

CJDSL001 Korpusová lingvistika (4)

Klára Osolsobě

osolsobe@phil.muni.cz

Experimentální a počítačová lingvistika

O čem budeme mluvit

- Krátký historický exkurz
- Definice korpusu v moderním slova smyslu
- Dva metodologické přístupy k vytěžování korpusu
- Dva pohledy na korpus (lingvista a informatik)
- Filologie a korpusy
- **Výuka jazyků a korpusy**

Korpusy a jejich využití ve výuce

- Metoda DDL (Tim Johnes)
- Žákovské korpusy (Learner Corpora)
- Učebnicové korpusy

DDL / Data Driven Learning

- Termín DDL byl zaveden britským korpusovým lingvistou Timem Johnsem, jenž jako první využíval korpusy ve výuce angličtiny. Tato metoda je založena na poznatcích studenta, který objevuje cizí jazyk na základě korpusových dat. Učitel je zde pouze koordinátorem studentovy práce, který zajišťuje vhodné prostředí, v němž bude student schopen dospívat k lingvistickým závěrům. Jedná se o rozvinutí studentových schopností učit se jak se učit (*learn how to learn*). Student je také vědcem, jenž je veden korpusovými daty k tomu, aby vyvodil závěry a dospěl k obecným gramatickým pravidlům.
- Na rozdíl od ve výuce běžného deduktivního přístupu je DDL přístupem induktivním, který se soustředí v první řadě na pozorování korpusových dat, na základě kterých student klasifikuje jazykové jevy a následně generalizuje gramatická pravidla. Učitel přesně neví, k jakým závěrům dospěje student, proto koordinuje studentovu práci a jeho pozice je spíše partnerská než autoritativní.

HARD/SOFT

- *DDL hard* – (též *hands-on*) kdy studenti pracují přímo s korpusovým výhledávačem
- *DDL soft* – (též *hands-off*) kdy studenti řeší úkoly připravené na základě korpusových dat, např. ve formě vytištěných konkordancí.
- Výhody/nevýhody obou postupů

Některé nástroje dostupné v českém prostředí

- **Slovo v kostce** (WaG) poskytuje rychlý a základní přehledu o tom, jak se používá zadané slovo. Informace z korpusů ČNK pomocí dostupných nástrojů podávají přehledný profil slova z různých perspektiv. (<https://www.korpus.cz/slovo-v-kostce/>)
- Portál UČNK pro školy (<https://korpus.cz/proskoly>)
- VŠ učebnice

WaG

- Co vás napadne, když se řekne *reprodukce*?
- Jak vypadá profil uvedeného slova nabízený nástrojem WaG a jak vypadá profil téhož slova nabízený nástroji Sketch Engine?
- Který z nástrojů byste preferovali ve výuce?

Typy a výhody korpusových cvičení













- vycházejí z autentických psaných i mluvených textů
- jsou vhodná pro žáky ZŠ i studenty SŠ ve výuce češtiny i cizích jazyků
- jsou ideálním domácím úkolem, který bude studenty bavit
- pracovní listy lze snadno stáhnout i vytisknout jako zadání
- více informací najdete [v manuálu](#)
- Jak citovat:
Lukešová, L. (ed.) (2017). Pro školy - repozitář korpusových cvičení. FF UK, Praha. Dostupný z WWW: <<http://korpus.cz/proskoly>>.

Možnost pracovat s připravenými cvičeními/ inspirace pro rozvoj vlastních nápadů

Hledejte pomocí štítků

Kliknutím vyberte, dalším kliknutím zrušte

ČJ ČČJ AJ NJ FJ papír počítač projektor slovotvorba tvarosloví slovní zásoba skladba pravopis stylistika mix A1
A2 B1 B2 C1 rodilý 6.-7. roč. 8.-9. roč. SŠ pozorování porovnávání doplňování spojování hledání mix psaní
mluvení zařazování

22. Česká přísloví	22. 10. 2018	ČČJ papír počítač slovní zásoba B2 C1 spojování hledání	  
21. Moderní elektronika - slovní zásoba	22. 10. 2018	ČČJ papír slovní zásoba B2 C1 pozorování doplňování mix	  
20. Česká přirovnání	22. 10. 2018	ČČJ počítač slovní zásoba B2 C1 doplňování hledání	  
19. Rozdíly ve významu slov	15. 5. 2018	ČČJ počítač slovní zásoba B2 C1 porovnávání hledání	  

Žákovské korpusy / Learner Corpora

- Žákovské korpusy (learner corpora) obsahují texty většinou nerodilých mluvčích, kteří se jazyk teprve učí. Povětšinou se jedná o [psané](#) (školní) práce, zřídka nacházíme [mluvené](#) žákovské korpusy.
- Význam žákovských korpusů vzrůstá nejen s jejich rozsahem ale také s jejich [anotací](#). Pokud je součástí jejich metadat také tzv. chybová anotace, je možné klasifikovat jednotlivé odchylky od běžného úzu, zjišťovat jejich četnost, příp. i motivaci. Součástí metadat bývá také rozšířená informace o pisateli, která je často obohacena o relevantní údaje vztahující se k jeho věku, mateřskému jazyku, úrovni jeho znalostí cílového jazyka apod.
- Žákovské korpusy zachycují jazyk mluvčích/pisatelů, kteří dosud nezvládají cílový jazyk dokonale; v tomto smyslu je možné je chápat jako specifický případ [korpusů akvizičních](#). V rámci ČNK jsou k dispozici žákovské korpusy [CzeSL-plain](#) a [CzeSL-SGT](#).

