



FACULTY OF ARTS
Charles University

Institute of Czech Studies

Language Corpora in Foreign Language Learning, Teaching and Research

Svatava Škodová



Language Corpora in Foreign Language Learning, Teaching and Research

The lecture presents the possibility of using language corpora in learning and teaching Czech as a foreign language. The first part is focused on basic information about Czech language corpora, esp. learner corpora (CzeSL, Merlin) and Data-Driven Learning (DDL), a method which uses corpus data in language teaching. The text briefly presents the application of this method in teaching Czech as a foreign language. The second part is devoted to the tools used in work with corpus data. There are applications for Czech such as SyD and EVALD being developed at the Institute of Formal and Applied Linguistics at Charles University.

The third part presents concrete types of exercises for students which can be used during the class.

In summary the author points out the wide use of language corpora in foreign language teaching and its importance for foreign language students without a direct contact with the target language.

Key words: Czech, Data-Driven Learning, foreign language teaching, language corpora, learner corpora, corpus linguistics,

OUTLINE

1. LC generally
2. CzeSL
3. Merlin
4. EVALD
5. Data Driven Learning
6. Examples of corpus based exercises

Language Learners' Data

- SLA (Second Language Acquisition), SLT (Second Language Teaching), SLR (Second Language Research),
- Controlled language tasks – artificial data
- not necessarily a reflection of what Ls do in more natural communication contexts
- Representativeness?
- Wish to produce more learner-aware/learner-focused pedagogical tools
- **LEARNER CORPORA** = electronic collections of natural data produced by FL or L2 learners and assembled according to explicit design criteria

Learner corpus research (LCR)

- emerged in the 1980s,
- an offshoot of corpus linguistics,
- The role of English in LCR!
- The number of NN-speakers far outnumbers N-speakers

Advantages of LC

- Quite large collection
- Collected from number of learners
- More representative
- Electronic format
- Analysed with software tool
- POS tagger,
Learners use of spec.gram.categories: prepositions, modal auxiliaries
- concordance programs
Lexis, phraseology
- Error editors – computer-aided error analysis

LC - general

- English – the major lingua franca
- Increasing number of different L2s in LR
- the website of University of Louvain
 - 37 learner corpora,
 - 82 (60%) representing L2 English,
 - the rest focusing on other languages
- Type of texts: essay writing
- Growing number: learner speech

CR - trends

- Dominant: cross-sectional studies
- Rise: longitudinal corpora
- Attention to: individual variability
- more sophisticated statistical techniques in LCR

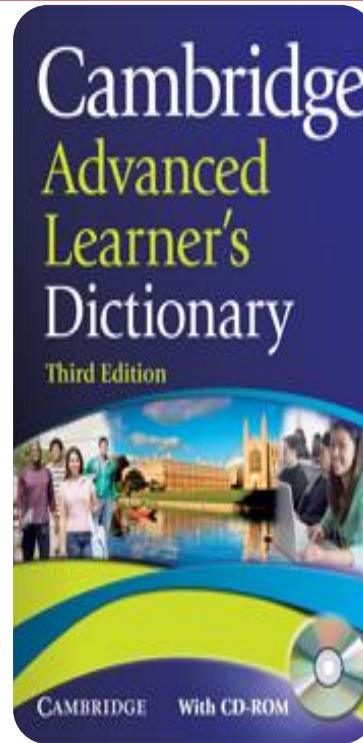
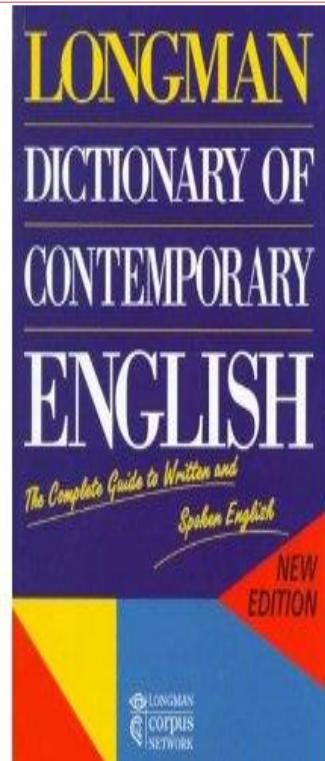
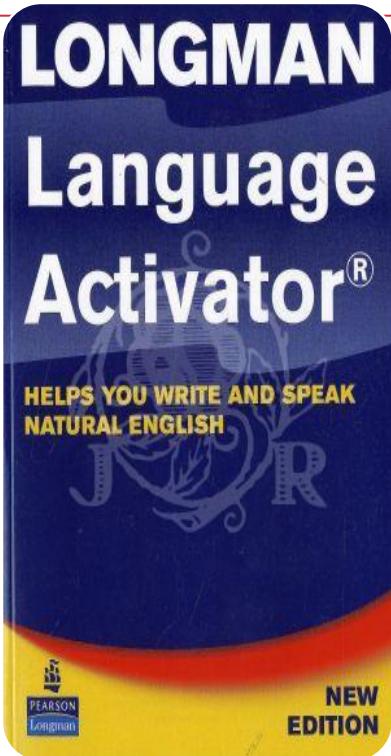
European corpora

CORPUS	words	Annotation
ICLE International Corpus of Learner English	3 mil.	Error annot, POS
LINDSEI Louvain International Database of Spoken English Interlanguage	100 000	Ortogr. transcription
LLC Longman Learners' Corpus	10 mil.	POS
PELCRA Polish-English Language Corpus Research and Applications	500 000	POS
ISLE The ISLE Corpus of non-native spoken English		error, pron.error
JPU Janus Pannonius University Corpus	400 000	
CLC Cambridge Learners Corpus	10 mil.	POS, error 2,5 mil.
USE Uppsala Student English Project	1 mil.	

Non-European Corpora

KORPUS	ROZSAH (slova)	ANOTACE
JEFLL Corpus	1 mil.	erron, POS
CEJL Corpus of English by Japanese Learners	1 mil.	Error partially
TAO Corpus Standard Speaking Test (SST) Corpus	1 mil.	Error partially
HKUST Corpus of Learner English (Hong Kong)	25 mil.	POS (1 mil.) error (100 000)
TeleNex/TELEC TELEC Student Corpus	3 mil.	Plain text

Dictionaries based on LCR



HELP BOXES

or Mr Gilard.

I am writing this letter because as you know we are exchanging jobs for a year in some days from now and I would like to inform you about what my job involves, what you have to deal with as soon as you arrive and some advice on how you could be a useful member of the working team. First of all I would like to inform you that my job here as a graphic designer env... and inside, saying how to suggest a good colour have to redesign the s... sometimes be given a



Dictionary Thesaurus Study Center Plus Dictionary Vocabulary Trainer

achieve

Go To Advanced Search

achieve verb
achievement noun
acid noun
acid rain noun
acknowledge verb
acne noun
acorn noun
acoustic adjective
acquaintance noun
acquainted adjective
acquire verb
actress noun
acrobat noun

a-chieve */ə'tʃiv/* verb *adj*

to succeed in doing or getting something you want

- She always wanted to be a lawyer and she was very goal.

