Seminar 7

Transforming language teaching with corpora – Bridging research and practice

Dr. Reka R. Jablonkai
Assistant Professor in Education and Applied Linguistics
(University of Bath, UK)

THE UNIVERSITY OF QUEENSLAND

International Perspectives on Corpus Technology for Language Learning

unesp

CAPES

Outline

- Corpus revolution in language teaching
- Approaches
- Teaching and research gap
- Principles for course design

Corpus revolution? Transformed language teaching?

- corpus revolution (e.g. Conrad, 2000)
- gradual transformative change in language teaching (Leech, 1997)

- There still seems to be a gap between corpus research and teaching practices using corpora (Chambers, 2019).

Bridging the gap between research and practice

Corpora and language teaching

- Main approaches

(1) corpus-informed language teaching,

(2) integrated corpus-supported teaching and learning,

(3) self-directed DDL.

Corpus-informed language teaching

- Researchers and materials writers access corpora directly
- make pedagogical decisions regarding the contents and sequencing of language teaching syllabi and materials based on the results of corpus analysis.

(Römer, 2011)

Corpus-informed language teaching

- great increase since the 1990s
- Indirect use of corpora

description of language

- reference works – dictionaries, grammar books
- e.g. Cobuild, Longman

syllabus design and materials development, evaluation of textbooks

- collocations, lexico-grammar
- E.g. Touchstone (CUP), Grammar and Beyond (Reppen, 2012)

language testing

- e.g. validate tests, generate test items, etc.

learner corpora

- analysis of interlanguage
- ICLE – International Corpus of Learner English

(Jablonkai & Csomay, 2022; McEnery, & Xiao, 2011; Römer, 2011)

Corpus-informed materials design in ESP

- **Corpus design and corpus building**

Needs analysis

- a) learner needs are a fundamental element in ESP course and materials design;
- b) subject-specialists can inform the corpus design process about the range and relevance of particular text categories, genres

English EU Discourse corpus

- 1,174,753 running words
- 241 texts published by several EU institutions
- 40 different genres
- whole texts
- balanced for 34 subject fields of EU activities

Jablonkai, 2020

Lexical analysis for ESP pedagogic purposes

- Word list (e.g. Nation, 2016; Coxhead, 2018)
- collocation list (e.g. Green & Lambert, 2018, 2019).
- lexical bundles (e.g. Biber, 2006; Cortes, 2013; Wright, 2019)

Word lists in language teaching

- General Service List of English Words (GSL) (West, 1953)
- analysis of general and specialised corpora resulted in word lists
- Academic Word List (Coxhead, 2000)
- New General Service List (Brezina & Gablasova, 2015)
- New Academic Vocabulary List (AVL) (Gardner & Davies, 2014)
- subject-specific academic word lists analysing textbooks and research articles: **business** (Hsu, 2011; Konstantakis, 2007), **agriculture** (Martínez, Beck, & Panza, 2009), **engineering** (Hsu, 2014; Ward, 2009), **medicine** (Hsu, 2013; Wang et al., 2008; Yang, 2015), **the environment** (Liu & Han, 2015) and **applied linguistics** (Vongpumivitch, Huang, & Chang, 2009).
- Student Engineering Word List (SEWL) 1260 word families
- Chemistry Academic Word List (Valipouri and Nassaji, 2013) of 1400 word families
- Medical Academic Word List (MAWL) of 623 word families

EU Word list

- 405 word families
- legal words: *REGULATE, TREATY*
- words in connection with funding: *FUND, RESOURCE*
- the main EU institutions: *COMMISSION, PARLIAMENT* and *PRESIDENCY*
- Abbreviations: *DG, EC, OJ, SME*
- geographical names, which include all member states and names of two cities: *BRUSSELS* and *LISBON*
- function words such as *PRIOR, BEHALF* and *VIA*
- *Specificity:*

Texts	Tokens	Text coverage
EEUD Corpus	1,076,460	14.06%
EU English Corpus	197,620	13.10%
20th century literary texts	105,578	0.88%
News texts	117,164	4.76%

Jablonkai, 2020, p. 13

EU collocation list

CRITERION noun	
BNC Written	EEUD
grammatical relation: object of	
Semantic set 1 meet collocates: *satisfy, fulfil, meet, match, fit* Semantic set 2 set collocates: *formulate, adopt, outline, define, establish* Semantic set 3 respect collocates: - Semantic set 4 list collocates: *list, specify* Semantic set 5 apply collocates: *apply, use, employ* Semantic set 6 evaluate collocates: *assess, judge, review*	Semantic set 1 meet collocates: *fulfil, fulfill, meet, satisfy* Semantic set 2 set collocates: *set, agree, establish, lay* Semantic set 3 respect collocates: *follow, respect* Semantic set 4 list collocates: *list, specify, give* Semantic set 5 apply collocates: *apply* Semantic set 6 evaluate collocates: -
Other collocates: *invoke, exemplify, propose, interpret, identify, derive, alter*	Other collocates: *need, see, base, propose*
Number of preferential semantic sets	
5	5
Number of identical preferential semantic sets: 4	

criterion (n. 370)
eligibility (21/11.42)
award (18/11.22)
selection (19/11.16)
set (34/10.46)
follow (23/9.34)

