Part 1
General information on ICLE

Thomas Zapounidis
Auth, February 10th 2023

What is ICLE ?

WHAT is ICLE

The abbreviation **ICLE** stands for

International **C**orpus of Learner **E**nglish

What is a learner corpus?

"Learner corpora are electronic collections of natural or near-natural foreign or second language learner texts assembled according to explicit design criteria". (Granger, 2017, p.429)

Natural or near-natural

The phrase *natural or near-natural* stresses the "need for data that reflects as closely as possible 'natural' language use (i.e., language that is situationally and interactionally authentic) while recognizing that the limitations facing the collection of such data often obligate researchers to resort to clinically elicited data (for example, by using pedagogic tasks)" *(Ellis and Barkhuizen 2005, p. 7).*

FL and SL learners

Learner corpora "can contain data from both foreign language (**FL**) learners, who learn a language in a country where they have **little exposure outside the classroom** (e.g., learning English in Germany or Japan), and second language (SL) learners, who acquire a language in a country where that language is the predominant language of communication (e.g., learning English in the United States)" *(Granger,2017,p.229).*

Texts

The term texts ''highlights the fact that learner corpora contain continuous stretches of oral or written discourse rather than decontextualized sentences. This makes it possible to study a much wider range of interlanguage features than in previous SLA studies, which have tended to focus on more local features like grammatical morphemes'' *(Granger, 2017,p. 229).*

Explicit criteria

"The requirement of explicit design criteria stems from the necessity to control the wide range of variables that affect learner language." (Granger, 2017,p. 229).

Table 1 *ICLE* design criteria

Learner variables	Task variables
Age	Medium
Learning context	Field
Proficiency level	Genre
Gender	Length
Mother tongue background	Topic
Region	Timing
Knowledge of other foreign languages	Exam
Amount of L2 exposure	Use of reference tools

Who is this Granger?

Sylviane Granger

►Professor Emerita of English Language and Linguistics (University of Louvain).

►Founder and Former Director of the Centre for English Corpus Linguistics.

►Pioneering work in the field of Learner Corpus Research

►She is known for her belief that learner language is best approached holistically and naturalistically.

► She is the one who launched the *International Corpus of Learner English* enterprise.

►Former President of the Learner Corpus Association.

In essence, ICLE is a corpus of written production by higher intermediate to advanced learners of English as a foreign language. The corpus (launched by the Centre for English Corpus Linguistics at the University of Louvain) includes a number of collaborating universities around the world (Granger et al., 2020)

According to its manual (Granger, Dupont, Meunier, Naets, and Paquot, 2020) the 3rd version of ICLE includes around **5 million words** with written production of English by learners with different L1 backgrounds (Brazilian, Bulgarian, Chinese, Czech, Dutch, Greek, Finnish, French, German, Hungarian, Italian, Japanese, Korean, Lithuanian, Macedonian, Norwegian, Pakistani, Persian (Iran), Polish, Portuguese, Russian, Serbian Spanish, Swedish, Turkish, and Tswana).

The platform offers rich **metadata** on each of the texts included in the corpus, pertaining to both the learners (e.g. mother tongue, age, time spent in an English-speaking country) and the writing tasks (e.g. topic, use of reference tools, conditions of production of the text).

Learners' age

NATIONAL SUBCORPUS	Average Age	Minimum Age	Maximum Age	Q1	Median	Q3	Without undefined	Number of undefined
BRAZILIAN	21.89	16	44	20	22	23	409	3
BULGARIAN	20.55	19	23	20	20	21	302	0
CHINESE	20.49	18	48	19	20	21	956	26
CZECH	22.07	20	29	21	22	23	239	4
DUTCH	20.75	18	43	19	20	21	257	6
FINNISH	22.73	18	49	20	22	23	385	5
FRENCH	21.70	20	38	21	21	22	344	3
GERMAN	23.39	18	36	22	23	25	414	23
GREEK	21.65	20	38	20	21	22	462	0
HUNGARIAN	23.57	19	41	21	22	24	382	54
IRANIAN	24.86	17	50	21	23	26	346	13
ITALIAN	24.59	18	42	23	24	26	382	10
JAPANESE	20.06	18	36	20	20	20	366	0
KOREAN	22.15	20	28	21	22	22	400	0
LITHUANIAN	21.76	20	33	21	22	22	375	34
MACEDONIAN	21.81	19	46	20	21	22	311	24
NORWEGIAN	23.94	18	55	20	22	25	287	30
PAKISTANI	19.79	18	28	19	20	20	306	0
POLISH	23.39	19	38	22	23	24	357	8
RUSSIAN	21.20	17	36	19	20	21	261	15
SERBIAN	22.17	18	38	21	22	23	325	0
SPANISH	21.72	17	54	19	21	22	243	8
SWEDISH	27.74	19	71	22	25	29.5	331	24
TSWANA	22.47	17	40	20	23	25	443	76
TURKISH	22.08	19	38	21	22	23	280	0
ICLEv3	22.27	16	71	20	21	23	9163	366

Taken from Granger, et al. (2020), p.23.

Learners' gender

NATIONAL SUBCORPUS	PERCENTAGE FEMALE	PERCENTAGE MALE	PERCENTAGE UNKNOWN
BRAZILIAN	61.89%	38.11%	
BULGARIAN	82.78%	17.22%	
CHINESE	63.95%	35.64%	0.41%
CZECH	72.02%	27.57%	0.41%
DUTCH	73.38%	26.62%	
FINNISH	85.13%	14.87%	
FRENCH	87.61%	12.39%	
GERMAN	78.03%	21.28%	0.69%
GREEK	95.67%	4.33%	
HUNGARIAN	78.67%	21.33%	
IRANIAN	73.82%	26.18%	
ITALIAN	91.84%	8.16%	
JAPANESE	72.95%	26.50%	0.55%
KOREAN	75.25%	23.75%	1.00%
LITHUANIAN	92.91%	7.09%	
MACEDONIAN	83.28%	16.72%	
NORWEGIAN	74.13%	25.87%	
PAKISTANI	65.69%	34.31%	
POLISH	80.00%	19.45%	0.55%
RUSSIAN	84.42%	15.22%	0.36%
SERBIAN	77.85%	22.15%	
SPANISH	86.45%	13.15%	0.40%
SWEDISH	76.62%	22.54%	0.85%
TSWANA	59.73%	36.61%	3.66%
TURKISH	81.43%	18.57%	
ICLEv3	77.20%	22.38%	0.42%

Taken from Granger, et al. (2020), p.24.

