

Využití korpusů

při výuce češtiny jako cizího jazyka

PLIN022

6.

Co je to chyba?

Hned vedle základní školi se nachází i gymnásium.

Kdo se rád věnuje domácím pracem?

Češi pořád jí smažený sír.

Studenti musejí zapsat předměty do informačního systému, po prvé s tím mívají problémy.

Rádi by jsme Vás a vaše rodiče pozvali na slavnostní zahájení.

Obě mé nejlepší kamarádky jsou lékařky, jedna je chirurgka, druhá psychiatryně.

Co je to chyba?

Rád piju pivo a rád jím eggs.

Musela jsem včera vyhodit chleba, byl úplně čerstvý.

Omlouvám se, že jsem pozdě.

Dobry den

Musim jet do Prahy. Budu tam od utery do sobotu a chtela bych se zeptat pokud máte volny pokoj a taky kolík ten pokoj stoj. Nevím jestly muzu snídat v hotelu. Taky chtěla jsem se zeptat pokud máte parkoviště protože přijdu s autem. Potřebuju najít lyžařskou škola. muzete my taky napsat pokud má takovou školu blizko hotelu.

seděla jsem a vzpomněla mému prvnímu přítelu

Co je to chyba?

L1

- nesprávné užití jazykového prostředku
- odchylka od platné kodifikace

L2

- odchylka od normy i intuice rodilých mluvčích

Co je to chyba?



Larry Selinker (1972): interlanguage

- chyba = **evidence o vývoji**
- bezchybný projev nepodává dostatek informací

Pit Corder (60. léta 20. století): **mistake** x **error**

- **mistake** (chyba performanční, nesystémová; rodilí i nerodilí mluvčí, oprava mluvčím možná)
- **error** (chyba kompetenční, systémová – nepochopení pravidla, nedostatečná automatizace; nerodilí mluvčí; oprava mluvčím není možná)

Co je to chyba?

M. Hrdlička

- negativní transfer (interference) – vliv mateřského jazyka na jazyk cílový



Chybová anotace

Chyby v korpusech:

mistakes and errors, metadata, interpretace, taxonomie

= **chybová anotace**

Typy anotací:

- žádná x částečná x celková
- manuální x automatická
- lineární model x víceúrovňový distanční model

CzeSL – chybová anotace

- kombinace manuální a automatické
- vlastní anotační program **FEAT**, korpusový manažer SeLaQ
- víceúrovňový model, **zásada postupné emendace** (R0, R1, R2)
R0: *dvá další musea* – R1: *dva další musea* – R2: *dvě (agr→musea) další musea*
- pravidlo kontextu a pravidlo minimálního zásahu
(někdy nestačí, např: *má hezkou tužky* – *má hezké (←) tužky (~~x má hezkou tužku~~)*)
- na základě **porovnání původní a opravené podoby tvarů**
je stanoven druh chyby → možné nepřesnosti

Anotační program FEAT

feat 201005242201

File View Tools Window Help

Kar_DA_001_2_ST.b KRIS_DB_003_ST.b Tou_H209_087_ST.b

2 / 4 Test Refresh Layout Export Spacing X: 100 Y: 100

R0: Býdlíme ve svém domě .
unkStem X X X X
HRANA

R1: Bydlíme ve svém domě .
X X X X
UZEL

R2: Bydlíme ve svém domě .

Máme hezkých psa a kočku Rása psa
X X X X X X
unkStem X
Máme hezkých psa a kočku Rása psa
X X X X X X
agr X
Máme hezkého psa a kočku Rása psa
X X X X X X
CHYBOVÝ KÓD **ODKAZ** **PAHÝL**

Moje rodina.

PŘEPIS

Má rodina není velká. Bydlím s otcem a mátkou. Jsem jeden syn v rodině. Ještě mám 2 nevlastné sestry. Jedna má manžela a dceru. Její dcera je má něteř. Studuje ve střední škole. Druhá sestra bydlí v inem městě. Me rodič jsou úředníci. Mají malý bu-biznes. Býdlíme v domě ve svém domě. Máme hezkých psa a kočku. Rása psa je „dvůrterjěr“. Tak jmenuje ho můj tatínek. Pes se jm(e)ě muje „Šarik“. Kočka se jmenuje „Bars(j)e k“. Máme ještě velkou zahrádu. Tam je malý proud – rybník. S zlatými rybami. Má Můj otec je Sjs, a proto býdlíme v domě i nebýdlíme v bytě. Má maminka vaří moc dobře. Chutné mi j jídlo.

