

Pojďte si hrát s korpusy a korpusovými nástroji (Treq a Skell – jak využít korpus při překladu a studiu lexika L2)¹

Klára Osolsobě

osolsobe@phil.muni.cz

Anotace

Cílem přednášky bude prezentace dvou korpusových nástrojů, které nabízejí uživatelům podporu při překládání, popřípadě při pokročilém studiu jazyka. První z nich je založen na paralelních korpusech budovaných v rámci projektu Intercorp Ústavu Českého národního korpusu (<http://ucnk.korpus.cz/intercorp/>)², druhý je produktem firmy Lexical Computing, rovněž využívá jazykové korpusy ke studiu L2 (<https://www.sketchengine.eu/skell/>).³

Úvod

Studium/výuku L2 a překlad si v dnešní době lze jen těžko představit bez využití nejrůznějších počítačových nástrojů a podpor. Překladaelé mohou využívat a využívají nejen množství elektronických slovníků a podpory opírající se o strojový překlad, ale existují a budují se i tzv. paralelní korpusy (<http://wiki.korpus.cz/doku.php/pojmy:paralelni>).

Krátký historický exkurz

Hovoříme-li o korpusové lingvistice, pak máme většinou na mysli poměrně mladé odvětví spojené s technickými možnostmi, které se objevují v 50. letech XX. století a rozšiřují se od 80. let téhož století.

¹ Text vznikl za podpory projektu **MUNI/A/1061/2018** Čeština v jednotě synchronie a diachronie – 2019.

² „Paralelní korpus slouží jako zdroj dat pro teoretické studie, lexikografii, studentské práce, výuku, zejména výuku cizích jazyků, počítačové aplikace, překladaele i veřejnost. Korpus se buduje s účastí řešitelských pracovišť a je dostupný on-line pro registrované uživatele.“ Viz také podrobně <http://wiki.korpus.cz/doku.php/cnk:intercorp>.

³ Sketch Engine je nejmodernější online nástroj pro budování, správu a prohledávání velkých textových kolekcí (tzv. korpusů) v desítkách jazyků. Tento nástroj využívají po celém světě jednotlivci i společnosti, kupříkladu Cambridge University Press, Oxford University Press nebo Macmillan.

Je ovšem také třeba si uvědomit, že myšlenka korpusu je mnohem starší. Již ve strukturalistických přístupech k jazyku, v myšlence toho, že systém (langue) poznáváme na základě pozorování jeho realizací na rovině parole (promluv, textů) je patrná idea korpusu (parolových dat) a procedury (pravidel, kterými jsou vydělitelné jednotky jako hláska, morf, ... a vztahy mezi nimi – např. dvojice lišící se distinktivními rysy).

Ještě starší jsou jazykové analýzy opřené o textové korpusy (**papírové korpusy**). Jeden ze základních termínů korpusové lingvistiky – **konkordance** je termín převzatý z **biblické kritiky**.⁴ V korpusové lingvistice se termínem konkordance označuje **soupis všech výskytů klíčového slova** (může jít ovšem i o klíčové spojení více slov) **ve všech kontextech**, v nichž se v korpusu vyskytuje.

Typické bylo užití korpusů (sbírek textů) v **lexikografické praxi**. Podíváme-li se na slavného slovníkáře **Samuela Johnsona (1709-1784)**, tak je známo, že ve výkladu používal doklady z pečlivě sestaveného korpusu. Podobně když se podíváme na **Příruční slovník jazyka českého (PSJČ)**⁵, tak vidíme, že u řady dokladů jsou zkratky odkazující ke jménům klasiků české obrozenecké literatury. Slovníkářská praxe se dnes zcela běžně opírá o korpusy, existuje tzv. korpusová lexikografie.⁶

Korpus a internet

V dnešní době představuje asi největší sbírku elektronicky uložených textů internet. V čem je jazykový korpus lepší?

