



Научни потенцијал ученичког корпуса *ICLE v3* на примеру његове србофоне компоненте

Јелена Марковић

Универзитет у Источном Сарајеву

Филозофски факултет

Предмет излагања

- Улога и значај ученичких корпуса у примењеној лингвистици данас
 - Дефиниција
 - Кратка историографија: ученички корпуси и контрастивна анализа међујезика
- Међународни ученички корпус *International Corpus of Learner English (ICLE)*
 - Формирање треће верзије корпуса
 - Варијабле контекста језичке продукције
 - Варијабле говорника
 - База метаподатака
 - Доступни интегрисани корпусни алати
- Научни потенцијал – резултати који су постигнути коришћењем његове србофоне компоненте, *ICLE-SE*
- Корпус *ICLE-SE* и његов значај за примењену лингвистику на српском говорном подручју

Ученички корпуси (енг. learner corpora)

- „Компјутерски ученички корпуси електронске су базе података формиране од аутентичне дискурсне продукције на страном/другом језику по експлицитним критеријумима, тако да се могу користити за истраживање у оквиру усвајања другог језика и/или наставе страног језика. Базе података бележе се стандардизовано и хомогено, уз информације о пореклу података.“ (Granger, 2002, 7, према Марковић, 2019, 13).
- “Computer learner corpora are electronic collections of authentic FL/SL textual data assembled according to explicit design criteria for a particular SLA/FLT purpose. They are encoded in a standardized and homogeneous way and documented as to their origin and provenance.” (Granger, 2002, 7)

Аутентичност

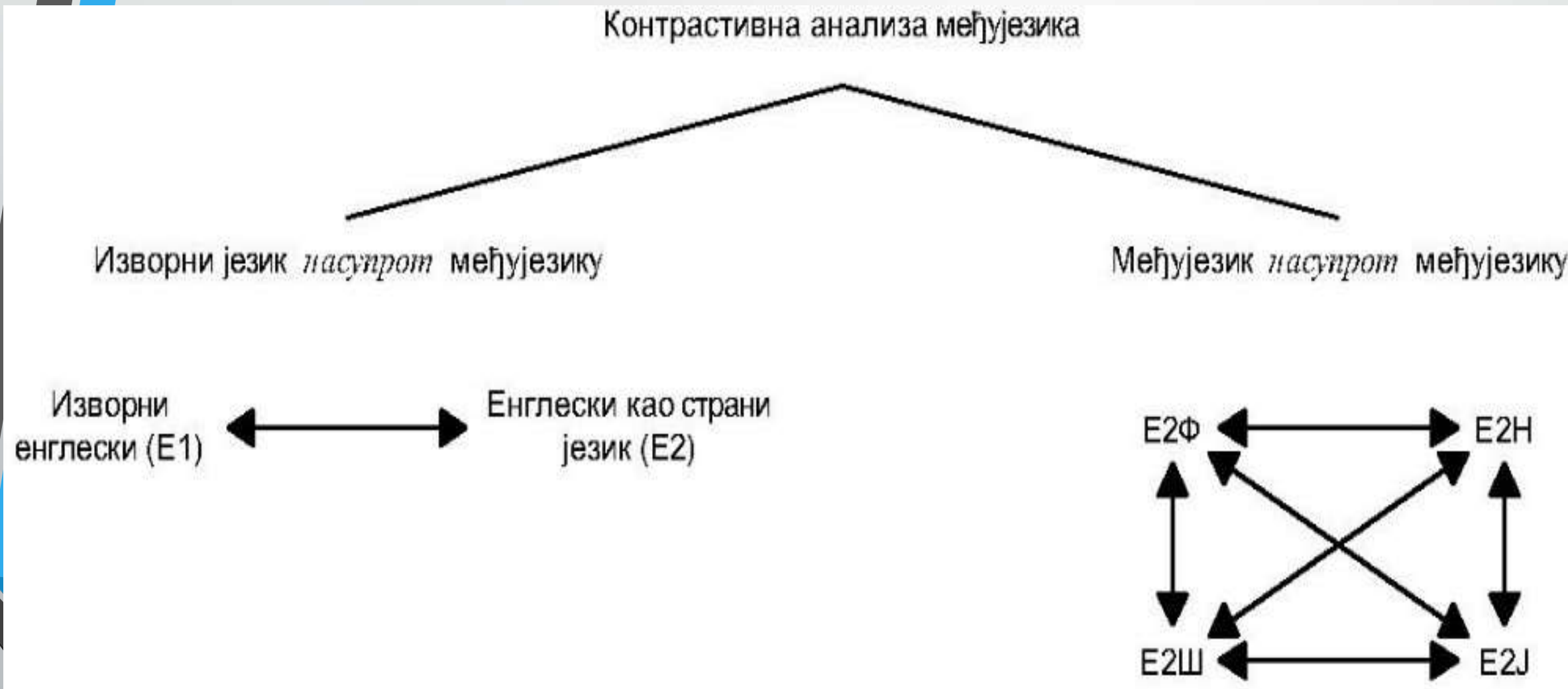
- Корпуси подразумевају **аутентичност** језика
- поставља се питање аутентичности **неизворног** језика
 - **Рedefинисање питања аутентичности**
 - Простор за апсолутну аутентичност у учењу страног језика у учионици је мали, ако уопште постоји (Sinclair, 1996; Granger, 2002, 8).
 - Стога се аутентичном продукцијом на страном језику може сматрати **језичка продукција која има за циљ комуникацију на страном језику, а може се одвијати у условима који су делимично контролисани**, како истиче и Род Елис (Ellis, 1994, 671).
- Неопходност **контекстуализоване** дискурсне продукције у ученичким корпусима: нпр. писање састава (не појединачних реченица)

