

ADRIAN JAN ZASINA

**KORPUSOVÁ CVIČEBNICE
PRO STUDENTY ČEŠTINY
JAKO CIZÍHO JAZYKA**

KAROLINUM

Korpusová cvičebnice pro studenty češtiny jako cizího jazyka

Adrian Jan Zasina

Recenzovaly:

Mgr. Svatava Škodová, Ph.D.

PhDr. Marie Koprřivová, Ph.D.



Financováno
Evropskou unií
NextGenerationEU



Národní
plán
obnovy



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

Publikace byla vydána za podpory Ministerstva školství, mládeže
a tělovýchovy a Národního plánu obnovy v rámci projektu
Transformace pro VŠ na UK (reg. č. NPO_UK_MSMT-16602/2022).

Vydala Univerzita Karlova
Nakladatelství Karolinum
Praha 2023
Redakce Eva Flanderková
Grafická úprava Jan Šerých
Sazba Stará škola
Vydání první

© Univerzita Karlova, 2023
© Adrian Jan Zasina, 2023

ISBN 978-80-246-5146-0
ISBN 978-80-246-5464-5 (pdf)



Univerzita Karlova
Nakladatelství Karolinum

www.karolinum.cz
ebooks@karolinum.cz

ÚVOD -----	7
JAK ZAČÍT S KORPUSEM? -----	9
ÚVODNÍ CVIČENÍ -----	10
KORPUSOVÉ NÁSTROJE -----	13
1. VÝZNAM SLOV NA ZÁKLADĚ KONTEXTU/ROZDÍLY VE VÝZNAMU -----	15
2. GRAMATICKÝ ROD -----	29
3. SKLOŇOVÁNÍ -----	33
NOMINATIV -----	33
GENITIV -----	37
DATIV -----	40
AKUZATIV -----	44
VOKATIV -----	48
LOKÁL -----	50
INSTRUMENTÁL -----	54
NOMINATIV VS. AKUZATIV PLURÁLU ŽIVOTNÝCH MASKULIN -----	57
4. SUBSTANTIVIZOVANÁ ADJEKTIVA -----	60
5. TVRDÁ VS. MĚKKÁ ADJEKTIVA -----	63
6. ZÁJMENA -----	66
7. PŘEDLOŽKY -----	70
8. PŘEDPONY S-/Z- -----	73
9. PŘÍČESTÍ MINULÉ -----	76
10. KOLOKACE -----	80
11. FRAZÉMY -----	82
12. STYLOVÉ VARIANTY: MLUVENÁ VS. PSANÁ ČEŠTINA -----	84
13. KONKURENCE KONCOVEK GSG MI A- A U- -----	88
14. KVANTITA -----	90
KLÍČ -----	95
SUMMARY -----	134
BIBLIOGRAFIE -----	135

TYPY CVIČENÍ



Přímá práce s korpusem (s počítačem)



Nepřímá práce s korpusem (bez počítače)



Nekorpusové doplňující cvičení

ÚVOD

Tato cvičebnice je věnována studentům češtiny jako cizího jazyka (ČCJ). Byla původně uspořádána jako skripta k předmětu *Gramaticko-lexikální cvičení založená na korpusu*, věnovanému korpusovému přístupu ve výuce češtiny. Metoda výuky založená na korpusovém přístupu je v anglosaské literatuře známa pod názvem *data-driven learning* (DDL), což v doslovném překladu znamená *výuka řízená daty*. Za zakladatele této metody je považován Tim Johns, vysokoškolský učitel angličtiny na Univerzitě v Birminghamu. Ve svém hojně citovaném článku z počátku devadesátých let 20. století jako jeden z prvních systematicky popisuje metodu DDL (Johns, 1991). Zdůrazňuje přitom, že studentovým úkolem je objevit s pomocí konkordance¹ jistá pravidla na základě předložených příkladů. Je to metoda zcela protikladná k tradiční deduktivní výuce, kde se studenti nejdříve seznamují s pravidly a poté procvičují látku. Základní procedurou, kterou využívá Johns v učení založeném na konkordanci, jsou tři etapy: identifikace, klasifikace a generalizace. V první fázi student identifikuje problém, následně na základě pozorování dat kategorizuje jazykový jev a potom generalizuje pravidla.

