MUNI
ARTS

Využití korpusů

při výuce češtiny jako cizího jazyka

PLIN022 2.

Opakování klíčových pojmů

1. Čeština jako mateřský jazyk (L1)
2. Čeština jako druhý jazyk (L2)

   Zděděný jazyk
3. Čeština jako cizí jazyk
4. Čeština pro cizince
5. SERR
6. Korpus
7. Data-driven learning (DDL)

Jazykový korpus

„**Rozsáhlý soubor autentických textů** (psaných nebo mluvených)
převedený do **elektronické podoby** v jednom formátu tak,
aby v něm bylo možné **jednoduše vyhledávat** jazykové jevy
(zejména slova a slovní spojení / kolokace) ...“

F. Čermák (2017): KORPUS. NESČ. URL: https://www.czechency.org/slovnik/KORPUS

Korpusy a výuka

1. **nepřímé využití**

   lingvisté, autoři učebních materiálů
   = analýza dat a aplikace na výukové materiály

2. **přímé využití**

   studenti, lektoři = vlastní aktivity s korpusy

3. **popularizace**

Hlavní motivace

MUNI ARTS

Výuka podložená daty

Jak se může korpusový výzkum projevit ve výuce?

Vědecká gramatika (popis; corpus-based, corpus-driven)

X

Pedagogická gramatika (užití)

Pedagogická gramatika

...je pro ni příznačný vědecký odborný styl?

... popisuje mluvnici daného jazyka v celé její šíři?

... je určena poučenému uživateli – filologovi?

... přináší netradiční pohled na mluvnici daného jazyka?

... by měla odrážet potřeby vyučovací praxe?

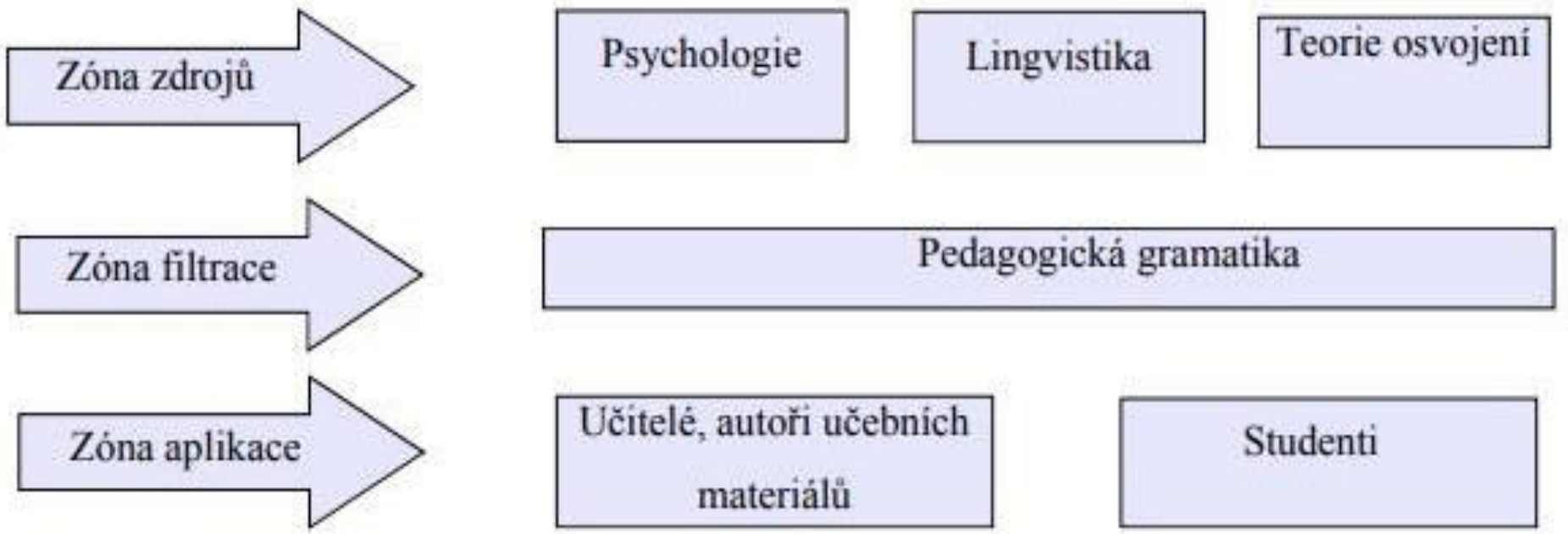
... popis pravidel je orientován především prakticky?

... mluvnice daného jazyka musí být zjednodušována?