Метаподаци

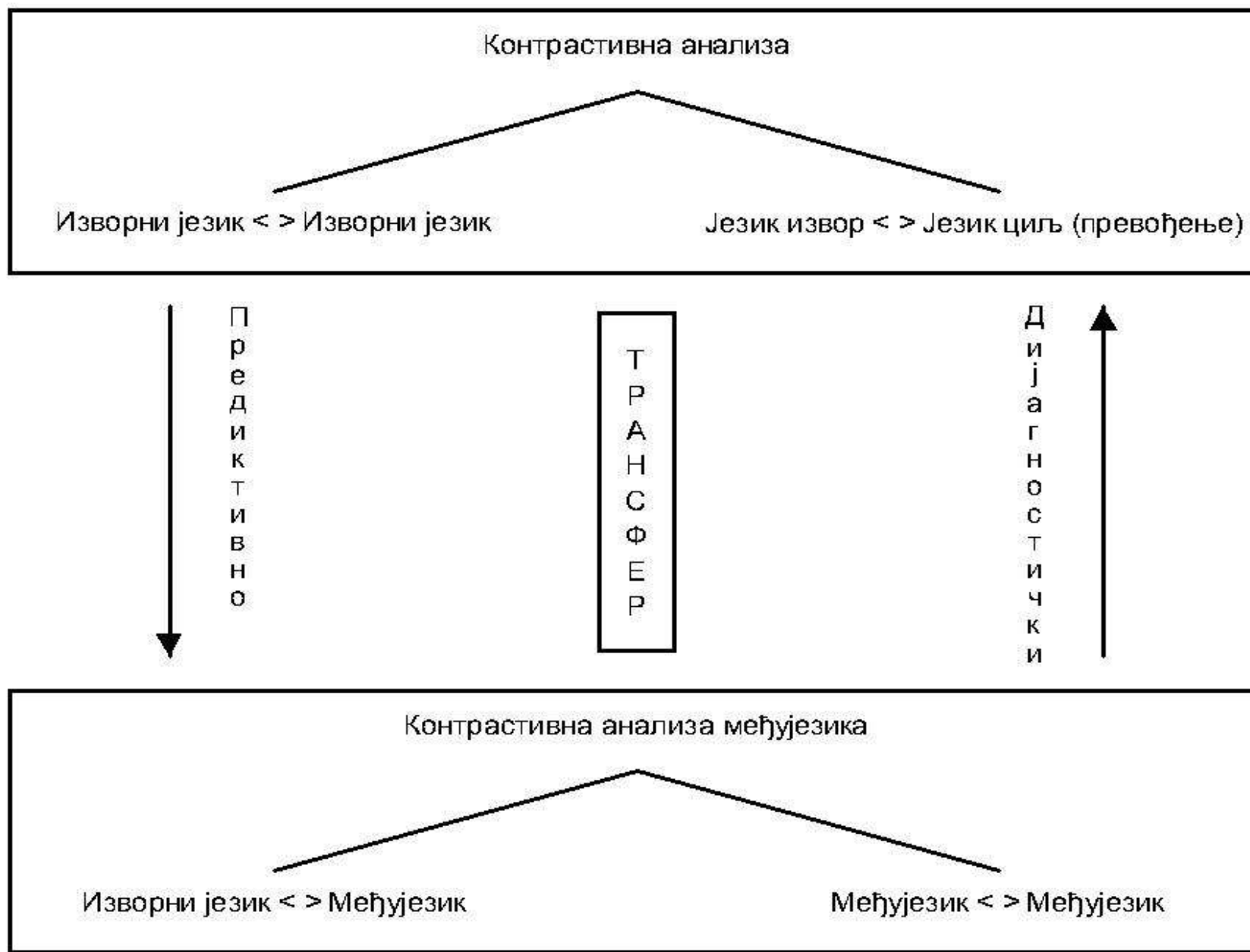
- **Постојање метаподатака** – од посебног значаја за ученичке корпусе
 - Варијабле говорника
 - Варијабле говорне ситуације
- Планирање метаподатака у складу са планирањем коришћења корпуса
- Значај метаподатака за истраживачке сврхе
- Недовољно коришћење и постојећих метаподатака (нпр. *ICLE v3*)

- Термин *контрастивна анализа међујезика* (енг. *Contrastive Interlanguage Analysis*) први пут употребила је Силвијана Грејнџер [Sylviane Granger] 1996. године у раду под називом "From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora".
 - *компаративна истраживања међујезика* (енг. *comparative interlanguage research*; Granger, 1993), преузет из извора где је коришћен у значењу контрастирања језичког понашања неизворних и изворних говорника у говорним ситуацијама које се могу поредити (Péry-Woodley, 1990, 143–144).
 - *контрастивна анализа међујезика* у научним публикацијама на српском језику поменута у раду под називом „О дискурсним маркерима у писању на енглеском језику као страном“ (Марковић, 2016).

Слика 1. Контрастивна анализа међујезика (Granger, 1996, 44)



Слика 2. Интегрисани контрастивни модел (Granger, 1996, 47)



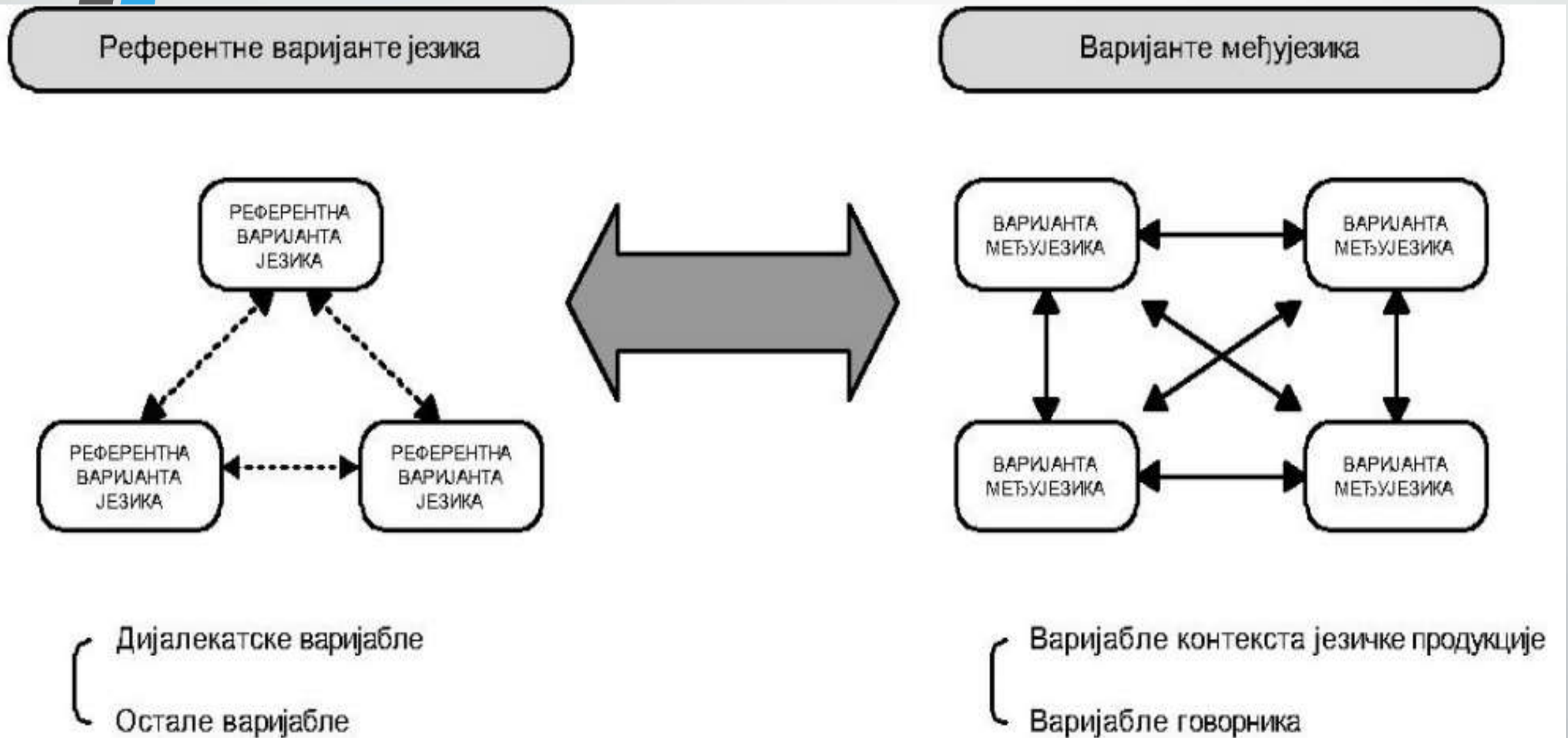
- централно место – **трансфер**, феномен који је једна од кључних тема у усвајању страног језика, присутан у свим фазама усвајања (Jarvis, 2016, 23).
 - методолошка утемељеност истраживања трансфера, јер се комплементарно користе описима изворних и описима неизворних језика (пример: Gilquin, 2008, 7–8).

