

Využití korpusů ve výuce češtiny jako cizího jazyka

FF MU, 7.12. 2020

Co se dnes dozvíte?

- jaké korpusy se ve výuce používají
- jakým způsobem je možné korpusy ve výuce využívat
- co je to metoda data-driven learning
- co je to SKeLL nebo SmartTool

Korpusy ve výuce

1



2



3



Korpusy ve výuce

1.

přímé využití - studenti pracují přímo s korpusy

2.

nepřímé využití - korpusy jsou využívány pro tvorbu studijních materiálů, učebnic aj.

3.

popularizace - workshopy, prezentace o možnostech korpusového

přístupu

Historie

80. léta 20. století

- John Sinclair: COBUILD projekt
- Tim Johns: data-driven learning
- Dave Willis: The new lexical syllabus
- Jane Willis: Collins COBUILD English Course

90. léta 20. století

- první MLD slovník
- Teaching and Language Conference (TaLC)

po roce 2000

- učebnice Touchstone, program TestBuilder
- práce s korpusy jako součást vzdělávání učitelů

Hlavní motivace



reálné užití
jazyka



autentická data



výzkum podložený
daty

Korpusová typologie

OBEČNÉ KORPUSY

SPECIALIZOVANÉ KORPUSY

Korpusová typologie

korpusy řady SYN

KSK, NET, ČAPEK

Korpusová typologie: výuka

KORPUSY PRO VÝUKU
(OBEČNÉ KORPUSY)

KORPUSY Z VÝUKY
(SPECIALIZOVANÉ KORPUSY)

Specializované korpusy pro L2

1. akviziční korpusy (žákovské korpusy)
2. korpusy jazyka pro specifické účely (např. akademický jazyk, profesní mluva)
3. pararelní korpusy
4. jiné (např. korpus vyučovacích hodin, učebnicový korpus)

Zajímavosti z učebnicových korpusů

- učebnicový jazyk
- obsah (nadužívání slov / stereotypizace)
- P. Vališová: Účko1, Účko2

Chutná vám české **pivo**, pane Nováku?

Ano, české **pivo** mi chutná.

Chutná ti české **pivo**, Samire?

Ne, české **pivo** mi nechutná.

Hana: Teď mám žízeň.

Jan: Ty máš štěstí! Mám tady **pivo**.

Hana: To je problém. Bohužel nemám ráda **pivo**.

Jan: No ne!!! Nemáš ráda **pivo**? Tak... To je tvůj problém!

Žákovské korpusy

G. Leech (1998):

Užitečný nástroj pro každého, kdo chce zkoumat, jak se lidé učí jazyky a jakým způsobem by se je mohli učit lépe.

Sběr dat pro výzkum L2

spojitost se sběrem dat pro L1

4 etapy (B. MacWhinney)

1. etapa impresionistických studií

2. etapa deníkových zápisů (od 70. let 19. století)

3. etapa transkriptů (od 50. let 20. století)

4. etapa moderních technologií vč. korpusů (od 80. let 20. století)

Child Language Data Exchange System (1984), TalkBank (2001)