-achievable *adjective*

able to be achieved

- an achievable goal

WORD FAMILY: achieve

achieve verb
achievable adjective
achievement noun

CzeSL

-
- Czech as a Second Language
 - <https://wiki.korpus.cz/doku.php/en:cnk:czesl-plain>
 - 2009*

← → C wiki.korpus.cz/doku.php/en:cenk:czeSL-plain

Aktualizovat

WaG KonText SyD Morfio KWords Treq | Wiki Support Biblio | ☰

Svatava Skodova Logout | česky Log In

WIKI
ČESKÝ NÁRODNÍ
KORPUS

The CzeSL-plain corpus

The learner corpus **CzeSL-plain** (Czech as a Second Language, plain = without annotation) is one of the results of the project *Innovation of Education in the Field of Czech as a Second Language*, a part of the operational program *Education for Competitiveness*, funded by the EU Structural Funds (ESF) and the Czech government.

The institutions involved in the creation of the corpus include: *Technical University of Liberec* as the beneficiary of the support, *Charles University in Prague* and *The Association of Teachers of Czech as a Foreign Language* as partners, as well as a number of elementary schools and high schools, civic associations, NGOs and other institutions as well as individual collaborators.

The corpus contains a total of about 2.3 million tokens (revised version 2; for more details see the table below). The revision was caused by presence of several tens of Czech pupils' essays in *ciz* part included by mistake. The original version 1 can be made available upon request. The corpus includes three subcorpora marked as three text types:

- **ciz** – transcripts of essays written by non-native speakers in language teaching classes of various types and levels;
- **kval** – academic texts obtained from non-native speakers of Czech studying at Czech universities in Masters or doctoral programmes;
- **rom** – transcripts of texts written at school by pupils and students with Romani background in communities endangered by social exclusion.

The corpus does not include any other data about the author or about the text itself.

The texts in all the three groups were produced by speakers who have not (yet) acquired the Czech linguistic skills of an adult native speaker. As an acquisition corpus the texts may serve both for research in the field of learning and teaching and for practical educational purposes. The first two datasets concern Czech as a second/foreign language, constituting an L2 acquisition (learner) subcorpus, while the third dataset is a subcorpus focusing on L1 acquisition (Czech is not considered to be a foreign language for students with Romani background monitored in this project). This is the very first publicly available corpus of this type for Czech.

The texts were collected in 2009–2012 mostly in schools, i.e. in a formal environment, and included in the corpus with the consent of the relevant institutions and individuals.

The essays and handwritten school exams were collected as manuscripts, scanned and transcribed into an electronic form. Academic texts by non-native speakers were obtained from the authors already in an electronic form. While these texts were not written in a class or with the aim to be included in a corpus, their final form may have been affected by an automatic spellchecker.

Texts of non-native speakers (the **ciz** part), extended by some newer texts, are available as the [czeSL-sgt](#) corpus, together with metadata and automatically

Read Show pagesource Old revisions

Available corpora

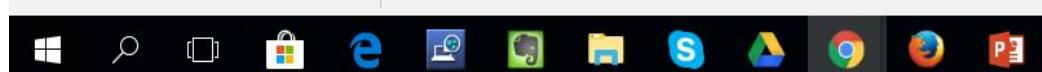
Keywords

[Reset](#) [SYN series](#) [ORAL series](#) [InterCorp](#) [synchronic](#) [diachronic](#) [spoken](#) [written](#) [web](#) [current version](#) [older version](#) [Czech](#) [non-Czech](#) [parallel](#) [comparable](#) [author](#) [representative](#) **learner** [specialized](#)

(Hold CTRL/Command to select multiple labels)

► [Advanced filter](#)

Name	Size (positions)	Labels			
czesl-plain v2	2M	synchronic written learner Czech		Details	
czesl-sgt	1M	synchronic written learner Czech		Details	
lindsei_cz	135k	synchronic spoken learner		Details	
schola2010	1M	synchronic spoken learner Czech		Details	
skript2012	709k	synchronic written learner Czech		Details	



Specifications of *CzeSL*

- Collected 2009–2012
- in the CR and abroad
- elicited during all types of situations
 - during the class,
 - as homework
 - in an examination
- Also academic texts, Bachelors', Masters' and doctoral theses

4 main groups of NN learners of Czech

1. Speakers of Slavic languages (mainly Russian, Polish)
2. Speakers of distant non-Indo-European languages Chinese, Vietnamese, Arabic
3. other Indo-European languages: German, French, English
4. Romani ethnolect

CEFR levels in CzeSL

- Covers all levels A1 – C1
- A1 and A2 prevail
- Recently: some efforts aiming at a more balanced mix of levels and L1s have been made

Výskytů: 2 239 | i.p.m. : 1 951,24 (vztaženo k celému cesel-sgt) | ARF : 1 044,41 | Výsledek je promichán