Jednou z podmínek, kterou musí korpus jako zdroj dat pro studium jazyka splňovat, je v moderní době elektronické uložení a elektronická přístupnost jazykových dat, která korpus tvoří, dále vymezený obsah a rozsah jazykových dat, která korpus tvoří a standardní anotace,

⁴ Šlo o alfabertický soupis slov Bible s odkazy na všechna místa, kde se vyskytují. Proč? Např. tzv. synoptická evangelia – Mk., Mt., Lk., líčí paralelně události života Kristova, podobně existují odkazy na biblické texty SZ v NZ. První biblickou konkordanci sestavili dominikáni k latinské Vulgátě (Hugo de Sancto Charo, vyšlo v roce 1230, později vznikly konkordance pro hebrejskou Bibli a pro Septuagintu). Pro angličtinu vznikly první konkordance v 16. stol. České konkordance jsou až z 20. stol. V dnešní době existují počítačové programy pro tvorbu biblických konkordancí.

⁵ Viz http://www.ujc.cas.cz/elektronicke-slovníky-a-zdroje/Přirucni_slovik_jazyka_ceskeho.html.

⁶ Ke konkrétním výstupům bylo zaměřeno několik projektů nakladatelských domů Longman a Collins Cobuild – *Longman Dictionary of Contemporary English* a *Collins Cobuild English Language Dictionary*). Za vrchol se pokládá *OED2*. V českém prostředí vyšla péčí Ústavu Českého národního korpusu celá řada slovníků (autorské slovníky Karla Čapka a Bohumila Hrabala, frekvenční slovníky psané i mluvené češtiny a řada dalších, např. *A Frequency Dictionary of Czech: Core Vocabulary for Learners*, viz <https://ucnk.ff.cuni.cz/cs/veda-a-vyzkum/publikace/>).

keré data v korpusu popisují. Tyto podmínky (vymezený rozsah a obsah a standardní anotace) nesplňuje internet. Pro češtinu existuje v současnosti velké množství jazykových korpusů, které tuto podmínku splňují.

Význam slov prozradí jejich okolí

Nejobvyklejším základním nástrojem pro práci s jazykovým korpusem je korpusový vyhledávač (manažer), který v zásadě umožňuje klást dotazy na jazykové jednotky a zobrazovat jejich výskyt ve formátu KWIC (**key word in context**).

V jazyce platí podobné zákonitosti, jaké sleduje psychologie osobnosti: „*Řekni mi, kým se obklopuješ, a já ti řeknu, jaký jsi*“, o významu slova ti leccos napoví jeho kontext.

Víte například, co je to **hořčák**? Zadáme-li lemma *hořčák* do vyhledávače google, pak na první stránce narazíme na obrázky, z nichž je patrné, že jde o druh houby, týž význam nabízí i česká Wikipedie.

Jak si ale poradíte s tím, když Vás někdo pozve na *hořčák*?

Na obrázku 1 vidíme, jak vypadá konkordance klíčového lemmatu *hořčák* z korpusu SYN2015⁷.

Obrázek 1

když sme pár dnů nato seděli v třešňovce a popijeli lenosti , protože byl nejbliž . Po prvním usrknutí ten jednu zvláštnost místního slovníku : Egri Bikaver se tu říká	hořčák	. „ Kdo jinej než ta svině Pačes . ”
houby vám může způsobit zažívací obtíže . Jedinou výjimkou jsou a tomu sme se zasmáli a zase si přihli teplýho	hořčáku	letěl do koše a další chybu už jsem neudělal .
když si poručil pivo , přidal k tomu ještě jeden a když dopis měl být kromobyčejně výrazný , dal si	hořčáky	, v té době nejlacinějšímu , příšernému ovocnému vínu .
Horší , než kdyby poslední česká hokynářka pozvala na skleničku , z jejichž slibů by se taky mohl stát pěkný	hořčáku	. Ty raději špičkou jazyka jso prověřte , jestli nejdou
psal dál , kolem úst se mu třpytily proužky od nepoživatelná . Tato houba , lidově také známá jako „	hořčák	a řekli sme : „ Na Zátopka ! ” Medvěd
že ten dopis vzbudí pohoršení , dal si ještě jeden	hořčák	, a když dopis měl být kromobyčejně výrazný , dal ještě jeden . Když pak zjistil , že je ve
	hořčáku	Rothschilda nebo britskou královnu . Malý český atašé a velký
	hořčák	. Ve svém městě na něj fakt nejsem zvědavý ,
	hořčáku	, jistě mu výčepní teď nese třetí dvoudecák , ach
	hořčák	“ , je mimochodem opravdu pěkná , oproti tzv. pravákům
	hořčák	, aby měl silu dopis zalepit a hodit do poštovní

Prostým pročtením 12 kontextů můžeme odlišit tři různé významy:

⁷ SYN2015 je stamilionový korpus synchronních psaných českých textů (viz <https://wiki.korpus.cz/doku.php/cnk:uvod>).