Learners' languages

FOREIGN LANGUAGE	NUMBER	PERCENTAGE
No other foreign language	3370	35.37%
German	1722	18.07%
French	1440	15.11%
Russian	578	6.07%
Spanish	452	4.74%
Italian	358	3.76%
Dutch	216	2.27%
Swedish	214	2.25%
Afrikaans	108	1.13%

NUMBER OF MONTHS	NUMBER OF LEARNERS	PERCENTAGE
0	5001	52.48%
1	600	6.30%
2	433	4.54%
3	280	2.94%
4	314	3.30%
5	95	1.00%
6	220	2.31%
7	51	0.54%
8	52	0.55%
9	63	0.66%
10	98	1.03%
11	43	0.45%
12	235	2.47%
13	11	0.12%
14	12	0.13%
15	31	0.33%
16	6	0.06%
17	11	0.12%
18	25	0.26%
19	1	0.01%
20	4	0.04%
21	1	0.01%
22	3	0.03%
23	2	0.02%
24	70	0.73%
More than 2 years	183	1.92%
Unknown	1684	17.67%

Taken from Granger, et al. (2020), pp.26-27.

Learners' CEFR level

MOTHER TONGUE	B2 (AND LOWER)	C1	C2	TOTAL
Brazilian Portuguese	8	9	3	20
Bulgarian	2	16	2	20
Chinese	19	1	0	20
Czech	11	9	0	20
Dutch	1	11	8	20
Finnish	3	8	9	20
French	3	11	6	20
German	1	12	7	20
Greek	6	13	1	20
Hungarian	2	13	5	20
Italian	10	9	1	20
Japanese	18	2	0	20
Korean	15	4	1	20
Lithuanian	7	8	5	20
Macedonian	4	4	12	20
Norwegian	8	7	5	20
Persian	7	11	2	20
Polish	1	12	7	20
Punjabi/Urdu	15	5	0	20
Russian	3	15	2	20
Serbian	4	11	5	20
Spanish	12	8	0	20
Swedish	0	14	6	20
Tswana	18	0	2	20
Turkish	16	4	0	20
ICLEv3	**194**	**217**	**89**	**500**

Taken from Granger, et al. (2020), p.30.

As the 3rd version of ICLE includes for the first time Greek students' contributions (GRICLE), we have been given free access to all contents and data of ICLE. The coordinators of the Greek sub-corpus are **Anna-Maria Hatzitheodorou** and **Marina Mattheoudakis**.

The Greek subcorpus comprises 462 essays for a total number of 264,260 words (see Table 24). All the essays are argumentative, and come from the same institution: Aristotle University of Thessaloniki.

BATCH	ESSAY CODES	TEXT TYPES			Number of ESSAYS	Number of WORDS
		A	L	O		
GRAR1	GRAR1001-1155	155	0	0	155	89,670
GRAR2	GRAR2001-2063	63	0	0	63	35,820
GRAR3	GRAR3001-3047	47	0	0	47	27,844
GRAR4	GRAR4001-4095	95	0	0	95	53,045
GRAR5	GRAR5001-5102	102	0	0	102	57,881
TOTAL		462	0	0	462	264,260

Table 24 taken from Granger, et al. (2020), p.41.

Module 2
WHY ICLEv3

It differs from previous ICLE versions

► **Larger** (5.7 million words) **and richer** (25 mother tongue backgrounds)

► **It is hosted on a web-based interface which allows** "not only for easier and more flexible access but also for the regular inclusion of new subcorpora as they are completed, thereby highlighting the fundamentally dynamic nature of the ICLE project" *(Granger et al., 2020, p.i)*

Version	1st	2nd	3rd
Year	2002	2009	2020
Size	2.5 million	3.7 million	5.7 million
Mother Tongue Backgrounds	11	16	25
Distribution	CD-ROM	CD-ROM	Web-based interface

Distribution of essays/words per national subcorpus

NATIONAL SUBCORPUS	Number of essays	Number of words	Average Length	Minimum Length	Maximum Length	Q1	Median	Q3
BRAZILIAN	412	206,024	500.06	206	976	439.75	511	562.25
BULGARIAN	302	201,925	668.63	216	2289	447	562.5	909.75
CHINESE	982	493,080	502.12	138	1151	436	502	553.75
CZECH	243	202,464	833.19	231	1514	591	844	1017.5
DUTCH	263	236,831	900.50	319	4239	532	639	1163.5
FINNISH	390	276,950	710.13	272	1797	558.25	651.5	824.25
FRENCH	347	226,896	653.88	220	2211	514	586	723
GERMAN	437	232,219	531.39	155	2101	350	486	665
GREEK	462	264,260	571.99	332	1058	522	553	608
HUNGARIAN	436	208,663	478.58	352	744	448	477	503
IRANIAN	359	227,287	633.11	228	1238	516.5	617	733.5
ITALIAN	392	226,043	576.64	189	1277	500	558	632
JAPANESE	366	200,958	549.07	399	1006	487	520	579.75
KOREAN	400	224,655	561.64	324	984	503.75	524.5	597
LITHUANIAN	409	236,027	577.08	212	1301	505	545	637
MACEDONIAN	335	218,789	653.10	467	1176	506	587	807
NORWEGIAN	317	213,701	674.14	321	1782	549	625	749
PAKISTANI	306	198,887	649.96	69	1851	535.5	598	710.75
POLISH	365	235,571	645.40	230	1109	539	626	727
RUSSIAN	276	230,045	833.50	107	3087	567.5	809	1068.75
SERBIAN	325	202,621	623.45	310	1144	521	605	715
SPANISH	251	199,948	796.61	224	3447	536	662	938.5
SWEDISH	355	201,467	567.51	235	1279	503	546	608.5
TSWANA	519	200,610	386.53	92	991	292.5	374	471.5
TURKISH	280	200,601	716.43	500	1423	580.75	716	826
ICLEv3	9529	5,766,522	605.16	233.41	69	4239	488	550

Table taken from Granger, et al. (2020), p.33.

It is POS tagged

ICLEv3 was POS tagged with the Constituent Likelihood Automatic Word-tagging System (**CLAWS**) C7 (Garside and Smith 1997)

Why is POS knowledge important?