1 / 56 ►►

<input type="checkbox"/> UJA_K9_012	V utery , středu , čtvrtek a pátek taky studuju	čeština	na univerzite . Ale večer v utery divam se na
<input type="checkbox"/> UJA_L4_003	je můj program v diáři ? V ponděle odpoledne studuju	čeština	. Večer posloucha rádio . V úterý odpoledne taky studuju
<input type="checkbox"/> SME_M1_007	budeš mít kamarádů a budeš - li mít příležitost používat	čeština	v běžné komunikaci (ačkoliv i na internetu) ve
<input type="checkbox"/> LES_DR_048	tam je moc težké business ® nevíš jak to v	češtině	Nikdy ne myslela , že budou tak žít . Mám
<input type="checkbox"/> UJA2_NR_002	svůj program na týden . Od pondělí do páteka studuju	čeština	ve škole . A v vikend můžu odpočívat . Ja
<input type="checkbox"/> UJA2_K4_006	ale jsem unaveny trouch . Každý pracovní den musím studovat	čeština	9 hodin až do 13 hodin . Můj příštý týden
<input type="checkbox"/> HRD_AS_008	který mám k dispozici , bude zdávat ještě kratší .	Čeština	a já Studuji čeština dva a půl roky . Nyní
<input type="checkbox"/> PJA_9V_001	Jsem studentka a taky učitelka - studuju angličtinu a učím	čeština	. Můj muž se jmenuje Adam . Je podnikatel .
<input type="checkbox"/> UJA_L4_003	. Večer posloucha rádio . V úterý odpoledne taky studuju	čeština	. Večer se dívám na televizi , Ve středu odpoledne
<input type="checkbox"/> UJA_SU_002	týden . V pondělí v 9 hodin ráno začínám studovat	čeština	a ají do 13 hodin . Pak jdu domů a
<input type="checkbox"/> ttt_J5_078	zije v Rusko . Ja bydlím v Praze protože studuju	čeština	. Moje maminky se jmenuje Eva . Ona je hezká
<input type="checkbox"/> TOU_H212_299	směšnými , protože jsem na ně zvykla . Zatím není	čeština	pro mě seriální , ale doufám , že se brzy
<input type="checkbox"/> UJA2_MB_002	bych tam dokončit . Jak se máš ? Studuju jsem	čeština	, a proto chci napsat český . Co jsi deláš
<input type="checkbox"/> UJA2_GS_002	napsát , co dělám celý týden . V pondělí mám	čeština	od 9 hodin do 13 hodin . V 18 hodin
<input type="checkbox"/> HRD_CV_258	, ale když jsem přijela do Prahy , rozuměla jsem	češtině	hrozně malo , a mohla jsem jenom předpokládat o vyznamech
<input type="checkbox"/> UJA_ZL_011	Mám hezký týden . od pondělí do pátek jsem studoval	čeština	od 9:00 - 13:00 . čeština je těžká ale je
<input type="checkbox"/> ZOM_MM_008	a pak studovat na univerzitě . Když jsem začala studovat	čeština	, jsem porozuměla , že to není tak snadno ,
<input type="checkbox"/> TOU_H117_295	Když jsem začala můj studium v Praze stávající znalosti o	češtině	z univerzity pomohli mi porozumět její zakladatelky . Díky našim
<input type="checkbox"/> TOU_H105_291	říkají . Jsem ráda , že se snažím se naučit	čeština	. Je to krásný jazyk . A kromě toho ,
<input type="checkbox"/> TOU_H115_426	a ruštiny . Co mají společného ? Co pro mě	čeština	? To jsou pro mě nejdůležitější otázky . Protože teď
<input type="checkbox"/> TOU_H203_113	Nashledanou . Co bych dělal , kdybym nestudoval v Praze	čeština	. Kdybym nestudoval v Praze čeština , studoval bych už
<input type="checkbox"/> HRD_TL_006	, není lehké . Ráda se učím jazyky ale ta	čeština	pro mě není jednoduchá . To mě vlastně neudivuje .
<input type="checkbox"/> DGD_E2_032	Nic ne snídám . Do škol jedu tramvajem . Studuju	čeština	, matematiku a angličtinu . Po vyučování obyčejně snídám v
<input type="checkbox"/> ZOM_GA_016	a můj nejoblibnější místo v celém Japonsku . Já a	čeština	Čeština je docela krásná , ale , určitě , arménština
<input type="checkbox"/> DTO_A8_290	omlouvam se , tento den budu mít státní zkoušku z	češtiny	. To je velmi důležité pro mě , protože pak
<input type="checkbox"/> DGD_C1_016	Ruska , ale teď žiju v Praze , protože studuju	čeština	na Karlově Univerzitě . Se učím česky v Hloubětíně .

<input type="checkbox"/> LES_DR_048	tam je moc težké business ® nevím jak to v	češtině	Nikdy ne myslela , že budou tak žít . Mám
<input type="checkbox"/> UJA2_NR_002	svůj program na týden . Od pondělí do pátku studuju	češtinu	ve škole . A v vikend můžu odpočívat . Ja
<input type="checkbox"/> UJA2_K4_006	ale jsem unaveny trouh . Každý pracovní den musím studovat	češtinu	9 hodin až do 13 hodin . Můj příštý týden
<input type="checkbox"/> HRD_AS_008	který mám k dispozici , bude zdávat ještě kratší .	Čeština	a já Studuji češtinu dva a půl roky . Nyní
<input type="checkbox"/> PJA_9V_001	Jsem studentka a taky učitelka - studuju angličtinu a učím	češtinu	. Můj muž se jmenuje Adam . Je podnikatel .
<input type="checkbox"/> UJA_L4_003	. Večer posloucha rádio . V úterý odpoledne taky studuju	češtinu	. Večer se dívám na televizi , Ve středu odpoledne
<input type="checkbox"/> UJA_SU_002	týden . V pondělí v 9 hodin ráno začínám studovat	češtinu	a ají do 13 hodin . Pak jdu domů a
<input type="checkbox"/> ttt_J5_078	zije v Rusko . Ja bydlím v Praze protože studuju	češtinu	. Moje maminky se jmenuje Eva . Ona je hezká
<input type="checkbox"/> TOU_H212_299	směšnými , protože jsem na ně zvykla . Zatím není	čeština	pro mě seriózní , ale doufám , že se brzy

 UJA2_MB_002 UJA2_GS_002 HRD_CV_258 UJA_ZL_011 ZOM_MM_008 TOU_H117_295 TOU_H105_291 TOU_H115_426 TOU_H203_113 HRD_TL_006 DGD_E2_032 ZOM_GA_016 DTO_A8_290 DGD_C1_016 UJA2_LB_008 HRD_1G_166 DTO_OA_100 UJA2_K3_005 UH_MY_001

doc.t_id:	UJA_K9_012	doc.t_datum:	2009-12-29
doc.t_medium:	rukopis	doc.t_limit_minut:	45
doc.t_pomucka:	ne	doc.t_zkouska:	ano prubezna
doc.t_limit_slova:	40-	doc.t_nazev:	Týdenní program
doc.t_tema_typ:	obecné	doc.t_aktivita:	jine
doc.t_tema_zadane:	zadane	doc.t_postup_zadany:	
doc.t_postup_prevazujici:	informace	doc.t_pocet_slov:	79
doc.t_rozsah_slov:	50-99	doc.s_id:	UJA_K9
doc.s_pohlavi:	m	doc.s_vek:	23
doc.s_vek_kat:	16-	doc.s_jazyk1:	ru
doc.s_jazyk1_skupina:	S	doc.s_dalsi_jazyky:	
doc.s_cj_SERR:	B1	doc.s_cj_v_rodine:	
doc.s_roky_v_CR:		doc.s_studium_cj:	VS
doc.s_studium_cj_mesice:	6-12	doc.s_studium_cj_hod_tyden:	15-
doc.s_ucebnice:	jine	doc.s_bilingvni:	
s.id:	5		



adat o vyznamech

dy . Díky našim

ože teď

oval bych už

divuje .

yčejně snídám v

nština

bětině .

na oboru

ně Čechů

MERLIN

-
- ODKAZ: <http://www.merlin-platform.eu/index.php>



Home | Search About MERLIN ?

Czech German English Italian

Using MERLIN

MERLIN for research

MERLIN corpus

MERLIN annotations

Documentation

Download whole corpus

[Simple search](#)

[Advanced search](#)

[Define a subcorpus](#)

[Statistics](#)

Search for words in the learner texts and display them in context (e.g., full text):

examples: [welchen](#), [gli](#), [rád](#)

Search in:

Corpus:

No subcorpus yet? Compile your own text collection » [here](#).

The MERLIN corpus

MERLIN provides access to 2.286 texts written by learners of [Czech](#), [Italian](#) and [German](#).