... je často kritizovaná pro neodbornost, přílišné zobecnění či výběrovost?

ANO/NE

Pedagogická gramatika



Graf 5: Pedagogická gramatika

MUNI ARTS

Pedagogická gramatika

popis gramatiky jazyka
pro účely výuky a učení druhého jazyka

jazyková úroveň kompetence
a znalosti žáka
výukový materiál
výzkum
tradice
výukové metody
vliv dalších vědních disciplín

pokyny, co musíme nebo
nesmíme v cizím jazyce dělat

x

stigma zjednodušené
vědecké gramatiky

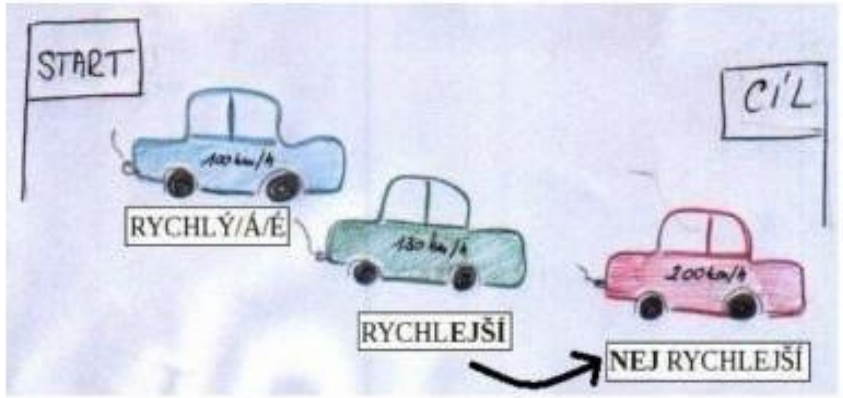
Zajímavosti ze světa pedagogické gramatiky

Imperfektum x perfektum

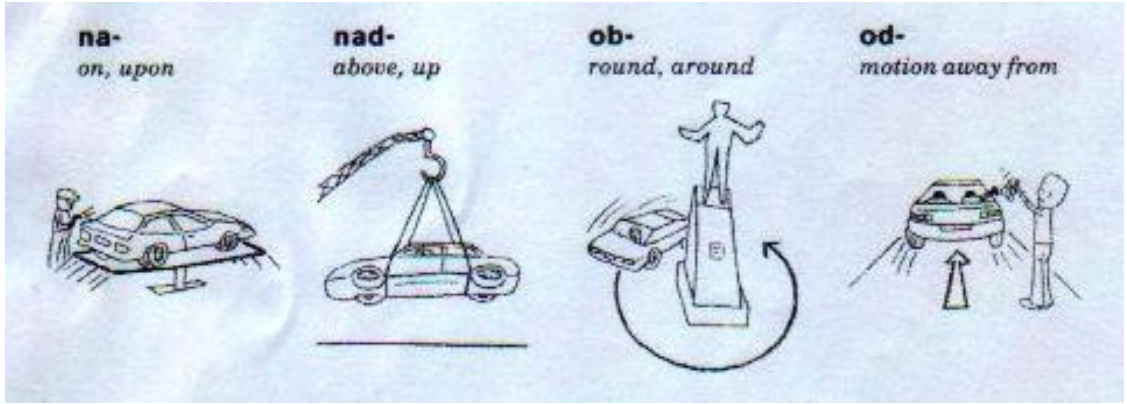
Vlastní vzory a modely

Signální gramatika

(Např. Signální slova jako *včera, dnes, zítra* signalizují použití správného gramatického času.)



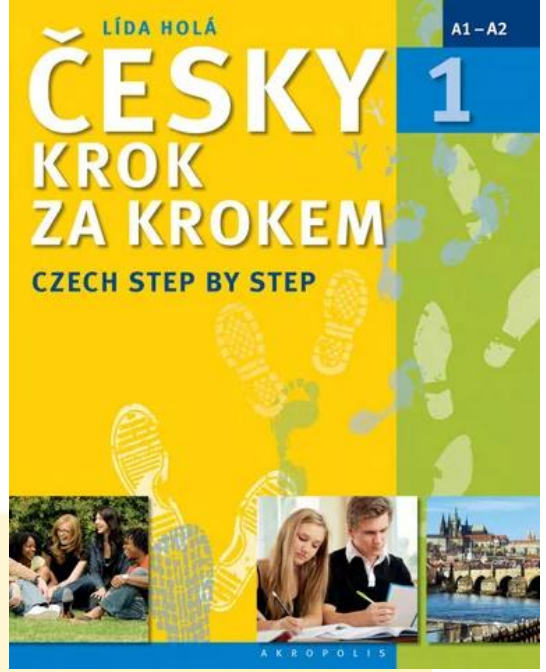
Obrázek 1: Vizuálně zpracované pravidlo, stupňování přídavných jmen, Strejcová, 2012, s. 80