Učebnicové korpusy

- Specializované korpusy vytvořené z učebnic
- Obsahuje soubor všech slov a frází, se kterými student přišel do styku při práci s učebnicí (může se k nim vracet) a nahrazuje a automatizuje klasickou analýzu učebnic
- Využití: a) výzkum slovní zásoby a komparace s autentickým územ v národních korpusech; b) studium osvojování jazyka (je možné vyhledat všechny kontexty, ve kterých se slovo či fráze vyskytují v učebnicích různých úrovní); c) výběr slovní zásoby pro jednotlivé úrovně a porovnání stavu v učebnicích a v SERR; d) porovnání inputu v podobě textů, poslechů i cvičení; e) porovnáním kolokací slov ve cvičeních v různých učebnicích; f) výzkum metajazyka /terminologie užívaného v učebnicích; g) zastoupení různých jazykových variant v učebnicích; h) zkoumání různých gramatických jevů (máme-li k dispozici tagovaný korpus)

UČKo – příklad učebnicového korpusu

1	Adamovičová, A. – Ivanovová, D. (2007): <i>Basic Czech I, II</i> . Praha: Karolinum.
2	Bischofová, J. et al. (2007): <i>Čeština pro cizince a azylanty B1</i> . Brno: SOZE.
3	Bořilová, P. – Holá, L. (2011): <i>Česky krok za krokem 2</i> . Praha: Akropolis.
4	Cvejnová, J. (2011): <i>Česky, prosím</i> . Praha.
5	Čechová, E. – Remediosová, H. (2005): <i>Chcete mluvit česky? 1, 2</i> Liberec: Harry Putz.
6	Froulíková, L. (2008): <i>Adam a Eva v Českém ráji</i> . Praha: Academia.
7	Hádková, M. (2005): <i>Čeština pro cizince a azylanty A1, A2</i> . Brno: SOZE.
8	Holá, L. (2006): <i>New Czech Step by Step</i> . Praha: Akropolis.
9	Holá, L. (2010): <i>Čeština Express 1, 2</i> . Praha: Akropolis.
10	Kestránková, M. et al. (2010): <i>Čeština pro cizince B1</i> . Brno: Cpress.
11	Matula, O. (2007): <i>Český den. Kurz českého jazyka pro azylanty navazující na Manuál pro učitele českého jazyka pro cizince bez znalosti latinky</i> . Praha: Člověk v tísni o.p.s., Projekt Varianty.
12	Nekovářová, A. (2006): <i>Čeština pro život - 15 moderních konverzačních témat</i> . Praha: Akropolis.
13	Parolková, O. (2004): <i>Czech for foreigner s 1, 2</i> . Praha: Bohemika.
14	Pintarová, M. – Režková, I. (1995): <i>Communicative Czech. Elementary Czech. Intermediate Czech</i> . Praha: Univerzita Karlova.
15	Rigerová, K. (2000): <i>Czech for Everyone</i> . Praha: K. Rigerová.
16	Štindl, O. (2008): <i>Easy Czech. Elementary</i> . Praha: Akronym.
17	Štindlová, B. (2008): <i>Česky v Česku 1, 2</i> . Praha: Ústav jazykové a odborné přípravy UK: Akropolis.

Korpus UčebKo

- Korpus učebnic češtiny pro cizince
- Součást projektu SLAN (Slovotvorný analyzátor pro výuku češtiny pro cizince) na PedF MU
- Zpracování české a ruské slovotvorby pro potřeby výuky češtiny jako druhého/cizího jazyka.
- Budování nástroje CZed

Úkol

Pravidla českého slovosledu, zejména postavení auxiliáru a významového slovesa v českém minulém čase, jsou dosti obtížná a cizinci, kteří se učí česky, v této oblasti chybují.

Pokuste se v Žákovském korpusu najít příklady chybného slovosledu uvedeného typu.

https://www.korpus.cz/kontext/view?maincorp=cze-sl-plain&viewmode=kwic&pagesize=40&attrs=word&attr_vmode=visible-kwic&base_viewattr=word&structs=doc&refs=doc%2C%3Ddoc.name%2C%3Ddoc.group&q=~Ro yWCuW8GwiY

Pozdě **napsal jsem** ředitelkě omluvenku , proto ne budu na univerzitě .

`<doc name=UJA2_LM_003_t_1 group=ciz>` Minulý víkend **měl jsem** narozeniny .

Odpoledne **pojela jsem** hořku polevku a potom divala se na televizi už do večera .