The learner texts stem from standardized language tests and they have been reliably related to the CEFR levels. » [read more](#)

Use MERLIN ...

... to better understand the levels of the Common European Framework of Reference (CEFR). » [read more](#)

Video tutorial ...



MERLIN: korpus

MERLIN:
anotace

Dokumentace

Ke stažení

Pocet výskytu
(celkem) 19

1 2

>>

Strana 1 z(e) 2 (15 počet výskytů na straně)

Vyhledávané slovo v kontextu

ještě musíte vědět? Učím **češtinu** asi tři semestry na

Informace o autorovi textu a hodnocení

univerzitě , také znám **češtinu** množi let „protože

Informace o autorovi textu a hodnocení

Čechách . Proto učím **češtinu** , abych mohl se

Informace o autorovi textu a hodnocení

,že zlepším moje češtinu a můžu dovědět víc

Informace o autorovi textu a hodnocení

můžu dovědět víc přes **češtinu** a české republiky .

Informace o autorovi textu a hodnocení

a tam chci studovat **češtinu** . A mimo to

Informace o autorovi textu a hodnocení

napad ,že učím češtinu tam . V létě

Informace o autorovi textu a hodnocení

je Brno? Jsou kurzy **češtinu** tam v leté? Chtěla

Informace o autorovi textu a hodnocení

se učit mnoho **češtinu** . V Brně ,

Informace o autorovi textu a hodnocení

Chci absolvovat tam kurz **češtinu** v leté! To je

Informace o autorovi textu a hodnocení



[Úvod | Vyhledávání](#) [O projektu MERLIN](#) [?](#)

[cesky](#) [nemecky](#) [anglicky](#) [italsky](#)

MERLIN: v praxi

MERLIN: výzkum

MERLIN: korpus

MERLIN: anotace

Dokumentace

Ke stažení

[Jednoduché vyhledávání](#) [Pokročilé vyhledávání](#) [Vytvořit subkorpus](#) [Statistika](#)

[Zpět na výsledky vyhledávání](#)

Tip: Pro vyhledávání konkrétního slova nebo jazykového jevu v kontextu použijte *Jednoduché* nebo *Pokročilé vyhledávání*.

Obecné::

ID textu: 1113 [Kopírovat ID textu](#)

Cílový jazyk: čeština

Úroven testu podle SERR: B1

Testová úloha: [odpoved agenture \(Tandem\)](#) ⓘ

Materský jazyk: nemčina

Vek: not reported

Pohlaví: muž

Hodnocení:

Celkové hodnocení podle SERR: A2+

Gramatická správnost: A2

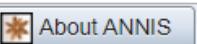
Pravopis: B1

Slovní zásoba: rozsah: B1

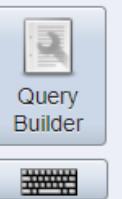
Slovní zásoba: ovládání: B1

Koherence/koheze: B1

Sociolingvistická primerenost: A2

[Upravit předchozí dotaz](#)[Vysvetlivky k tomuto výstupu](#)[Zkratky](#)

learner="čeština"

[Search](#)[More ▾](#)[History ▾](#)19 matches
in 16 documents[Corpus List](#)[Search Options](#)

Visible: All

[Help us to make ANNIS better!](#)not logged in [Login](#)[Help/Examples](#)[Query Result](#)

Base text Token Annotations

1 / 2 Displaying Results 1 - 10 for: learner="čeština"

(learner 69 - 79)

TH/EA (grid)

Brno	?	Jsou		kurzy	čeština	tam	v	letě
Brno	?	Jsou	tam	kurzy	češtiny		v	léte
			MOVT		CHA	MOVS		CHA
Brno	?	Jsou	tam	kurzy	češtiny		v	létě
			MOVT		CHA	MOVS		CHA
				G_Wo				
					G_Morphol_Wrong			

telltxt

	case
	womaincl

transcript (grid)

automatic (grid)

dependencies (arcs)

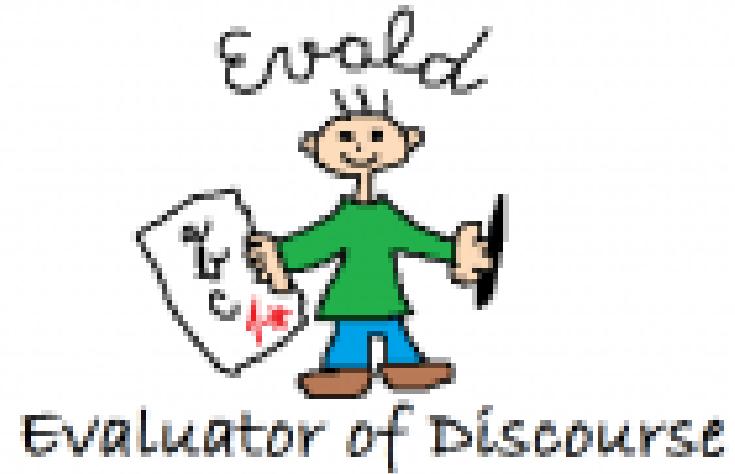
full text

4 Path: MERLIN_Czech > 0929 left context: 5 right context: 5
(learner 65 - 75)

EVALD Automatic EVALUATOR OF DISCOURSE

<https://lindat.mff.cuni.cz/services/evald-foreign/>

-
- Institute of Formal and Applied Linguistics, CU Prague
 - Monitors: 200 language features in mutual interaction
 - language fields: spelling, morphology, lexicology, syntax, text syntax (sentence information structure)
 - and text structure (semantic discourse relations, anaphora and coreference).



LINDAT CLARIN Repository TreeQuery Treex More Apps About CLARIN

**EVALD**
Evaluator of Discourse





Evald 2.0 for Foreigners

EVALD 2.0 for Foreigners is a software for automatic evaluation of surface coherence (cohesion) in Czech texts written by non-native speakers of Czech ([click here](#) for EVALD 2.0 that is designed for native speakers of Czech). For more information, visit [the project web pages](#).