Obdobný přístup je využíván i ve výuce v českém prostředí a tato cvičebnice se snaží předat jak učitelům, tak i studentům inspiraci ke cvičením, jež procvičují problematické oblasti češtiny nerodilých mluvčích. Jádro práce pochází z detailně provedené analýzy žákovského korpusu (Zasina, 2019b), dříve publikovaných odborných prací (Zasina, 2018a, 2018b, 2019a, 2020, 2022a, 2022b, 2022c) a vlastních učitelských zkušeností. Cvičení jsou uspořádána do samostatných kapitol s ohledem na tyto obtížné jevy: rozdíly ve významu a význam slov, gramatický rod, skloňování, substantivizovaná adjektiva, tvrdá vs. měkká adjektiva, zájmena, předložky, předpony *s-/z-*, přičestí minulé, kolokace, frazémy, stylové varianty, konkurence genitivních koncovek *a-* a *u-*, kvantita. Jednotlivá cvičení jsou opatřena příslušným grafickým symbolem označujícím, zda dané cvičení vyžaduje přímou práci s počítačem, nebo ho lze vypracovat v papírové formě. Drtivá většina úkolů využívá data nejnovějšího korpusu současné psané češtiny SYN2020 (Křen et al., 2020), ale také korpusů SYN2015 (Křen et al., 2015), ORTOFON verze 1 (Kopřivová et al., 2017), InterCorp verze 12 (Rosen et al., 2019) a CZeSL-SGT (Šebesta et al., 2014) a používá korpusové nástroje SyD (Cvrček & Vondříčka, 2011), Slovo v kostce (Machálek, 2019) a KonText (Machálek, 2014).

Tato publikace je koncipována jako klasická cvičebnice, ze které učitelé a studenti mohou vybírat jednotlivá cvičení podle potřeby. Ačkoli některá z nabízených cvičení na sebe bezprostředně navazují, není nutné, aby studenti všechna cvičení v knize řešili od začátku jedno po druhém (případná návaznost cvičení je v zadání vždy signalizována). Cvičení, jež vyžadují přímou práci s korpusovými daty, jsou často propojena příslušným odkazem na daný nástroj. Sama cvičení neobsahují postup práce s korpusem; předpokládá se totiž, že student pracující s korpusovým materiálem se již dříve seznámí se základními funkcemi

1 Konkordance je způsob zobrazování výsledku v korpusu s klíčovým slovem uprostřed a jeho pravým a levým kontextem (viz McEnery & Hardie, 2012, s. 35–37). Viz též pojmy:konkordance. (2016, Jun 7), in Příručka ČNK (online), <http://wiki.korpus.cz/doku.php?id=pojmy:konkordance&rev=1465291976> (29. března 2019).

korpusových nástrojů a prací s korpusem obecně. Nástroje jsou stručně popsány spolu s příslušnými odkazy na podrobnější popis funkcionalit korpusu v další části úvodu. Ke všem cvičením je k dispozici klíč, který v mnoha případech kromě odpovědí uvádí i postup řešení.

Předkládaných cvičení lze využít na různých úrovních znalosti češtiny, doporučujeme však s cvičebnicí pracovat od úrovně B1 a vyšší, na nichž jsou studenti do určité míry jazykově samostatní a nebudou mít větší potíže s pochopením zadání a prací s korpusovými nástroji. Nicméně posouzení vhodnosti daného cvičení je na učiteli, který sám nejlépe dokáže odhadnout schopnosti svých studentů a podle potřeb přizpůsobit danou látku.

Doufáme, že tento učební materiál pomůže jak učitelům, tak studentům systematizovat problematické jevy v češtině, nahlížet na češtinu z jiného úhlu pohledu a zároveň bude inspirativní pro další zájemce o korpusové metody, kteří by rádi připravili vlastní cvičení nebo prozkoumali oblasti zatím detailně nepopsané, případně objevili ještě neobjevené zákonitosti češtiny.