Jablonkai, 2020, p. 14; 15

Presentation of collocational profile

IMPLEMENT verb	
construction	**semantic groups**
IMPLEMENT + noun	1. legislation collocates: *measure, rule, regulation, provision, directive, legislation, recommendation, decision, convention* The Commission shall **implement** this Regulation in accordance with the Financial Regulation. 2. plans collocates: *reform, strategy, programme, project, policy, commitment, budget, plan* Many European policies and programmes are **implemented** at regional and local levels. 3. approach collocates: *approach, principle* The forthcoming proposal for a new Directive **implementing** the principle of equal treatment outside employment will be addressed. 4. activity collocates: *action, tool, operation* By way of derogation from paragraph 1 , in-kind contributions , depreciation costs and overheads may be treated as expenditure paid by beneficiaries in **implementing** operations under the following conditions:
IMPLEMENT + adverb	1. positive collocates: *properly, effectively, fully, successfully, actively* The Commission, in its role of guardian of the Treaty, is responsible for ensuring that Community legislation is properly transposed into national law and properly **implemented** and enforced by national authorities in the Member States. 2. negative *not* However, Albanian legislation does not yet protect these rights sufficiently and is not fully **implemented**. Other collocates: *systematically, as, directly*

Table A1. Extract from the collocational profile of the verb IMPLEMENT

(Jablonkai, 2020, p. 17)

Activity type

Aim: to raise learners' awareness of collocates of particular lexical items

A.1 Instruction: Study the collocational profile of the verb IMPLEMENT underline the nouns in the table that are likely to be used with it

the accession criteria	function
opinion	the acquis
a reform	measures
the internal market	a directive
legislation	a summit
a timetable	policies
a programme	a debate

Key to Task type A.1:

the accession criteria	function
opinion	the acquis
a reform	**measures**
the internal market	**a directive**
legislation	a summit
a timetable	**policies**
a programme	a debate

(Jablonkai, 2020, p. 17)

Lexical bundles in the EEUD Corpus

- Most frequent ones: *in accordance with the, of the European Union, referred to in article*
- Least frequent ones: *of the internal market, the procedure referred to* and *within the scope of*
- Lexical bundles are very frequent in the EEUD Corpus – indicates a formulaic language use
- More lexical bundle types than in other registers
- Four-word lexical bundles are frequently part of longer word combinations
  - E.g. *in accordance with the*
  - *in accordance with the principle of proportionality*

Corpus research findings

Klett

3 Collocations

Decide which verbs and nouns can go together. Sometimes the same verb can combine with several nouns and likewise, the same noun may go with more than one verb. Write your solutions in the table below.

Verbs
check • devise • discuss • draft • draw up
enforce • exercise • fulfil • make • manage
make • implement • prepare • represent
serve • take

Nouns
decision • executive • power
initiative • interest • law
policy • responsibility

- What can you both make and take? _____
- What can you draft, devise, check, discuss or draw up? _____
- What can you represent and serve? _____
- What can you manage and implement? _____

possible verb + noun combinations	

focus on the most important collocations and word partnerships in EU English

a data-base of over one million words of English language EU texts was used to design the vocabulary tasks in this book

5 Study the concordance sample for *white paper* and *green paper*.

List the verbs that follow them using the space below. Can you think of other verbs that you could use to talk about how a written document introduces an idea?

concordance lists help you study the words in context

1	and freedom of expression,	**the White Paper**	proposes drawing up a European charter
2	exchanging best practice,	**the White Paper**	puts forward two options: setting up a
3	public opinion. Lastly,	**the White Paper**	again underlines the need to 'do the job
4	as part of its review of	**the 2001 White Paper**	on transport, in the form of a
5	standards for consultation).	**The White Paper**	aims to involve the media more effectively
6	subscribe to on a voluntary basis.	**The White Paper**	speaks of 'empowering citizens', by
7	Jacques Delors, published	**a White Paper**	setting out a timetable for completing the
8	Jacques Delors, published	**a White Paper**	seeking to abolish, within seven years,
9	monetary claims in Europe.	**The Green Paper**	describes the problems involved in the
10	conducted on the basis of	**a Green Paper**	put out in May. The June European Council
11	the publication in March of	**a Green Paper**	entitled 'Confronting demographic change

Integrated corpus-supported teaching and learning

Origins

- Tim Johns 1990s Classroom concordancing
- Learners as researchers
- Autonomous learning

Theoretical underpinning

- Noticing hypothesis (Schmidt, 1990)
- Discovery learning
- corpus application in language teaching, DDL (Boulton & Cobb, 2017; Johns,1991)
- corpora enhance learners' language awareness (Chen & Flowerdew, 2018; Römer, 2009), confidence and autonomous learning (Yoon, 2008)

Direct use of corpus data and corpora

to submit, a grant application for this project to	a European institution	or to another community body?
previously obtained directly or indirectly from	a European institution	or Community body during the
Financial support obtained directly or indirectly from	a European Institution	or Community body during the
to submit, a grant application for this project to	a European institution	or to another community body?
Financial support obtained directly or indirectly from	a European Institution	or Community body during the
instruments to ensure the rapid development of	a European internal	market for energy-efficient goods
as a priority matter in the establishment of	a European judicial-	area. The programme of measures
crew would serve to take account of the objective of	a European labour	market for cabin crew. Amendment
integrated space science programme and producing	a European launcher.	The national agencies of certain
numerous technical provisions regarding safety at	a European level.	These comprise, in addition to
national interests and aims to develop synergies at	a European level;	- The way cooperation and
of transport. Actions of this kind are innovative at	a European level	in terms of logistics, equipment,
measures, these lifetimes will be harmonised on	a European level.	The Commission, assisted by the
initiative to date has stimulated the development of	a European-level	world-class institution with the aim of
of the codecision procedure. (e) LRIT: Setting-up of	a European Long	Range Identification and Tracking of
11 (e) LRIT: Setting-up of	a European Long	Range Identification and Tracking of
and of the Council of 27 June 2002 establishing	a European Maritime	Safety Agency , and in particular
for a revision of the Regulation establishing	a European Maritime	Safety Agency and on a
a European Maritime Safety Agency and on	a "European Maritime	Transport Space without
of the various national market organisations; (c)	a European market	organisation. 2. The common
for human and veterinary use and establishing	a European Medicines	Agency already lay down a
competent authorities, the possibility of setting up of	a European Medicines	Information System for patients
of the Community's external assistance, in particular	a European Neighbourhood	and Partnership
Year's actions; 7. the establishment of Equinet,	a European network	of national equality bodies, which
those involving further training, in order to develop	a European network	which can receive Community
either by an individual European operator or by	a European network	of national public or private