- Појава и развој **ученичких корпуса** нераскидиво су повезани са појавом **методологије контрастивне анализе међујезика**, односно „корпусном револуцијом“ у лингвистици уопште (Meunier, 2002, 135; Hasselgard & Johansson, 2011, 57),
 - парадигма типова података коришћених у истраживањима у периоду који је претходио корпусној револуцији: **језичка продукција готово није присутна**

Ревидирани модел контрастивне анализе међујезика

- друга верзија, 2015. године
- “Contrastive interlanguage analysis: A reappraisal”, *International Review of Learner Corpus Research* (Granger, 2015, 7–24)
- питање **варијабилности**
- кључне термиолошке иновације: употреба термина
 - **референтне варијанте језика** (енг. *reference language varieties*) и
 - **варијанте међујезика** (енг. *interlanguage varieties*),

Слика 3. Контрастивна анализа међујезика (ревидирани модел) (Granger, 2015, 17)



Референтна варијанта

- референтна варијанта, термин од кључног значаја
- одабир референтног корпуса мора бити оправдан циљевима истраживања, с тим што он не подразумева истовремено и језички узор
 - Пример: Ли и Чен (Lee & Chen, 2008) ; Making a bigger deal of the smaller words: Function words and other key items in research writing by Chinese learners. *Journal of Second Language Writing*, 18, 281–296.
 - два референтна корпуса:
 - корпус научних радова из лингвистичких и примењенолингвистичких часописа који зову „експертски корпус“, „радови експерата коришћени као узор у писању којима треба тежити“ (Lee & Chen, 2008, 284)
 - корпус завршних радова изворних говорника у оквиру корпуса *BAWE (British Academic Written English)*, јер он представља „неку врсту средине између дискурса писаног на страном језику и објављених радова експерата, те ће стога бити реалнија референтна мера“ (2008, 284)

Варијанте међујезика

- Матерњи језик – основна варијабла поређења међујезика, али не и једина у овом моделу КАМ
- поређење варијанти међујезика и по другим варијаблама
- *варијабле говорника* (енг. *learner variables*),
 - узраст,
 - период учења језика или
 - познавање других језика,
- *варијабле задатка* (енг. *task variables*), односно *варијабле контекста језичке продукције*:
 - жанр,
 - могућност коришћења извора и/или речника,
 - испитни/други контекст или
 - (не)постојање временског ограничења.

• **метаподаци**

Различити приступи

- У истраживањима ученичких корпуса неопходно је препознати разлику између два методолошка приступа: **истраживања која су вођена корпусом** (енг. *corpus-driven*) и **истраживања која су заснована на корпусу** (енг. *corpus-based*).
 - Критеријум разликовања два методолошка приступа – порекло истраживачког питања.
 - у првом случају истраживачко питање резултат је проучавања корпуса
 - у другом случају истраживачко питање унапред је дефинисано.
- Једино у истраживањима која су вођена корпусом користе се сви његови потенцијали
 - оваква корпусна истраживања мање су заступљена (Granger, 2012, 12).

Термини *overuse/underuse*

- термини *overuse/underuse*, често коришћени термине у истраживањима у оквиру контрастивне анализе међујезика, често и критиковани
- квантитативна поређења учесталости речи у ученичким корпусима и изворним (или другим референтним) варијантама језика обично укажу на значајне разлике, које се затим именују терминима ***overuse*** (учестала употреба) и ***underuse*** (ретка употреба).
- предмет погрешне интерпретације, будући да се некад схватају као крајњи резултати истраживања
- квантитативни резултати поређења учесталости неког језичког феномена могу имати научну вредност тек у каснијим интерпретацијама истраживача, односно у тумачењу компетенције говорника у светлу првенствено функционалне дистрибуције учесталог или ретког феномена у неизворном језику.

International Corpus of Learner English v3

- **ICLE**, Centre for English Corpus Linguistics, Université catholique de Louvain
- Верзија 3: око 5,7 милиона речи, 25 корпуса који су одређени матерњим језиком говорника
 - Међу њима *ICLE-SE*, корпус србофоних говорника
- Референтни изворни корпус формиран на британским и америчким универзитетима, *LOCNESS* (*Louvain Corpus of Native English Essays*)

International Corpus of Learner English v 1-3

- припремљени по строго поштованим критеријумима одабира учесника и обезбеђивања метаподатака.
- 11 матерњих језика (бугарски, чешки, холандски, фински, француски, немачки, италијански, пољски, руски, шпански и шведски)
- 16 матерњих језика (поновљено, уз кинески, јапански, норвешки, сетсвана и турски)
- 25 матерњих језика (поновљено, уз неколико нових међу којима је и србофона компонента)
- Сваки појединачни поткорпус одређен матерњим језиком минималног је обима 200.000 речи
- Величина састава који су укључени у корпус јесте између 500 и 1.000 речи, а састави су писани на опште теме.

International Corpus of Learner English v 1-3

- карактеристике учесника и жанра унапред су задате, тако да је добијени материјал снажне унутрашње саобразности, што је од почетка био један од прокламованих циљева пројекта.
- Подаци о променљивим варијаблама (учесника и контекста писања састава) брижљиво су сакупљени и сачувани.
- Неке теме предложила је матична институција, а неке теме одабране су у оквиру националних тимова.
- Пожељно је било да један студент напише највише један састав. Од овог правила одступило се у малом броју случајева, где је један студент учествовао са два састава чија дужина свакако није прешла максималних 1.000 речи по учеснику. То значи да је број састава готово идентичан укупном броју учесника.
- У случају да поткорпуси нису могли да дођу до минималних 200,000 речи, дозвољено је укључивање малог броја састава на књижевне теме (до око 25% укупног броја састава у оквиру појединих поткорпуса).