"The POS-tagging of the learner essays makes it possible to search [(a)] **for all the occurrences of a lemma** (e.g. the lemma USE used as a noun), [(b)] a **POS-tag** (e.g. all the adverbs used in the corpus) or [(c)] **a sequence of POS tags** (e.g. a plural noun followed by a lexical verb) in ICLEv3's concordance" (Granger et al., 2020, p.20)

Concordancer included

'The ICLEv3 concordancer makes it possible to search for **word forms**, **lemmas, multiword units and part of-speech tags**. Concordances generated through the ICLEv3 interface **can be exported in a number of formats,** thus facilitating further analysis and treatment of the data outside the interface" (Granger, et al., 2020, p.22)

International project

► ICLE is an International and well known project

► Much research has been conducted using the ICLE

You can use it for comparison purposes

► with e.g. primary/secondary/tertiary writings
► with other corpora e.g. BNC

(cf. Gabrielatos, 2013 If-conditionals in ICLE and the BNC: A success story for teaching or learning?)

Do you know what CIA is?

Contrastive Interlanguage Analysis is a method of comparing :
 (a) a learner data with native speaker data (L2 vs L1) *rendering qualitative differences (misuse) and quantitative differences (over- and underuse)*
 (b) different types of learner data (L2 vs L2). *This comparison establishes whether the differences are developmental or transfer related.*

You can use it to examine academic written language in many ways

► According to Granger (2003a) the computer aided contrastive approach has enabled researchers "to uncover a wide range of **patterns of under-, over-, and misuse on learner lexis, (lexico-) grammar, and discourse**. Among the many topics that have been analyzed so far on the basis of ICLE data are high-frequency words, Romance words, recurrent combinations, collocations and formulae, prefabricated language, lexical profiling, lexical variation, adjective intensification, the verb make, progressives, passives, modality, noun phrase complexity, demonstratives, contractions, logical connectors, causal links, conjunctions, participle clauses, direct questions, tense errors, lexical errors, part-of-speech tagging and parsing"(p.242).

You can use it to examine academic written language in many ways

► Virtanen, T. (1997). The **progressive in NNS and NS** student compositions : Evidence from the International Corpus of Learner English.

► Svenfelt, M*. (1997). **Hedging in NS and NNS** Student Writing: A cross-cultural Study of Hedges in Argumentative Essays*

► Virtanen, T. (1998). ***Direct questions*** *in argumentative student writing.*

► Virtanen, T. & Lindgrén, S.A. (1998). ***British or American English?*** *Investigating what EFL students say and what they do.*

► Ringbom, H. (1998). **Vocabulary frequencies** in advanced learner English: A cross-linguistic approach.

► Ringbom, H. (1999). ***High-frequency verbs*** *in the ICLE corpus.*

► Agerström, J. (2000). **Hedges** in argumentative writing: A comparison of native and non-native speakers of English

You can use it to examine academic written language in many ways

► Bolton, K., Nelson, G., & Hung, J. (2002). *A corpus-based study of **connectors** in student writing : Research from the International Corpus of English in Hong Kong (ICE-HK).*

► Granger, S. (2003b). A multi-contrastive approach to the use of **link words** by advanced learners of English : Evidence from the International Corpus of Learner English

► Waibel, B. (2005). Corpus-based approaches to **learner interlanguage** : Case studies based on the International Corpus of Learner English

► Kobayashi, T. (2008*). Usage of **countable and uncountable nouns** by Japanese learners of English: two studies using the ICLE error-tagged Japanese sub-corpus.*

The studies presented do not constitute a comprehensive list but only serve as a sample of the kind of research that one can engage in by using the ICLE and its subcorpora.

Summary

The latest ICLE version

(a) Larger in size
(b) International with a variety of mother tongue backgrounds
(c) Web based
(d) Concordancer
(e) POS tagged
(f) Wide scope of research

LIMITATIONS

It only includes a set number of texts at a specific point of time.

Suggestions

► More DELEAR projects

► Why not a diachronic Tertiary Corpus of Learner English (AuThCLE)

Part 3
ICLEv3 HOW

3.1
Accessing ICLE

As mentioned in the 1st part due to the contribution of the Greek subcorpus (**GR**ICLE) any registered user of the Aristotle University of Thessaloniki is granted free access to the 3rd version of the ICLE.

Initially a set of snapshots will showcase the detailed steps while in the end an actual logging in will take place.

STEP 1

Use the following URL to access the ICLEv3 web interface:
https://corpora.uclouvain.be/cecl/icle/home

ICLE

About

International Corpus of Learner English

The **International Corpus of Learner English** is a **corpus of writing** by **higher intermediate to advanced learners of English as a foreign language**. The corpus collection was initiated by the *Centre for English Corpus Linguistics* at the *University of Louvain* and is the result of close collaboration with a large number of partner universities around the world.

A first version of the ICLE corpus appeared in 2002 and contained c. 2.5 million words produced by learners from 11 mother tongue backgrounds (L1). The second version, released in 2009, contained c. 3.7 million words of written texts representing sixteen language backgrounds. This interface gives access to the **third version** of the ICLE, which is larger than the previous versions in terms of both number of words (c. 5 million words) and mother tongue backgrounds. In addition to those already present in ICLEv2 (Bulgarian, Chinese, Czech, Dutch, Finnish, French, German, Italian, Japanese, Norwegian, Polish, Russian, Spanish, Swedish, Turkish, Tswana), ICLEv3 also contains English data produced by learners from the following L1s: Brazilian Portuguese, Greek, Hungarian, Persian (Iran), Korean, Lithuanian, Macedonian, Pakistani and Serbian.

The corpus offers **rich metadata** on each of the texts included in the corpus, pertaining to both the learners (e.g. mother tongue, age, time spent in an English-speaking country) and the writing tasks (e.g. topic, use of reference tools, conditions of production of the text)

Registered access

Login | Login

Password | Password

SIGN IN

Forgotten password — Forgotten login

OR

1 Institutional access (Shibboleth)

Login or password problem?

You enter your login and password above and the message *"You must enter a valid login and/or password (delete your browser cookies if the problem persists)"* appears, even though you are sure of your login and password?

This may be a problem with Google Chrome and all Chrome-based browsers, such as Microsoft Edge.

There are 3 solutions:

1. delete all cookies from your browser (procedure for Chrome, procedure for Edge);
2. use your browser's private browsing (procedure for Chrome, procedure for Edge);
3. use a browser such as Firefox.

Trial version and getting started videos

You can try the demo version which includes 100 texts: trial version

The getting started video offers a quick overview of the main functionalities afforded by the ICLE web interface.

STEP 3

Choose our institution
(*Aristotle University of Thessaloniki*)
from the drop down menu.

