resources.“ *Proceedings of the 7th International Scientific Conference Technics and Informatics in Education*. Čačak: Faculty of Technical Sciences.
- Radovanović, Miroslav. 1988. „Terminologija – terminografija – terminogradnja.“ Уредник Nadežda Vinaver. *Terminologija – načela, normiranje, usaglašavanje rada* (Prevodilac) 9–16.
- Repar, Andraž, Vid Podpečan, Anže Vavpetič, Nada Lavrač, и Senja Pollak. 2019. „TermEnsembler An ensemble learning approach to bilingual term extraction and alignment.“ *Terminology International Journal of Theoretical and Applied Issues in Specialized Communication* 25 1: 93–120.
- Rey, Alain. 1995. *Essays on terminology*. John Benjamins B.V.
- Sager, Juan C. 1990. *A practical course in terminology processing*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Šandrih, Branislava. 2020. *Impact of text classification on natural language processing applications, Doctoral dissertation*. Belgrade: University of Belgrade, Faculty of Mathematics.
- Šandrih, Branislava, Cvetana Krstev, и Ranka Stanković. 2020. „Two Approaches to Compilation of Bilingual Multi-Word Terminology Lists from Lexical Resources.“ *Natural Language Engineering* 26 (4): 455–479.
- Šandrih, Branislava, и Ranka Stanković. 2019. „Extraction of Bilingual Terminology using Graphs, Dictionaries and GIZA++.“ *Infotheca* 19 (2): 119–138.
- Schaer, Phillip. 2012. „Better than their reputation? On the reliability of relevance assessments with students.“ *Third international conference on Information Access Evaluation: multilinguality, multimodality, and visual analytics (CLEF'12)*. Berlin-Heidelberg: Springer-Verlag. 124–135.
- Semmar, Nasredine. 2018. „A Hybrid Approach for Automatic Extraction of Bilingual Multiword Expressions from Parallel Corpora.“ *Proceedings of the 11th International Conference on Language Resources and Evaluation*. European Language Resources Association. 311–318.
- Silaški, Nadežda. 2012. *Srpski jezik u tranziciji: o anglicizmima u ekonomskom registru*. Beograd: CID Ekonomskog fakulteta.
- Silberztein, Max. 1993. „Dictionnaires Electroniques et Analyse Automatique de Textes.“ *Le systeme INTEX*.
- Sinclair, John M. 1990. „Looking Up: An Account of the COBUILD Project in Lexica lComputing.“ *Computational Linguistics* 16 (3): 184–186.

- Šipka, Danko. 2006. *Osnovi leksikologije i srodnih disciplina*. drugo, izmijenjeno i dopunjeno izdanje. Novi Sad: Matica srpska.
- Spasić, Irena, Mark Greenwood, Alun Preece, Nick Francis, и Glyn Elwyn. 2013. „FlexiTerm: a Flexible Term Recognition Method.“ *Journal of Biomedical Semantics* 4 (1).
- Stanković, Ranka. 2009. *Modeli ekspanzije upita nad tekstuelnim resursima, dokt. disertacija*. Beograd: Univerzitet u Beogradu, Matematički fakultet.
- Stanković, Ranka, Cvetana Krstev, Duško Vitas, Nikola Vulović, и Olivera Kitanović. 2017. „Keyword-Based Search on Bilingual Digital Libraries.“ *Semantic Keyword-Based Search on Structured Data Sources – Second COST Action IC1302 International KEYSTONE Conference*. Cluj-Napoca, Romania: Springer.
- Stanković, Ranka, Ivan Obradović, и Cvetana Krstev. 2009. „Proširivanje upita zasnovano na leksičkim resursima.“ Уредник Đuro Kutlača. *SNTPI '09 – Naučno-stručni skup Sistem naučnih, tehnoloških i poslovnih informacija*. Beograd: Fakultet informacionih tehnologija. 77–80.
- Stanković, Ranka, Ivan Obradović, и Miloš Utvić. 2014. „Developing termbases for expert terminology under the TBX standard.“ *Natural language processing for Serbian: resources and applications* 12–26.
- Steinberger, Ralf, Bruno Pouliquen, Anna Widiger, Camelia Ignat, Tomaž Erjavec, Dan Tufiş, и Dániel Varga. 2006. „Joint Research Centre (JRC): A Multilingual Aligned Parallel Corpus with 20+ Languages.“ *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*. Genoa, Italy: European Language Resources Association (ELRA). 2142–2147.
- Stevens, S. S. 1946. „On the Theory of Scales of Measures.“ *Science, New Series* 103 (2684): 677–680.
- Stubbs, Michael. 1993. „British traditions in text analysis: From Firth to Sinclair.“ Уредник Baker Mona, Francis Gill и Elena Tognini-Bonelli. *Text and technology: in honour of John Sinclair (John Benjamins B.V.)* 1–33.
- Taylor. 2008. „What is corpus linguistics? What the data says.“ *ICAME Journal* 179–200.
- Taylor, Lita, Geoffrey Leech, и Steven Fligelstone. 1991. „A survey of English machine-readable corpora.“ У *English Computer Corpora: selected papers and research guide*, 319–354. Berlin, New York: Mouton de Gruyter.
- Temmerman, Rita. 2000. *Towards New Ways of Terminology Description: the sociocognitive approach*. T. 3. Philadelphia: John Benjamins publishing company.
- Terminology Coordination Unit of the European Parliament. 2019. *Rita Temmerman*. Directorate-General for Translation. Последњи приступ August 2019. <https://termcoord.eu/events/seminar-a-term-is-born-neologisms-in-the-digital-age/rita-temmerman/>.
- Teubert, Wolfgang. 2007. „Corpus linguistics and lexicography.“ Уредник Wolfgang Teubert. *Text Corpora and Multilingual Lexicography (John Benjamins Publishing Co.)* 109–133.
- Teubert, Wolfgang. 2005. „My version of corpus linguistics.“ *International Journal of Corpus Linguistics (John Benjamins Publishing Company)* 10 (1): 1–13.
- Thompson, Geoff, и Susan Hunston. 2006. „System and corpus: two traditions with a common ground.“ У *System and Corpus – Exploring Connections*, 1–14. London.
- Tien, Ha Nguyen, Quyen Ngo The, Huyen Nguyen Thi Minh, и Linh Ha My. 2019. „Rule Based English-Vietnamese Bilingual Terminology Extraction from Vietnamese Documents.“ *Proceedings of*