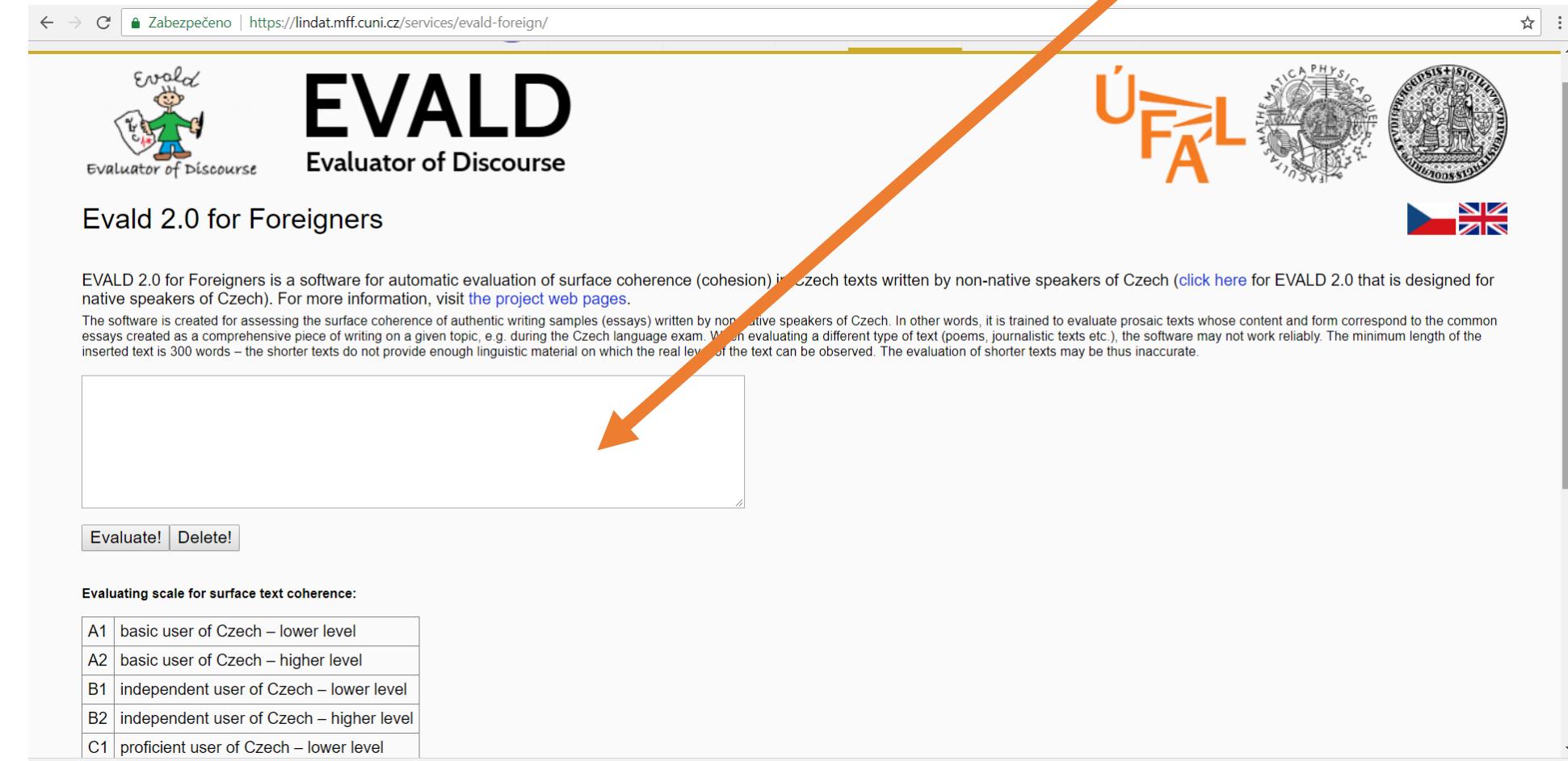
The software is created for assessing the surface coherence of authentic writing samples (essays) written by non-native speakers of Czech. In other words, it is trained to evaluate prosaic texts whose content and form correspond to the common essays created as a comprehensive piece of writing on a given topic, e.g. during the Czech language exam. When evaluating a different type of text (poems, journalistic texts etc.), the software may not work reliably. The minimum length of the inserted text is 300 words – the shorter texts do not provide enough linguistic material on which the real level of the text can be observed. The evaluation of shorter texts may be thus inaccurate.

Enter a Czech text here (the maximum length is limited to 8,192 characters) and click on the button "Evaluate!" below. The evaluation may take some time (tens of seconds or more, depending on the length of the text).

Evaluating scale for surface text coherence:

Anotace.docx 

Here you place a text



Zabezpečeno | https://lindat.mff.cuni.cz/services/evald-foreign/

 **EVALD**
Evaluator of Discourse

Evald 2.0 for Foreigners

EVALD 2.0 for Foreigners is a software for automatic evaluation of surface coherence (cohesion) in Czech texts written by non-native speakers of Czech ([click here](#) for EVALD 2.0 that is designed for native speakers of Czech). For more information, visit [the project web pages](#).

The software is created for assessing the surface coherence of authentic writing samples (essays) written by non-native speakers of Czech. In other words, it is trained to evaluate prosaic texts whose content and form correspond to the common essays created as a comprehensive piece of writing on a given topic, e.g. during the Czech language exam. When evaluating a different type of text (poems, journalistic texts etc.), the software may not work reliably. The minimum length of the inserted text is 300 words – the shorter texts do not provide enough linguistic material on which the real level of the text can be observed. The evaluation of shorter texts may be thus inaccurate.

Evaluate! Delete!

Evaluating scale for surface text coherence:

A1	basic user of Czech – lower level
A2	basic user of Czech – higher level
B1	independent user of Czech – lower level
B2	independent user of Czech – higher level
C1	proficient user of Czech – lower level