Fédération eduGAIN

🏛 Select your organisation

In order to access the service **corpora.uclouvain.be** please select or search the organisation you are affiliated with.

1

2 | Aristotle University of Thessaloniki | ✕ ▾ |

3 | Select |

☐ Remember selection for this web browser session.

🔑 FÉDÉRATION
par RENATER

Είσοδος Single Sign-On

Πρόσβαση **με ένα βήμα** σε όλες τις υπηρεσίες που χρησιμοποιούν ιδρυματικούς λογαριασμούς ΑΠΘ

Έχετε ζητήσει πρόσβαση στον ιστοχώρο corpora.uclouvain.be

username

password

Είσοδος

STEP 5

Finally, you should see the following picture which means that you are logged in and ready to use ICLEv3.

ICLE International Corpus of Learner English

The **International Corpus of Learner English** is a **corpus of writing** by **higher intermediate to advanced learners of English as a foreign language**. The corpus collection was initiated by the *Centre for English Corpus Linguistics* at the *University of Louvain* and is the result of close collaboration with a large number of partner universities around the world.

A first version of the ICLE corpus appeared in 2002 and contained c. 2.5 million words produced by learners from 11 mother tongue backgrounds (L1). The second version, released in 2009, contained c. 3.7 million words of written texts representing sixteen language backgrounds. This interface gives access to the **third version** of the ICLE, which is larger than the previous versions in terms of both number of words (c. 5 million words) and mother tongue backgrounds. In addition to those already present in ICLEv2 (Bulgarian, Chinese, Czech, Dutch, Finnish, French, German, Italian, Japanese, Norwegian, Polish, Russian, Spanish, Swedish, Turkish, Tswana), ICLEv3 also contains English data produced by learners from the following L1s: Brazilian Portuguese, Greek, Hungarian, Persian (Iran), Korean, Lithuanian, Macedonian, Pakistani and Serbian.

The corpus offers **rich metadata** on each of the texts included in the corpus, pertaining to both the learners (e.g. mother tongue, age, time spent in an English-speaking country) and the writing tasks (e.g. topic, use of reference tools, conditions of production of the text)

Getting started

The getting started video ↗ offers a quick overview of the main functionalities afforded by the ICLE web interface.

More details on ICLE are available in the ICLEv3 manual ↗.

International
Corpus of
Learner English

Version 3

Sylviane Granger, Maïté Dupont, Fanny Meunier,
Hubert Naets & Magali Paquot (eds)

PUL PRESSES
UNIVERSITAIRES
DE LOUVAIN

A pdf file of 277 pages

1. Description of the corpus (pages 3-22)
2. Project teams (pages 23-32)
3. ICLEv3 corpus breakdown (32-52)
4. ICLEv3 user manual (pages 53-81)
5. Status of English (pages 84-241)
6. Appendices (242-277)

Part 3.2
ICLEv3 Interface

International Corpus of Learner English ▣ Text selection▾   ⬇ Text download   ☰ Concordances   📖 Help ▾   ⓘ About ▾   👤 Account ▾

ICLE International Corpus of Learner English

The **International Corpus of Learner English** is a **corpus of writing** by **higher intermediate to advanced learners of English as a foreign language**. The corpus collection was initiated by the *Centre for English Corpus Linguistics* at the *University of Louvain* and is the result of close collaboration with a large number of partner universities around the world.

A first version of the ICLE corpus appeared in 2002 and contained c. 2.5 million words produced by learners from 11 mother tongue backgrounds (L1). The second version, released in 2009, contained c. 3.7 million words of written texts representing sixteen language backgrounds. This interface gives access to the **third version** of the ICLE, which is larger than the previous versions in terms of both number of words (c. 5 million words) and mother tongue backgrounds. In addition to those already present in ICLEv2 (Bulgarian, Chinese, Czech, Dutch, Finnish, French, German, Italian, Japanese, Norwegian, Polish, Russian, Spanish, Swedish, Turkish, Tswana), ICLEv3 also contains English data produced by learners from the following L1s: Brazilian Portuguese, Greek, Hungarian, Persian (Iran), Korean, Lithuanian, Macedonian, Pakistani and Serbian.

The corpus offers **rich metadata** on each of the texts included in the corpus, pertaining to both the learners (e.g. mother tongue, age, time spent in an English-speaking country) and the writing tasks (e.g. topic, use of reference tools, conditions of production of the text)

Getting started

1

The getting started video ↗ offers a quick overview of the main functionalities afforded by the ICLE web interface.
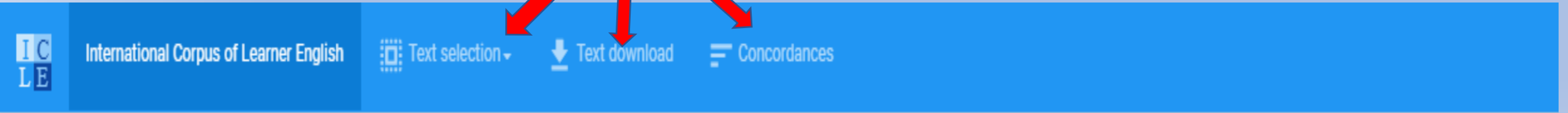
2

More details on ICLE are available in the ICLEv3 manual ↗.

There are 3 tabs in the welcome menu

International Corpus of Learner English

International Corpus of Learner English

Text selection ▾   Text download   Concordances

International Corpus of Learner English

The **International Corpus of Learner English** is a **corpus of writing** by **higher intermediate to advanced learners of English as a foreign language**. The corpus collection was initiated by the *Centre for English Corpus Linguistics* at the *University of Louvain* and is the result of close collaboration with a large number of partner universities around the world.

What do you think each tab is for?