- **Data-driven learning (DDL)**
- learners discover patterns in language
- concordance lines
- KWIC – key word in context
- teacher-directed or learner-led

(Warren, 2016)

Jablonkai, 2011

Chain of activities

Research – Practise – Improvisation (Johns, 1991)

Learners

- LOOK at concordance lines of the new vocabulary item and words surrounding it, thinking of meaning
- FAMILIARISE themselves with the patterns of language surrounding the word by referring to the concordances as you complete the tasks
- PRACTISE words without referring to concordances
- CREATE their own piece of writing using the words studied to fulfil particular functions in writing, speaking

(Thurstun & Candlin, 1998)

Activity type
LOOK

of this Regulation and every four years thereafter, **the Commission shall** report to the European Parliament
level experts appointed by the Commission. 3. **The Commission shall** have regard to the balance
of all denunciations of such conventions, which **the Commission shall** publish in the Official Journal of the
contractual obligations to which they are Parties. **The Commission shall** publish the list in the Official
that, while selecting the projects to be co-funded, **the Commission shall** have special regard to
will therefore investigate this finding further. 23. **The Commission shares** the Court's view that steady
calculating minimum requirements. Nevertheless, **the Commission shares** the Court's view that Member
into account the low value exports. Nevertheless, **the Commission shares** the Court's view that Member
," said Commissioner Reding. "On this solid basis, **the Commission should** be able to present a legislative
and film industries have expressed the view that **the Commission should** be prepared to take legislative
the principle that major policy proposals by **the Commission should** be accompanied by its appraisal
no longer includes nationals of all Member States, **the Commission should** pay particular attention to the
in relations with all Member States. Accordingly, **the Commission should** liaise closely with all Member
States. The Conference also considers that **the Commission should** take all the necessary measures
undertakings made in this respect shall be sent to **the Commission should** it so request. ARTICLE II. 4 –

training programmes, surveys and studies which **were implemented throughout** the EU to contribute to
by the Member States and the Commission **when implementing and** operating electronic customs
activities performed by the Commission **when implementing and** operating electronic customs
should follow the principles of this Declaration **when implementing bilateral** development aid programmes,
and procedures concerned will be voluntary; (b) **when implementing the** OMC, special attention will be paid
training programmes shall be taken into account **when implementing the** Programme. 4. Member States
and the ownership of secondary rights. (50) **When implementing the** provisions of Article 4 of Directive
be a primary consideration for Member States **when implementing the** provisions of this Directive that
for the operational programme carried out **when implementing the** communication plan; (b) the
clarity and legal security for better regulation **when implementing the** directive in order to avoid any
together with the Mediterranean partners **when implementing this** Common Strategy, in particular
region. Global initiatives may also be financed **when implemented through** a regional or international
the formulation of the scientific advice. 82. **Where implementing powers** have been conferred on the
resources, also prohibit the completion of acts **which implement contracts** concluded before the entry into
concerning restrictive measures against Iran **which implemented the** United Nations Security Council
European Community Development Policy, **which implements the** European vision on development set
to reinforce cooperation between officials **which implement the** political measures described in this
effects in the market, meaning that the market **will implement a** measure automatically without any
use increased aid flows more effectively. The EU **will implement and** monitor its commitments on Aid
aid to film production. The Hungarian authorities **will implement any** changes that may be required after
decision Evaluation by experts The EIB **will implement RSFF** in close collaboration with all major
law enforcement services. The Presidencies **will implement the** new Customs Action Plan 2008-2009
Schengen acquis covered by this Part, whether it **will implement this** measure in its national law. If it
assistance is justified by the progress made **with implementing a** project, the Commission shall request
assistance is justified by the progress made **with implementing an** action, the Commission shall request
, the Commission may deem costs directly linked **with implementing the** activity supported to be eligible,
supply. There has been some progress **with implementing the** action plan related to the national

Instruction:
Study the concordance lines and list (words) nouns, verbs, adjective, adverbs used with ‘implement’, ‘commission’

Activity type
FAMILIARISE
Instruction: Compare the use of the word ‘commission’ in BNC and EEUD