1) *hořký alkoholický nápoj* (obrázek 2)

Obrázek 2

„Někdo mi tu občanku musel ukrást ve škole.“ uvažoval nahlas Aleš, když sme pár dnů nato seděli v třešňovce a popíjeli hořák .
Po prvním uskrtnutí ten hořák letěl do koše a další chybu už jsem neudělal.
Egri Bikaver se tu říká hořáku , v té době nejlacnějšímu, přišerému ovocnému vínu.
Příští tejdén má přijít na besedu, ale vzhledem k tomu, že tenhle chlap nechodí, nýbrž pouze a pouze běhá, tak by to moh koupit za Pačesa,“ dodal Venca Popelka a tomu sme se zasmáli a zase si přihlí teplého hořáku a řekli sme : Napřed učinil zápis do deníku, potom se uchýlil do ústraní, nejráději v hospodě, když si poručil pivo, přidal k tomu ještě jeden hořák , a když dopis měl být kromobyčejně výrazný, dal si hořák ještě jeden.
Napřed učinil zápis do deníku, potom se uchýlil do ústraní, nejráději v hospodě, když si poručil pivo, přidal k tomu ještě jeden hořák , a když dopis měl být kromobyčejně výrazný, dal si hořák ještě jeden.
Horší, než kdyby poslední česká hokynářka pozvala na skleničku hořáku Rothschilda nebo britskou královnu.
Vladimír zůlívě psal dál, kolem úst se mu típly proužky od hořáku , jistě mu výčepní teď nese třetí dvoudecák, ach, to bude dopis, to bude dopis, pro milosrdenství, čtvrtá stránka a pořád Vladimír ještě nejde z kopce, ale do kopce, litovat
Když dopsal, zapálil si obřadně cigaretu, uvažoval a sledoval, co asi ten dopis vyvolá, když cítil, že ten dopis vzbudí pohoršení, dal si ještě jeden hořák , aby měl sílu dopis zalepit a hodit do poštovní schránky.

2) *hořká houba* (obrázek 3)

Obrázek 3

houby vám může způsobit zažívací obtíže. Jedinou výjimkou jsou hořáky .	. Ty raději špičkou jazyka jso prověřte, jestli nejdou
nepoživatelná. Tato houba , lidově také známá jako „ hořák “	“, je mimochodem opravdu pěkná, oproti tzv. pravákům

3) *hořkost* (obrázek 4)

Obrázek 4

choval bych se stejně. Straně, která odmítá zveřejnit výsledky prověrek svých členů, bych se obloukem vyhnul. V komunálkách speciálně chci volit konkrétní lidi, jimž lze věřit nebo jejich činy kontrolovat, a ne anonymní partaj, z jejichž slibů by se taky mohl stát pěkný **hořák**. Ve svém městě na něj fakt nejsem zvědavý, a zvláště ne, když by trval čtyři roky a nestál jen dvacku. Jiří Needrle, editor MF DNES Severní Čechy Praha má unikát, díky firmě z kraje ▶

Podobný postup můžeme zopakovat například se substantivem *svízel*. Na obrázku 5 vidíme část konkordance lemmatu *svízel* z korpusu SYN2015.

Pročítáme-li konkordance, narazíme na doklad, z jehož okolí se dozvíme, že slovo se v češtině vyskytuje ve dvou různých významech, které se navíc liší gramatickým rodem (obrázek 6). První význam ilustrují řádky 1, 3, 4, 5, 7, 8, 9, druhý význam řádky 2, 6, 10, 11 na obrázku 5.