Неки примери истраживања других *ICLE* компоненти

- Метадискурс, Adel, 2006
- Глагол + партикула: Deshors, 2016
- Језичке иновације од говорника другог до говорника страног језика: Deshors, Götz, Laporte, 2018
- Говорници страног језика и говорници другог језика: Gilquin & Granger, 2011
- Језичке формуле, колокације: Granger, 2018
- Колокације у немачком корпусу: Nesselhauf, 2005

Наслови састава које су студенти најчешће бирали у србофоној компоненти

Is reading a book better than watching its movie adaptation?	84
One should never judge a person by external appearances.	33
Most university degrees are theoretical and do not prepare students for the real world.	26
Are we too dependent on computers?	24
In the words of the old song "Money is the root of all evil." Agree or disagree.	23
Conflicts are necessary for healthy relationships.	20
Does age matter in relationships?	19
Discuss the advantages and disadvantages of banning smoking in restaurants.	14
In our modern world, dominated by science, technology and industrialization, there is no longer a place for dreaming and imagination. What is your opinion?	14
Discuss the necessity of make-up.	11

Пример коришћења функције KeyWord (кључне разлике учесталости речи) (WordSmith Tools v.8): BG,NS:ES,BL

C:\Users\Jelena52\Desktop\2021 NAUKA I STVARNOST\BG_NS_complete.kws

N	Key word	Freq.	%	Texts	RC. Freq.	%	BIC	Log_L	Log_R	P	Lemmas	Set
1	SMOKING	137	0.13	1	35	0.04	49.30	61.52	1.92	0.0000000000		
2	SMOKERS	95	0.09	1	18	0.02	42.96	55.18	2.36	0.0000000000		
3	RESTAURANTS	84	0.08	1	22	0.02	24.56	36.78	1.89	0.0000000000		
4	BANNING	56	0.05	1	10	0.01	21.72	33.94	2.44	0.0000000028		
5	RESTAURANT	46	0.04	1	6		21.44	33.66	2.89	0.0000000036		
6	SMOKE	70	0.07	1	17	0.02	20.81	33.03	2.00	0.0000000061		
7	ALTHOUGH	70	0.07	1	19	0.02	17.30	29.52	1.84	0.0000000525		
8	NON	53	0.05	1	11	0.01	16.48	28.70	2.22	0.0000000816		
9	EARLY	18	0.02	1	65	0.07	17.52	-29.74	-1.90	0.0000000465		
10	SO	258	0.25	1	386	0.39	17.53	-29.75	-0.63	0.0000000463		
11	LIFE	185	0.18	1	297	0.30	17.64	-29.86	-0.73	0.0000000434		
12	DIFFERENCE	16	0.02	1	62	0.06	18.20	-30.42	-2.00	0.0000000318		
13	LOVE	68	0.07	1	170	0.17	36.17	-48.39	-1.37	0.0000000000		
14	MARRIAGE	7		1	63	0.06	41.07	-53.29	-3.21	0.0000000000		
15	WE	961	0.93	1	1,384	1.39	78.22	-90.44	-0.57	0.0000000000		

KWs plot links clusters filenames source text notes

15 entries Row 1 0% T S tag file: 0 entries found

EN 11:18 AM 5/13/2021

Пример коришћења функције KeyWord (кључне разлике учесталости речи) (WordSmith Tools v.8): ES,BL:BG,NS

C:\Users\Jelena52\Desktop\2021 NAUKA I STVARNOST\ES_BL_complete.kws

N	Key word	Freq.	%	Texts	RC. Freq.	%	BIC	Log_L	Log_R	P	Lemmas	Set
1	MARRIAGE	63	0.06	1	7		41.07	53.29	3.21	0.0000000000		
2	DIFFERENCE	62	0.06	1	16	0.02	18.20	30.42	2.00	0.0000000318		
3	EARLY	65	0.07	1	18	0.02	17.52	29.74	1.90	0.0000000465		
4	NON	11	0.01	1	53	0.05	16.48	-28.70	-2.22	0.0000000816		
5	THE	4,060	4.07	1	4,700	4.56	16.80	-29.02	-0.17	0.0000000688		
6	ALTHOUGH	19	0.02	1	70	0.07	17.30	-29.52	-1.84	0.0000000525		
7	SMOKE	17	0.02	1	70	0.07	20.81	-33.03	-2.00	0.0000000061		
8	HOWEVER	67	0.07	1	155	0.15	20.96	-33.18	-1.17	0.0000000055		
9	RESTAURANT	6		1	46	0.04	21.44	-33.66	-2.89	0.0000000036		
10	BANNING	10	0.01	1	56	0.05	21.72	-33.94	-2.44	0.0000000028		
11	RESTAURANTS	22	0.02	1	84	0.08	24.56	-36.78	-1.89	0.0000000000		
12	SMOKERS	18	0.02	1	95	0.09	42.96	-55.18	-2.36	0.0000000000		
13	SMOKING	35	0.04	1	137	0.13	49.30	-61.52	-1.92	0.0000000000		

KWs plot links clusters filenames source text notes

13 entries Row 1 0% T S tag file: 0 entries found

EN 11:22 AM 5/13/2021

Пример коришћења функције KeyWord (кључне разлике учесталости речи) (WordSmith Tools v.8): ICLE-SE: AE (1)