COMMISSION n

BNC Written			EEUD Corpus		
Grammatical relation	**Frequency per million words**	**Example**	**Grammatical relation**	**Frequency per million words**	**Example**
modifier	55.85	***Audit*** *Commission,* ***Monopolies*** *and* ***Mergers*** *Commission*	object_of	10.42	***inform*** *the European Commission*
					be ***charged*** *a commission of 6%*
subject_of	19.52	*Royal Commission* ***recommended***	possession	10.42	*The Commission 's* ***powers*** *of inspection*
and/or	15.02	*new regulatory commissions and consumer* ***councils***	subject_of	10.42	*the Commission* ***is****, as a rule, exempt from all taxes*
object_of	14.06	*A Truth Commission was* ***established***	modifier	5.95	*be provided by the* ***joint*** *Commission, the* ***European*** *Commission*
		banks may ***charge*** *a commission*			
pp_obj_of	11.95	***President*** *of the Cuban Commission of Human Rights*	modifies	5.21	*Commission* ***approval***
modifies	7.19	*Special Commission* ***report***	and/or	3.72	*the commission and other* ***charges***
pp_obj_by	6.11	***proposed*** *by the European Commission*			
pp_on	5.64	*the Royal Commission on* ***Pollution***			
possession	5.39	*The Royal Commission 's* ***proposals***			
pp_of	5.24	*commission of* ***inquiry***			
pp_for	4.28	*the Commission for Racial* ***Equality***			
pp_obj_to	3.50	*submitted* ***evidence*** *to the Royal Commission*			

Activity type FAMILIARISE

B.1 Instruction: Study the concordance lines and add an appropriate adverb to the sentences below.

storage, transport and retail sale, will be	**implemented**	systematically and in a consistent manner
RTD activities. An ethical review will be	**implemented**	systematically for proposals dealing with
to ensure that Albania can properly	**implement**	the new public procurement legislation
fight against money laundering if properly	**implemented**	. Some progress has been made on upgrading
provisions of the agreement are being properly	**implemented**	. Where the Agency and/or the Commission
laws. (2) For the purposes of effectively	**implementing**	the relevant provisions of the Amsterdam
The new strategy needs to be effectively	**implemented**	and monitored. The existing legislation
transposed in a timely manner and effectively	**implemented**	by Member States. The process of evaluating
capacity building to design and effectively	**implement**	sound trade and integration policies, as
relationship, the Union will strive to fully	**implement**	the Joint EU-Africa Strategy as well as
Presidencies will make every effort to fully	**implement**	the Action Plan on Simplification, for
guarantees on freedom of expression are not yet	**implemented**	fully, particularly regarding the print
these rights sufficiently and is not fully	**implemented**	. As regards children's rights , the government
on a business permit system is not fully	**implemented**	, because, mainly, of the lack of implementing
determined to pursue reforms and to fully	**implement**	the work programme, in particular by
partner in ensuring that they are successfully	**implemented**	. This plan will be presented before the
represents a key instrument for successfully	**implementing**	EPAs, and whereas regional integration

a. This reform could significantly enhance the fight against money laundering if _____ **implemented**.

b. Member States are determined to pursue reforms and to _____ **implement** the work programme.

c. The 2006 law on a business permit system is not _____ **implemented**, because, mainly, of the lack of implementing regulations.

d. The new strategy needs to be _____ **implemented** and monitored.

e. In this way, this policy covering all sectors of the food chain will be **implemented** _____ and in a consistent manner.

(Jablonkai, 2011)

Activity type FAMILIARISE

Aim: to familiarise with frequent collocates of a particular lexical item

groups according to a ranking based on the **above criteria. The** first group shall be composed of five
2007. CTF meetings focused on political **accession criteria and** rule of law issues. Each CTF meeting
) and (4)(b) of the Convention by the targeted **activity criterion, already** present in Article 15 of the Brussels
inspection services. It will be based on **agreed criteria for** the performance of these systems, and
, objective and transparent resource **allocation criteria based** on needs and performance will guide
explanation of how it had regard to the **allocation criteria established** in accordance with paragraph 2
, considers that a single set of EC aid **allocation criteria should** be targeted at the poorest countries
and practices, on the basis of common principles **and criteria. [17]** See: Commission Staff Working
and practices, on the basis of common principles **and criteria. [17]** See: Commission Staff Working
sectoral targets and measures to meet them; **and - criteria and** provisions to ensure sustainable
application A uniform application of the rules **and criteria established** by the Dublin Regulation is
progress to respond to the other issues **and criteria for** membership included in the Commission's
any other controls and may set out the methods **and criteria for** use in any later evaluation, in particular
Directive, but is directly linked to the definition **and criteria of** economic activity under national income tax
shall select projects on the basis of the objectives **and criteria set** out in Articles 1, 3 and 4 and in Annexes I
6 of this Article, while respecting the objectives **and criteria set** out in Articles 1, 3 and 4. 10. The
for Community funding in relation to the objectives **and criteria set** out in Articles 1, 3 and 4. Member States
the Commission in this field should be increased **and criteria should** be established which allow the
interests. Direct tendering has been abolished **and criteria to** identify abnormally low bids have been
out the Millennium Development Goals (MDGs) **as criteria established** jointly by the international
and energy savings as a possible **assessment criterion in** competitive tendering for public contracts.
projects will be assessed against the following **award criteria: 1)** Compliance with the objectives of the call
FINANCIAL CAPACITY 13 7.3. AUDIT 13 8. **AWARD CRITERIA 14** 8.1. EUROPEAN DIMENSION OF THE
Authority. See Appendix 2, Form I.4. 4. **AWARD CRITERIA 4.1.** Generalities Proposals will be
It will verify the eligibility, the selection and the **award criteria and** propose to the LIFE+ Committee a list of
to each proposal on the basis of the following 7 **award criteria applicable** to all three branches of LIFE+