the 10th International Symposium on Information and Communication Technology. Association for Computing Machinery. 56–62.

- Tognini Bonelli, Elena. 2010. „Theoretical overview of the evolution of corpus linguistics.“ У *The Routledge Handbook of Corpus Linguistics*, уредник Anne O’Keeffe and Michael McCarthy, 14–27. New York: Routledge.
- Tognini-Bonelli, Elena. 2001. *Corpus linguistics at work*. John Benjamins Publishing Co.
- Universal Decimal Classification. 2005. *Universal Decimal Classification*. Том. 1 – Systematic tables. London: British Standards Institution.
- Urbanová, Markéta. 2012. *The life and work of Bohumil Trnka*. Последњи приступ децембар 31, 2018. <https://otik.zcu.cz/bitstream/11025/4552/1/The%20life%20and%20work%20of%20Bohumil%20Trnka%20-%20PDF.pdf>.
- Utvić, Miloš. 2011. „Annotating the Corpus of Contemporary Serbian.“ *Информека* 12 (2): 36а–47а.
- Utvić, Miloš. 2013. *Izgradnja referentnog korpusa savremenog srpskog jezika, dokt. disertacija*. Beograd: Filološki fakultet Univerziteta u Beogradu.
- Vitas, Duško. 1981. „Generisanje imeničkih oblika u srpskohrvatskom jeziku.“ *Informatica* 3: 49–55.
- Vitas, Duško. 2007. „Lokalne gramatike srpskog jezika.“ *Zbornik Matice srpske za slavistiku* 71–72: 305–317.
- Vitas, D., S. Koeva, C. Krstev, и I. Obradović. 2008. „Tour du monde through the dictionaries.“ L’Aquila: 27th conference on Lexis and Grammar.
- Vitas, Duško, и Cvetana Krstev. 2005. „Derivational Morphology in an E-Dictionary of Serbian.“ *Proceedings of 2nd Language and Technology Conference*. Poznań, Poland: Poznań: Wydawnictwo Poznańskie Sp. z.o.o. 139–143.
- Vitas, Duško, и Cvetana Krstev. 2007. „Literature and Aligned Texts.“ *Readings in Multilinguality* 148–155.
- Vitas, Duško, и Cvetana Krstev. 2012. „Processing of Corpora of Serbian Using electronic dictionaries.“ *Prace Filologiczne* 279–292.
- Vitas, Duško, и Cvetana Krstev. 2009. „Srpski jezik i SNTPI.“ *SNTPI '09 – Naučno-stručni skup Sistem naučnih, tehnoloških i poslovnih informacija*. Beograd: Fakultet informacionih tehnologija. 87–90.
- Williams, Geoffrey. 2006. „La linguistique de corpus: une affaire prépositionnelle.“ *Corpus en lettres et sciences sociales: des documents numériques à l’interprétation*. (Texte) 151–158.
- Xiao, Richard. 2010. „Corpus Creation.“ У *Handbook of Natural Language Processing*, уредник Nitin Indurkha & Fred J. Damerau, 147–165. Taylor and Francis Group.
- Xiao, Richard. 2008. *Well-known and influential corpora*. Т. 1, у *Corpus Linguistics: An International Handbook*, уредник Anke Lüdeling and Merja Kytö, 383–457. Berlin: Walter de Gruyter.