C:\Users\Jelena52\Desktop\2021 NAUKA I STVARNOST\SERBICLECOMPLETE vs AE.kws

N	Key word	Freq.	%	Texts	RC. Freq.	%	BIC	Log_L	Log_R	P	Lemmas	Set
1	WE	2,345	1.16	1	116	0.21	522.54	535.00	2.43	0.0000000000		
2	MOVIE	826	0.41	1	0		378.05	390.51	141.06	0.0000000000		
3	YOU	1,494	0.74	1	63	0.12	362.64	375.10	2.66	0.0000000000		
4	BOOK	1,022	0.50	1	18	0.03	345.08	357.54	3.92	0.0000000000		
5	BOOKS	628	0.31	1	1		272.67	285.13	7.39	0.0000000000		
6	MOVIES	399	0.20	1	0		176.18	188.64	140.01	0.0000000000		
7	US	619	0.31	1	23	0.04	153.56	166.02	2.84	0.0000000000		
8	YOUR	489	0.24	1	12	0.02	142.85	155.31	3.44	0.0000000000		
9	COMPUTERS	323	0.16	1	0		140.25	152.70	139.71	0.0000000000		
10	READING	428	0.21	1	15	0.03	105.58	118.04	2.93	0.0000000000		
11	ADAPTATION	243	0.12	1	0		102.43	114.88	139.29	0.0000000000		
12	MONEY	494	0.24	1	25	0.05	98.57	111.03	2.40	0.0000000000		
13	CHARACTERS	221	0.11	1	0		92.03	104.48	139.16	0.0000000000		
14	RELATIONSHIP	338	0.17	1	11	0.02	83.91	96.37	3.03	0.0000000000		
15	READ	339	0.17	1	12	0.02	80.60	93.06	2.91	0.0000000000		
16	IMAGINATION	192	0.09	1	1		68.91	81.37	5.68	0.0000000000		
17	SMOKING	172	0.08	1	0		68.86	81.32	138.80	0.0000000000		
18	THINGS	369	0.18	1	20	0.04	66.65	79.11	2.30	0.0000000000		
19	WATCHING	224	0.11	1	4		65.64	78.09	3.90	0.0000000000		
20	WORLD	473	0.23	1	36	0.07	63.24	75.69	1.81	0.0000000000		
21	EVERYTHING	249	0.12	1	7	0.01	62.88	75.34	3.25	0.0000000000		
22	LOT	240	0.12	1	7	0.01	59.13	71.59	3.19	0.0000000000		
23	LOVE	238	0.12	1	7	0.01	58.30	70.76	3.18	0.0000000000		
24	APPEARANCE	165	0.08	1	1		56.45	68.91	5.46	0.0000000000		
25	ALWAYS	301	0.15	1	16	0.03	52.97	65.42	2.33	0.0000000000		
26	WATCH	169	0.08	1	3		46.55	59.01	3.91	0.0000000000		
27	CONFLICTS	141	0.07	1	1		45.42	57.87	5.23	0.0000000000		
28	TECHNOLOGY	186	0.09	1	5		44.76	57.22	3.31	0.0000000000		

KWs plot links clusters filenames source text notes

161 entries Row 1 0% T S tag file: 0 entries found

EN 11:26 AM 5/13/2021

Пример коришћења функције KeyWord (кључне разлике учесталости речи) (WordSmith Tools v.8): ICLE-SE: AE (2)

C:\Users\Jelena52\Desktop\2021 NAUKA I STVARNOST\SERBICLECOMPLETE vs AE.kws

N	Key word	Freq.	%	Texts	RC. Freq.	%	BIC	Log_L	Log_R	P	Lemmas	Set
134	BY	647	0.32	1	332	0.61	74.04	-86.50	-0.94	0.0000000000		
135	BLACK	13		1	46	0.09	74.81	-87.27	-3.73	0.0000000000		
136	RESEARCH	20		1	53	0.10	76.44	-88.89	-3.31	0.0000000000		
137	CLAIM	55	0.03	1	79	0.15	78.29	-90.75	-2.43	0.0000000000		
138	BIOLOGICAL	4		1	38	0.07	81.44	-93.89	-5.15	0.0000000000		
139	VALUES	36	0.02	1	68	0.13	82.31	-94.76	-2.82	0.0000000000		
140	REASONING	3		1	38	0.07	85.92	-98.37	-5.57	0.0000000000		
141	OF	5,191	2.56	1	1,825	3.37	86.08	-98.54	-0.40	0.0000000000		
142	SEX	10		1	51	0.09	96.77	-109.23	-4.26	0.0000000000		
143	RELIGIOUS	10		1	51	0.09	96.77	-109.23	-4.26	0.0000000000		
144	SUPPORT	42	0.02	1	79	0.15	97.34	-109.80	-2.82	0.0000000000		
145	PUNISHMENT	21	0.01	1	63	0.12	99.33	-111.78	-3.49	0.0000000000		
146	RELIGION	16		1	59	0.11	101.22	-113.68	-3.79	0.0000000000		
147	NUCLEAR	3		1	45	0.08	106.75	-119.21	-5.81	0.0000000000		
148	GENETIC	3		1	47	0.09	112.73	-125.19	-5.88	0.0000000000		
149	DRUG	8		1	55	0.10	114.77	-127.22	-4.69	0.0000000000		
150	SCHOOL	56	0.03	1	97	0.18	115.32	-127.78	-2.70	0.0000000000		
151	PENALTY	15		1	72	0.13	139.02	-151.48	-4.17	0.0000000000		
152	PUBLIC	40	0.02	1	107	0.20	167.81	-180.27	-3.33	0.0000000000		
153	EVIDENCE	8		1	75	0.14	172.42	-184.87	-5.14	0.0000000000		
154	THIS	988	0.49	1	562	1.04	175.86	-188.31	-1.09	0.0000000000		
155	SCHOOLS	28	0.01	1	105	0.19	191.10	-203.55	-3.81	0.0000000000		
156	STATES	16		1	93	0.17	194.00	-206.45	-4.45	0.0000000000		
157	SUICIDE	3		1	74	0.14	194.22	-206.67	-6.53	0.0000000000		
158	OPPONENTS	4		1	78	0.14	200.54	-213.00	-6.19	0.0000000000		
159	DEATH	35	0.02	1	133	0.25	246.62	-259.07	-3.83	0.0000000000		
160	ARGUMENT	42	0.02	1	181	0.33	355.67	-368.13	-4.01	0.0000000000		
161	THE	8,760	4.32	1	3,544	6.55	398.99	-411.44	-0.60	0.0000000000		

KWs plot links clusters filenames source text notes

161 entries Row 1 0% T S tag file: 0 entries found

EN 11:27 AM 5/13/2021

Варијабле говорника

- Матерњи језик: српски
- Држава: Босна и Херцеговина, Србија
- Узраст: 22.17
 - ICLEv3 22.27, 23
- Пол: 77.85%, 22.15%
 - ICLEv3 77.20% 22.38%
- Страни језици, осим енглеског
- Језици који се говоре код куће, осим матерњег
- Године учења енглеског језика: 0.62 (4), 99.38 (6+)
- Процена нивоа владања језиком (CEFR) : B2 4, C1 11, C2 5 (од укупно 20)
 - ICLEv3 194 217 89 (500)

Период проведен на енглеском говорном подручју

Пример снимка екрана: варијабле говорника 1

ICLE online

corpora.uclouvain.be/cecl/icle/

selected variables: Country × Native language × RESET VARIABLES

selected texts: 325 (202,621 words)

Native language ?