Obrázek 5

Vyskytů: 217 i.p.m.: 1,8 (vztaženo k celému korpusu) ARF: 71,87 Výsledek je promíchán		
Vyběr rádků: základní ▾		
<input type="checkbox"/>	... I když i toto poznávání pro mě mělo velké	svízele ... Teprve když jsem dostala díky Františku Pavlíčkovi možnost hrát
<input type="checkbox"/>	... , rád zasahuje i na písčiny . Oblibuje spolu se	svízelem ... bílým též liniový výskyt na mezích podél polních cest či
<input type="checkbox"/>	... švy , tím vše bylo čitelnější . A divadelnější .	Svízele ... rychle přešuměvší inscenace asi tkvěla ve výkladu - žádoucí byla
<input type="checkbox"/>	... ; na klienta působí příznivě vědomí , že ve své	svízeli ... není sám a že je členem skupiny lidí , kteří
<input type="checkbox"/>	... případě může hokejový svaz žádat IIHF o disciplinární řízení .	Svízele ... nastala ve chvíli , kdy vedení extraligy odsouhlasilo Vítkovicím přeložení
<input type="checkbox"/>	... Na sušších stránkách v teplejších oblastech se vyskytuje zvláštní poddruh	svízele ... bílý chlupatý (Gallium album subsp . pycnotrichum) ,
<input type="checkbox"/>	... vedení klubu nepodařilo Jágra obklopit vhodnými spoluhráči , byla největší	svízele ... v posledních dvou letech . Ta je zažehnána . Stejně
<input type="checkbox"/>	... sebou . Jak správně ručnickovat Jenže to je právě ta	svízele Kdo chce opravdu krásné místo , stín pod borovicí
<input type="checkbox"/>	... , a to jak českými , natož cizojazyčnými , velká	svízele Možná si taky vybavíš slavné čtvrtě fronty , když
<input type="checkbox"/>	... , že Halas říká " svízeli zlatý " , kdežto	svízele ... je přece ženského rodu . " ... Svízeli
<input type="checkbox"/>	... kaje za zpěvu minnesāngrů kajicnā smrtelnost - Žal kosou dějiny	svízele ... s hnědozelenou barvou očí Mlčí před panty ZAKÁZANO Právě teď

Obrázek 6

Závratí struny už pomalu znám jen v kolících ne se vyznati najednou zaskřípám Někomu nebylo nápadné nic , jinému spojení svízele s epitetem zlatý , tedy snad jakýsi halasovský paradox . . . ? Zvědavější se pozastavili konečně nad tím , že Halas říká " svízeli zlatý " , kdežto **svízele** je přece ženského rodu . . . Svízeli zlatý Křemeni Závratí . . . " Věcný výklad je mnohem jednodušší . Je míněn ten svízele , svízele syřišťový , Gallium verum , Halas jej nazývá " mateřským znamínkem našeho kraje " , jsou to ty žluté chuchvalce s

Treq nástroj pro překladatele

Kromě jednojazyčných korpusů se budují i korpusy vícejazyčné, tzv. paralelní korpusy. Ve Wiki ÚČNK (<https://wiki.korpus.cz/doku.php/pojmy:paralelni>) je paralelní korpus definován jako „Jeden z typů korpusů, jehož účelem je rozšířit metodologii, propracovávanou v rámci korpusové lingvistiky původně pouze pro jeden jazyk, také na výzkum kontrastivní (mezijazykové srovnání). Paralelní korpusy se dělí na srovnatelné (angl. comparable) a překladové (angl. translational).“ Překladové paralelní korpusy (dvou jazyků) nemusí obsahovat pouze originály a překlady, ale mohou obsahovat též dvojici překladů stejného textu ze třetího jazyka. Jako projekt ČNK vznikl vícejazyčný překladový paralelní korpus InterCorp (<https://wiki.korpus.cz/doku.php/cnk:intercorp>). Nad korpusem InterCorp byla automaticky vytvořena databáze česko-cizojazyčných a anglicko-cizojazyčných ekvivalentů *Treq* (<http://treq.korpus.cz/>). Na obrázku 7 vidíme příklad anglických ekvivalentů slova *svízele*.