Sémantika prefixů (Holá 2006, s. 137)

MUNI ARTS

Signální gramatika



ASES

8. Podívejte se na obrázky. Odpovězte na otázky.

1. Kam obvykle v pátek večer chodí Eva? Kam jde dneska večer?
2. Kam obvykle v pátek večer jezdí Marina? Kam jede dneska večer?

obvykle chodit/jezdit
dneska jít/jet

MUNI ARTS

Vysvětlete studentům češtiny

Gramatický rod v češtině

Akuzativ podstatných jmen

Slovesný vid v češtině

L. Holá: New Czech Step by Step (2004)

a) Nouns with majority endings (about 75 %):

student, kamarád, tygr	banán, sýr, jogurt	káva, voda, majonéza	auto, pivo, rádio
consonant	*consonant*	*– a*	*– o*

b) Nouns with minority endings (about 25 %):

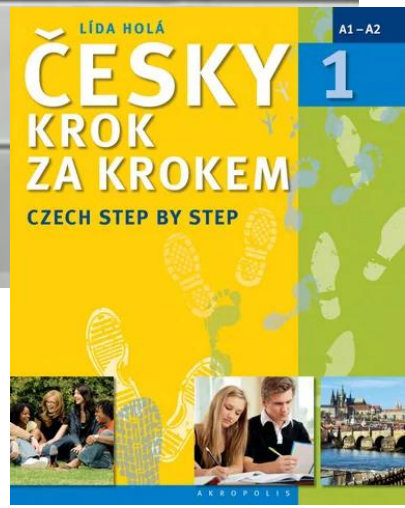
kolega, soudce	chleba	televize, policie, kancelář	nádraží, dítě, muzeum
-a *(rarely)* **-e** *(rarely)*	**-a** *(an exception)*	**-e** *(very often)* **consonant**	**-i, -e /-ě** *(not very often)* **-um** *(loanwords)*

L. Holá: Česky krok za krokem (2016)

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

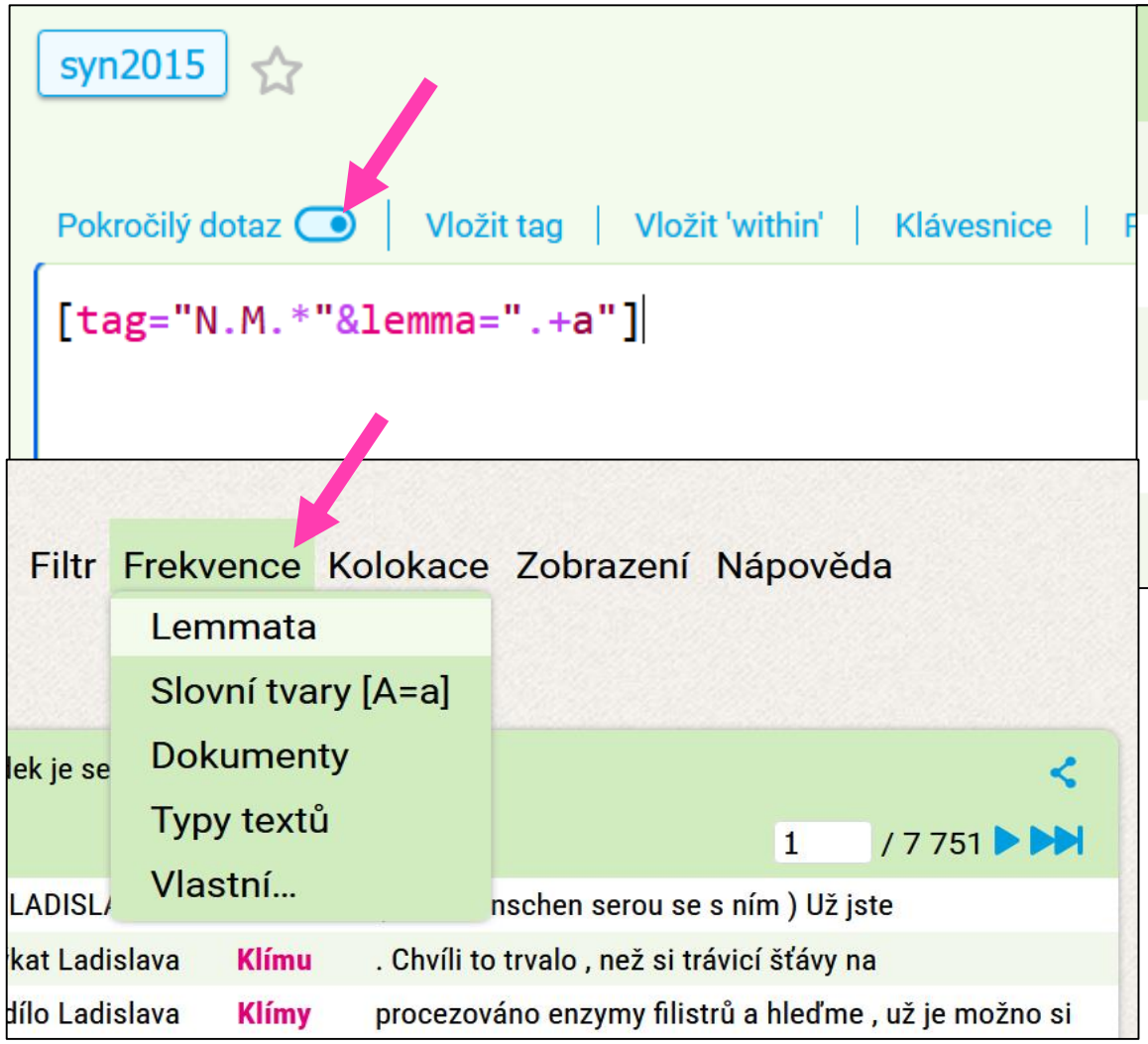
*Nouns ending in -ire and -yně are mostly feminine. Words ending in -iště are always neuter.