TEXT SELECTION TAB

You select a text according to a set of learner and task variables

The drop down menu of the TEXT SELECTION tab renders 4 choices:
(a) Learner variables 1
(b) Learner variables 2
(c) Task variables
(d) Filtering

Learner variables 1 allows you to select the corpus with respect to the profile of the learner

selected variables: RESET VARIABLES   selected texts: 9435 (5,707,759 word

Native language ⑦	select: all none ⑊ ⇕
☐ Albanian	6
☐ Arabic	2
☐ Aromanian (Vlach)	1
☐ Bosnian	2
☐ Bulgarian	300
☐ Chinese	160
☐ Chinese-Cantonese	814
☐ Chinese-Mandarin	8
☐ Czech	241

Country ⑦	select: all none ⑊ ⇕
☐ Austria	70
☐ Belgium	473
☐ Bosnia Herzegovina	144
☐ Botswana	161
☐ Brazil	412
☐ Bulgaria	302
☐ China-Hong Kong	800
☐ China-Mainland	179
☐ Czech Republic	241

Gender ⑦	select: all none ⑊ ⇕
☐ Female	7285
☐ Male	2110
☐ Unknown	40

Age ⑦	reset ⑊
min:	16
max:	66

☐ Unknown (365)

1st other foreign language ⑦	select: all none ⑊ ⇕
☐ Afrikaans	108
☐ Albanian	1
☐ Arabic	14
☐ Aromanian (Vlach)	1
☐ Bulgarian	8
☐ Chinese	33
☐ Chinese-Mandarin	7
☐ Croatian	4

2nd other foreign language ⑦	select: all none ⑊ ⇕
☐ Afrikaans	38
☐ Albanian	1
☐ Ancient Greek	1
☐ Arabic	13
☐ Belorusian	1
☐ Bulgarian	14
☐ Chinese	8
☐ Chinese-Mandarin	2

3rd other foreign language ⑦	select: all none ⑊ ⇕
☐ Afrikaans	27
☐ Albanian	4
☐ Arabic	5
☐ Bosnian	1
☐ Bulgarian	7
☐ Chinese	1
☐ Croatian	10
☐ Czech	4

Learner variables 2 allows you to further refine your searches (e.g. years of exposure to English, institution)

International Corpus of Learner English | Text selection | Text download | Concordances | Help | About

selected variables: RESET VARIABLES

⊘ selected texts: 9435 (5,707,759 words)

Years of English at school ⊘ — select: all none

☐ 0	1
☐ 1	8
☐ 2	46
☐ 3	99
☐ 4	676
☐ 5	495
☐ 6 or more	7914
☐ Unknown	196

Years of English at university ⊘ — select: all none

☐ 0	1034
☐ 1	915
☐ 2	1867
☐ 3	2158
☐ 4	2319
☐ 5	663
☐ 6 or more	305
☐ Unknown	174

Months in English-speaking country ⊘ — select: all none

☐ 0	4926
☐ 1	600
☐ 2	433
☐ 3	280
☐ 4	314
☐ 5	95
☐ 6	219
☐ 7	51
☐ 8	52

Institution ⊘ — select: all none

☐ Austria - Paris-Lodron-Universität Salzburg	55
☐ Austria - Wirtschafts- und Fremdsprachenakademie Salzburg	24
☐ Belgium - Katholieke Vlaamse Hoogeschool, Antwerpen	166
☐ Belgium - Université catholique de Louvain	210
☐ Belgium - Université de Liège	63
☐ Belgium - Université libre de Bruxelles	74
☐ Bosnia and Herzegovina - University of Banja Luka, Faculty of Philology	73
☐ Bosnia and Herzegovina - University of East Sarajevo, Faculty of Philosophy	75
☐ Brazil - Fulminense Federal University	45

Task variables allows you to select your text on the basis of the text type, conditions use of reference tools and if it was part of examination or not.

selected variables: RESET VARIABLES   ⑦ selected texts: 9435 (5,707,759 words)

File name ⑦	
name:	File name

Title ⑦	reset
words:	Type the words here

Length in words ⑦	reset ⊞
min:	17
max:	4239

Type ⑦	select: all none ⊞
☐ Argumentative	8871
☐ Literary	364
☐ Other	200

Conditions ⑦	select: all none ⊞
☐ No timing	5700
☐ Timed	3379
☐ Unknown	356

Reference tools ⑦	select: all none ⊞
☐ No	4772
☐ Unknown	469
☐ Yes	4194

Examination ⑦	select: all none ⊞
☐ No	5894
☐ Unknown	898
☐ Yes	2643

The filtering page allows you to select the texts you want to include in your search and sort and filter the texts according to a range of variables

selected variables: File name ✕ RESET VARIABLES   ⊘ selected texts: 9529 (5,766,522 words)

This page allows you to select the texts that you want to include in your corpus search. It gives you the opportunity to sort and filter the texts in the corpus according to a range of variables and (de)select any texts individually.

Show variables: SELECT ALL VARIABLES DESELECT ALL VARIABLES

☑ Native language ☑ Country ☑ Gender ☑ Age ☐ 1st other foreign language ☐ 2nd other foreign language ☐ 3rd other foreign language ☐ 1st language at home ☐ 2nd language at home ☐ 3rd language at home ☐ Years of English at school ☐ Years of English at university ☐ Months in English-speaking country
☐ Institution ☐ Title ☐ Length in words ☐ Type ☐ Conditions ☐ Reference tools ☐ Examination ☐ Date (yyyy-mm-dd) ☐ Comments ⑦

⬚	⬥ File name	⬥ Native language	⬥ Country	⬥ Gender	⬥ Age
✓	BGSU1001	Bulgarian	Bulgaria	Female	20
✓	BGSU1002	Bulgarian	Bulgaria	Female	20
✓	BGSU1003	Bulgarian	Bulgaria	Female	20
✓	BGSU1004	Bulgarian	Bulgaria	Female	20
✓	BGSU1005	Bulgarian	Bulgaria	Female	21
✓	BGSU1006	Bulgarian	Bulgaria	Female	21
✓	BGSU1007	Bulgarian	Bulgaria	Female	21
✓	BGSU1008	Bulgarian	Bulgaria	Female	21
✓	BGSU1009	Bulgarian	Bulgaria	Female	21
✓	BGSU1010	Bulgarian	Bulgaria	Female	21
✓	BGSU1011	Bulgarian	Bulgaria	Male	20
✓	BGSU1012	Bulgarian	Bulgaria	Male	20

Dynamic display of selected text

The interface offers different types of dynamic display of the selected texts.

Firstly, all **grey numbers** that appear next to each selection criterion indicate the number of texts that correspond to that criterion in the corpus selection.

The total number of texts corresponding to the selection is also provided in the **top right-hand corner** of the page, together with the number of words that these selected texts represent.