C1. missing word:_____

shall be based on objective and _____ **criteria** , including those listed in Annex III, taking
programmes. 2.2 Objective and _____ **criteria** for resource allocation 64. Within global
objective and _____ resource allocation **criteria** based on needs and performance will guide
consistent basis and in line with _____ **criteria** , as part of integrated impact assessments
must be granted on the basis of _____ **criteria** applied in a non-discriminatory way to
be selected on the basis of _____ **criteria** . Information provided by the UN Task Force

C2. missing word:_____

work programme, to check the _____ **criteria** and any other additional conditions that
your proposal eligible? The _____ **criteria** are given in the work programme. See also
consortium. Have any additional _____ **criteria** been set for this call? Check that you
that your proposal meets the _____ **criteria** that apply to this call and funding scheme
measures which fall within the _____ **criteria** and main scope of, or receive assistance
applications that fulfil the _____ **criteria** will be considered for a grant. If an application
financial requirements, such as _____ **criteria** and financial capacity, with regard to
SELECTION CRITERIA The following _____ **criteria** define the scope of the call and apply

C3. missing word:_____

situation in Albania in terms of the political **criteria** for _____; - analysis the situation
in Albania on the basis of the economic **criteria** for _____; - reviews Albania's capacity
progress to respond to the other issues and **criteria** for _____ included in the Commission

C4. missing word:_____

no longer _____ the regional eligibility **criteria** of the Convergence objective and which
by LIFE+ if they _____ the eligibility **criteria** set out in Article 3: (a) operational activities
eligible provided that they _____ the **criteria** set out in the previous paragraph

Key to Task type C: 1. transparent, 2. eligibility, 3. membership, 4. satisfy, meet

(Jablonkai, 2011)

Concordance lines of multi-word sequences

is a crucial factor to achieve	**a high level of**	social cohesion. The Community will
out by Member States should add	**a high level of**	credibility to the annual accounts of
safety framework, which can deliver	**a high level of**	public health and consumer
The Union shall endeavour to ensure	**a high level of**	security through measures to prevent
animal feed production, establish	**a high level of**	consumer health protection and
Member State concerned, it may be decided,	**in accordance with the**	same procedure, that the Community financial
established in an implementing Regulation	**in accordance with the**	procedure laid down in Article 18. 6. In the
of the financial contribution may be approved,	**in accordance with the**	same procedure, depending on the outcome of
abovementioned period may be extended,	**in accordance with the**	same procedure, if examination of the
an implementing Regulation may specify,	**in accordance with the**	procedure laid down in Article 18, cases in
and the Commission are in agreement.	**In accordance with the**	procedure laid down in Article 18, this
may not be made out more than 14 days before	**the date on which**	the plants, plant products or other objects leave
shall not cover expenditure incurred before	**the date on which**	the Commission received the application.
obligations that were concluded or arose before	**the date on which**	those accounts became subject to restrictive
tasks may under no circumstances begin before	**the date on which**	the specific contract enters into force. I.2.3
may implementation commence before	**the date on which**	the Contract enters into force. Execution of the
obligations that were concluded or arose before	**the date on which**	those accounts became subject to restrictive

Activity type LOOK

Aim: to identify patterns of use of particular lexical bundles

Instruction: Study the concordance lines containing the lexical bundle *in accordance with the*. Notice the words and phrases that precede and follow this lexical bundle. Can you identify patterns? What are possible positions of this lexical bundle in a sentence?

Activity type FAMILIARISE

Aim: to practise identifying the function of lexical bundles in EU documents

Instruction: Study the concordance lines of the lexical bundles and identify what they make reference to. Choose from the list below:
Time, Place, Purpose, Quality, Reference to another document

E1) Reference to: _____

COUNCIL OF THE EUROPEAN UNION,	**having regard to the**	Treaty on the European Union and in
Union and in particular Article 15 thereof,	**having regard to the**	conclusions of the European Councils
COUNCIL OF THE EUROPEAN UNION,	**having regard to the**	Treaty on European Union and in particular
regard to Rule 45 of its Rules of Procedure, –	**having regard to the**	report of the Committee on Culture and
of the Diversity of Cultural Expressions, –	**having regard to the**	Commission communication on a European
the Diversity of Cultural Expressions, 2005,	**having regard to the**	Council conclusions of 13 and 14 November

E2) Reference to: _____

may not be made out more than 14 days before	**the date on which**	the plants, plant products or other objects leave
shall not cover expenditure incurred before	**the date on which**	the Commission received the application.
obligations that were concluded or arose before	**the date on which**	those accounts became subject to restrictive
tasks may under no circumstances begin before	**the date on which**	the specific contract enters into force. I.2.3
may implementation commence before	**the date on which**	the Contract enters into force. Execution of the
obligations that were concluded or arose before	**the date on which**	those accounts became subject to restrictive

E3) Reference to: _____

is a crucial factor to achieve	**a high level of**	social cohesion. The Community will
out by Member States should add	**a high level of**	credibility to the annual accounts of
safety framework, which can deliver	**a high level of**	public health and consumer
The Union shall endeavour to ensure	**a high level of**	security through measures to prevent
animal feed production, establish	**a high level of**	consumer health protection and

Self-directed DDL and DIY corpora

- corpus as a tool of self-directed learning, **DIY corpora with PhD students** (Charles, 2012; 2014)
- corpus-based EAP course, attitude to corpus use - anxiety, large general corpora – BNC, COCA (Leńko-Szymańska, 2014; Lee & Swales, 2006; Yoon & Hirvela, 2004)
- **corpus approach for English for various specialisms (ESP) at undergraduate level** (Ackerley, 2017; Flowerdew, 2002)
- **do-it-yourself (DIY) corpora** as discipline-specific models for future professions (Hyland, 2017; Charles, 2017; Frankenberg-Garcia, 2015)
- DDL courses apply direct corpus use & CALL technology (Hubbard, 2013; Friginal 2018)