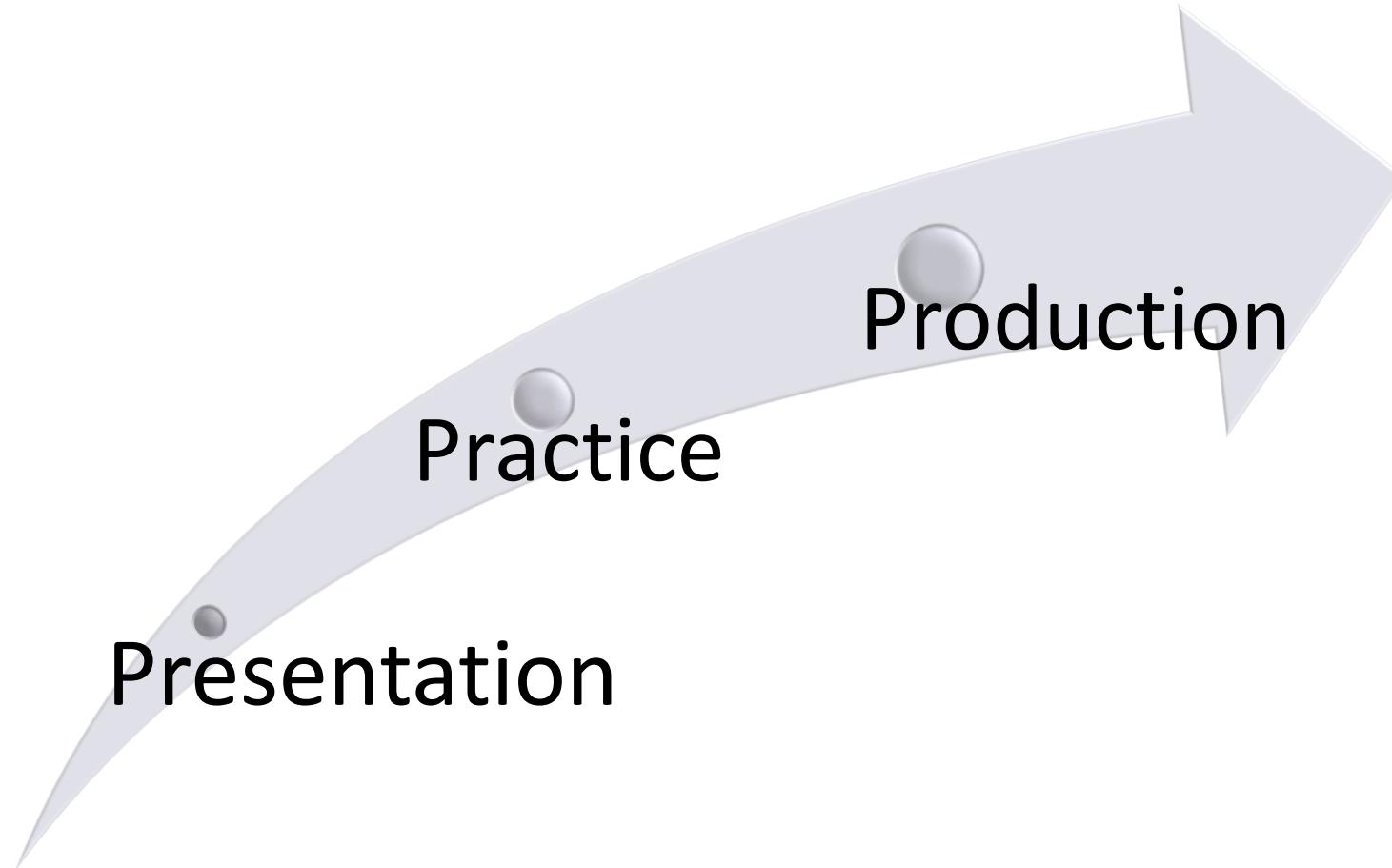
CHANGING EDUCATIONAL PARADIGMA

(Granger 2002 , 21)

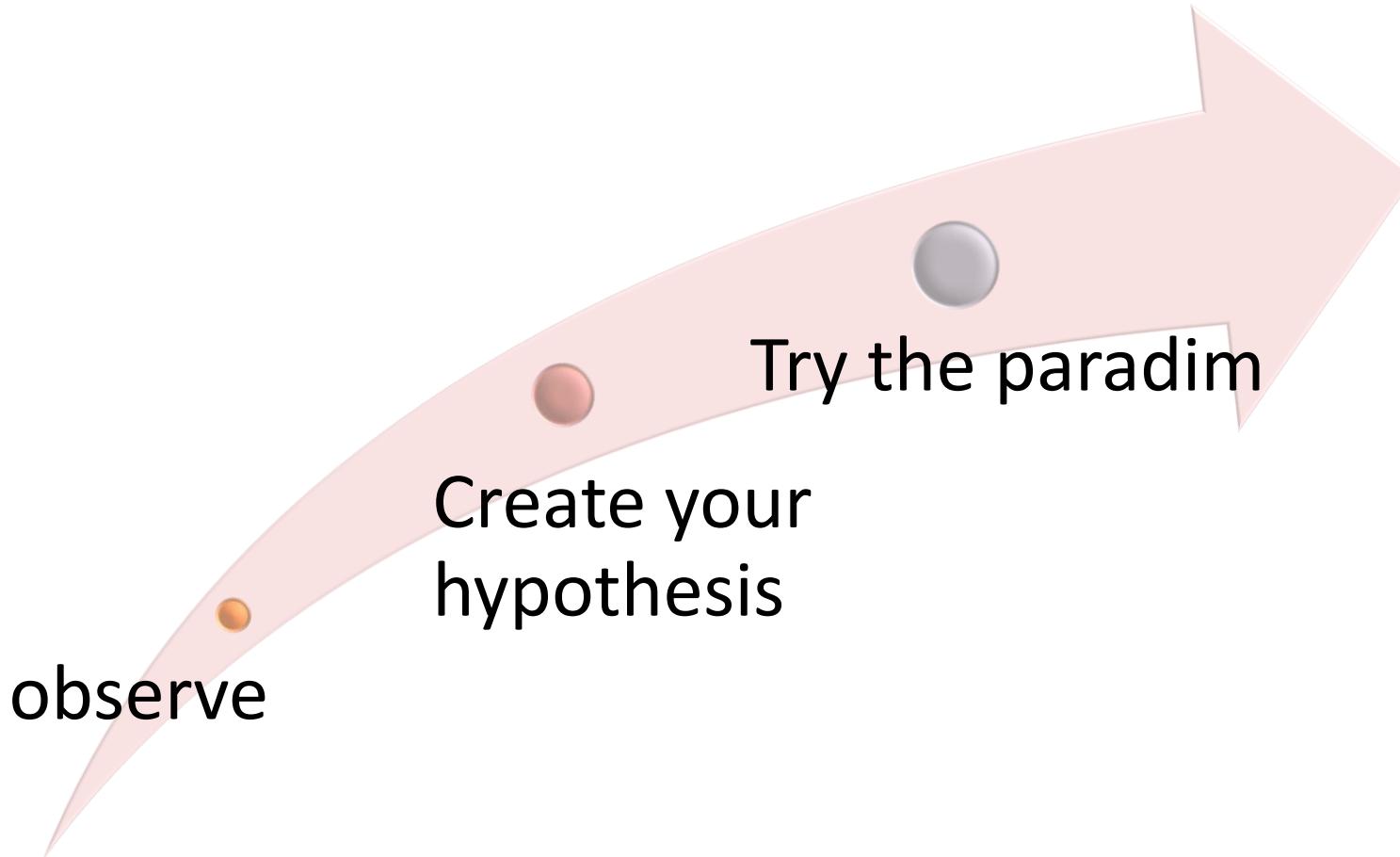
A Bird's-eye view of learner corpus research.

Granger, S. at al. (eds) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*.
Amsterdam/Philadelphia : Jonh Benjamins Publishing Company.

Up to now paradigm



Emerging paradigm



Data Driven Learning

„Research is too serious to be left to the researchers.“

(1991)

Tim Johns, University of Birmingham



Examplex of Corpora Based Exercises

1. Exercise - Noun forms analysis

Read the sentences:

1. Jsem žák druhé **třídy**, ale tolik už přece vím.
2. Na hlavní **třídě** si všiml obchodu s pánskými oděvy.
3. Objednal si tedy sedadlo v první **třídě**.
4. My, Angličané střední **třídy**.
5. Zabočil s autem za hlavní **třídu** a rychle jel k mostu.
6. Ten nejspíš letí první **třídou**.
7. Poslouchala jsem pod oknem **třídy** asi deset minut.
8. Za jejich vilou byla široká **třída**, po které jezdily dvě tramvaje.
9. Martin vlastně s námi do **třídy** nechodil.
10. Sedíte v kupé druhé **třídy** s dvojicí neznámých.
11. Specialista první **třídy**.
12. Postoupila do poslední **třídy** gymnázia.