<https://jakobson.korpus.cz/~rosen/public/anotace.pdf>

R1

- výsledkem opravy je posloupnost správných tvarů existujících spisovných českých slov
- opravy týkající se jednotlivých slov a slovních tvarů
- chyby hláskové, slovotvorné, tvarotvorné

*kostích – kostech, chlapca – chlapce,
dědečka – dědečka, kadeřníčka – kadeřnice,
napisovat – napsat, dobrej – dobrý, mam – mám*

*on mi ho **dál** – **dal**, jeli jsme na **jich** – **jih**,
ted' **mužů** říct – **můžu***

<https://jakobson.korpus.cz/~rosen/public/anotace.pdf>

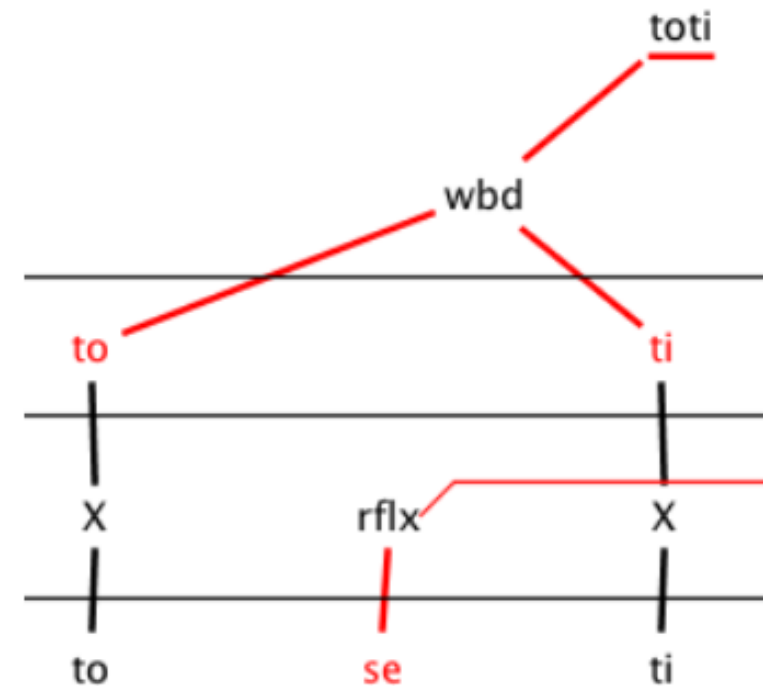
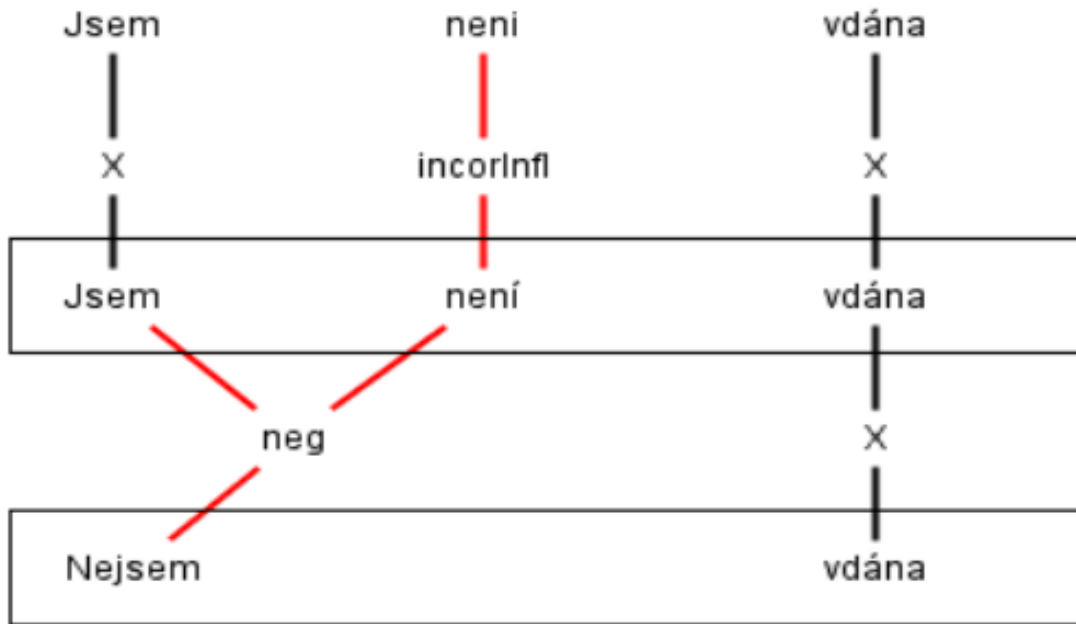
R2