Tak **zůstala jsem** tam a viděla jsem mnoho hudebních představeních a taky jsem uviděla slovinský prezident .

Ale když šla jsem na vyučování **rozuměla jsem** že teď musíme hodně studovat `</doc>`

Fakt , ČR mě překvapila , protože je velmi krásná a co se mi líbí moc " česká příroda " , že moc mám ráda stromy , řeky a přírodní přehledy a opravdu tady **nechala jsem** téty věci .

Tam **dělali jsme** piknik a večer se vrátili do kempingu .

Mám velkou rodinu , tak **nemohla jsem** mít naději , že něco dostanu .

Trochu **pracoval jsem** jako budovatel , noční hlídač a nakladatel , ale to netrvalo příliš dlouho .

Jsem moc šťastná , že jsem se mohla zúčastnit tohoto kurzu , díky kterému **byla jsem** v místě , kam je docela těžko jet .

Dofam , že budu spím brzo ale každý večer **měla jsem** moc práce potom nikdy neshela jsem `</doc>`

Na konci prazdnin **začala jsem** být našťavaná , protože se přiblížilo termín cestování , protože vím dobře , že to bude velmi těžký pro mě , že žiju na zemi a moje rodinu na jině .

Minulý víkend **měla jsem** velkou procházku po Praze se svými kamaradami .

Bohudík byla sleva ve obchodech , tak **mohla jsem** jít do jednoho z nich a mimochodem jsem se schovala před deštěm .

Potom **jeli jsme** do zábavného parku a nastoupili jsme mnoho houpaček .

Kamarádi **darovali jsme** mi velký obraz , který chtel ja .

Tam , v budoucnosti , už jsem vystudoval na vysoké škole a začal pracovat jako novinář , nebo **studoval jsem** kromě toho ještě na psychologa a pak psál články o psychology a namahal líde zvládat s jejich vnitřními problémy .

Nástroj SKELL

(<https://www.sketchengine.eu/skell/>)

Jste zvyklí užívat SKELL?

Máte zkušenosti s případy, kdy je třeba vysvětlit, jak funguje a co vlastně zprostředkuje?

Podívejte se do nástroje SKELL na slova jako svízel, modrák, hořčák, bledule, ...

Příklady užití



svízel 0.74 hits per million

1. **Svízel** – v tomto případě v podobě finanční újmy má podobu nejružnějších oprav v domě či bytě.
2. Ovšem předcházela tomu **svízel** v podobě nedostatečného kapitálu pro vstup do elitní španělské ligové soutěže.
3. Přes veškeré tyto počáteční **svízele** s provozem nové požární stanice musela být zachována plná pohotovost.
4. Jaké tedy byly vůbec možnosti záchrany posádky v případě havárie rakety nebo jiných **svízeli** ?
5. Už zmíněný zákon o sociálním bydlení v tomto ohledu představuje jen malou **svízel** .
6. Je to dnes s hledáním práce opravdu **svízel** a autorce seriálu článků fandím.
7. Podstata veškerých **svízeli** se dokonce datuje už o třináct roků dříve.
8. V krátkosti nelze zdaleka postihnout všechny komplikace a **svízele** průzkumné akce.
9. Žena zdaleka není jediná, kdo nyní řeší podobnou **svízel** .
10. Kdykoli v případě nějakých **svízeli** tedy mohli zasáhnout sami kosmonauti.
11. Jenže tato činnost ve velkých městech sebou nese nejednu **svízel** .
12. Přes různé **svízele** se dostanou až k jezeru s polodem

WS

svízel

Čeština

Examples Word sketch Similar words

svízel

noun



Show context



modifiers of "svízel"

- [syřišťový](#) svízel syřišťový
- [vonný](#) svízel vonný (Galium odoratum
- [sudetský](#) . Svízel sudetský

prepositional phrases

- ["svízel" s ...](#)
- [... na "svízel"](#)

Podobná slova

svízel noun ▾



strážeň protivenství violka hvozdík příkoří strast ústrk nesnáz soužení nebezpečnoství běda prvosenka nepohodlí vstavač strádání
náprstník zábělník těžkost vachta lopuch trampota útrapa hrachor kruštík starček přeslička trabl kakost kostival patálie



Závěr

- Využití korpusů ve výuce
- Využití korpusů pro zlepšení výuky
- Tvorba specializovaných nástrojů spojených s výukovými potřebami
- Tvorba specializovaných korpusů spojených s výukovými potřebami

Vytvořte skupiny po třech a pokuste se

- navrhnout jednoduchý výukový materiál, v němž byste využili korpusové nástroje, o nichž jsme mluvili v průběhu kurzu
- najít žákovské korpusy jazyka, který byste mohli (máte aprobaci) vyučovat
- najít učebnicové korpusy jazyka, který byste mohli (máte aprobaci) vyučovat

Seminární práce

Jak souvisí korpusová lingvistika s tématem disertačního projektu.
Jaké korpusy užívám/buduji a jaké problémy řeším.

Děkuji vám za pozornost