Exercise 2

Use the sentences from Ex. 1, find the forms of TŘÍDA for each case in the paradigm.

1. pád věta č.:

2. pád věta č.:

3. pád věta č.:

4. pád věta č.:

5. pád věta č.:

6. pád věta č.:

7. pád věta č.:

.

Exercise 3

Write sentences using the forms of TŘÍDA which are not in Exercise 1. Check the forms in corpus.

Exercise 3 polysemy

Guess which sentences belong to which meaning.

- i. místo, kde se učí žáci
- ii. ročník ve školním vyučování
- iv. velká ulice ve městě
- v. společenská vrstva
- vi. označení kvality

Collocation Vlak

- jede, přijede, zastaví, pojede, vyjíždí, vjíždí, jezdí, stojí, projíždí, odjíždí, rozjíždí se, pohně se, přejede, zmizí, vyjede, rozjede se, vykolejí, vezl, zůstane, uhání, zastavuje, uhání, blíží se, přepravuje, usmrtí
- plný, naložený, půlnoční, směřující, jedoucí, tažený, označený, odjíždějící, jedoucí, projíždějící, odvážející, převážející
- [autobus](#)(224) [cestovat](#)(87) [jedoucí](#)(95) [jet](#)(538) [jezdit](#)(293) [motorový](#)(67) [nákladní](#)(209) [odjíždět](#)(81) [odjezd](#)(120) [osobní](#)(253) [ostře](#)(60) [parní](#)(90) [projíždějící](#)(55) [projíždět](#)(72) [rozjetý](#)(55) [sledovaný](#)(60) [srazit](#)(73) [srážka](#)(64) [ujet](#)(95) [zpoždění](#)(69)

Exercise 4 – Meaning based on the context

Find the most frequent collocations of the verb ZTRATIT.

Write your own sentences with these collocations.

Slovní tvary

frekvence celkem 16732

psaný

ztratit 16192

Ztratit 510

ZTRATIT 30

**Gram. kat.: Sloveso: infinitiv ,
dok. 100,00% (1673)**

Kolokace

co(353)

nic(37)

minuta(14)

tvář(14)

vědomí(12)

kontakt(10)

vteřina(10)

víra(10)

ZTRATIT (Loose)

- Jaká kolokace je nejčastější?
- Dosadte místo zájmene **CO** konkrétní slova. Co všechno můžete ztratit?
- Je něco **CO** nemůžete ztratit? Zkontrolujte v korpusu – jaké jsou kolokace pro NEZTRATIT.

Můžu ztratit

Nemůžu ztratit

- Pozorujte příklady kolokací **ZTRATIT + CO**:
- Všimněte si, **KDE** stojí slovo **CO**. Před slovesem? Nebo za slovesem?
- S jakými slovesy se pojí infinitiv **ZTRATIT** v těchto větách?

Ve frázi ***Hleděla jsem se co nejdříve ztratit*** má ztratit jiný význam než ve všech ostatních větách. Odhadněte co znamená:

- i. *rychle jsem chtěla nastoupit,*
 - ii. *rychle jsem chtěla odejít,*
 - iii. *rychle jsem hledala ztracenou jízdenku.*

Exercise 6 Text correction

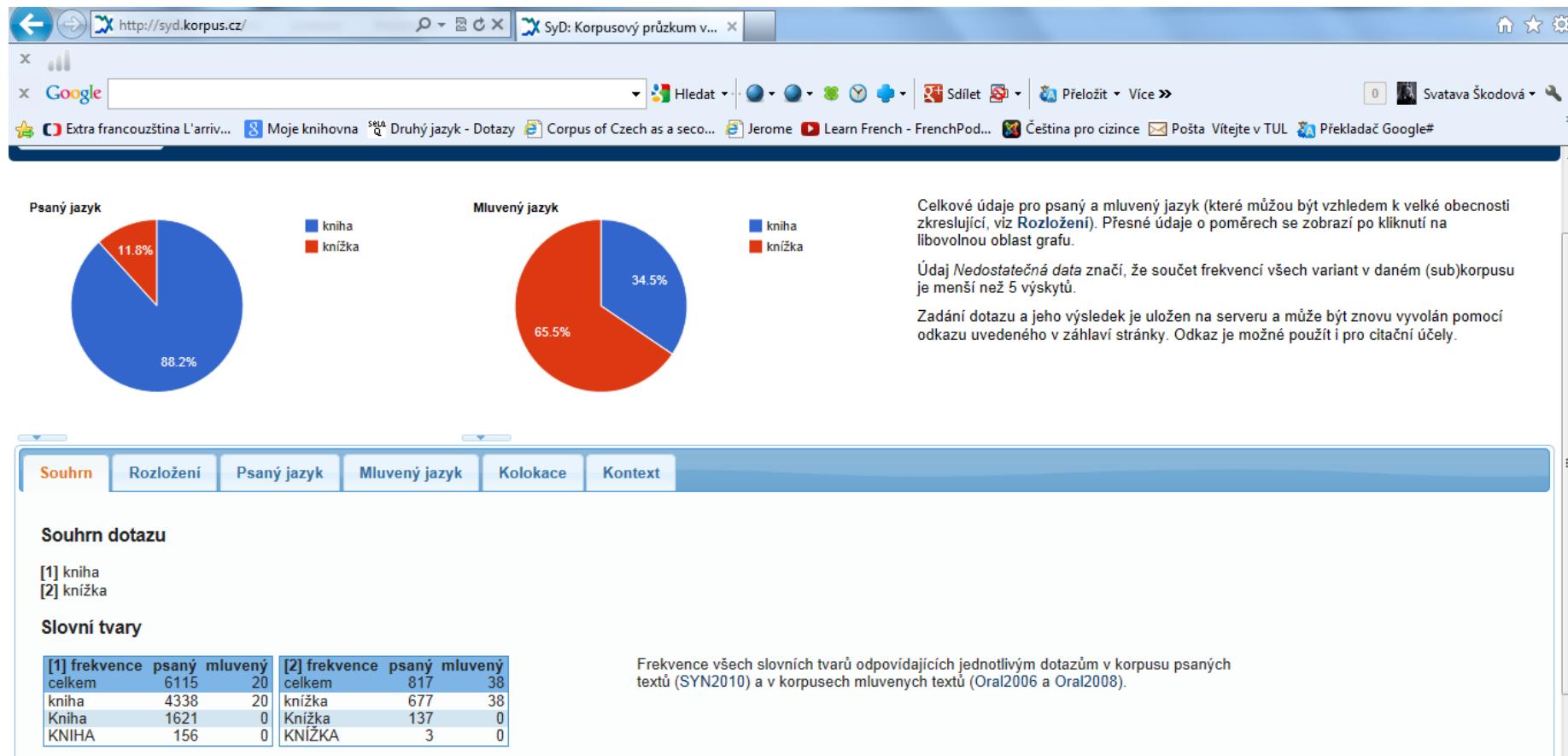
Přečtěte si text, který napsal žák z Ukrajiny. Čísla v textu jsou vždy za slovem, kde udělal chybu. Všechna slova mají společný jeden typ chyby – jaký? Pokuste se chyby opravit.