- výsledkem je gramaticky správná věta
- opravy týkající se více slov, anotují se chyby plynoucí ze vztahů mezi (izolovaně) správnými tvary na R1
- důvodem chyby je narušení vztahu se slovem jiným (shoda, valence, slovosled, dlouhá forma osobního zájmena v příklonkové pozici ap.)

pro maminka – pro maminku
včera pršeli – včera pršelo
až mi to dal, půjdeme – až mi to dá, půjdeme
my by jsme rádi – my bychom rádi
Petr bude dělá – Petr bude dělat

<https://jakobson.korpus.cz/~rosen/public/anotace.pdf>

- složitější vztah, tzv. pavouk
- jednomu (více slovům) odpovídá jedno (více slov) na rovině sousední



<https://jakobson.korpus.cz/~rosen/public/anotace.pdf>

Manuální anotace

Tabulka 7: a. Chyby značkové na rovině 1 (R1)

TYP CHYBY	POPIS	PŘÍKLAD
<i>incorInfl</i>	nesprávná flexe	<i>spám málo; tři měsícu; bydlím s matkoj</i>
<i>incorBase</i> ¹⁹⁶	nesprávný slovní základ	<i>kočka se jmemuje; libila se mi; musíš to posvětlit</i>
<i>fwFab</i>	neemendovatelné, „vymyšlené“ slovo	<i>je tam hodně jinaků; slyšet smášky</i>
<i>fwNC</i>	cizí slovo	<i>jím rád eggs; byla v hangu</i>
<i>flex</i>	doplňující příznak u chyb <i>fwFab</i> a <i>fwNC</i> značící přítomnost flexe	<i>jdu do shopa</i>
<i>wbdPre</i>	prefix oddělený mezerou a předložka bez mezery	<i>musíš to při pravít; jede dolesa</i>
<i>wbdComp</i>	neoprávněně rozdělená kompozita	<i>český anglický slovník</i>
<i>wbdOther</i>	jiná chyba týkající se hranice slova	<i>mochezky; atak; kdy koli</i>
<i>stylColl</i>	obecněčeský tvar	<i>dobrej film</i>
<i>stylOther</i>	knižní, nářeční, slangový, hyperkorektní výraz	<i>holka s hnědými očimi</i>
<i>problem</i>	problémová chyba (doplňkový příznak)	

b. Chyby značkové na rovině 2 (R2)