Библиографија публикована на ћирици

- Андоновски, Јелена. 2019. *Мрежа отворених података и језички ресурси у процесу изградње српско-немачког литерарног корпуса*. Београд: Докт. дисертација, Филолошки факултет Универзитета у Београду.

- Анђелковић, Јелена. 2017. *Терминологија менаџмента – у прилог развоју термилошке и језичке политике и планирања, докторска дисертација*. Београд: Универзитет у Београду, Филолошки факултет.
- Анђелковић, Јелена, Даница Сеничић, и Ранка Станковић. 2018. „Паралелизовани корпус за домен менаџмента: поступак припреме и могућности примене.“ *Инфотека 2*: 7–31.
- Бјелаковић, Исидора. 2017. „Друштво српске словесности и проблем израде терминологије.“ Уредник Злата Бојовић. *Књижевност и језик у Друштву српске словесности* (Српска академија наука и уметности) 61–76.
- Бјелаковић, Исидора. 2017а. *Терминологија код Срба у 18. и 19. веку (математичка географија и картографија)*. Нови Сад: Два пера.
- Богдановић, Димитрије. 1981. „Почеци српске књижевности.“ У *Историја српског народа*, 212–229. Београд: Српска књижевна задруга.
- Витас, Душко, и Љубомир Поповић. 2003. „Конспект за изградњу референтног корпуса српског стандардног језика.“ *Научни састанак слависта у Вукове дане 221–227*.
- Гортан-Премк, Даринка. 2004. *Полисемија и организација лексичког система у српскоме језику*. Београд: Завод за уџбенике и наставна средства.
- Грицкат, Ирена. 1970. „Језик књижевности и књижевни језик – на основу српског писаног наслеђа из старијих епоха.“ *Јужнословенски филолог XXVIII* св. 1–2: 1–35.
- Грицкат, Ирена. 1964. „Покушаји стварања српске научне терминологије средином прошлог века.“ *Наш језик*, 130–140.
- Драгићевић, Рајна. 2010. *Лексикологија српског језика*. Београд: Завод за уџбенике.
- Дудок, Мирослав. 1996. „Термин и текст.“ *Стандардизација терминологије*, 17–20.
- Ерјавец, Томаж. 2010. „Смернице иницијативе за кодирање текста и њихова локализација.“ *Инфотека 3–15*.
- Ивановић, Тања. 2016. „Лексичка анализа двочланих термилошких синтагми *електродистрибутивног* система.“ *Инфотека 1–2*: 49–72.
- Ивић, Павле. 1981. „Језик у немањихкој епоси.“ У *Историја српског народа – прва књига: од најстаријих времена до маричке битке (1371)*, 617–640. Београд: Српска књижевна задруга Београд.
- Ивић, Павле. 1998. *Преглед историје српског језика*. Нови Сад: Издавачка књижарница Зорана Стојановића.
- Ивић, Павле. 1986. *Српски народ и његов језик*. Београд: Српска књижевна задруга.
- Институт за стандардизацију Србије. 2018. *Међународна класификација стандарда*. Треће издање. Институт за стандардизацију Србије (ИСС).
- Караца-Гарић, Мевлида. 1986. „Актуелна питања српскохрватског језика.“ *Летопис Матице српске* 158–161.
- Кедрова, Галина Евгеньевна, Е. Б. Омелъянова, и Анатолий Михайлович Егоров. 2002. *Трубецкой Николай Сергеевич*. Последњи приступ октобар 2019. <http://fonetica.philol.msu.ru/bibl/tru.htm>.
- Костић, Ђорђе. 2009. *Квантитативни опис структуре српског језика: српски језик од XII до XVIII века*. 1. Београд: Институт за експерименталну фонетику и патологију говора.

- Крстев, Цветана, Бојана Ђорђевић, Сања Антонић, Невена Ивковић-Берчек, Зорица Зорица, Весна Црногорац, и Љиљана Мацура. 2008. „Кооперативан рад на доградњи Српског Wordnet-а.“ *Инфотека* 57–75.
- Митровић, Јелена Д. 2018. *Електронски језички ресурси и алати за обраду српског језика и њихово унапређивање путем модела групне расподеле рада*. Београд: Докт. дисертација, Универзитет у Београду, Филолошки факултет.
- Младеновић, Миљана. 2016. *Евалуација семантичких ознака*. Последњи приступ децембар 2020. http://jerteh.rs/wp-content/uploads/2017/01/seminarJerteh_122016.pdf.
- Прћић, Твртко. 2005. „И ми речи за трку имамо: српски на супрот англосрпском.“ *Свет речи, часопис за српски језик и књижевност* (19–20): 18–24.
- Скерлић, Јован. 1914. *Историја нове српске књижевности*. друго издање. Београд: Издавачка књижарница Геце Кона.
- Станковић, Ранка, Бранислав Тривић, Оливера Китановић, Бранислав Благојевић, и Велизар Николић. 2011. „Развој геолошког термилошког речника ГеолИССТерм.“ *Инфотека* 53–67.
- СТИЈОВИЋ, Рада, и Јелена Лекић. 2016. „О неким термилошким проблемима у језику медицине.“ *Годишњак за српски језик*, 257–262.
- СТИЈОВИЋ, Рада, Олга Сабо, и Ранка Станковић. 2017. „Речник САНУ као база термилошких речника (на примеру речника кулинарства).“ *Словенска терминологија данас* 109–123.
- Тртовац, Александра С. 2016. *Дескриптори метаподатака и дескриптори садржаја у проналажењу информација у дигиталним библиотекама*. Београд: Докторска дисертација, Универзитет у Београду, Филолошки факултет.
- Хорват, Радослав. 1996. „Примена општих принципа у раду на југословенској електротехничкој терминологији.“ *Стандардизација терминологије*, 229–232.