Source: Granger et al. (2020). *The International Corpus of Learner English version 3*. Louvain-la-Neuve: Presses universitaires de Louvain,p.57

Dynamic display of selected text

When you first enter the text selection tab and select learner variables you will see on the top right area the total selected texts and words

International Corpus of Learner English | Text selection | Text download | Concordances | Help | About | Account

selected variables: File name ✕ | RESET VARIABLES | **3** | **1** | ⑦ selected texts: 9529 (5,766,522 words)

Native language ⑦	select: all none
☐ Albanian	6
☐ Arabic	2
☐ Aromanian (Vlach)	1
☐ Bosnian	2
☐ Bulgarian	300
☐ Chinese	160
☐ Chinese-Cantonese	814
☐ Chinese-Mandarin	8

Country ⑦	select: all none
☐ Austria	70
☐ Belgium	473
☐ Bosnia Herzegovina	144
☐ Botswana	161
☐ Brazil	412
☐ Bulgaria	302
☐ China-Hong Kong	800
☐ China-Mainland	179

Gender ⑦	select: all none
☐ Female	**2** 7356
☐ Male	2133
☐ Unknown	40

Age ⑦	reset
min:	16
max:	71

Once you have selected your text and variables you can download it in a zip file.

International Corpus of Learner English   ⬚ Text selection ▾   ⬇ **Text download**   ☰ Concordances

selected variables: Country ✕   Native language ✕   Gender ✕   File name ✕   RESET VARIABLES

Download the selected texts

This page allows you to download the texts matching the set of variables that you selected in the 'text selection' tab. The interface will generate a *.zip* fil

- one folder including separate *.txt* files for each text in your corpus;   **1**
- one file grouping all the texts in your corpus in a single *.txt* file;   **2**
- and a database containing the learner profile information (i.e. metadata) for each text in your corpus in *.csv* and *.xlsx* format   **3**

(Optional) Choose a file name:

my corpus test

🔄 **GENERATE ZIP FILE**

1 📁 texts

3 📊 metadata

📊 metadata

2 📄 subcorpus

Why is it useful to download the texts?

selected variables: Country × | Native language × | Gender × | File name × | RESET VARIABLES

Simple search   Advanced search

Enter your query

---

selected variables: Country × | Native language × | RESET VARIABLES   ⓘ selected texts: 298 (195,359 wo

Simple search   **Advanced search**

— | ➕   🗑 🔍 ⓘ ▮▮▮ ▮▮▮ ▮▮▮ ›

Form (case insensitive) ⓘ	equals ▾	Form (case insensitive)
Form (case sensitive) ⓘ	equals ▾	Form (case sensitive)
Lemma ⓘ	equals ▾	Lemma
Part of speech ⓘ	equals ▾	Part of speech
Simplified Part of speech ⓘ	equals ▾	Simplified part of speech
Any word ⓘ	☐

← MOVE TO THE LEFT   → MOVE TO THE RIGHT   🗑 DELETE THIS WORD   🗑 DELETE THE QUERY   🔍 SEARCH

On the top right corner of the CONCORDANCE tab you will see the following icons

selected texts: 9529 (5,766,522 words)
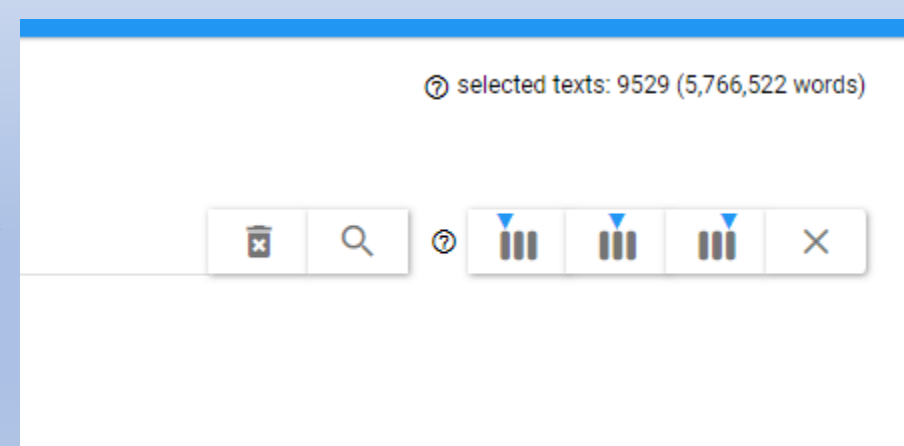
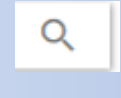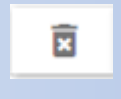| Icon	Description
(sort left)	Sort the concordance lines according the words that appear **to the left of the search string**
(sort middle)	Sort the concordance lines according the **search string**
(sort right)	Sort the concordance lines according the words that appear **to the right of the search string**
×	Restore the default order (sort according to texts)
(search)	Perform the search
(trash)	Restore the search
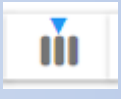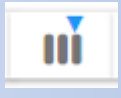
Simple search   **Advanced search**

+

🗑 🔍 ⑦ ⬛ ⬛ ⬛

Form (case insensitive) ⑦	equals ▾	Form (case insensitive)
Form (case sensitive) ⑦	equals ▾	Form (case sensitive)
Lemma ⑦	equals ▾	Lemma
Part of speech ⑦	equals ▾	Part of speech
Simplified Part of speech ⑦	equals ▾	Simplified part of speech
Any word ⑦		☐

← MOVE TO THE LEFT   → MOVE TO THE RIGHT   🗑 DELETE THIS WORD   🗑 DELETE THE QUERY   🔍 SEARCH

Simplified tag	CLAWS7 tag	
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)
	RT	Quasi-nominal adverb of time (e.g. now, tomorrow)
CCO	CC	Coordinating conjunction (e.g. and, or)
	CCB	Adversative coordinating conjunction (but)