TYP CHYBY	POPIS	PŘÍKLAD
<i>agr</i>	narušení shody	<i>máme hezkých psa; Jana čtu</i>
<i>dep</i>	chyba ve vyjádření syntaktické závislosti	<i>věřím učitelku; otázka čas; bojím se jí zavolám</i>
<i>ref</i>	chyba v zájmeném odkazu	<i>paní, jenž jsem potkal</i>
<i>vbx</i>	chyba v analytickém slovesném tvaru a složeném přísudku	<i>Jana bude dělá; guláš bylo chutná mi; kluci jsou běhali</i>
<i>rflx</i>	chyba v reflexivním výrazu	<i>smála si; narodila jsem v Petrohradu</i>
<i>neg</i>	chyba v negaci	<i>mám žádný čas; on ne velký</i>
<i>lex</i>	chyba v lexiku a frazeologii	<i>jsem Vietnam; dopadlo to přírodně</i>
<i>use</i>	chyba v užití gramatické kategorie	<i>tričko je nejvíc nejhezčí; celé dopoledne uvařím oběd</i>
<i>sec</i>	sekundární, „zavlečená“ chyba (doplňkový příznak)	<i>dívá se na americkém filmu</i>
<i>stylColl</i>	obecněčeský tvar	<i>viděli jsme hezký holky</i>
<i>stylOther</i>	knižní, nářeční, slangový výraz	<i>zvedl se mi kufr</i>
<i>stylMark</i>	výplňkové slovo jako „diskurzivní marker“	<i>no, teda, jo</i>
<i>disr</i>	rozvrácená konstrukce	<i>kratka jakost vyborné ženy</i>
<i>problem</i>	problémová chyba (doplňkový příznak)	

Manuální anotace

Štindlová, B. (2013), s. 130.

CzeSL-SGT

(korpus češtiny nerodilých mluvčích s **automaticky provedenou anotací**)

<https://jakobson.korpus.cz/~rosen/public/2014-czesl-sgt-cs.pdf>

2 sady atributů:

- *word, lemma a tag*
- *word1, lemma1 a tag1*
 - pokud původní tvar *word* byl vyhodnocen jako chybný
 - vztahují se k cílové **hypotéze** – tj. tomu tvaru, který se pravděpodobně pokusil nerodilý mluvčí vytvořit

CzeSL-SGT

<https://jakobson.korpus.cz/~rosen/public/2014-czesl-sgt-cs.pdf>

- **word, lemma a tag**
- **word1, lemma1 a tag1**
- **gs** (pravopisná nebo gramatická chyba)
- **err** (typ chyby)

word	lemma	tag	word1	lemma1	tag1	gs	err
Tén	Tén	X@-----	Ten	ten	PDYS1-----	S	Quant1
pes	pes	NNMS1-----A----	pes	pes	NNMS1-----A----		
míluje	míluje	X@-----	miluje	milovat	VB-S---3P-AA---	S	Quant1
svécho	svécho	X@-----	svého	svůj	P8MS4-----	S	Voiced
kamarada	kamarada	X@-----	kamaráda	kamarád	NNMS4-----A----	S	Quant0
-	-	Z:-----	-	-	Z:-----		
člověka	člověk	NNMS2-----A----	člověka	člověk	NNMS4-----A----		
.	.	Z:-----	.	.	Z:-----		