Obrázek 7

ver. 2.0

Výchozí jazyk: Čeština → Cílový jazyk: Angličtina

Omezit na: Kolekce: 6

svízel Hledej

Lemmata Víceslovné Regulární A = a

▲ Frekvence ▼	▲ Procenta ▼	▲ Čeština ▼	▲ Angličtina ▼
3	20.0	svízel	nuisance
2	13.3	svízel	quandary
1	6.7	svízel	arduous
1	6.7	svízel	trouble
1	6.7	svízel	dilemma
1	6.7	svízel	ahead
1	6.7	svízel	embarrassment
1	6.7	svízel	discomfiture
1	6.7	svízel	predicament
1	6.7	svízel	troubles
1	6.7	svízel	difficulty
1	6.7	svízel	tough
15			

Na obrázku 8 vidíme naopak ekvivalenty anglického *trouble*. Nejobvyklejším ekvivalentem v příslušné databázi je slovo *problém*. Povšimněme si, že slovo *svízel* se mezi frekventovanými ekvivalenty nevyskytuje.

Obrázek 8

Ver. 2.0

Výchozí jazyk: Angličtina → Cílový jazyk: Čeština Omezit na: Kolečce: 6

trouble Hledej

Lemmata Víceslovné Regulární A = a

▲ Frekvence ▼	▲ Procenta ▼	▲ Angličtina ▼	▲ Čeština ▼
3712	38.9	trouble	problém
2122	22.2	trouble	potíž
407	4.3	trouble	malér
396	4.1	trouble	průšvih
185	1.9	trouble	obtěžovat
174	1.8	trouble	trable
157	1.6	trouble	trápit
145	1.5	trouble	starost
123	1.3	trouble	průser
110	1.2	trouble	nesnáz

Můžeme se podívat i na kontexty, v nichž se jako ekvivalent anglického *trouble* (obrázek 9) slovo *svízeľ* (obrázek 10) objeví.

Obrázek 9

[InterCorp v9 - English](#)

Capote-Snidane_Tiffany	" That ' s one of the troubles .
Capote-Snidane_Tiffany	And there shouldn ' t be too much trouble on the man front .
hailey-konecna_diag	The trouble is , we ' ve had so many capital expenditures . "
kohout-snezim	" He could n't bear to have his private troubles aired in government circles like a juicy morsel for opportunists .
Mandelstamova-DveKnihy	I understand so clearly , and ache from the pain of it , that those winter days with all their troubles were the greatest and last happiness to be granted us in life .
Steinbeck-Hrozny_hnevu	" I been in the hills , thinkin ' , almost you might say like Jesus went into the wilderness to think His way out of a mess of troubles . "
Tolkien-Hobit	Bombur slept on with a smile on his fat face , as if he no longer cared for all the troubles that vexed them .
Toole-Spolceni_hlupcu	He says you gave him more trouble than any vendor he ever had . "
Urban-Lord_Mord	A defensive action on the part of a nobleman who senses danger and trouble .
_SUBTITLES	There may be trouble ahead ...

Obrázek 10

Capote-Snidane_Tiffany	" Jenže to je svízel .
Capote-Snidane_Tiffany	A co se týká mužských , taky s ní asi nebude velká svízel .
hailey-konecna_diag	Svízel je , že jsme měli tolik velikých výdajů . "
kohout-snezim	Zmiňoval sice ještě jednoho starého přítele (koho ?) , ale ten má vézet v obdobném kolotoči jako on , a Viky je ke všemu šileně hrdý (ani na to jsem nevzala zřetel !) , nesnesl by , aby jeho privátní svízel prosákla do vládních kruhů jako tučné sousto pro intrikány .
Mandelstamova-DveKnihy	Pochopila jsem , až to bolelo , že ta zima , ty dny a ty svízele bylo to nejlepší a poslední štěstí , které nám osud dal .
Steinbeck-Hrozny_hnevu	" Byl jsem v horách , tam jsem špekuloval , řekli byste , že to bylo skoro , jako když Ježíš odešel na poušť a tam uvažoval o cestách z těch svých svízelů . "
Tolkien-Hobit	Bombur spal s úsměvem v tlustém obličejí , jako by mu už nezáleželo na žádných svízelích , které sužovaly jeho druhy .
Toole-Spotceni_hlupcu	Povídal , že takovýhle svízele jaktěživo s žádným párkařem neměl . "
Urban-Lord_Mord	Obranné gesto slechtice , který cítí nebezpečí a svízel .
_SUBTITLES	Možná nás čeká svízel ...