Žákovské korpusy

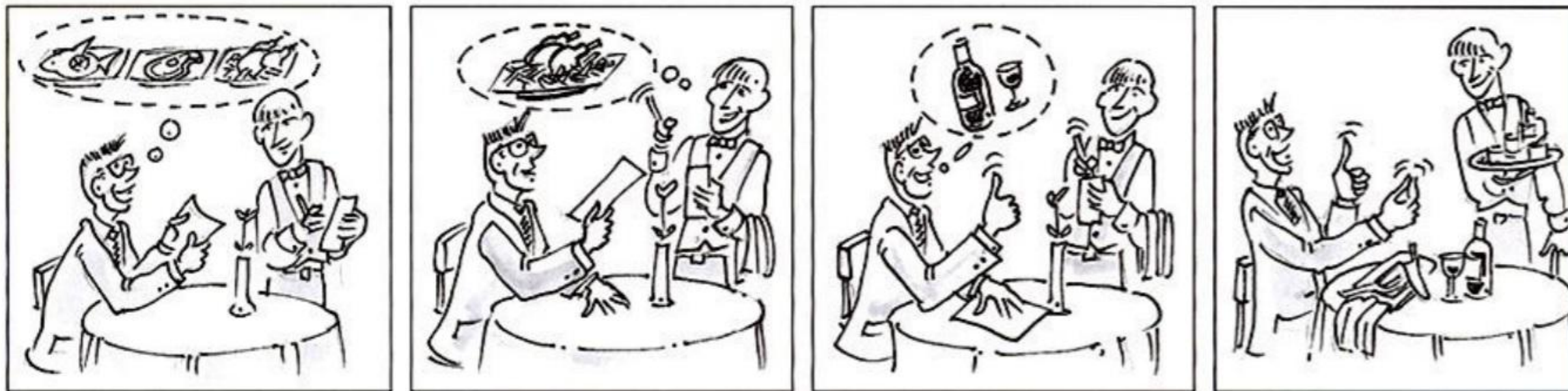
–angl. learner corpus

–studijní korpus (Čermák) = žákovský korpus

–snaha shromáždit objektivní data pro

1. popis žákovského jazyka (interlanguage /mezijazyk)
2. popis odchylek a rozdílů (contrastive interlanguage analysis (S. Granger); computer aided error analysis)
3. popis mimojazykových vlivů na L2 (metadata)

Jak vznikají data pro žákovské korpusy?



Obrázek 1 Příběh v obrázcích. Zdroj: Španělský mluvený korpus, online.

Jak vznikají data pro žákovské korpusy?

1. řízené projevy
2. strukturované projevy
3. elicitace
4. improvizace

Jak vznikají data pro žákovské korpusy?

1. certifikované zkoušky
2. jazykové školy
3. univerzity

!!!

minimum přirozených promluv

Jak vznikají data pro žákovské korpusy?

doc.t_id	UJA3_TT_001	doc.t_datum	2011-10-22
doc.t_medium	rukopis	doc.t_limit_minut	45
doc.t_pomucka	ne	doc.t_zkouska	ano prubezna
doc.t_limit_slova	20	doc.t_nazev	Popis pokoje
doc.t_tema_typ	obecne	doc.t_aktivita	
doc.t_tema_zadane	zadane	doc.t_postup_zadany	volne
doc.t_postup_prevazujici	popis	doc.t_pocet_slov	32
doc.t_rozsah_slov	-50	doc.s_id	UJA3_TT
doc.s_pohlavi	z	doc.s_vek	17
doc.s_vek_kat	16-	doc.s_jazyk1	fi
doc.s_jazyk1_skupina	nIE	doc.s_dalsi_jazyky	sv
doc.s_cj_SERR	A1	doc.s_cj_v_rodine	
doc.s_roky_v_CR		doc.s_studium_cj	VS
doc.s_studium_cj_mesice		doc.s_studium_cj_hod_tyden	15-
doc.s_ucebnice	NCSS	doc.s_bilingvni	ne
s.id	2		

Žákovské korpusy

- typologie podle různých kritérií
 - velikost, dostupnost
 - povaha dat a povaha sběru dat
 - L1, cílový jazyk
 - metadata a anotace
- Štindlová (2013): přehled 57 žákovských korpusů
 - dominují psané korpusy střední velikosti (do 1 mil slov) pokročilejších studentů s angličtinou jako cílovým jazykem
- Learner Corpus Association (2018): 167 žákovských korpusů

Žákovské korpusy ve světě

–Komerční korpusy

- CLC: Cambridge Learner Corpus, část přístupná přes Sketch Engine

–Lovaňské korpusy

- ICLE: International corpus of learner English
- LINDSEI: Louvain International Database of Spoken English Interlanguage
- LONGDALE: Longitudinal Database of Learner English
- VESPA: Varieties of English for Specific Purposes Database

–Korpusy slovanských jazyků

- RLC: Russian Learner Corpus; CroLTeC: Croatian Learner Text Corpus

Žákovské korpusy u nás

–projekt AKCES (2005)

- Ústav českého jazyka a teorie komunikace FF UK + další pracoviště
- prof. PhDr. Karel Šebesta, CSc.
- osvojování češtiny různými skupinami mluvčích (L1 i L2)
- žákovský korpusy – CZESL–PLAIN, CZESL–SGT, CZEFL–LONG

–mezinárodní žákovský korpus MERLIN (2012-2014)

–databáze CHRUP (2013)

- prof. Ph.Dr. Hana Gladkova, CSc., FF UK
- didaktický cíl: specifika osvojování L2, který je blízký L1

Žákovské korpusy

- angl. learner corpus
- studijní korpus (Čermák) = žákovský korpus
- snaha shromáždit objektivní data pro
 1. popis žákovského jazyka (interlanguage /mezijazyk)
 2. popis odchylek a rozdílů (contrastive interlanguage analysis (S. Granger); computer aided error analysis)
 3. popis mimojazykových vlivů na L2 (metadata)

Výzkum žákovského jazyka



Jak probíhá osvojování druhého/cizího jazyka?