Compiled by SYN2015



5 nejfrekventovanějších substantiv **mužského rodu životného** (**Ma**) na **–a:**

[tag="N.M.*"&lemma=".+a"]



	Filtr	lemma	Freq ▼	i.p.m.
1	p / n	táta	13 022	107,84
2	p / n	kolega	12 321	102,04
3	p / n	policista	11 970	99,13
4	p / n	předseda	11 414	94,53
5	p / n	starosta	10 250	84,89

5 nejfrekventovanějších substantiv

mužského rodu životného (**Ma**) na **-e**:

[tag="N.M.*"&lemma=".+e"]

středního rodu (**N**) na **-e**:

[tag="N.N.*"&lemma=".+e"]

Nebo jsem viděla zajímávý věc, taky jsem fotografovala (A2)

Mobil je nedaležity věc pro mě . (A2)

Jíný věc je že v Finsku všichní zujou se boty , když vystoupíme na domu . (B1)

Ale každý věc má druhou stranu , taký je tam nějaká falešná reklama. (B1)

Nepřímé využití

nové přístupy

evidence-based koncept (založený na „tvrdých“ datech, v praxi ověřených)

frekvence

ustálené souvýskyty slov

Frekvence

co je v jazyce časté?

hlavním zdrojem o frekvenci jsou korpusy

je základní stavební kámen korpusové lingvistiky

lze ji sledovat na všech jazykových úrovních

bychom x bysme; televizor x televize;

měli bychom si koupit televizi, protože televize jsou teď za dobrou cenu (…)

Frekvence ve výuce jazyků

aplikace především v oblasti **slovní zásoby** a **gramatiky**

předpoklad: častěji používaná slova jsou součástí všech (?) typů konverzací,
jejich znalost tedy usnadní porozumění

v angličtině velice populární seznamy nejčastěji používaných slov

General Service List (West 1953), British National Corpus List (Nation 2006),
new-General Service List (Brezina; Gablasova 2015), Essential Word List (Dang; Webb 2016)

frekvence gramatických struktur a jejich revize pro učební materiály
(např. pasivum přesunuto na vyšší jazykové úrovně a zasazeno do patřičného kontextu)

Akalex

slovní zásoba akademické češtiny

seznam nejčastějších akademických slov a víceslovných jednotek

databáze (materiál pro další výzkum)