Tabulka 1: Ukázka anotace jedné věty

Automatická anotace

Korpus CzeSL – automaticky identifikované chyby

Typ chyby	Popis chyby	Příklad
Podrobně rozlišené chyby		
CK0	chybně zaměněné <i>c</i> → <i>k</i> , mimo palatalizaci (<i>c</i>)	<i>Atlantic</i> → <i>Atlantik</i>
CK1	chybně zaměněné <i>k</i> → <i>c</i> , mimo palatalizaci (<i>k</i>)	<i>UNESKO</i> → <i>UNESCO</i>
Cap0	chybně použité malé písmo	<i>evropě</i> → <i>Evropě</i> ; <i>štedrý</i> → <i>Štedrý</i>
Cap1	chybně použité velké písmeno	<i>Staré</i> → <i>staré</i> ; <i>Rodině</i> → <i>rodině</i>
Caron0	chyba v diakritice – chybí háček	<i>vecí</i> → <i>věcí</i> ; <i>sobe</i> → <i>sobě</i>
Caron1	chyba v diakritice – háček navíc	<i>břečel</i> → <i>brečel</i> ; <i>bratřem</i> → <i>bratrem</i>
DiaE	chyba v diakritice – <i>ě</i> → <i>é</i> , popř. <i>é</i> → <i>ě</i>	<i>usměvavé</i> → <i>usměvavé</i> ; <i>poprvě</i> → <i>poprvé</i>
DiaU	chyba v diakritice – <i>ú</i> → <i>ů</i> , popř. <i>ů</i> → <i>ú</i>	<i>nemůžeš</i> → <i>nemůžeš</i> ; <i>úkoly</i> → <i>úkoly</i>
Dtn	chyba v psaní <i>dě</i> / <i>tě</i> / <i>ně</i> , <i>di</i> / <i>ti</i> / <i>ni</i>	<i>ňikdo</i> → <i>nikdo</i> ; <i>ješťerka</i> → <i>ještěrka</i>
EpentE0	chyba v epentezi <i>e</i> (chybí <i>e</i>)	<i>najdnou</i> → <i>najednou</i> ; <i>domček</i> → <i>domeček</i>
EpentE1	chyba v epentezi <i>e</i> (chybně <i>e</i> navíc)	<i>rozeběhl</i> → <i>rozběhl</i> ; <i>účety</i> → <i>úcty</i>
EpentJ0	chybí <i>j</i> po <i>i</i> před vokálem	<i>napie</i> → <i>napije</i>
EpentJ1	chybně vložené <i>j</i> po <i>i</i> před vokálem	<i>dijamant</i> → <i>diamant</i>
GH0	chybně zaměněné <i>g</i> → <i>h</i> , mimo palatalizaci (<i>g</i>)	<i>glavní</i> → <i>hlavní</i>
GH1	chybně zaměněné <i>h</i> → <i>g</i> , mimo palatalizaci (<i>h</i>)	<i>johurt</i> → <i>jogurt</i>

<https://jakobson.korpus.cz/~rosen/public/SeznamAutoChybR0R1.html>

seděla jsem a vzpomněla mému prvnímu příteli

Problémy

1

chybová analýza, kontrastní analýza = cílení na jeden typ chyby

X

korpusový výzkum mezijazyka = komplexní pohled

2

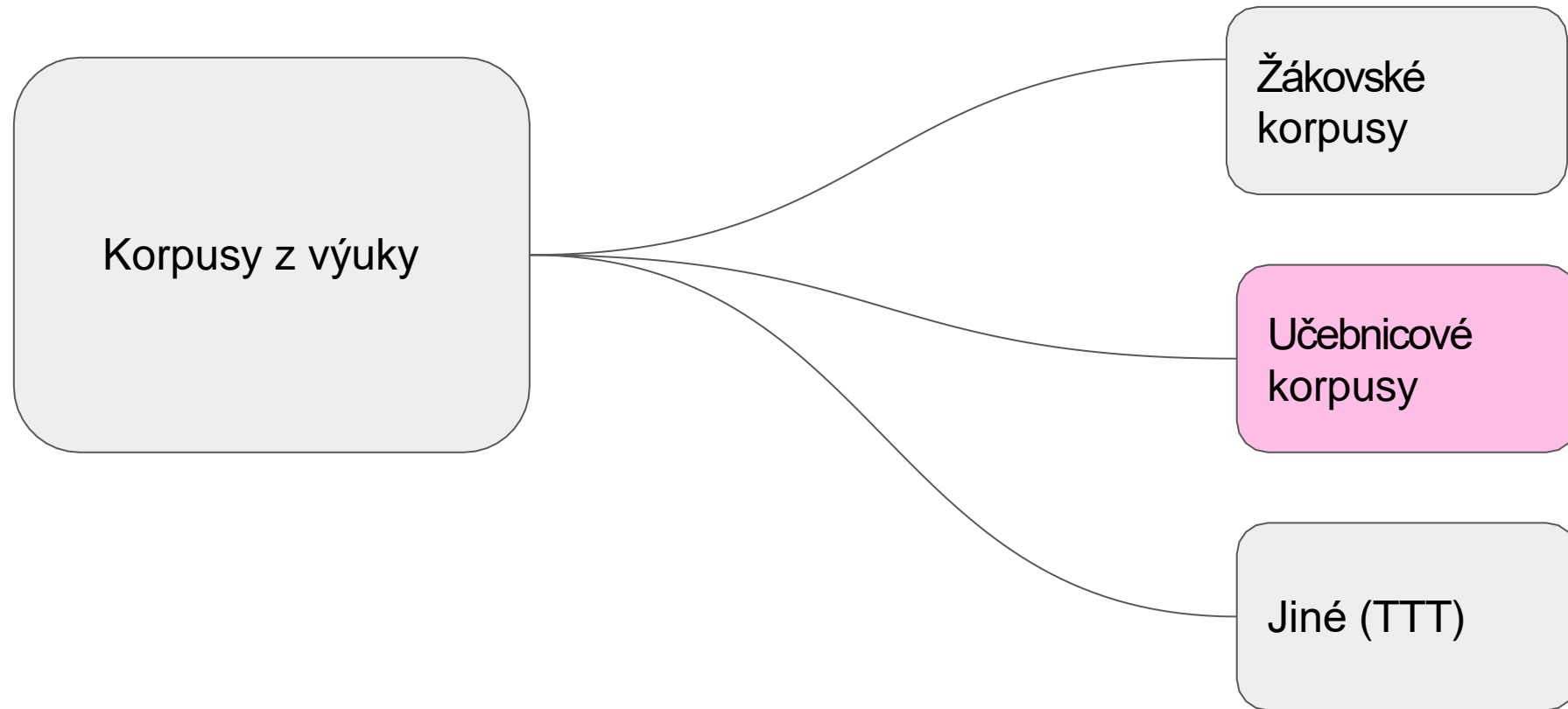
Oprava chyby = interpretace, pohled jiného mluvčího