- ☐ Macedonian
- ☐ Norwegian
- ☐ Persian
- ☐ Polish
- ☐ Portuguese
- ☐ Punjabi
- ☐ Russian
- ☒ Serbian
- ☐ Spanish
- ☐ Swedish

Country ? select: all none

- ☒ Montenegro 2
- ☐ Netherlands
- ☐ Norway
- ☐ Pakistan
- ☐ Poland
- ☐ Romania
- ☐ Russia
- ☒ Serbia 179
- ☐ South Africa

Gender ? select: all none

Gender	Count
Female	253
Male	72
Unknown	0

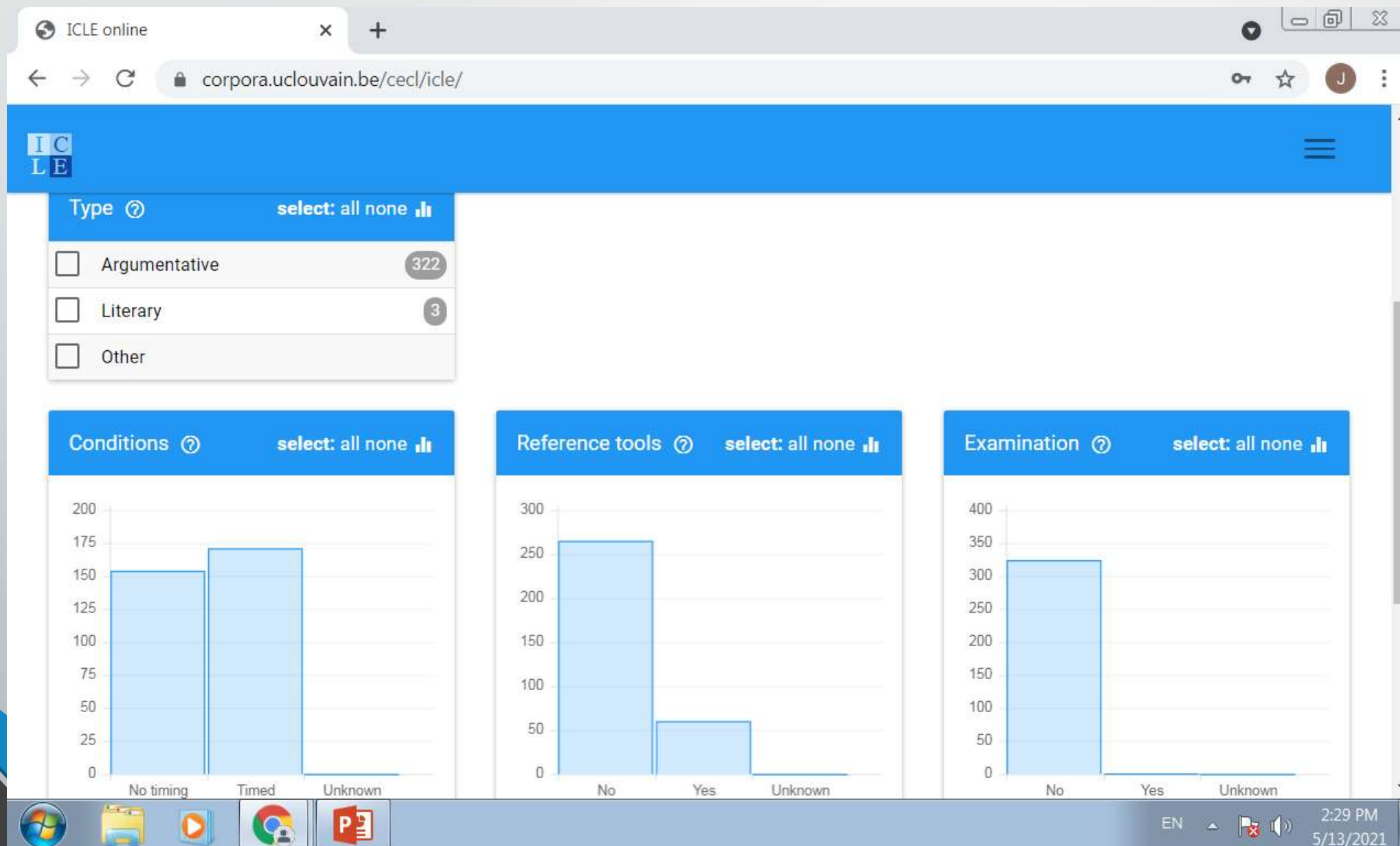
- ☐ Female 253
- ☐ Male 72

EN 2:17 PM 5/13/2021

Варијабле контекста језичке продукције

- Жанр: аргументативни састав/ књижевни састав: 99.08%/ 0.92%
 - ICLEv3 94.08%/ 3.82% 2.10%
- Дужина састава 623.45 речи (најкраћи 310, најдужи 1144 речи)
 - ICLEv3 605.16
- Временски ограничени/ временски неограничени: 47.38%/ 52.62%
- Коришћење референтних издања (не/да): 81.54%/18.46%
- Испитни услови/ ван испита: 0.31%/ 99.69%

Пример снимка екрана: варијабле говорне ситуације



Пример снимка экрана: simple search: things

The screenshot displays the ICLE online search interface in a web browser. The browser's address bar shows the URL `corpora.uclouvain.be/cecl/icle/`. The page features a blue header with the ICLE logo and a hamburger menu icon. Below the header, the 'selected variables' section includes 'Country' and 'Native language' filters, along with a 'RESET VARIABLES' button. A status indicator shows 'selected texts: 325 (202,621 words)'. The search mode is set to 'Simple search', and the search term 'things' is entered in the search bar. The results section shows 'Results: 367 hit(s)' and provides options to download concordance in .XLSX or .CSV format. A list of search results is displayed, with the word 'things' highlighted in blue in each snippet. The Windows taskbar at the bottom shows the system time as 4:24 PM on 5/13/2021.

ICLE online

corpora.uclouvain.be/cecl/icle/

selected variables: Country Native language RESET VARIABLES

selected texts: 325 (202,621 words)

Simple search Advanced search

things

things

Results: 367 hit(s) DOWNLOAD CONCORDANCE IN .XLSX (MAXIMUM: 10,000 HITS)

DOWNLOAD CONCORDANCE IN .CSV (MAXIMUM: 10,000 HITS)

1 SEBG1001 It cuts all unnecessary things out .

2 SEBG1002 Little do we know about each other , our emotions and things that affected us most profoundly .

3 SEBG1002 trust that person more than someone whose thoughts are just of video games useless things .

4 SEBG1002 they have different desires and goals , but they probably do not have many things in common .

5 SEBG1004 Nowadays , talking about sex , drugs , tattoos and many other things is not " forbidden " anymore .

EN 4:24 PM 5/13/2021

Доступни интегрисани корпусни алати

- Конкорданце
 - Једноставна претрага (*simple search*)
 - Напредна претрага (*advanced search*), са детаљним описом тагова у приручнику:
 - По нашој процени, највећа вредност ове верзије ученичког корпуса, осим обима, управо јесте морфосинтасичко тагирање корпуса које пружа бројне истраживачке могућности
 - Није увек и искључиво засновано на форми упита – комбиновање формалних и семантичких/функционалних карактеристика.