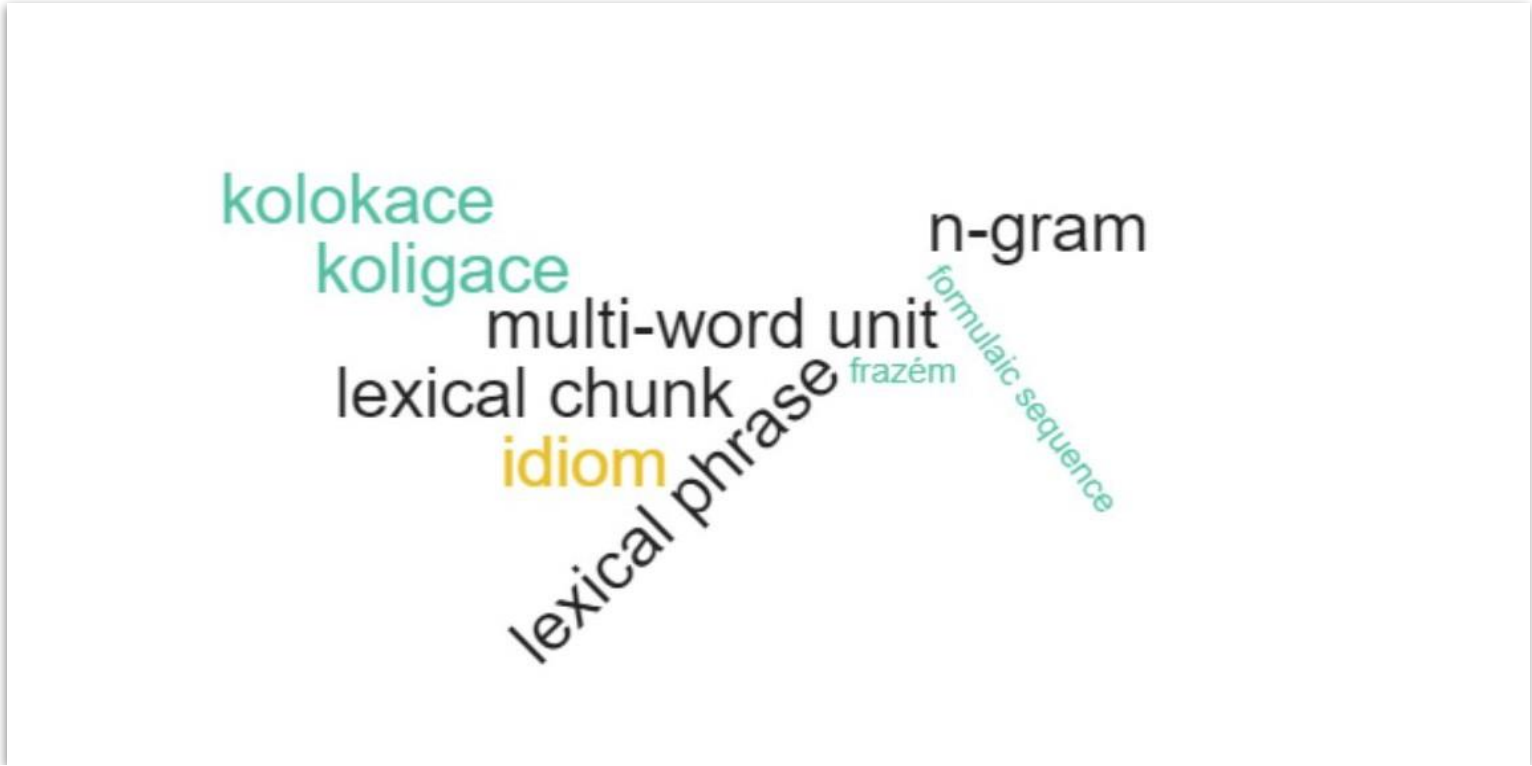
pomůcka při výuce či při psaní akademických textů (diplomních prací apod.)

https://www.korpus.cz/akalex/

Ustálené souvýskyty slov

údaje o frekvenci slouží k identifikaci souvýskytu slov

opakování/opakovatelnost X nahodilost



Korpusová lingvistika

kolokace (slovní spojení)

tratoliště krve, Červený kříž, praní špinavých peněz

koligace (typ kolokace, princip předurčeného výběru)

jezdit do školy každý den X teď; stojí to 100 korun X koruny;
mluvit česky X čeština

n-gram(y) (víceslovné výrazy, řetězce)

to je v pořádku; jak se ...?; myslím, že; na druhou stranu

Kolokace a kolokační paradigma

seznam kolokátů jednoho slova

kolokační paradigma slova *hlavní*:

Lemma	Frekvence	T-score	MI-score	logDice
město	4554	66,969	7,037	10,091
role	2068	45,284	7,895	9,931
hrdina	769	27,636	8,191	8,874
důvod	1105	32,84	6,372	8,795
cíl	942	30,351	6,492	8,703
vchod	575	23,894	8,133	8,495
nádraží	546	23,258	7,748	8,386
postava	553	23,27	6,579	8,193
příčina	494	22,05	6,983	8,149
silnice	527	22,632	6,143	8,026
úkol	487	21,791	6,316	7,99
téma	481	21,654	6,303	7,974
líčení	282	16,757	8,882	7,575