Korpusy pro výuku



Korpusy z výuky



Učebnicové korpusy



- po roce 2000, Ute Römer
- nahrazení klasické analýzy učebnic
- obsah
 - struktura
 - design
 - čtení, psaní, mluvení, poslech
 - metajazykové informace
- učebnicový jazyk

Učebnicové korpusy – využití

- 1) **deskriptivní** – učebnicový jazyk jako specifický žánr
- 2) **didaktické** – zkoumají typy cvičení

German English as a Foreign Language Textbook Corpus (GEFL TC, 2004)

Ute Römer(ová) (2004, 2006)

srovnává tento učebnicový jazyk s autentickým jazykem z britských nár. korpusů

Učebnicové korpusy – využití

1) deskriptivní – učebnicový jazyk jako specifický žánr

- German English as a Foreign Language Textbook Corpus (GEFL TC, 2004)

Ute Römer(ová) (2004, 2006):

srovnává učebnicový jazyk s autentickým jazykem z britských nár. korpusů

- metajazyk – instrukce ke cvičením, lingv. terminologie
- podíl mluvenosti/psanosti (resp. v č.j. spis.j./obecná čeština/dialekty)

Učebnicové korpusy – využití

2) didaktické – zkoumají typy cvičení (podle jevu, tématu, referenční úrovně)

- Gouverner(ová) (2008):
studie ukázala, že učebnice obsahují málo pozorovacích cvičení

Jazyk pod lupou: Subjekt x objekt
Doplňte z textu:

	<i>subjekt (nominativ singuláru):</i>	<i>objekt (akuzativ singuláru)</i>	<i>koncovky</i>
<i>Maskulinum inanimatum</i>	Černý čaj je dobrý.	Dám si černý čaj.	---
<i>Femininum</i>	Černá káva je dobrá. Dušená ryže je dobrá.	Dám si _____ Dám si _____	-á > _____ -a > _____ -e > _____
<i>Neutrum</i>	Černé pivo je dobré.	Dám si _____	---