Adrian Jan Zasina
Praha, 30. června 2021

JAK ZAČÍT S KORPUSEM?

Hovoříme-li o publikaci obsahující cvičení na práci s korpusem, nemůžeme opomenout vůbec první korpusovou uživatelskou příručku *Jak vyžívat Český národní korpus* (Čermák et al., 2005), která byla doporučena studentům druhého stupně základních škol a středních a vysokých škol. Další uživatelskou příručkou pro školy byla publikace Michala Šulce (2007), v níž je stručně, krok za krokem popsán způsob práce s korpusovými daty. Během dalších let vzniklo několik studií zaměřených na využití korpusu ve výuce ČCJ (srov. Lukšija, 2010; Osolsobě, 2010; Vališová, 2009), v nichž jsou prezentovány konkrétní aplikace korpusu s ohledem na obtíže při studiu určitých jazykových jevů. V poslední době se rovněž objevily dvě publikace (Šmejkalová & Kopřivová, 2019; Šormová et al., 2019) věnované učitelům a studentům učitelství češtiny na základních a středních školách; ty nicméně neberou v potaz využití korpusu ve výuce cizojazyčných mluvčích. Zlom v oblasti korpusového přístupu ve výuce ČCJ přináší zevrubná analýza Adriana Jana Zasin (2019b) identifikující nejvýznamnější problematické oblasti češtiny nerodilých mluvčích na základě žákovského korpusu. Zasinova studie rovněž nabízí metodologický rámec pro tvorbu korpusových cvičení spolu s příklady.

Podívejme se nyní na definici jazykového korpusu. Korpus je zdroj elektronicky uložených jazykových dat, která odrážejí skutečnou jazykovou situaci. Na jazyk lze tak nahlížet prostřednictvím korpusu jako jeho autentického vzorku. Poněvadž ke každému slovu v korpusu je přidána jeho gramatická interpretace, lze pomocí různých nástrojů statisticky vyhodnocovat jazykové jevy. Můžeme se tak vžít do role badatele a zkoumat jazykové zákonitosti, na jejichž základě lze nejrůznější jazykové jevy kategorizovat a následně sestavit konkrétní pravidla, která se uplatní ve výuce cizinců.

Aby byl uživatel schopen plně využívat všechny dostupné funkce portálu <https://www.korpus.cz>, je nutná registrace, která je bezplatná. Registrace je velmi jednoduchá; návod je k dispozici zde: <https://wiki.korpus.cz/doku.php/kurz:zaciname?rev=1533556421>. Potřebujete-li se blíže seznámit se způsobem vyhledávání v korpusu, můžete projít on-line kurzem práce s korpusovým rozhraním v sedmi lekcích (<https://wiki.korpus.cz/doku.php/kurz:uvod>). Nemusíte se bát, pokud něčemu nerozumíte – vždy se můžete obrátit na podporu a položit libovolný dotaz (<https://podpora.korpus.cz/>). K dispozici je rovněž repozitář korpusových cvičení Pro školy (<https://korpus.cz/proskoly>), který je sice zaměřen na rodilé mluvčí, nicméně najdete zde i některé pracovní listy vhodné pro cizince.

ÚVODNÍ CVIČENÍ

Jako nejlepší start pro práci s korpusem jsme připravili několik úvodních cvičení, jež mají uvést uživatele do základní teorie a následné praxe. Díky této přípravě bude mít uživatel představu o tom, co lze v korpusu hledat a jak ho v praxi využívat. Práce s korpusovými cvičeními bude tak pro vás snazší a příjemnější.

Zadání: V rámci samostudia si přečtěte definici jazykového korpusu (<https://wiki.korpus.cz/doku.php/pojmy:korpus?rev=1634547326>). Pak odpovězte na níže uvedené otázky.

Jaké jsou hlavní přednosti korpusu?

Co znamená lemmatizace a tagování?

Co je to specializovaný korpus? Uveďte příklady.