L. Selinker (70. léta 20. století)

žákovský jazyk, mezijazyk, interlangue

svébytný jazykový útvar

proměnlivý a dynamický systém, který obsahuje řadu nestandardních prvků

- míšení kódů
- chyby
- jazykový transfer, pozitivní (kognáty) a negativní (false friends)
- naučené fráze versus improvizace

Výzkum je zaměřen na popis procesů a utváření mezijazyka či mimojazykové vlivy (např. věk).



A proč bych vám já měl sdělovat všechny moje příjmy? Sorry jako.

Čeština Andreje Babiše

- 2014: žádost o tlumočnicka při projevech Andreje Babiše ve sněmovně
- Karel Oliva: Babišova úroveň je C2, žádost zamítnuta
- Vít Benešovský (2018): Analýza Babišova mezijazyka
 - analýza přepisu rozhovorů z let 2011, 2014, 2017
 - případová studie projevu nerodilého mluvčího češtiny v českojazyčném prostředí
 - interferenční chyby (výslovnost, morfologie), slovakismy a střídání jazykového kódu
 - mírné zlepšení po roce 2014 (mimojazykové vlivy)

Nepřímé využití: žákovské korpusy

Grammar: information

Do not say "an information" or "informations." Say, for example: some information, a lot of information, or a piece of information:

- Do you have any information about tickets?
- I need a couple more pieces of information before I can finish my report.

Zdroj: The Longman Corpus Network, 2015, online.

Nepřímé využití: žákovské korpusy

👁 Spelling mistakes are among the most frequent errors made by candidates in the exam. Find and correct the spelling mistakes made by candidates in the sentences below.

1 As you can see in the advertisement, the holiday is quite cheap.

Zdroj: Brook-Hart – Haines, 2014–2015, s. 21.

Nepřímé využití: žákovské korpusy

Zadání:

Přečtěte si text, který napsal žák z Ukrajiny. Čísla v textu jsou vždy za slovem, kde udělal chybu. Řekněte, jaké jsou to chyby a opravte je.

Mám moc radá (1) sport. Hrajeme (2) moc venkovní sporty. Ráno jezdím na kole do škóly (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 hrajeme volejbal a basketbal. Golf a baseball nejsou na Ukrajině populární. Hrajeme taky kulečnik (13), šachy a stolní tenis. Myslím že, šach je dobrý duševní sport. Můžu (14) trochu hrát šachy, ale nemůžu (15) bruslit. Na podzim nesportujeme moc, protože velmi moc prší. Pak, dívám se sport na televízi (16).

Obrázek 6 Cvičení s textem z žákovského korpusu, Zdroj: Škodová a Šindelářová, 2012, online.

Nepřímé využití: obecné korpusy

<i>a) Nouns with majority endings (about 75 %):</i>			
student, kamarád, tygr consonant	banán, sýr, jogurt consonant	káva, voda, majonéza - a	auto, pivo, rádio - o
<i>b) Nouns with minority endings (about 25 %):</i>			
kolega, soudce -a (rarely) -e (rarely)	chleba -a (an exception)	televize, policie, kancelář -e (very often) consonant	nádraží, dítě, muzeum -í, -e / -ě (not very often) -um (loanwords)