Course outline

	Corpus Topic	Aspect of EU Discourse
1	Definition of a corpus, types of corpora	English EU Discourse Corpus
2	Online available English corpora (e.g. BNC, COCA, BAWE)	EU topics
3	• Online corpora and corpus analysis tools 1 • Mark Davies's COCA, BNC • Frequency lists	EU institutions
4	• Online corpora and corpus analysis tools 2 • Sketch Engine (Kilgarriff, Baisa, Bušta, Jakubíček, Kovář, Michelfeit, Rychlý, Suchomel, 2014) • POS tag and lemma list	Most frequent nouns, verbs and adjectives in general corpora and EU topic corpora
5	• Designing and creating a corpus • Using WebBootCat (Kilgarriff et al., 2014) to create a subject-specific corpus	Criteria for text selection for an EU topic corpus
6	• AntConc (Anthony, 2014) • Keywords, keyword analysis	EU keywords
7	Concordancing	usage of words in EU texts
8	• Collocations • Semantic preference and prosody	General and EU-specific collocations
9	Clusters and lexical bundles	lexical bundles in EU discourse
10	• Range (Heatley, Nation, & Coxhead, 2002) • Text coverage	EU word list

Core aims of the corpus-based ESP course

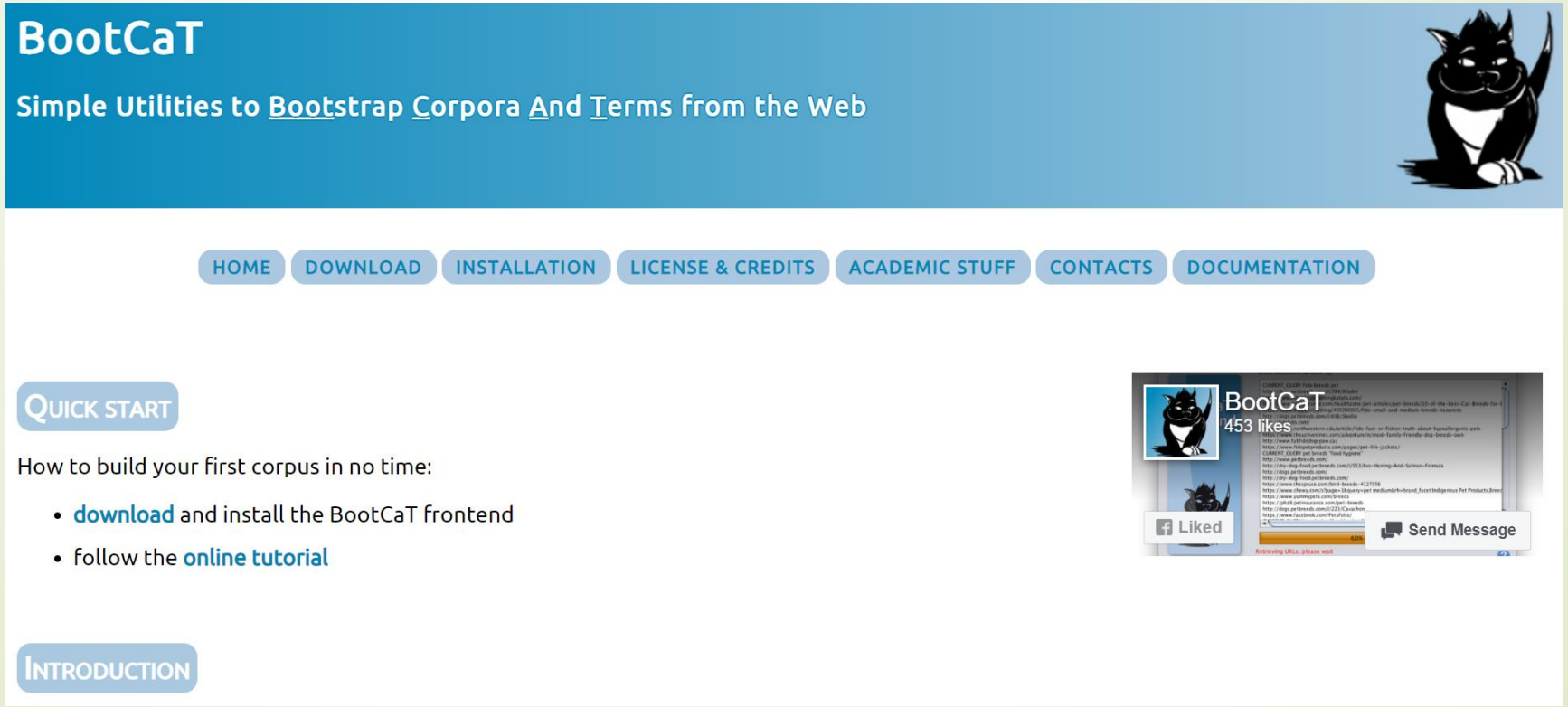
- develop students' language skills for their future professions
- sensitize students to variation in lexico-grammatical features of subject-specific language use
- raise awareness of the formulaic quality of naturally occurring language
- enhance autonomous language learning
- provide a tool for students' long-term professional use