Kdy vznikl Český národní korpus a kdo je jeho zakladatelem?

Zadání: Na uvedených stránkách (<https://wiki.korpus.cz/doku.php/cnk:uvod>) se podívejte na nabídku korpusů v rámci Českého národního korpusu. Doplněte do tabulky čtyři základní typy korpusů a uveďte ke každému příklad korpusu, který obsahuje nejvíce slov.

Typ korpusu				
Příklad korpusu				

Zadání: Nyní se zkuste zamyslet a odpovědět na otázku, který typ korpusů nebo konkrétní korpus bude pro účely vaší práce nejvhodnější a proč.

Zadání: Seznamte se s vyhledáváním v korpusu v první lekci on-line kurzu (https://wiki.korpus.cz/doku.php/kurz:prvni_dotaz?rev=1615195905). Kurz vás provede základním vyhledáváním a typy dotazů v nástroji KonText.

Zadání: Nyní zkuste v rozhraní KonText (<https://www.korpus.cz/kontext>) vyhledat seznam kolokací doporučeným způsobem. K substantivům *věk*, *muž*, *nemocnice* najděte jejich tři nejčastější adjektivní kolokáty na základě dat z korpusu SYN2020.

_____ věk

_____ muž

_____ nemocnice

K tomu postačí zadat dotaz a následně specifikovat kontext (viz obrázek), kde v rámci slovního druhu zvolíte rozsah kontextu **-1** – adjektivum se nachází na první pozici vlevo od KWIC (klíčového slova v kontextu) – a označíte přídavné jméno. Následně stiskněte *Enter*, nebo klikněte na *Hledat*.

The screenshot shows the KonText search interface. At the top left, a button labeled "Specifikovat kontext" is circled in red. Below it, the "Lemmata v kontextu" section has a "Rozsah kontextu:" field with a range of numbers from -5 to 5, where "-1" is highlighted in red. The "Seznam lemmat oddělených mezerami:" field is empty, and a dropdown menu is set to "musí být přítomny všechny položky".

The "Slovní druhy v kontextu" section has a "Rozsah kontextu:" field with a range of numbers from -5 to 5, where "-1" is highlighted in red. Below it, the "PoS list:" section has a dropdown menu set to "musí být přítomny všechny položky" and a list of parts of speech. The "přídavné jméno" option is checked and circled in red.

- částice
- číslovka
- citoslovce
- cizí slovo
- interpunkce
- neznámý
- podstatné jméno
- předložka
- přídavné jméno
- příslovce
- segment
- sloveso
- spojka
- zájmeno
- zkratka

Abyste získali seznam nejčastějších adjektiv, vyberte v dalším kroku z nabídky *Frekvence* > *Vlastní...* Zobrazí se vám okno, v němž zvolíte atribut *lemma* (základní slovníkový tvar) a pozici *1L* (adjektivum se nachází na první pozici vlevo).

Frekvencní distribuce

Běžná Podle typů textů Závislost dvou atributů

Frekvencní limit: 1

Úroveň	Atribut	Nerozlišovat velikost	Pozice	(Node) začít od
1.	lemma	<input type="checkbox"/>	1L	slova KWIC nejvíce vlevo

Vytvořit frekvencní seznam

Zadání: V korpusovém rozhraní KonText (<https://www.korpus.cz/kontext>) existují dva typy dotazu: jednoduchý a pokročilý. V jednoduchém typu dotazu můžete vyhledávat jednotlivá slova nebo slovní spojení. V pokročilém typu dotazu můžete např. vyhledávat podle specifikace gramatických kategorií. Zkuste vyhledat v korpusu SYN2020 předložky a vypsat pět nejčastějších na základě frekvencního seznamu.

Nečastější předložky v korpusu SYN2020

1. _____
2. _____
3. _____
4. _____
5. _____

Ve vyhledávacím formuláři přepněte dotaz na pokročilý. Hned vedle se vám zobrazí funkce *Vložit tag*, kterou použijete pro vložení tagu označujícího předložku. Stačí vybrat z nabídky slovní druh a následně kliknout na *Vložit*.