Mám moc radá (1) sport. Hrájeme (2) moc venkovní sporty. Ráno jezdím na kole do školy (3) s mym (4) kamaradem (5). Nemůžeme lyzovat (6), sáňkovat nebo běhat na lých (7), protože tu není sníh. Ale, v letě (8), když je teplo, plaváme (9) v bázeňe (10) (11), na řece nebo na moře. Mám ráda fotbal a obvykle jdu na fotbalové hřiště (12), ale nehraju fotbal dobře. Ještě hrájeme volejbal a basketbal. Golf a baseball nejsou na Ukrajině populární. Hrajeme taky kulečník (13), šachy a stolní tenis. Myslím že, šach je dobrý duševní sport. Múžu (14) trochu hrát šachy, ale nemužu (15) bruslit. Na podzim nesportujeme moc, protože velmi moc prší. Pak, dívám se sport na televizi (16).

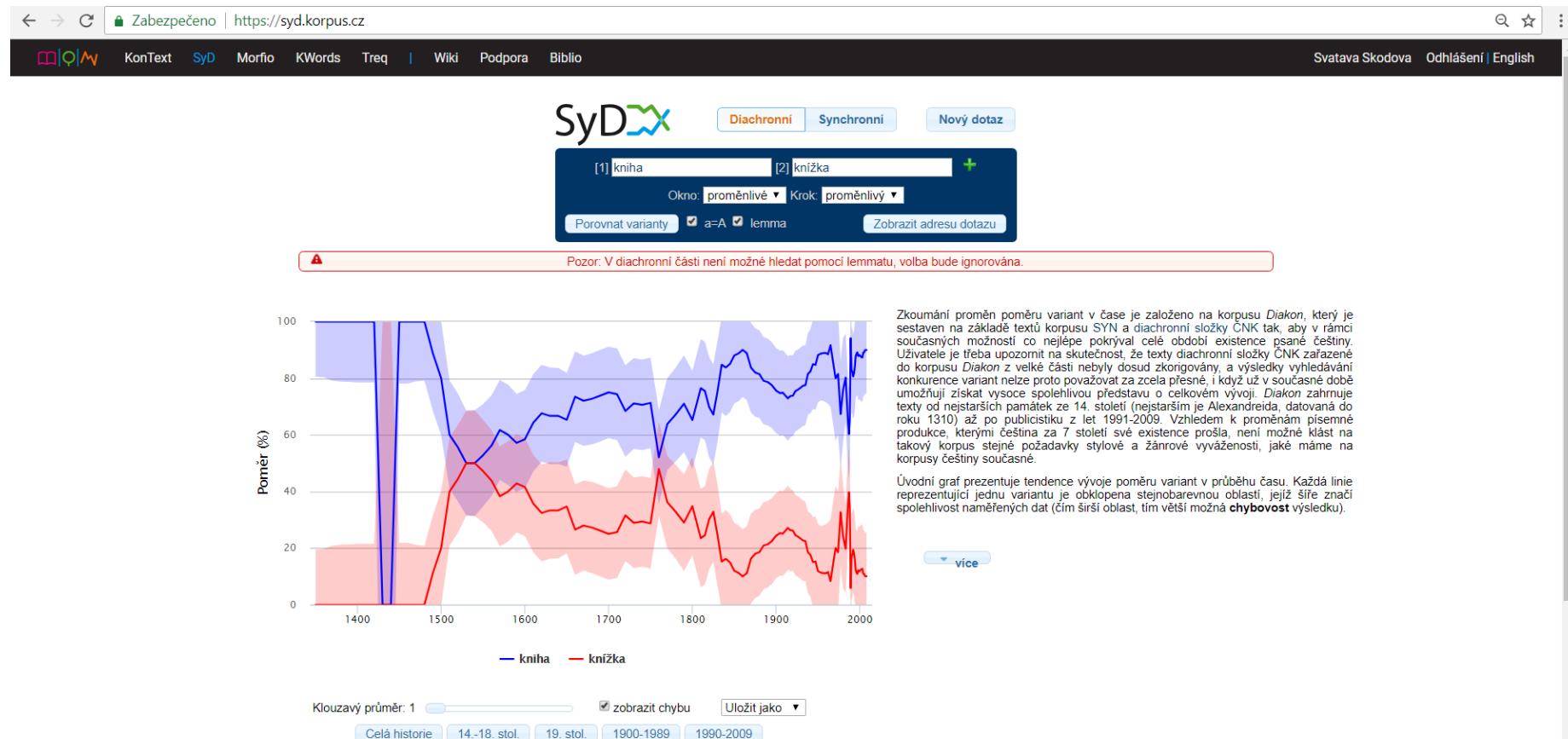
Exercise 7 STYLE/REGISTER

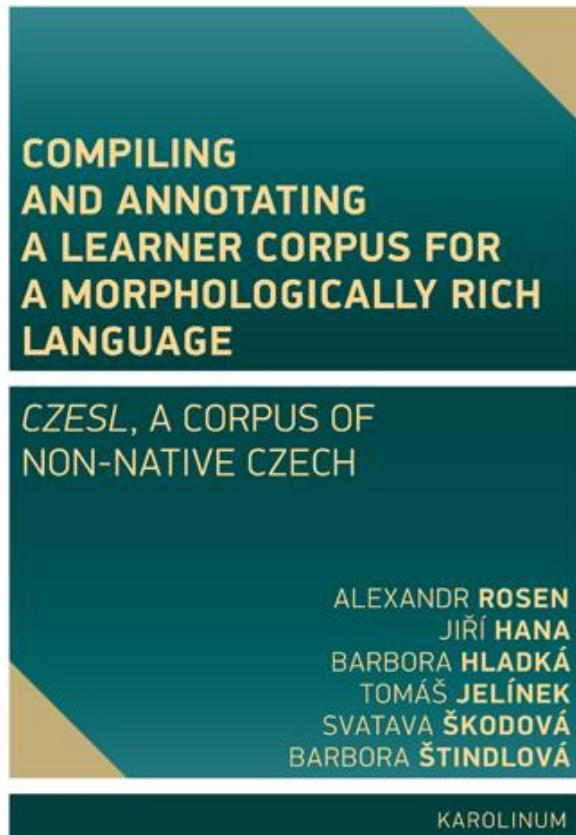
use of words in context

What are the differences in the use of **KNIHA** a **KNÍŽKA**?



Changes in KNIHA – KNÍŽKA through the history. What does it say to you?





A new volume a corpus of non-native Czech, against the background of theoretical and practical issues in the current learner corpus research.

<https://karolinum.cz/en/books/rosen-compiling-and-annotating-a-learner-corpus-for-a-morphologically-rich-language-23802>