Course outline

Corpus Topics	***Lab practice***
Module 1: Core Concepts	
Language corpora	on-line dictionaries vs corpus data
Collocations	on-line corpora in English and Slovene
Wordlists	frequency data, Academic Wordlist
Concordances	formulating queries
Module 2: Corpus Compilation	
DIY ESP corpora	purpose, type and size
Genres and registers	selecting documents
Harmonisation	corpus structure
Corpus consultation	pre-compiled corpora
Module 3: Corpus Analysis	
Corpus analysis tools	AntConc, SketchEngine
Corpus research	concordance manipulation
Analysing corpus output	word sketches, keyness
ESP terminology	BootCat: key words list, multi-word expressions
Corpus expansion	WebBootcat
Module 4: Autonomous Corpus Use	
Practicing corpus skills	Writing and translating
Presentations	Presentations of students' projects

BootCat corpus building tool

- https://bootcat.dipintra.it/



Analysis of students‘ essays

- In order to investigate students’ reflections on the benefits and challenges of their corpus use, the reflection and conclusion sections of the written reports was compiled into a small corpus of 2587 running words (Keyword Tool on the SketchEngine platform)
- The following categories emerged from the analysis of the concordance lines of the identified keywords:
  - Advantages of using a language corpus
  - Affordances of corpus tools and resources
  - Enhanced language awareness
  - Understanding the limitations of DIY corpora
  - Future plans to use language corpora

Understanding the limitations of DIY corpora

However, students also realized the limitations and problems of corpus-based investigation of language. 44% of students pointed out such disadvantages by using keywords such as *important, text, collocation, corpus* to express these concepts. Some point out that "we should be careful when collecting a **corpus**" and "select relevant **texts** "(S2), but also that "a too small **corpus** [will not] give accurate information" (S6).

Future plans to use language corpora

While all the students expressed their positive perception of corpus consultation, only 33% of them hinted at their future plans to engage in autonomous corpus use. The keywords used to express such plans were *academic, corpus, vocabulary, language, new, study* and *use*. One student planned that "In the future I would like to focus on compiling an **academic corpus**, […]in order to expand my **academic vocabulary** and […] as an aid for translation studies" (S8). Another claimed that "**studying a corpus** or compiling a new one will move us step by step to a better knowledge and **use** of **language**" (S5).

Summary of findings

Findings suggest that students considered this approach useful.

- engagement with corpus compilation seemed to help undergraduate students develop a deeper understanding of corpus structure, potential scopes of investigation and interpretation of output of corpora in general (also Google Translate)
- most importantly, help them achieve a deeper understanding of how language works, i.e., how we make meaning through language.

Aspects to reconsider

- the required minimum size of DIY corpora should be increased to at least 50,000 running words in order to yield relevant data.
- the tasks for exploration of various lexico-grammatical aspects of language should also include comparisons based on online corpora in students' L1.
- such an approach may especially help enhance undergraduate students' language awareness. (the concept of collocation in L1 and L2)

Challenges of DDL

- Learners might find it technically challenging
- Time-consuming
- Learners might feel overwhelmed by the amount of data
- Better suited for learners with proficiency from an intermediate level (BUT lower level learners benefit as well) (Boulton, 2009)
- Scaffolding exercises, guidance from the teacher
- Gradually introducing DDL and direct corpus consultation
- technology in language learning – computer-anxiety (Ortega, 1997; Sullivan & Pratt,1996)
- […] computer anxiety is a concept-specific anxiety because it is a feeling that is associated with a person's interaction with computers (Kira & Saade, 2006, p. 32).

Corpus anxiety Phase 2

	Mean
I feel confident using a corpus.	4
I feel calm when asked to search a corpus.	3.8
I become nervous when asked to use a corpus.	2.78
I get nervous when searching a corpus.	2.85
I feel anxious as I use a corpus.	2.5
I worry that I am not able to do well when using a corpus.	3.14
I worry that my computer skills are not good enough to complete corpus searches.	2.28
Overall, I feel comfortable using a corpus.	3.9
I am anxious that the other students in class do better than I when we are using a corpus.	2.5

Tentative principles of designing DDL-based language teaching
(Jablonkai & Cebron, 2020)

1. introduce students to basic concepts of corpus linguistics (Leńko-Szymańska, 2017).
2. include components that focus on technical skills
3. include both direct and indirect corpus use , guided, scaffolded corpus consultation (Ackerley, 2017; Yoon & Jo, 2014; O'Keefe & Farr, 2003)
4. Allow sufficient time and repetition for hands-on corpus querying in class (Leńko-Szymańska, 2014)

5. include compiling and analysing a DIY corpus (Charles, 2012; 2014)
6. Promote autonomous corpus use outside of the classroom (Charles, 2012; Lee & Swales, 2006; Jablonkai & Čebron, 2017)
7. Contextualise the demonstration of corpus resources and tools (Charles, 2018; Poole, 2016)
8. include components of reflection (Rogers, 2001; Schön, 1987)

Contextualisation

- Materials should be pedagogically relevant and effective (Ackerley, 2017, p.195)
- what is taught should make sense to learners and be useful and adaptable to their interests and levels (Granger and Meunier, 2008, p. 250)

Reflection task with learners

- Corpora are good for …
- Corpora are good for language learning, because …
- Corpora are challenging to use, because …
- Corpora are not good for language learning, because …

Normalisation of DDL as CALL
Lessons from CALL and TEL

- "the stage when the technology becomes invisible, embedded in everyday practice and hence normalised" (Bax, 2003: 23)
- impeding factors for normalisation, that is, (a) logistics; (b) stakeholders' conceptions, knowledge and abilities; (c) syllabus and software integra- tion; and (d) training, development and support. (Chambers & Bax; 2006; Pérez-Paredes, 2019)