Речници

- Arsenijević, Nada S. 1971. *Elektrotehnički rečnik : nemačko-srpskohrvatski*. Beograd: Beogradski izdavačko-grafički zavod.
- Dabac, Vlatko. 1952. *Elektrotehnički rječnik: njemačko-hrvatski i hrvatsko-njemački*. Zagreb: Školska knjiga.
- Dragović, Ivan, Milan Pavićević, и Petar Vujačić. 1987. *Rečnik industrijske elektrotehnike, nemačko-srpskohrvatski*. Drugo izdanje. Beograd: Privredni pregled.
- Dragović, Ivan, Milan Pavićević, и Petar Vujačić. 1971. *Rečnik industrijske elektrotehnike: nemačko-srpskohrvatski*. Beograd: Interpres.
- Duncan, Katarina, и Mirko Ivković Ivandekić. 1997. *Statistička terminologija korišćena u elektroprivredi*. Beograd: Elektroenergetski koordinacioni centar.
- Ilić, Evgenija. 1969. *Elektrotehnički rečnik: rusko-srpskohrvatski, srpskohrvatsko-ruski*. Zagreb: Tehnička knjiga.
- Ilingvort, Valeri, ур. 1990. *Oksfordski rečnik računarstva*. Beograd: Nolit.
- Marković, Jelica. 1992. *Englesko-srpski tehnički rečnik sa izgovorom, preko 30.000 pojmova: elektrotehnika, elektronika, informatika, fizika, akustika, radio-tehnika, televizija, energetika, automatika, telekomunikacije, električna merenja*. Beograd: NIP Tehnička knjiga.

- Marković, Jelica. 1986. *Englesko-srpskohrvatski elektrotehnički rečnik*. Beograd: Jugoslovenski centar za tehničku i naučnu dokumentaciju.
- Mlakar, France, up. 1957–1970. *Slovenski elektrotehniški slovar*. 10 том. Ljubljana: Elektrotehniška zveza Slovenije.
- Rašović, Miljan M. 1991. *Definicije pojmova iz elektrotehnike na osnovu IEC (CEI): Pojmovnik – rečnik elektrotehnike na pet jezika: srpski, ruski, francuski, engleski, nemački*. Beograd: Sfairros.
- Savez mašinskih i elektrotehničkih inženjera i tehničara Jugoslavije. 1965. *Elektrotehnički terminološki rečnik, grupa 07 – elektronika*. Уредник Тома Јовановић. Beograd: Novinsko-izdavačko preduzeće Tehnička knjiga.
- Savezni zavod za standardizaciju. 1996. *Međunarodni elektrotehnički IEC rečnik sa terminima na srpskom jeziku: elektroenergetika, elektronika i telekomunikacije, knjiga 1, Engleski, francuski, ruski, nemački, španski, italijanski, holandski, poljski, švedski, srpski*. Уредник Radoje Horvat. Т. 1. Beograd: Savezni zavod za standardizaciju.
- Savezni zavod za standardizaciju. 1996. *Međunarodni elektrotehnički IEC rečnik sa terminima na srpskom jeziku: elektroenergetika, elektronika i telekomunikacije, knjiga 2, francuski, engleski, ruski, nemački, španski, italijanski, holandski, poljski, švedski, srpski*. Т. 2. Beograd: Savezni zavod za standardizaciju.
- Savezni zavod za standardizaciju. 1996. *Međunarodni elektrotehnički IEC rečnik sa terminima na srpskom jeziku: elektroenergetika, elektronika i telekomunikacije, Knjiga 3, srpski, francuski, engleski*. Уредник Radoje Horvat. Т. 3. 5 том. Beograd: Savezni zavod za standardizaciju.
- Savezni zavod za standardizaciju. 1997. *Međunarodni elektrotehnički IEC rečnik sa terminima na srpskom jeziku: elektroenergetika, elektronika i telekomunikacije, Knjiga 4, rusko/srpski, srpsko/ruski jezik*. 4. Уредник Radoje Horvat. Т. 4. 5 том. Beograd: Savezni zavod za standardizaciju.
- Savezni zavod za standardizaciju. 1997. *Međunarodni elektrotehnički IEC rečnik sa terminima na srpskom jeziku: elektroenergetika, elektronika i telekomunikacije, Knjiga 5, nemačko/srpski, srpsko/nemački jezik*. Beograd: Savezni zavod za standardizaciju.
- Tankosić, Slobodan B. 2006. *Rečnik elektronike i elektrotehnike: englesko-srpski od 25.000 terminoloških jedinica*. Beograd: Građevinska knjiga.
- Мирковић, Мирко Добр. 1927. *Мали српско-француско-немачки електротехнички речник са скраћеницама и практичним забелешкама*. Beograd: Народна мисао.