ADJ	JJ	General adjective
	JJR	General comparative adjective (e.g. older, better, stronger)
	JJT	General superlative adjective (e.g. oldest, best, strongest)
	JK	Catenative adjective (able in 'be able to', willing in 'be willing to')
ADV	RA	Adverb, after nominal head (e.g. else, galore)
	REX	Adverb introducing appositional constructions (namely, e.g.)
	RG	Degree adverb (very, so, too)
	RGQ	Wh- degree adverb (how)
	RGQV	Wh-ever degree adverb (however)
	RGR	Comparative degree adverb (more, less)
	RGT	Superlative degree adverb (most, least)
	RL	Locative adverb (e.g. alongside, forward)
	RP	Prep. adverb, particle (e.g. about, in)
	RPK	Prep. adv., catenative (about in 'be about to')
	RR	General adverb
	RRQ	Wh- general adverb (where, when, why, how)
	RRQV	Wh-ever general adverb (wherever, whenever)
	RRR	Comparative general adverb (e.g. better, longer)
	RRT	Superlative general adverb (e.g. best, longest)



N	ND1	Singular noun of direction (e.g. north, southeast)
	NN	Common noun, neutral for number (e.g. sheep, cod, headquarters)
	NN1	Singular common noun (e.g. book, girl)
	NN2	Plural common noun (e.g. books, girls)
	NNA	Following noun of title (e.g. M.A.)
	NNB	Preceding noun of title (e.g. Mr., Prof.)
	NNL1	Singular locative noun (e.g. Island, Street)
	NNL2	Plural locative noun (e.g. Islands, Streets)
	NNO	Numeral noun, neutral for number (e.g. dozen, hundred)
	NNO2	Numeral noun, plural (e.g. hundreds, thousands)
	NNT1	Temporal noun, singular (e.g. day, week, year)
	NNT2	Temporal noun, plural (e.g. days, weeks, years)
	NNU	Unit of measurement, neutral for number (e.g. in, cc)
	NNU1	Singular unit of measurement (e.g. inch, centimetre)
	NNU2	Plural unit of measurement (e.g. ins., feet)
	NPD1	Singular weekday noun (e.g. Sunday)
	NPD2	Plural weekday noun (e.g. Sundays)
	NPM1	Singular month noun (e.g. October)



Vdo	VDD	Did
	VDG	Doing
	VDI	Do, infinitive (I may do... To do...)
	VDN	Done
	VDZ	Does
Vhave	VH0	Have, base form (finite)
	VHD	Had (past tense)
	VHG	Having
	VHI	Have, infinitive
	VHN	Had (past participle)
	VHZ	Has
Vlex	VV0	Base form of lexical verb (e.g. give, work)
	VVD	Past tense of lexical verb (e.g. gave, worked)
	VVG	-Ing participle of lexical verb (e.g. giving, working)
	VVGK	-Ing participle catenative (going in 'be going to')
	VVI	Infinitive (e.g. to give... It will work...)
	VVN	Past participle of lexical verb (e.g. given, worked)
	VVNK	Past participle catenative (e.g. bound in 'be bound to')
	VVZ	-S form of lexical verb (e.g. gives, works)
Vmod	VM	Modal auxiliary (can, will, would, etc.)
	VMK	Modal catenative (ought, used)



Преглед досадашњих научних резултата

- Заменички облици, заменица *one* – веома значајне разлике између изворних и неизворних говорника када је у питању употреба заменичких облика првог и другог лица (**степен формалности дискурса**) (Radić Bojanić 2019, Марковић 2017, Марковић 2019)
- Коришћење глагола *take* као примера вишефункционалног глагола у савременом енглеском језику: најчешће функционалне улоге код изворних говорника међу ретким примерима код србофоних говорника (Марковић 2018)
- Концесивни конектори *though* и *however* у писању на енглеском језику (Марковић 2020)
- Интерактивни метадискурс у саставима србофоних говорника: неуједначена употреба појединих категорија интерактивног метадискурса (Radonja 2019)
 - Дисертација у изради под називом „Трансфер у фразеолошкој компетенцији у писању на енглеском језику као страном код србофоних говорника“ кандидата Срђана Шућура

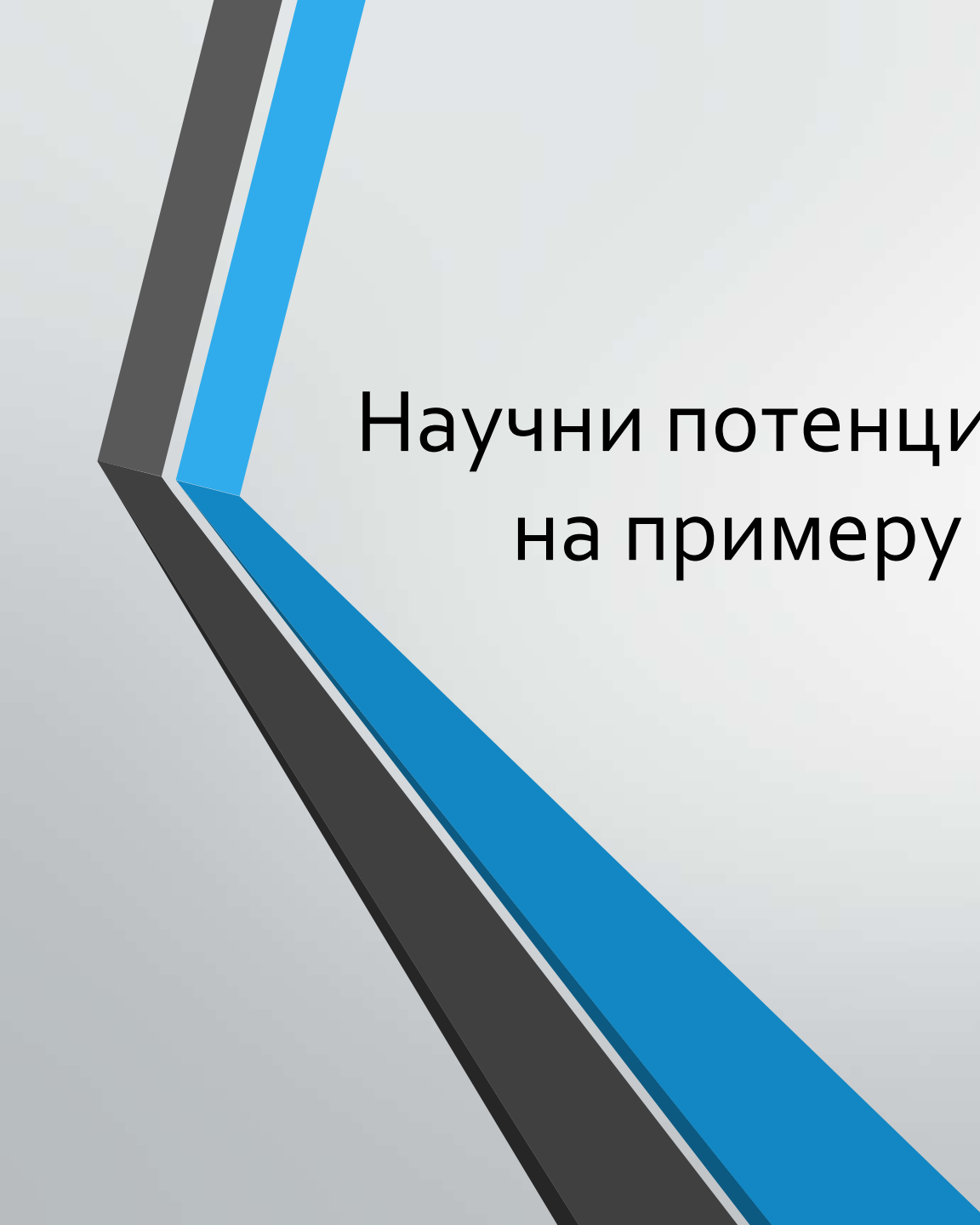
- Чарлс Филмор [Charles Fillmore] на симпозијуму о корпусној лингвистици у Стокхолму, у августу 1991. године, између осталог, навео је следеће:
 - I have two main observations to make. The first is that I don't think there can be any corpora, however large, that contain information about all of the areas of English lexicon and grammar that I want to explore; all that I have seen are inadequate. The second observation is that **every corpus that I've had a chance to examine, however small, has taught me facts that I couldn't imagine finding out about in any other way.**
(Fillmore, 1992, 35)