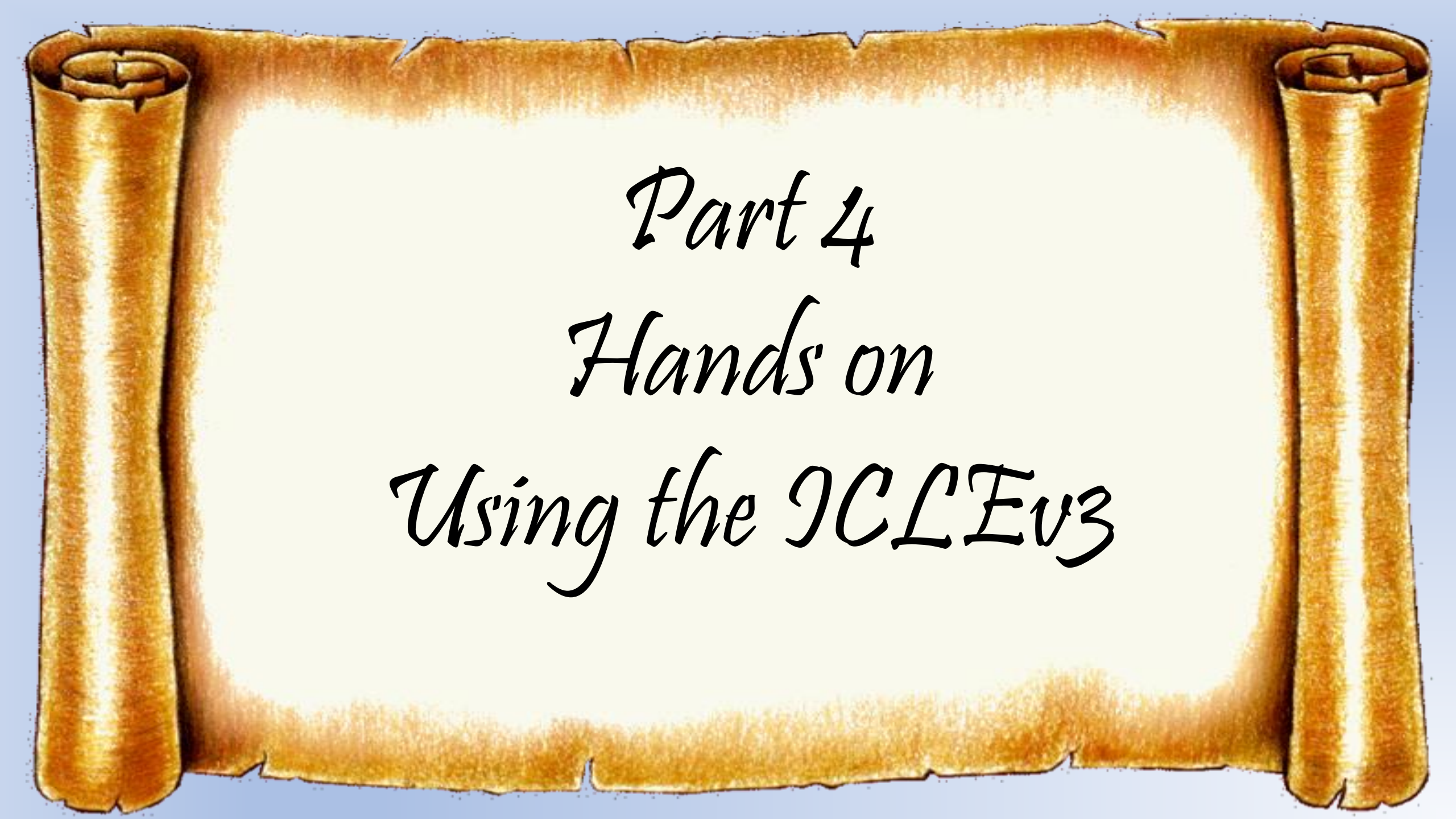|	VB0	Be, base form (finite, i.e. imperative, subjunctive)
Vbe	VBDR	Were
	VBDZ	Was
	VBG	Being
	VBI	Be, infinitive (To be or not... It will be …)
	VBM	Am
	VBN	Been
	VBR	Are
	VBZ	Is
Vdo	VD0	Do, base form (finite)
	VDD	Did
	VDG	Doing
	VDI	Do, infinitive (I may do... To do...)
	VDN	Done
	VDZ	Does
Vhave	VH0	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)

The corpus annotation procedure for CLAWS7 necessitated the separation of words containing an apostrophe by means of a space.

This means:

-Words with **apostrophes** e.g. "I'm" will be found as I 'm (a space before the apostrophe)

- If you are interested in searching for **Negative contacted forms** (can't, shouldn't, won't, etc.) you should search them considering the negation as n't so ca n't, should n't, wo n't)

- **Agglutinated forms** (cannot) may be searched by entering the non-agglutinated form (can not)

In part 3 we...

► learned how to log in

► familiarized ourselves with the three main tabs of the interface

Part 4
Hands on
Using the ICLEv3

Assignment 1
Selecting particular text and using advanced search in concordance

International Corpus of Learner English

The International Corpus of Learner English

Assignment 1

1. Log in to the ICLEv3 platform through your institutional account

2. Select the texts of Greek students

3. Concordance the word *'become'* in the simple search. How many hits did it return?

4. Using the advanced search how many adverbs precede and how many follow the word *'become'*?

SEARCH ITEM	HITS
become	
Lemma of become	
Lemma of Become + ADV	
ADV + lemma of become	

Suppose you wish to examine the use of a particular word (e.g. "**have**") in the essays of **male and female** learners with **Czech as L1 background** with no other influence from other languages or even further exposure (e.g. stay in English speaking country)

(a) You wish to examine differences between female and male learners

(b) You wish to exclude any other possible interference through the knowledge of other foreign languages.