Стандарди

- ISO 1087. 2020. *Терминолошки рад и наука о терминологији – Вокабулар*. Институт за стандардизацију.
- ISO 12620. 2020. *Управљање терминолошким ресурсима – Спецификација категорије података*. Институт за стандардизацију.
- ISO 16642. 2017. *Примена рачунара у терминологији – Оквир за терминолошко обележавање*. Институт за стандардизацију.
- ISO 18587. 2018. *Услуге превозиња – Захтеви за накнадно уређивање резултата машинског превозиња*. Институт за стандардизацију.
- ISO 30042. 2020. *Управљање терминолошким ресурсима – TermBase eXchange (TBX)*. Институт за стандардизацију.

- ISO. 1997. „Friendship among equals – Recollections from ISO's first fifty years.“ ISO Central Secretariat. Последњи приступ септембар 2019. https://www.iso.org/files/live/sites/isoorg/files/about%20ISO/docs/en/Friendship_among_equal_s.pdf.
- ISO/TC 37. 2009. *ISO 704:2009 "Terminology work — Principles and methods"*.
- SRPS EN 13237. 2013. *Потенцијално експлозивне атмосфере — Термини и дефиниције за опрему и заштитне системе намењене за употребу у потенцијално експлозивним атмосферама.*
- SRPS EN 60276. 2010. *Дефиниције и номенклатура угљених четкица, држача четкица, колектора и клизних колутова.*
- SRPS EN 60383-1. 2011. *Изолатори за надземне водове називног напона већег од 1 kV – Део 1: Керамичке или стаклене ланчане јединице за мреже наизменичне струје - Термини и дефиниције, методе испитивања и критеријуми за пријем.*
- SRPS EN 60383-2. 2011. *Изолатори за надземне водове називног напона већег од 1 kV – Део 2: Керамичке или стаклене ланчане јединице за мреже наизменичне струје - Термини и дефиниције, методе испитивања и критеријуми за пријем.*
- SRPS EN 60672-1. 2011. *Керамички и стаклени изолациони материјали — Део 1: Дефиниције и класификација.*
- SRPS EN 60674-1. 2011. *Спецификације за пластичне фолије за електричне сврхе — Део 1: Дефиниције и општи захтеви.*
- SRPS EN 61067-1. 2011. *Спецификација за траке тканине од стаклених влакана и стаклених полиестерских влакана — Део 1: Дефиниције, класификација и општи захтеви.*
- SRPS EN 80000-6. 2010. *Величине и јединице – Део 6: Електромагнетизам.*
- SRPS EN ISO 17100. 2017. *Услуге превођења – Захтеви за услуге превођења.* Институт за стандардизацију.
- SRPS EN ISO 17100:2017/A1. 2019. *Услуге превођења – Захтеви за услуге превођења – Измена 1.* Институт за стандардизацију.
- SRPS EN ISO 20108. 2019. *Симултано превођење – Квалитет и пренос звука и уноса слике – Захтеви.* Институт за стандардизацију.
- SRPS EN ISO 20109. 2017. *Симултано превођење – Опрема – Захтеви.* Институт за стандардизацију.
- SRPS EN ISO 2603. 2017. *Симултано превођење – Сталне кабине – Захтеви.* Институт за стандардизацију.
- SRPS EN ISO 4043. 2017. *Симултано превођење – Покретне кабине – Захтеви.* Институт за стандардизацију.
- SRPS IEC 60050-121. 2018. *Међународни електротехнички речник – Део: 121: Електромагнетизам.* Београд.
- SRPS IEC 60050-151. 2019. *Међународни електротехнички речник – Део 151: Електрични и магнетни уређаји.* Институт за стандардизацију Србије.
- SRPS IEC 60050-161. 2014. *Међународни електротехнички речник – Део 161: Електромагнетска компатибилност.*
- SRPS IEC 60050-191. 1997. *Међународни електротехнички речник – Део 191: Sigurnost funkcionisanja i kvalitet usluge.*