Analýza učebnic

Gramatika	Drliková (2008)	vid	4	A1-B2	bakalářská práce
	Hrdlička (2009)	česká mluvnice	33	A1-B2	monografie
	Lukšija (2010)	místní předložky	2	A1-A2	bakalářská práce
	Vališová (2011)	slovesné třídy	23	A1-A2	diplomová práce
	Bičišťová (2011)	číslovky	-	-	bakalářská práce
	Lukšija (2012)	deklinace substantiv	10	A1-A2	diplomová práce
	Kodajková (2012)	imperativ, kondicionál, pasivum	9	A1-B1	diplomová práce
	Kočařová (2013)	rod	5	A1	bakalářská práce
	Koláčková (2013)	imperativ	8	A1-B1	bakalářská práce
	Stočková (2015)	syntax	7	A1	bakalářská práce
	Lexikum	Kovaříková (2009)	obecná čeština	-	-
Nezvalová (2010)		slovotvorba	3	A1-A2	bakalářská práce
Nováková (2010)		obecná čeština	4	B1-B2	diplomová práce
Ocztošová (2012)		SZ pro německy mluvící	-	-	bakalářská práce
Metody	Hrdlička (2009)	komunikační metoda	33	A1-B2	monografie
	Hladíková (2011)	signální gramatika	4	A1	diplomová práce
	Šimová (2012)	komunikační metoda	-	-	diplomová práce
	Valková (2014)	komunikační metoda	4	A1-A2	monografie
Metajazyk	Hradilová (2010)	terminologie jmenné flexe	3	A1-A2	odborný článek
	Osolsobě a Vališová (2010)	lingvistická terminologie	17	A1-A2	odborný článek
Text	Volnogradská (2012)	typy textů	3	A1	diplomová práce
	Sedláček (2013)	vyváženost textů	10	A1-B1	diplomová práce

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!

A pink rounded rectangle with a thin black border, containing the text "Korpusy pro výuku".

Korpusy pro výuku

A gray rounded rectangle with a thin black border, containing the text "Korpusy z výuky".

Korpusy z výuky

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.

Pozorování

Pedagogové skoro ani nevěděli, že ho ve **třídě** mají.
Mercedes **třídy** B je mnohem příjemnější auto než předchůdce.
Na Budějovickou nebo na **třídu** 9 . května by se třeba hodil.
Na **třídě** Maršála Malinovského přecházeli ke kinu Hvězda chodci.
V 1. **třídě** zvyšují komfort polohovatelná sedadla se zvětšeným prostorem a elektrické zásuvky.

Jazyk pod lupou: Subjekt x objekt

Doplňte z textu:

	<i>subjekt (nominativ singuláru):</i>	<i>objekt (akuzativ singuláru)</i>	<i>koncovky</i>
<i>Maskulinum inanimatum</i>	Černý čaj je dobrý.	Dám si černý čaj .	---
<i>Femininum</i>	Černá káva je dobrá. Dušená ryže je dobrá.	Dám si _____ Dám si _____	-á > _____ -a > _____ -e > _____
<i>Neutrum</i>	Černé pivo je dobré.	Dám si _____	---

řízené objevování

Obrázek 14 Gramatická tabulka v učebnici *Čeština expres A1/1* (Bořilová a Holá, 2010)

Převzato z Vališová, P. (2017), dizertační práce.

Doplňování

MÝT X ČISTIT

Po nocích nádobí, abych mohl ve dne chodit do školy.

Přisedící hleděl před sebe a nervózně si brýle.

Když si obličej, vchází Jody a podává jí ručník.

Maritza mi rány a dávala mě dohromady.

Sekám trávník a okna, až se svítí.

Když zazvonil telefon, Denis si zuby.

Produkce

dům	rodinný, bytový, panelový, kulturní, obchodní, pasivní, obytný, cihlový, nízkoenergetický, činžovní, rodný, řadový, parkovací, zděný, měšťanský, patrový
	stavět, postavit, prodat, koupit, zateplit, rekonstruovat, vytápět, pronajmout, zbourat, opravovat, zdědit, obývat
	ulice, lokalita, centrum, část, čtvrť, město, zástavba, vesnice, styl, stav, sousedství, standard, budova
	výtah, zahrada, střecha, garáž, podkroví, pozemek, fasáda, okno, dispozice, bazén, prostor, výhled, číslo, spotřeba, nábytek, vchod, terasa, dvůr, novostavba

Literatura

Návod k anotaci chybového korpusu, verze 5.1 (9.3.2012).

Dostupné z: <https://jakobson.korpus.cz/~rosen/public/anotace.pdf> (cit. 10. 9. 2024)

Štindlová, B. (2013). Žákovský korpus češtiny a evaluace jeho chybové anotace. Univerzita Karlova v Praze, Filozofická fakulta, Praha.