Hledat v korpusu

syn2020 ☆

Pokročilý dotaz | Vložit tag | Vložit 'within' | Klávesnice | Předchozí dotazy

[tag="R.*"]

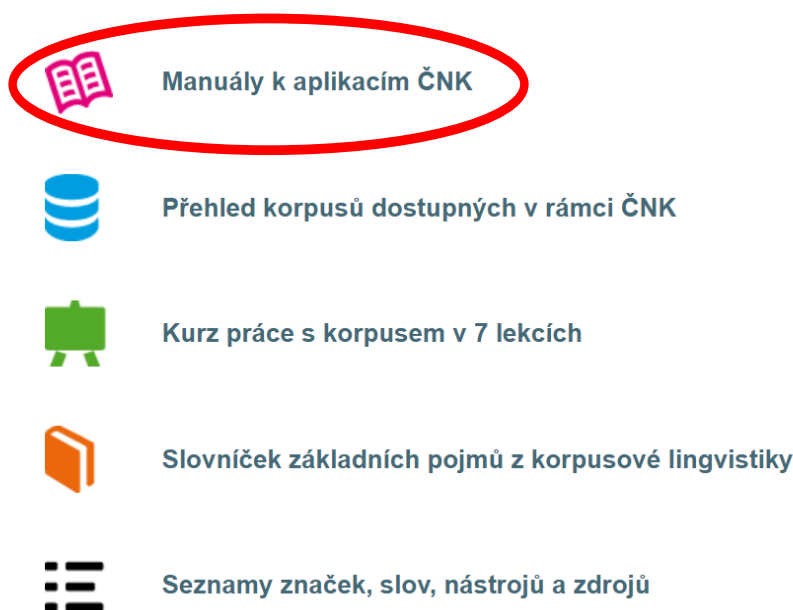
Nyní můžete vyhledat všechny předložky. Pro sestavení frekvencního seznamu vyberte z nabídky *Frekvence* > *Lemmata*. Takto získáte seznam nejčastějších předložek v korpusu.

KORPUSOVÉ NÁSTROJE

Pro práci s korpusovými daty lze využít několik různých nástrojů. Každý nástroj je vhodný pro jiné účely. V rámci Českého národního korpusu existuje několik aplikací, jejichž nabídka stále roste. Aktuální přehled všech nástrojů najdete na stránkách korpusové Wiki <https://wiki.korpus.cz> (viz obrázek). Aplikace jsou také aktualizovány, proto je vždy nejlepší seznámit se s aktuální verzí na webových stránkách.

Jaké informace zde najdete?

Tato wiki slouží uživatelům ČNK nejen jako zdroj informací o ČNK (popis zveřejněných korpusů a dokumentace k nim, manuály k aplikacím), ale také jako průběžně doplňovaná báze znalostí o korpusové lingvistice. Hlavní části wiki tvoří tyto oddíly:



S korpusovým vyhledávačem KonText jste se seznámili v předchozí části. Nyní se zaměříme na další aplikace, jež jsou vhodné pro využití ve výuce. Krátká zadání vám přiblíží základní funkcionalitu. Všechny aplikace jsou dostupné na portále korpusu z horní lišty aplikací; stačí kliknout na název příslušné aplikace.

Slovo v kostce (WaG): přehled informací o slově
(<https://wiki.korpus.cz/doku.php/manualy:wag>)

Zadání: Zadejte do nástroje **Slovo v kostce** dotaz *oko* a zodpovězte níže uvedené otázky.

Jaký je nejčastější slovní tvar lexému *oko*? _____

V jakém typu textu se lexém *oko* nejčastěji vyskytuje? _____

Jaká je nejčastější kolokace lexému *oko* v psaném jazyce? _____

SyD: porovnávání variant slov či slovních spojení v psané i mluvené češtině
(<https://wiki.korpus.cz/doku.php/manualy:syd>)

Zadání: Podívejte se do nástroje **SyD**, která varianta *sýra* × *sýru* je častější v psaném a mluveném jazyce.