V. Cvrček (2017): KOLOKACE. NESČ
https://www.czechency.org/slovnik/KOLOKACE

Ustálené souvýskyty slov

1. Před použitím léku si pečlivě přečtěte příbalový _____.
2. Hliník se odstěhoval do _____.
3. jarní _____
4. Tato změna je dána _____*, jak se mění poloha planety Země.
5. Na závěr přidáme sůl a _____.
6. Kdo jinému jámu kopá, _____
7. Chtěl jsem, _____se na to podívala ještě jednou.

Ustálené souvýskyty slov ve výuce

podpora dalších vědeckých disciplín

psycholingvistika: když mluvíme využíváme kapacitu krátkodobé paměti, obměňujeme struktury, netvoříme struktury nové

osvojování L1/L2: dítě si osvojuje celé struktury,
s postupující znalostí je schopno strukturu rozklíčovat

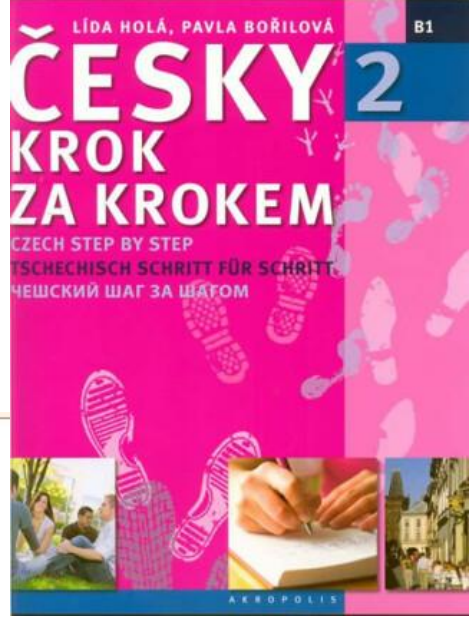
= konec izolování jednotlivých slov; doplňování nejen gramatického kontextu pro jednotlivá slova

look forward (to) *Have you ever .... ? (been / seen / had / heard / tried)*

modrý, modrá, modré = modré křeslo, modrý koberec, modrá deka

Ustálené souvýskyty slov ve výuce

→ konec izolování jednotlivých slov



!!! Model pro D pl.

Pamatujte si model pro dativ plurálu (frekventované formy). Více v lekci 16 na str. 160.

Pomáhám kamarádům, kamarádkám a kolegyním. *(ALE:* lidem, dětem*)*

1. Tvořte otázky a odpovídejte. V levém sloupci si všimněte forem dativu plurálu.

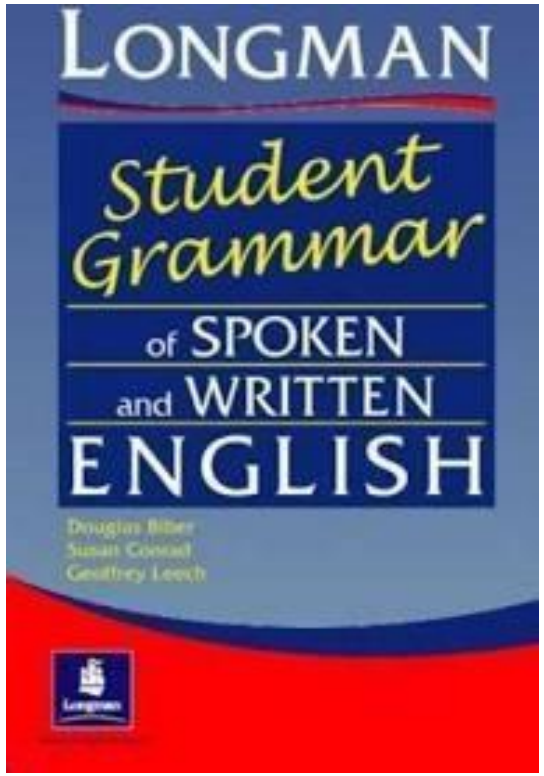
Například: Myslíte, že některým lidem vadí kouření na ulici?

některým cizincům v ČR mladým holkám a klukům
rockovým hudebníkům slavným umělkyním
dobrým sportovcům malým dětem starým lidem
vysokým modelkám