- SRPS IEC 60050-195. 2008. *Međunarodni elektrotehnički rečnik – Deo 195: Uzemljenje i zaštita od električnog udara.*
- SRPS IEC 60050-212. 2017. *Međunarodni elektротехнички речник – Део 212: Чврсти, течни и гасовити електрични изолациони материјали.*
- SRPS IEC 60050-300. 2007. *Međunarodni elektротехнички речник – Део 300, 311, 312, 313, 314. Београд.*
- SRPS IEC 60050-321. 1995. *Međunarodni elektротехнички речник – Део 321: Мерни трансформатори.*
- SRPS IEC 60050-411. 2013. *Međunarodni elektротехнички речник — Део 411: Обртне машине.*
- SRPS IEC 60050-421. 1996. *Međunarodni elektротехнички речник – Део 421: Енергетски трансформатори и пригушилице.*
- SRPS IEC 60050-426. 2003. *Međunarodni elektrotehnički rečnik – Deo 426: Električni uređaji za eksplozivnu atmosferu.*
- SRPS IEC 60050-436. 2016. *Međunarodni elektротехнички речник – Део 436: Енергетски кондензатори.*
- SRPS IEC 60050-441. 2014. *Međunarodni elektротехнички речник – Део 441: Расклопне апаратуре и осигурачи.*
- SRPS IEC 60050-442. 2007. *Međunarodni elektrotehnički rečnik – Deo 442: Elektroinstalacioni pribor.*
- SRPS IEC 60050-448. 2015. *Međunarodni elektротехнички речник – Део 448: Заштита електроенергетског система.*
- SRPS IEC 60050-461. 2014. *Međunarodni elektротехнички речник – Део 461: Електрични каблови.*
- SRPS IEC 60050-466. 1995. *Međunarodni elektротехнички речник – Део 466: Надземни водови.*
- SRPS IEC 60050-482. 2009. *Međunarodni elektrotehnički rečnik – Deo 482: Primarne i sekundarne ćelije i baterije.*
- SRPS IEC 60050-551. 2014. *Međunarodni elektротехнички речник – Део 551.*
- SRPS IEC 60050-601. 2015. „*Međunarodni elektротехнички речник – Део 601: Производња, пренос и дистрибуција електричне енергије.*“
- SRPS IEC 60050-602. 2015. *Međunarodni elektротехнички речник – Поглавље 602: Производња, пренос и дистрибуција електричне енергије – Производња.*
- SRPS IEC 60050-603. 2015. *Međunarodni elektротехнички речник – Поглавље 603: Производња, пренос и дистрибуција електричне енергије – Планирање и управљање електроенергетским системом.*
- SRPS IEC 60050-605. 2015. *Međunarodni elektротехнички речник – Поглавље 605: Производња, пренос и дистрибуција електричне енергије – Разводна постројења и подстанице.*
- SRPS IEC 60050-614. 2018. *Međunarodni elektротехнички речник – Део 614: Производња, пренос и дистрибуција електричне енергије – Погон (експлоатација).*
- SRPS IEC 60050-617. 2015. *Međunarodni elektротехнички речник – Део 617: Организација/тржиште електричне енергије.*
- SRPS IEC 60050-811. 2015. *Međunarodni elektротехнички речник – Поглавље 811: Електрична вуча.*
- SRPS IEC 60050-826. 2008. *Međunarodni elektrotehnički rečnik – Deo 826: Električne instalacije.*

- SRPS IEC 60050-841. 2016. *Међународни електротехнички речник – Део 841: Индустриска електротермија.*
- SRPS IEC 60050-903. 2014. *Међународни електротехнички речник – Део 903: Оцењивање ризика.*
- SRPS ISO 10241-1. 2018. *Терминолошки чланци у стандардима – Део 1: Општи захтеви и примери представљања.* Институт за стандардизацију.
- SRPS ISO 10241-2. 2018. *Терминолошки чланци у стандардима – Део 2: Усвајање стандардизованих терминолошких чланака.* Институт за стандардизацију.
- SRPS ISO 1087-1. 2003. *Терминолошки рад – Вокабулар – Део 1: Теорија и примена.* Институт за стандардизацију.
- SRPS ISO 1087-2. 2005. *Терминолошки рад – Вокабулар – Део 2: Примене рачунара.* Институт за стандардизацију.
- SRPS ISO 12199. 2018. *Алфабетски редослед вишејезичких терминолошких и лексикографских података дат латиничним алфабетом.* Институт за стандардизацију.
- SRPS ISO 12615. 2018. *Библиографске референце и идентификација извора за потребе терминолошког рада.* Институт за стандардизацију.
- SRPS ISO 12616. 2018. *Терминографија која се односи на превођење.* Институт за стандардизацију.
- SRPS ISO 15188. 2018. *Смернице за управљање пројектима терминолошке стандардизације.* Институт за стандардизацију.
- SRPS ISO 1951. 2009. *Приказивање/представљање одредница у речницима – Захтеви, препоруке и информације.* Институт за стандардизацију.
- SRPS ISO 22128. 2011. *Терминолошки производи и услуге – Преглед и упутство.* Институт за стандардизацију.
- SRPS ISO 22274. 2018. *Систем за управљање терминологијом, знањем и садржајем – Концептуални аспекти развоја и интернационализације класификационих система.* Институт за стандардизацију.
- SRPS ISO 23185. 2010. *Оцењивање и утврђивање референтних вредности (бенчмаркинг) терминолошких ресурса – Општи појмови, принципи и захтеви.* Институт за стандардизацију.
- SRPS ISO 24156-1. 2018. *Графичка нотација за моделовање појмова у терминолошком раду и његова веза са UML-ом – Део 1: Смернице за коришћење UML нотације у терминолошком раду.* Институт за стандардизацију.
- SRPS ISO 24610-1. 2018. *Управљање језичким ресурсима – Структуре својстава– Део 1: Представљање структура својстава.* Институт за стандардизацију.
- SRPS ISO 24610-2. 2018. *Управљање језичким ресурсима – Структуре својстава– Део 2: Декларација система својстава.*
- SRPS ISO 24611. 2018. *Управљање језичким ресурсима – Оквир за морфосинтаксичку анотацију (Morpho-syntactic annotation framework – MAF).* Институт за стандардизацију.
- SRPS ISO 24612. 2018. *Управљање језичким ресурсима – Оквир за лингвистичку анотацију (LAF).* Институт за стандардизацију.
- SRPS ISO 24613. 2018. *Управљање језичким ресурсима – Оквир за лексичко обележавање (LMF).* Институт за стандардизацију.