Skell (Sketch Engine for Language Learning) nástroj pro studenty L2

Skell (<https://csskell.sketchengine.co.uk/run.cgi/skell>) je jednoduchý nástroj pro studenty a učitele cizích jazyků, prostřednictvím kterého mohou zjistit, zda se nějaké slovo, či fráze užívají a jak je užívají rodilí mluvčí. Umožňuje automatickou extrakci zadaných slov v kontextech, v nichž byla užita rodilými mluvčími. Automatickými metodami založenými na statistickém zpracování jazyka vybírá a nabízí uživateli podobná slova. Nástroj pracuje nad databází obsahující 1,4 miliardy slov ve více než 90 milionech vět. Texty pocházejí z webové domény archivované Českým Webarchivem, z článků a diskusí české Wikipedie a z webového korpusu novinek Institutu Jožefa Štefana v Lublani.⁸

Význam adjektiva *jízlivý* pomocí vhodných příkladových vět vidíme na obrázku 11.

Word sketch = kolokační profil, tj. nejčastější vazby zachycující kolokační a gramatické chování slova *jízlivý* vidíme na obrázku 12.

Na obrázku 13 jsou blízká slova, tj. nejen synonyma, ale i slova vyskytující se v podobných kontextech jako adjektivum *jízlivý*.

Obrázek 11

⁸ Prohledávané texty nemusely projít jazykovou korekturou, mohou tudíž obsahovat pravopisné a stylistické chyby, jichž se dopouštějí rodilí mluvčí (skýtají tak reálný obraz jazyka).

jízlivý 0.91 v milionu

- 1 Jednou jsem nechal na jejich stránkách jakýsi **jízlivý** komentář no a od té doby jsem "v hledáčku".
- 2 Mluvili o jeho talentu a nadšení stejně jako o jeho někdy "**jízlivém**" smyslu pro humor.
- 3 Po kliknutí na náhled fotografie se fotografie zvětší a objeví se někdy trochu **jízlivý** komentář.
- 4 Bez ohledu na tvé **jízlivé** poznámky já jsem spokojená s tím jak jsem to vyřešila.
- 5 Zatím jsou tady od tebe jenom **jízlivý** řeči na českou repre a na trenéry.
- 6 Já jsem tam byl poměrně **jízlivý** a to je většinou celkem vděčná poloha autorská.
- 7 Výraz jeho obličeje jde od "pevného zaujetí" až po **jízlivý** úsměv.
- 8 Dnešní **jízlivý** článek mi ovšem přišel už trochu přes čáru a krokem vedle.
- 9 Dneska jsem se probudila se zvlášť **jízlivou** náladou a toho je potřeba využít.
- 10 Ale také hrou neobyčejně vtipnou - byť je občas humor **jízlivý** a zlý.
- 11 Takže vaše **jízlivá** reakce bez znalosti těch potřebných souvislostí je zcela mimo.
- 12 Článek jako za staré školy - spousta pekla a **jízlivých** poznámek.
- 13 Tentokrát si ale tuto kritiku v dost **jízlivé** formě vysloužil.
- 14 Již krátce po jeho smrti se uši plní **jízlivými** drby.
- 15 Přitom nalevo našli autoři víc příležitostí k **jízlivému** humoru.
- 16 Třeba by pak tato **jízlivá** reakce tu vůbec nebyla.
- 17 I když to už bylo příliš **jízlivé** si myslím.
- 18 Mně tady ty diskuze nepřijdou blbý a články **jízlivý** .

Obrázek 12

The image shows a digital interface for the word "jízlivý". At the top left is the word "jízlivý" in a large, dark font. To its right is a green button labeled "Kontext" with a circular icon. Below this, the text "'jízlivý" - přívlastek shodný (atribut)" is displayed in a teal color. Underneath, there is a horizontal row of six rounded rectangular buttons, each containing a related word: "rozpuk", "poznámka", "komentář", "úsměv", "humor", and "tón".