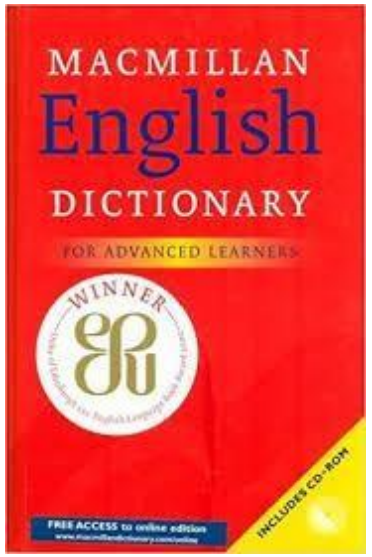
vadí/nevadí
se líbí/se nelíbí
chutná/nechutná
chybí/nechybí
sluší/nesluší
jde/nejde

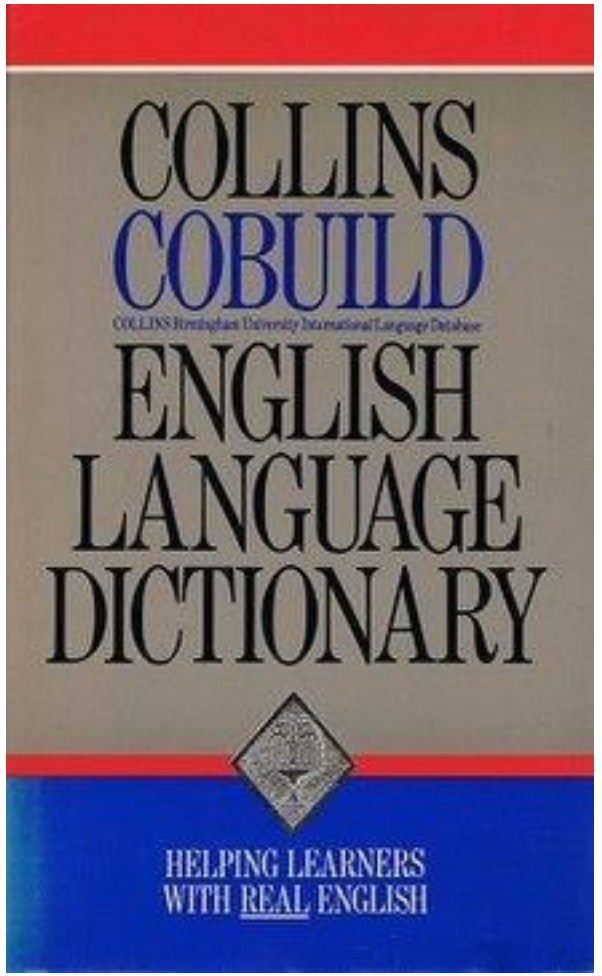
zelené nebo modré vlasy hlasitá hudba vysoké hory a moře
nízké platy tetování a piercing arogantní prodavači
české jídlo česká gramatika drahé šperky luxusní auta
hamburgery a hranolky extravagantní oblečení
vysoké platy ostré jídlo rychlá auta

(Korpusová) revoluce v ELT



“people who have never heard of a corpus are using the products of corpus research.” (McEnery/Xiao/Tono, 2006)





ELT – English Language Teaching

Korpusové gramatiky a slovníky češtiny

https://www.slovnikafixu.cz/

SMARTOOL: nástroj na výuku lexika a morfologie

TWIRLL: Targeting wordforms in russian language learning (2018)

Spolupráce:
The Artic University of Norway, **Tromsø**; Higher school of Economics, Moskva; Norwegian Agency for International Cooperation and Quality Enhancement in Higher Education

Laura A. Janda (slavistika, korpusová lingvistika, kognitivní lingvistika)
SlaviCorp 2018: navázání spolupráce s ÚČNK a jinými zahraničními institucemi

Janda, L. A.; Tyers, F. (2018): **Less is More**: **Why All Paradigms are Defective, and Why that is a Good Thing**. Corpus Linguistics and Linguistic Theory 14(2).