L. Holá: Czech Step by Step, 2004

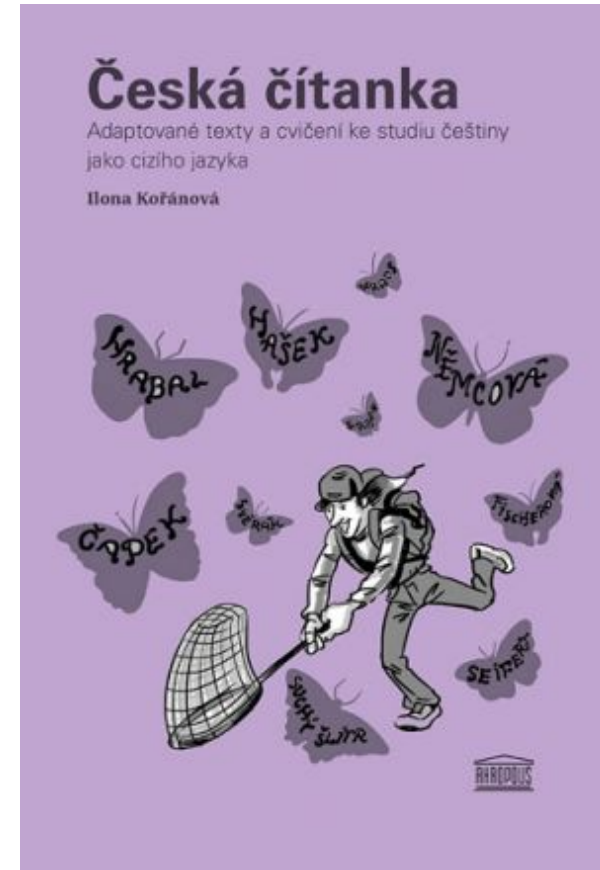
Nepřímé využití: obecné korpusy

When a word ends in a consonant, it is	76.9 % <i>masculine</i> (doktoř doctor, lékař medical doctor, obchod shop, počítač computer) 20.5 % <i>feminine</i> (kancelář office, tramvař tram) 2.6 % <i>neuter</i> (centrum centre, muzeum museum)
When a word ends in -a, it is	6.9 % <i>masculine</i> (kolegyně colleague, turista tourist) 92 % <i>feminine</i> (škola school, kavárna cafe) 1.1 % <i>neuter</i> (only loan words, e.g. téma theme)
When a word ends in -o, it is	0.7 % <i>masculine</i> (only loan words, e.g. gigolo gigolo, macho macho) 99.3 % <i>neuter</i> (divadlo theatre, kino cinema)
When a word ends in -e/-ě*, it is	3 % <i>masculine</i> (průvodce guidebook) 89 % <i>feminine</i> (restaurace restaurant, kolegyně colleague) 8 % <i>neuter</i> (letišťe airport, parkovišťe car park, kuře chicken)
When a word ends in -í, it is	0.5 % <i>masculine</i> (recepční receptionist) 0.5 % <i>feminine</i> (recepční receptionist, paní Mrs., Madam) 99 % <i>neuter</i> (nádraží station, náměstí square)

*Nouns ending in -ice and -yně are mostly feminine. Words ending in -iště are always neuter.
Compiled by SYN2015

L. Holá: Český krok za krokem, 2016.

Nepřímé využití: obecné korpusy





[Search by topic](#)

[Search by analysis](#)

[Search by dictionary](#)

[List of abbreviations](#)

[About](#)

Level

A1



Word

babička (grandma)

Show translation female voice

Byli s námi doma **babička** s dědečkem. (Nom.Sing )



Zůstanu u dědy a u **babičky** na vesnici. (Gen.Sing )



Pozdravuj **babičku**. (Acc.Sing )



Korpusy pro školy



- 2006, ÚČNK
- repozitář různých korpusových cvičení
- popularizace korpusových nástrojů, workshopy na ZŠ i SŠ
- listopad 2016 – červenec 2017: cca 3000 stáhnutí od 244 unikátních IP

Zadání

Zkuste v korpusu psaného jazyka (např. SYN2015) v rozhraní KonText (kontext.korpus.cz) vyhledat následující části idomů, doplňte jejich celé znění a na základě okolního kontextu se pokuste vysvětlit jejich význam.

a) od srdce

význam:.....

b) srdce na dlani

význam:.....

c) kámen ze srdce.

význam:.....

Metoda DDL

- studenti sami hledali v korpusu
- vybírali vhodné příklady
- sami určili význam

 učení řízené daty (data-driven learning)

Metoda DDL

1. řešení problémů
2. práce s autentickými daty
3. generalizace a kategorizace výsledků

 autonomní učení, indukativní přístup

Metoda DDL

- Tim Johns, 1991
- u nás po roce 2000 - J. Thomas, P. Vališová, A. Zasina
- student jako “Sherlock Holmes”, student jako výzkumník
- korpus jako autorita
- 2 typy: hands-on / hands-off