Suppose you wish to examine the texts written by **Greek male learners** with other software (e.g. AntConc).
Then you may download the selected text in a number of formats and further analyze them. Let's see how you can do this.
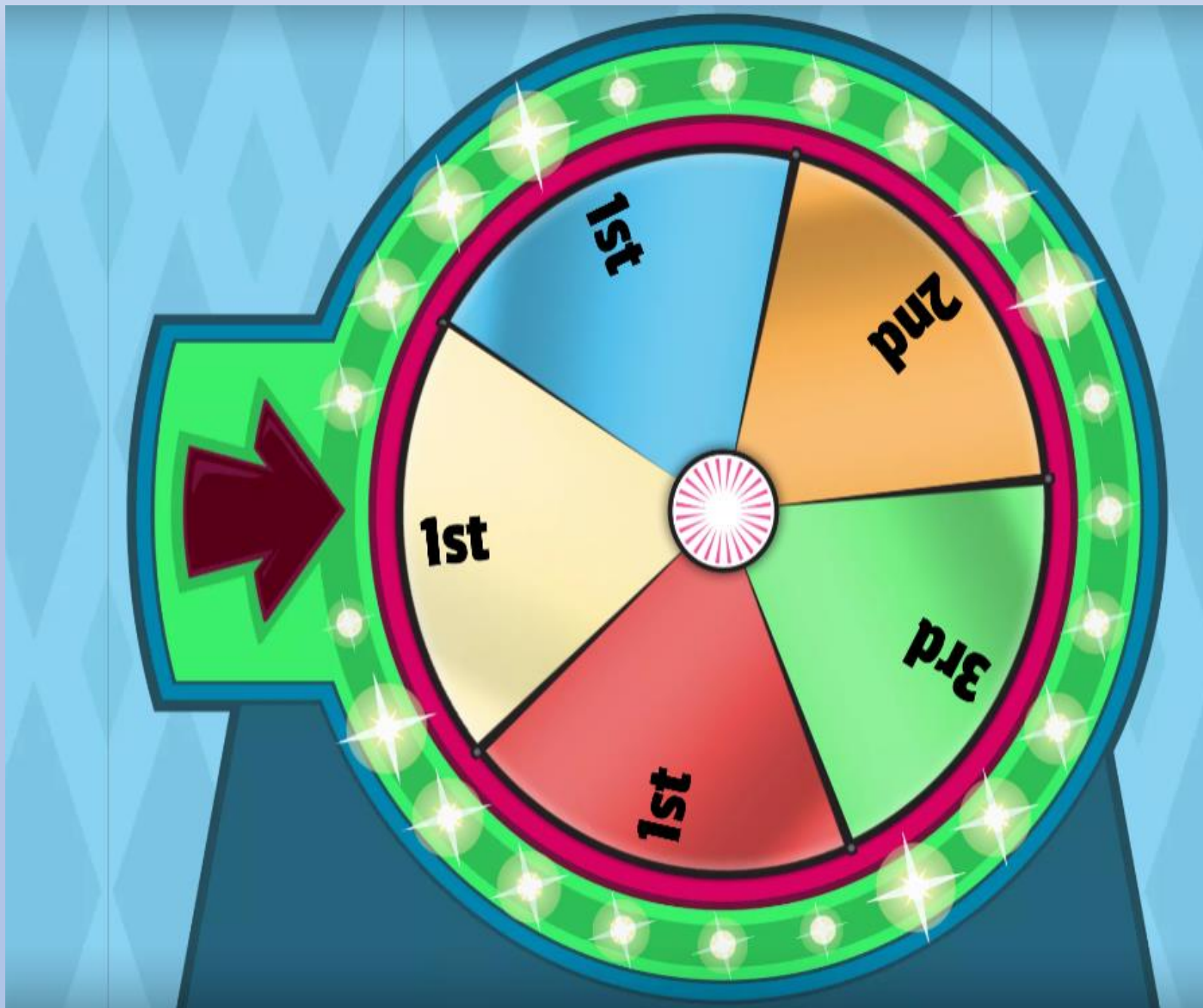
1st
triple chance of winning

2nd

3rd

https://wordwall.net/resource/52432680

ONLY 1 WINNER

Hands on activities through an ESCAPE ROOM game

https://view.genial.ly/607b277a33338f0d2474db30/interactive-content-escape-from-308b-lab

15:00

https://wordwall.net/resource/52432680

THANK YOU !!

Agerström, J. (2000). Hedges in argumentative writing: A comparison of native and non-native speakers of English. In Tuija Virtanen & J. Agerström (Éds.), *Three studies of learner discourse. Evidence from the International Corpus of Learner English* (Vol. 10, p. 5-42). Rapporter Från Växjö Universitet.

Bolton, K., Nelson, G., & Hung, J. (2002). A corpus-based study of connectors in student writing : Research from the International Corpus of English in Hong Kong (ICE-HK). *International Journal of Corpus Linguistics*, *7*(2), 165-182. https://doi.org/10.1075/ijcl.7.2.02bol

Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language.* Oxford: Oxford University Press

Gabrielatos, C. (2013). If-conditionals in ICLE and the BNC: A success story for teaching or learning? In S. Granger, G. Gilquin & F. Meunier (eds) Twenty Years of Learner Corpus Research: Looking back, Moving ahead. Corpora and Language in Use – Proceedings 1, Louvain-la-Neuve: Presses universitaires de Louvain, 155-166.

Granger, Sylviane. (2003a). *A multi-contrastive approach to the use of link words by advanced learners of English : Evidence from the International Corpus of Learner English*. Pragmatic Markers in Contrast, Brussels. http://hdl.handle.net/2078.1/75895

Granger, S. (2003b). The International Corpus of Learner English:A New Resource for Foreign Language Learning and Teaching and Second Language Acquisition Research. *TESOL Quarterly, 37*(3), 538-546.

Granger, S. (2017). *Learner corpora in foreign language education*. In Thorne, S. & May, S.(Eds.), Language and Technology. Encyclopedia of Language and Education. Springer International Publishing p. 427-440

Granger, S., Dupont, M., Meunier, F., Naets, H., & Paquot, M. (2020). *The International Corpus of Learner English.* Version 3. Presses universitaires de Louvain : Louvain-la-Neuve.

Kobayashi, T. (2008). Usage of countable and uncountable nouns by Japanese learners of English : Two studies using the ICLE error-tagged Japanese sub-corpus. Gakuen, 816(10), 73-82.

Ringbom, H. (1998). Vocabulary frequencies in advanced learner English: A cross-linguistic approach'. In: Sylviane Granger (ed.) *Learner English on Computer.* London: Longman: 41-52

Ringbom, H. (1999). High-frequency verbs in the ICLE corpus. In Antoinette Renouf (ed.), *Explorations in Corpus Linguistics*. Amsterdam: Rodopi: 191-200.

Svenfelt, M. (1997). Hedging in NS and NNS Student Writing: A cross-cultural Study of Hedges in Argumentative Essays. Unpublished MA thesis.

Virtanen, Tuija. (1997). The progressive in NNS and NS student compositions : Evidence from the International Corpus of Learner English. In M. Ljung (Éd.), *Corpus-based Studies in English. Papers from the seventeenth International Conference on English Language Research on Computerized Corpora (ICAME 17), Stockholm, May 15-19, 1996* (p. 299-309). Rodopi. https://research.abo.fi/converis/portal/Publication/12685894?auxfun=&lang=en_GB

Virtanen, T. (1998). Direct questions in argumentative student writing. In: Sylviane Granger (ed.) *Learner English on Computer*. London: Longman: 94-118.

Virtanen, T. & Lindgrén, S.A. (1998). British or American English? Investigating what EFL students say and what they do. In: Hans Lindquist, Staffan Klintborg, Magnus Levin & Maria Estling (eds.) *The Major Varieties of English. Papers from MAVEN 97*. Växjö: Växjö University: 273-281

Waibel, B. (2005). Corpus-based approaches to learner interlanguage : Case studies based on the International Corpus of Learner English. *Arbeiten aus Anglistik und Amerikanistik*, *30*(1-2), 143-176