- SRPS ISO 24614-1. 2018. *Управљање језичким ресурсима – Подела писаног текста на речи – Део 1: Основни појмови и општи принципи*. Институт за стандардизацију.
- SRPS ISO 24614-2. 2018. *Управљање језичким ресурсима – Подела писаног текста на речи – Део 2: Подела на речи у кинеском, јапанском и корејском језику*. Институт за стандардизацију.
- SRPS ISO 24615-1. 2018. *Управљање језичким ресурсима – Оквир за синтаксичку анотацију (SynAF) – Део 1: Синтаксички модел*. Институт за стандардизацију.
- SRPS ISO 24616. 2018. *Управљање језичким ресурсима – Оквир за вишејезичке информације (Multilingual Information Framework – MLIF)*. Институт за стандардизацију.
- SRPS ISO 24617-1. 2018. *Управљање језичким ресурсима – Оквир за семантичку анотацију (SemAF) – Део 1: Време и догађаји (SemAF-Time, ISO-TimeML)*. Институт за стандардизацију.
- SRPS ISO 24617-2. 2018. *Управљање језичким ресурсима – Оквир за семантичку анотацију (SemAF) – Део 2: Дијалошке форме*. Институт за стандардизацију.
- SRPS ISO 24617-4. 2018. *Управљање језичким ресурсима – Оквир за семантичку анотацију (SemAF) – Део 4: Семантичке улоге (SemAF-SR)*. Институт за стандардизацију.
- SRPS ISO 24617-6. 2018. *Управљање језичким ресурсима – Оквир за семантичку анотацију – Део 6: Принципи семантичке анотације (SemAF принципи)*. Институт за стандардизацију.
- SRPS ISO 24617-7. 2018. *Управљање језичким ресурсима – Оквир за семантичку анотацију – Део 7: Просторне информације (ISOspace)*. Институт за стандардизацију.
- SRPS ISO 24619. 2018. *Управљање језичким ресурсима – Трајна идентификација и одрживи приступ (PISA)*. Институт за стандардизацију.
- SRPS ISO 24622-1. 2018. *Управљање језичким ресурсима – Предефинисане компоненте метаподатака (CMDI) (енг. Component Metadata Infrastructure – CMDI) – Део 1: Модел компоненти метаподатака*. Институт за стандардизацију.
- SRPS ISO 24624. 2018. *Управљање језичким ресурсима – Транскрипција говорног језика*. Институт за стандардизацију.
- SRPS ISO 26162. 2018. *Систем за управљање терминологијом, знањем и садржајем – Пројектовање, примена и одржавање система управљања терминологијом*. Институт за стандардизацију.
- SRPS ISO 29383. 2012. *Терминолошка политика – Развој и имплементација*. Институт за стандардизацију.
- SRPS ISO 30042. 2018. *Систем за управљање терминологијом, знањем и садржајем – TermBase eXchange (TBX)*. Институт за стандардизацију.
- SRPS ISO 639-1. 2018. *Кодови за представљање назива језика – Део 1: Алфа 2 код*. Институт за стандардизацију.
- SRPS ISO 639-2. 2018. *Кодови за представљање назива језика – Део 2: Алфа 3 код*. Институт за стандардизацију.
- SRPS ISO 639-3. 2018. *Кодови за представљање назива језика – Део 3: Алфа 3 код за потпуни обухват језика*. Институт за стандардизацију.
- SRPS ISO 639-4. 2018. *Кодови за представљање назива језика – Део 4: Општи принципи кодирања и представљање назива језика и повезаних ентитета, са смерницама за примену*. Институт за стандардизацију.