Obrázek 13

The image displays a word cloud for the word "jízlivý". At the top left, the word "jízlivý" is written in a small, dark font. Below it, there are two rows of small, dark text listing various related words: "sarkastický", "uštěpačný", "kousavý", "posměšný", "břítký", "výsměšný", "škodolibý", "szíravý", "ironický", "pohrdavý", "vybíravý", "potouchlý", "zlomyslný", "satirický", "přihlouplý", "trefný", "pobavený", "duchaplý", "urážlivý", "provokativní", "štiplavý", "opovrhlivý", "příbtlý", "trpký", "cynický", "peprný", "sebeironický", "pomlouvačný", "morbidní", "pesimistický", "přezíravý", "ironizující", "žertovný", "dvojsmyslný", "mířený", "vulgární", "hanlivý", "pohrdlivý", "nadnesený", "dehonestující". Below these rows is a large, colorful word cloud where the most prominent words are "sarkastický", "pohrdavý", "ironický", "peprný", "břítký", "kousavý", "szíravý", "satirický", "stiplavý", "opovrhlivý", "přihlouplý", "pobavený", "duchaplý", "urážlivý", "provokativní", "štiplavý", "opovrhlivý", "příbtlý", "trpký", "cynický", "peprný", "sebeironický", "pomlouvačný", "morbidní", "pesimistický", "přezíravý", "ironizující", "žertovný", "dvojsmyslný", "mířený", "vulgární", "hanlivý", "pohrdlivý".

Závěr

Naším cílem bylo ukázat, jak fungují v praxi dva nástroje založené na korpusech, které lze používat jako pomůcku při učení se češtině i při překládání do/z češtiny.

V rámci přednášky bude prostor na další praktické ukázky práce s oběma nástroji i na diskusi.

Ti, kteří se přednášky účastní, si mohou dopředu připravit jedno až tři slova, jejichž význam přibližně znají. Pokusí se popsat jeho význam, přičemž se zaměří na slova významově blízká.

Součástí přednášky pak bude ukázka analýzy korpusových konkordancí uvedených slov i užití nástrojů *Treq* a *Skell*.

Bibliografie

Cvrček, V. – Čermáková, A. – Křen, M. (2016): Nová koncepce synchronních korpusů psané češtiny. *Slovo a slovesnost*, 77 (2), 83–101.

Charciarek, A. (2018): Možnosti využití korpusu InterCorp v česko-polské překladové lexikografii. *Časopis pro moderní filologii*, 100 (2), 206–222.

Milena Hnátková, Michal Křen, Pavel Procházka, Hana Skoumalová (2014): The SYN-series corpora of written Czech. In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, s. 160–164. Reykjavík: ELRA. http://www.lrec-conf.org/proceedings/lrec2014/pdf/294_Paper.pdf

Adam Kilgarriff, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovvář, Jan Michelfeit, Pavel Rychlý, Vít Suchomel (2014): The Sketch Engine: ten years on. *Lexicography*, 1, s. 7–36.

Křen, M. – Cvrček, V. – Čapka, T. – Čermáková, A. – Hnátková, M. – Chlumská, L. – Jelínek, T. – Kovářiková, D. – Petkevič, V. – Procházka, P. – Skoumalová, H. – Škrabal, M. – Truneček, P. – Vondříčka, P. – Zasina, A. (2016): SYN2015: Representative Corpus of Contemporary Written Czech. In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 2522–2528. Portorož: ELRA. http://www.lrec-conf.org/proceedings/lrec2016/pdf/186_Paper.pdf

Škrabal, M. – Vavřín, M. (2017): Databáze překladových ekvivalentů Treq. *Časopis pro moderní filologii*, 99 (2), 245–260.

Elektronické korpusy a korpusové nástroje

Křen, M. – Cvrček, V. – Čapka, T. – Čermáková, A. – Hnátková, M. – Chlumská, L. – Jelínek, T. – Kovářiková, D. – Petkevič, V. – Procházka, P. – Skoumalová, H. – Škrabal, M. – Truneček, P. – Vondříčka, P. – Zasina, A.: *SYN2015: reprezentativní korpus psané češtiny*. Ústav Českého národního korpusu FF UK, Praha 2015. Dostupný z WWW: <http://www.korpus.cz>

Nástroj *KonTetxt* dostupný z: WWW: https://kontext.korpus.cz/first_form.

Nástroj *SkELL* dostupný z: WWW: <https://www.sketchengine.co.uk/skell/>.

Průruční slovník jazyka českého (1935–1957) dostupný z: WWW: <https://psjc.ujc.cas.cz/>.