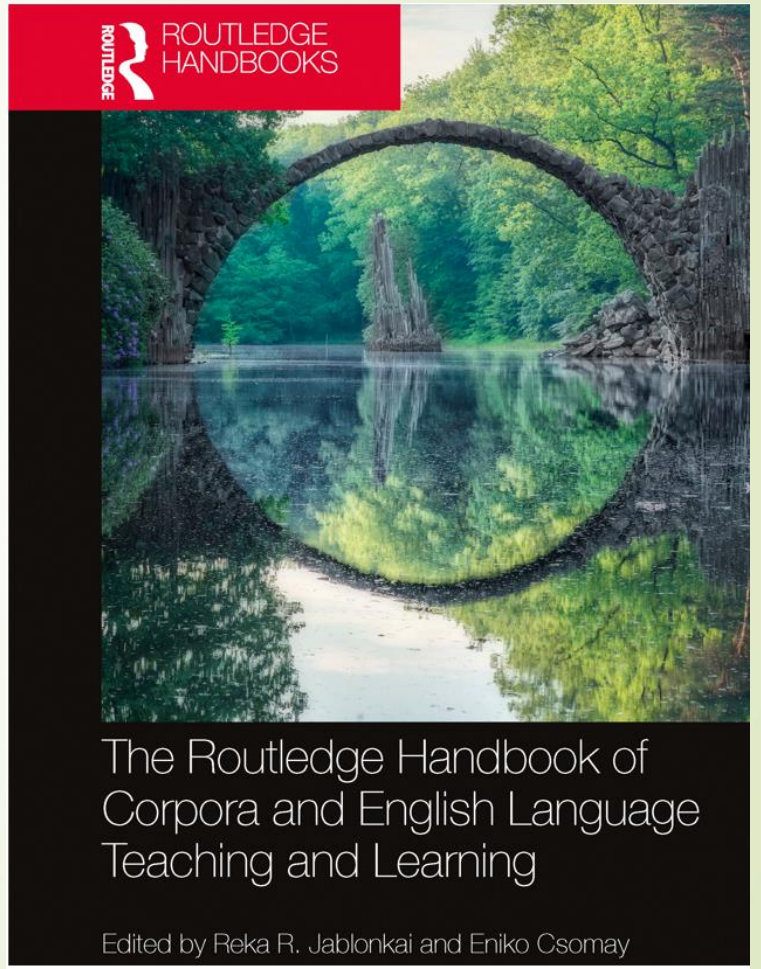
Models of integration and acceptance of technology

- TPACK (Crosthwaite et al.,2021)
- Technology Acceptance Model (Davies, 1998; Jablonkai, 2019)
- Learner training (Hubbard 2013)
- Usability evaluation of educational software, e.g. Nielsen (1995) (Hendry & Sheepy, 2022)

Future directions to help us cross the research-practice bridge

- Learning-driven and teaching-oriented corpora are needed
- Corpus analysis tools need to be more user-friendly
- Pedagogic approaches to integrate of corpus data and resources with overall teaching objectives
- Research on sociocultural and/or cognitive mechanisms of corpus applications, corpus consultation

Thank you for your attention!



References

- Boulton, A. (2010a). Learning outcomes from corpus consultation. In M. Moreno Ja´en, F. Serrano Valverde & M. Calzada P´erez. *Exploring new paths in language pedagogy: Lexis and corpus-based language teaching* (pp.129-144). London: Equinox.
- Boulton, A. (2010b). Data-driven learning: taking the computer out of the equation. *Language Learning, 60 (3),* 534-573.
- Charles, M. (2012). Proper vocabulary and juicy collocations: EAP students evaluate do-it-yourself corpus-building. *English for Specific Purposes, 31*, 93-102.
- Charles, M. (2014). Getting the corpus habit: EAP students' long-term use of personal corpora. *English for Specific Purposes*, *35*, 30-40.
- Flowerdew, L. (2013) Needs analysis and curriculum development in ESP. In B. Paltridge & S. Starfield (eds.) *Handbook of English for specific purposes* (pp. 325-346). Oxford: Blackwell.
- Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System, 32*, pp. 301–319.
- Johns, T. (1991). Should you be persuaded: two examples of data-driven learning. In: T. Johns & P. King (Eds.), *Classroom Concordancing. ELR Journal, 4 (Special Issue),* 1-16.
- Johns, T. (2002). Data-driven learning: the perpetual challenge. In: B. Kettemann & G. Marko (Eds.). *Teaching and learning by doing corpus analysis* (pp. 107-117). Amsterdam: Hunston: Rodopi.
- Lee, D., & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available corpora to self-compiled corpora. *English for Specific Purposes, 25*(1), *56-75.*
- Römer, Ute. 2009. Corpus research and practice: What help do teachers need and what can we offer? In: Aijmer, Karin (ed.). *Corpora and Language Teaching (pp.* 83-98). Amsterdam: John Benjamins.
- Römer, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics, 31*, 205–225.
- Yoon, H., & Hirvela, A. (2004). ESL student attitudes towards corpus use in L2 writing. *Journal of Second Language Writing*, *13*(4), 257–283.
- Yoon, H. (2008). More than a linguistic reference: The influence of corpus technology on L2 academic writing. *Language Learning and Technology, 12,* 31–48. Retrieved from http://llt.msu.edu/vol12num2/yoon.pdf.