- SRPS ISO 639-5. 2018. *Кодови за представљање назива језика – Део 5: Алфа 3 код за породице и групе језика*. Институт за стандардизацију.
- SRPS ISO 704. 2018. *Терминолошки рад – Принципи и методе*. Институт за стандардизацију.
- SRPS ISO 860. 2018. *Терминолошки рад – Усаглашавање појмова и термина*. Институт за стандардизацију.
- SRPS ISO/TS 11669. 2018. *Преводилачки пројекти – Опште смернице*. Институт за стандардизацију.
- SRPS ISO/TS 24617-5.: 2018. *Управљање језичким ресурсима – Оквир за семантичку анотацију (SemAF) – Део 5: Структура дискурса (SemAF-DS)*. Институт за стандардизацију.
- SRPS ISO/TS 24620-1. 2018. *Управљање језичким ресурсима – Контролисани природни језик (CNL) – Део 1: Основни појмови и принципи*. Институт за стандардизацију.
- SRPS L.G7.101. 1982. *Микроталасни инструменти – Термини и дефиниције*.

Биографија аутора

Тања Ивановић рођена је 1981. године у Београду. Дипломирала је на Филолошком факултету Универзитета у Београду 2005. године. Мастер студије завршила је на Факултету организационих наука Универзитета у Београду 2008. године. Школске 2012/2013. године уписала је докторске студије на Филолошком факултету Универзитета у Београду.

Од 2009. године ради у Канцеларији за европске интеграције Владе Републике Србије која 2017. године постаје Министарство за европске интеграције, и то на припреми српске верзије правних тековина Европске уније. Пре ангажовања у некадашњој Канцеларији односно данашњем Министарству за европске интеграције, радила је у ЈП Електромрежа Србије на пословима односа са јавношћу и превођењу службених докумената.

Члан је Друштва за језичке ресурсе и технологије (Јертех) Универзитета у Београду.

Говори енглески и шпански језик.

Изјава о ауторству

Име и презиме аутора: **Тања Д. Ивановић**

Број досијеа: **11134/Д**

Изјављујем

да је докторска дисертација под насловом:

Развој терминологије из области електроенергетике коришћењем метода обраде природних језика

- резултат сопственог истраживачког рада;
- да дисертација ни у целини ни у деловима није била предложена за стицање дипломе студијских програма других високошколских установа;
- да су резултати коректно наведени и
- да нисам кршила ауторска права и користила интелектуалну својину других лица.

Потпис аутора

У Београду, _____

Изјава о истоветности штампане и електронске верзије докторског рада

Име и презиме аутора: **Тања Д. Ивановић**

Број досијеа: **11134/Д**

Студијски програм: **модул Култура**

Наслов рада: **Развој терминологије из области електроенергетике коришћењем метода обраде природних језика**

Ментор: **Проф. др Цветана Крстев**

Изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предала ради похрањивања у **Дигитални репозиторијум Универзитета у Београду**.

Дозвољавам да се објаве моји лични подаци за добијање академског назива доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.

Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

Потпис аутора

У Београду, _____

Изјава о коришћењу

Овлашћујем Универзитетску библиотеку „Светозар Марковић“ да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом: **Развој терминологије из области електроенергетике коришћењем метода обраде природних језика**, која је моје ауторско дело.

Дисертацију са свим прилозима предала сам у електронском формату погодном за трајно архивирање.

Моју докторску дисертацију похрањену у Дигиталном репозиторијуму Универзитета у Београду, и доступну у отвореном приступу, могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучила:

1. Ауторство (CC BY)
2. Ауторство – некомерцијално (CC BY-NC)
3. Ауторство – некомерцијално – без прерада (CC BY-NC-ND)
4. Ауторство – некомерцијално – делити под истим условима (CC BY-NC-SA)
5. Ауторство – без прерада (CC BY-ND)
6. Ауторство – делити под истим условима (CC BY-SA)

(Молимо да заокружите само једну од шест понуђених лиценци. Кратак опис лиценци је саставни део ове изјаве).

Потпис аутора

У Београду, _____

1. **Ауторство.** Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце, чак и у комерцијалне сврхе. Ово је најслободнија од свих лиценци.

2. **Ауторство – некомерцијално.** Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела.

3. **Ауторство – некомерцијално – без прерада.** Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела. У односу на све остале лиценце, овом лиценцом се ограничава највећи обим права коришћења дела.

4. **Ауторство – некомерцијално – делити под истим условима.** Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца не дозвољава комерцијалну употребу дела и прерада.

5. **Ауторство – без прерада.** Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца дозвољава комерцијалну употребу дела.

6. **Ауторство – делити под истим условима.** Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца дозвољава комерцијалну употребу дела и прерада. Слична је софтверским лиценцама, односно лиценцама отвореног кода.