SMARTOOL – principy a teoretická východiska

rodilý mluvčí (osvojené paradigma) X nerodilý mluvčí (naučené paradigma)

podstatná jména –14 tvarů

1–3 slovní formy pro každý lexém

distribuce se mění lexém od lexému

kompletní paradigma = nefunkční model

x **usage-based model**

výuka L2 by měla cílit na tyto formy, nikoli na zvládnutí celého paradigmatu

korpus jako vhodný zdroj dat pro usage-based model

SMARTool for English-speaking learners of Russian

language SMARTool

Search by topic | Search by analysis | Search by dictionary | List of abbreviations | About

Level: A1

Word: бабушка (grandmother)

☐ Show translation ◉ male voice ○ female voice

Моя бабушка живёт в другом городе. (Nom.Sing)

У моей бабушки дома много интересных картин. Она рисует их сама. (Gen.Sing)

Я научила бабушку пользоваться айфоном, и теперь мы каждый день разговариваем с ней по скайпу. (Acc.Sing)

https://smartool.github.io/smartool-rus-eng/

language SMARTool

Search by topic | Search by analysis | Search by dictionary | List of abbreviations | About

Level: A1

Topic: rodina (family)

babička

☐ Show translation ○ female voice

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

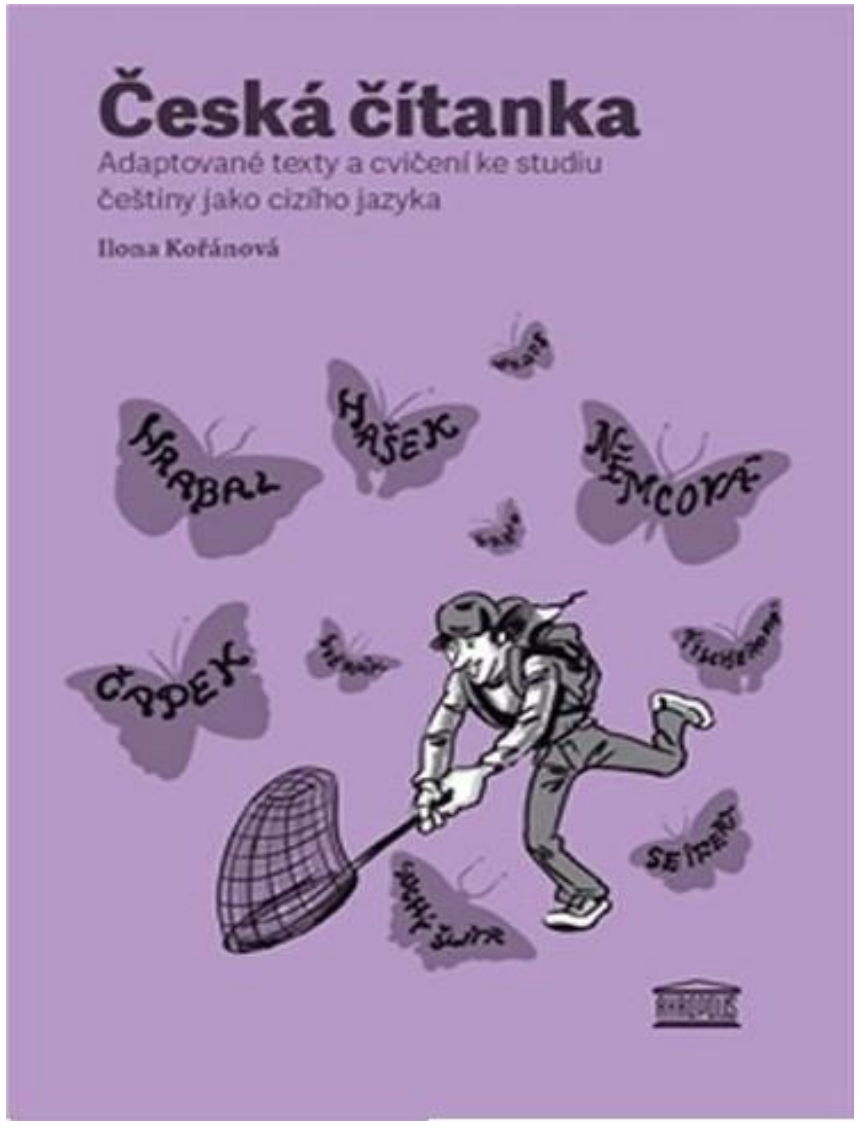
Zůstanu u dědy a u babičky na vesnici. (Gen.Sing)

Pozdravuj babičku. (Acc.Sing)

CZECH NATIONAL CORPUS | Diku | THE ARCTIC UNIVERSITY OF NORWAY UIT | HIGHER SCHOOL OF ECONOMICS NATIONAL RESEARCH UNIVERSITY

https://zasinaa.github.io/smartool_cs/

Nepřímé využití obecných korpusů



Česká čítanka

prezentace lexikálních a gramatických prostředků se opírá o statistická data ČNK

MUNI ARTS

Nepřímé využití obecných korpusů

Aplikace

Ten Ta To