Корпус *ICLE-SE*: Научни потенцијали

- Корпусна лингвистика као методологија на српском говорном подручју у области примењене лингвистике користи се недовољно интензивно
 - Постојање корпуса састава писаних на енглеском као страном
 - Постојање референтног корпуса састава на српском као матерњем
 - Постојање паралелизованих корпуса превода са српског и са енглеског
- Могућности проширења корпуса
- Могућност интегрисања нових софтверских алатки
- Широк опсег могућих тема истраживања, првенствено многобројних и разнородних дискурских феномена

Одабрана библиографија

- Ädel, A. (2006). Metadiscourse in L1 and L2 English (Studies in Corpus Linguistics ed., Vol. 24). Amsterdam/Philadelphia: John Benjamins.
- Deshors, S. (2016). Inside phrasal verb constructions: A co-varying collexeme analysis of verb-particle combinations in EFL and their semantic associations. *International Review of Learner Corpus Research*, 2(1), 1-30.
- Deshors, S. C., Götz, S., & Laporte, S. (2016). Linguistic innovations in EFL and ESL: Rethinking the linguistic creativity of non-native English speakers. *International Journal of Learner Corpus Research*, 2(2), 131-150.
- Ellis, R. (1994). *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
- Gilquin, G., & Granger, S. (2011). From EFL to ESL: Evidence from the International Corpus of Learner English. In *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a paradigm gap* (pp. 55-78). Amsterdam / Philadelphia: John Benjamins.
- Granger, S. (2018). Formulaic Sequences in Learner Corpora: Collocations and Lexical Bundles. In A. Siyanova-Chanturia, & A. Pellicer-Sanchez (Eds.), *Understanding Formulaic Language: A Second Language Acquisition Perspective* (pp. 228-247). London and New York: Routledge.
- Granger, S., Dagneaux, E., & Meunier, F. (Eds.). (2002). *The International Corpus of Learner English. Handbook and CD-ROM*. Louvain-la-Neuve: Presses universitaires de Louvain.
- Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (Eds.). (2009). *The International Corpus of Learner English. Handbook and CD-ROM (Version 2)*. Louvain-la-Neuve: Presses universitaires de Louvain.
- Granger, S., Dupont, M., Meunier, F., Naets, H., & Paquot, M. (Eds.). (2020). *International Corpus of Learner English v3*. Louvain-la-Neuve: Presses universitaires de Louvain.
- Nesselhauf, N. (2005). Collocations in a Learner Corpus (Studies in Corpus Linguistics ed., Vol. 14). Amsterdam and Philadelphia: John Benjamins.
- Petch-Tyson, S. (1998). Writer/reader visibility in EFL written discourse. In S. Granger (Ed.), *Learner English on Computer*. London and New York: Longman.
- Radić Bojanić, B. (2019). Neodređena zamenica one u pisanju kod neizvornih govornika engleskog jezika. *Godišnjak Filozofskog fakulteta u Novom Sadu*, 44(2), 39-52.
- Radonja, M. (2019). The Use of Interactive Metadiscourse in Serbian Students' Writing in English. *Радови Филозофског факултета: часопис за хуманистичке и друштвене науке*, 21.
- Scott, M. (2018). *WordSmith Tools Manual. Version 7*. Stroud, Gloucestershire, UK: Lexical Analysis Software Ltd. Retrieved from www.lexically.net
- Sinclair, J. (1996). EAGLES. Preliminary recommendations on Corpus Typology. Retrieved November 5, 2018, from www.ilc.pi.it/EAGLES/corpus typ/corpus typ.html
- Sinclair, J. (2004). Introduction. In J. Sinclair (Ed.), *How to Use Corpora in Language Teaching* (Studies in Corpus Linguistics ed., Vol. 12, pp. 1-10). Amsterdam and Philadelphia: John Benjamins.
- Tomović, N., & Stefanović, S. (2019). UTICAJ L2 I LEKSIČKI I LEKSIČKO-SINTAKSIČKI KALKOVI U PREVODU. *STUDIJA SLUČAJA. Komunikacija I Kultura Online*, 10, 144-154. doi:doi:10.18485/kkonline.2019.10.10.8
- Марковић, Ј. (2017). Лични метадискурс у писању код неизворних и изворних говорника енглеског језика [Personal metadiscourse in writing of native and non-native speakers of English]. *Филолог*, VII(15), 44-60.
- Марковић, Ј. (2018). Употребе глагола make у писању на енглеском језику као страном код изворних говорника српског језика (корпуснолингвистичка анализа). *Зборник матице српске за филологију и лингвистику*, LXI(1), 165-180.
- Марковић, Ј. (2019). *Кроз призму контрастивне анализе међујезика*. Пале: Филозофски факултет.
- Марковић, Ј. (2020). Концесивни конектори though и however у писању на енглеском језику код изворних и неизворних говорника. *Филолог*, 21, 13-35.



Научни потенцијал ученичког корпуса *ICLE v3*
на примеру његове србофоне компоненте

ХВАЛА