SkELL

- Sketch Engine - nástroj pro budování, správu a analýzu korpusových dat
- for Learning Languages: adaptace hlavních funkcí pro studium jazyka dostupných bez registrace
- slovníkové příklady, slovní profil, podobná slova



Á á Č č Ď d' É é ě Í í Ň ň Ó ó Ř ř Š š ť Ú ú ů Ý ý Ž ž

čeština 54.92 v milionu

- 1 **Čeština** už dnes nedělá takové problémy jako před lety.
- 2 Část studentů se letní školy **češtiny** účastní už poněkolkáté.
- 3 **Čeština** by měla být z logiky věci vlajkovou lodí.
- 4 Ústav jazykové a odborné přípravy se však soustředí především na výuku **češtiny** jako cizího jazyka.
- 5 Účastníci dostali k propagaci nové služby materiál i v **češtině** .
- 6 **Čeština** je zastoupena ve 3 % časopisů.
- 7 **Čeština** je totiž pro cizince velmi těžký jazyk.
- 8 **Čeština** je na tom z pohledu cizince určitě obdobně.
- 9 Úspěch má u něj hraní s **češtinou** .
- 10 **Čeština** umožňuje hrát si s termínem cena ve dvou směrech.
- 11 **Čeština** jde použít i pro novější verze.
- 12 **Čeština** je totiž pro takový slovník poměrně složitý jazyk.
- 13 Žádná z jejích knih ale zatím nebyla přeložena do **češtiny** .

čeština noun

angličtina

jazyk

němčina

literatura

překlad

podoba

styl

řeč

ruština

formát

dějiny

text

věda

obor

matematika

slovník

kultura

verze

průběh

umění

rámec

současnost

knihovna



Popularizace



Korpus do každé třídy

Přeneste platformu MERLIN do vaší třídy:

Můžete nechat své ([pokročilé](#)) studenty vyhledávat jazykové jevy v subkorpusu MERLIN, aby se s tímto nástrojem seznámili a zároveň podpořili svou samostatnost v učení se cizího jazyka. Mohou společně analyzovat jak [příklady chyb](#) nalezené v databázi MERLIN, tak své vlastní texty. Můžete je také nechat porovnat výsledky [vyhledávání v platformě MERLIN s národním korpusem](#), aby tak získali představu o rozdílech v užívání různých jazyků.

Současný stav

obecně přijímané výhody korpusového přístupu X nevyužívání korpusů ve výuce, kritika ze strany pedagogů
možné řešení:

- U. Römer (2006): user-friendly software, spolupráce mezi lingvisty a učiteli, popularizace
- S. Braun (2007): pedagogicky orientovaný korpus
- A. Boulton (2010): DDL součástí kurikula/učebnic

Kvíz



1.

Korpus je ve výuce nejčastěji využíván:

a. přímo

b. nepřímo

1.

Korpus je ve výuce nejčastěji využíván:

a. přímo

b. nepřímo

2.

Korpus, který zachycuje projevy nerodilých mluvčích, se nazývá:

- a. vyučovací korpus
- b. studentský korpus
- c. žákovský korpus

2.

Korpus, který zachycuje projevy nerodilých mluvčích, se nazývá:

- a. vyučovací korpus
- b. studentský korpus
- c. **žákovský korpus**

3.

Data pro žákovské korpusy nejčastěji vznikají:

- a. v přirozených a autentických podmínkách
- b. v uměle vytvořených podmínkách

3.

Data pro žákovské korpusy nejčastěji vznikají:

- a. v přirozených a autentických podmínkách
- b. v uměle vytvořených podmínkách**

4.

Zkratka DDL znamená:

- a. deep learning
- b. data driven learning
- c. data and definitions learning

4.

Zkratka DDL znamená:

- a. deep learning
- b. data driven learning**
- c. data and definitions learning

5.

První propagátoři korpusového přístupu John Sinclair a Tim Johns působili

- a. na univerzitě v Birminghamu
- b. na univerzitě v Bostonu
- c. na Masarykově univerzitě

5.

První propagátoři korpusového přístupu John Sinclair a Tim Johns působili

a. na univerzitě v Birminghamu

b. na univerzitě v Bostonu

c. na Masarykově univerzitě

Kam dál?

CJJ31: Tvorba elektronických výukových materiálů

PLIN022: Využití korpusů při výuce češtiny jako cizího jazyka

PLIN054: Projekt z češtiny jako cizího jazyka

Děkuji za pozornost!

