

4. lekce: Paralelní korpus

- Během této lekce se budete seznamovat s různými aplikacemi při multilingvální práci:
 - TreQ
 - Sloleks
 - Glosbe
 - InterCorp

1. úkol: aplikace TreQ

- Jděte na stránky <https://treq.korpus.cz/>
- TreQ je aplikace, se kterou můžete hledat překladové ekvivalenty mezi dvěma jazyky pomocí dat z ČNK
- Výchozí jazyk bude dále „čeština“ a cílový jazyk nastavte jako „slovinština“. Hledejte slovesa z předchozí lekce:
 - lhát
 - mlčet
 - mluvit
 - odpovědět
 - omlouvat
 - zavolat
- Kopírujte do souboru Word pro každé sloveso 5 prvních výsledků, které dostanete.

2. úkol: aplikace Sloleks

- Na stránkách <http://www.slovenscina.eu/sloleks> hledejte původní tvar každého rezultátu.
- Napište ho do souboru spolu s jeho slovním druhem do tabulky, jak uvedeno v příkladu níže.

Příklad tvarů „dela“ a „delava“:

SLOLEKS: Slovenski oblikoslovni leksikon

Kako pregibamo besede v slovenskem jeziku?

delava

Išči

příklad tvarů „dela“ a „delava“

- Sloleks je tvaroslovní lexikon slovinštiny, pomocí této aplikace můžete zjistit všechny tvary jednoho lemmatu.
- Píšeme v aplikaci tvar slova, které nás zajímá a klikneme na „išči“ (hledat).
- Pokud se daný tvar objevuje v jazyce v několika lemmatech, dostanete seznam všech možností. To je případ tvaru „dela“.
- V tomto případě napište všechny možnosti spolu s jejich slovním druhem

dela

del samostalnik, občno ime, moški spol

delati glagol, nedovršni

delo samostalnik, občno ime, srednji spol

Delo samostalnik, lastno ime, srednji spol

deti glagol, dovršni

příklad tvarů „dela“ a „delava“

dela

Išči

del samostalnik, občno ime, moški spol

delati glagol, nedovršni

delo samostalnik, občno ime, srednji spol

Delo samostalnik, lastno ime, srednji spol

deti glagol, dovršni

U toho výsledku vidíme, že tvar „dela“ může původně být:

- podstatné jméno (samostalnik) mužského rodu (moški spol): „del“
- sloveso (glagol) nedokonavé (nedovršni): „delati“
- podstatné jméno středního rodu: „delo“ nebo „Delo“ („Delo“ je název slovinského časopisu)
- sloveso dokonavé (dovršni): „deti“

příklad tvarů „dela“ a „delava“

- Pokud se daný tvar objevuje v jazyce jenom v rámci jednoho lemmatu, máte hned přístup ke všem tvarům. To je případ tvaru „delava“, který může pouze být tvar slovesa „delati“ v duálu. V slovinštině totiž ještě existuje duálové číslo, tj. plurál pro dva elementy.

delati glagol, nedovršni; 513.024 pojavitev

SEDANJIK

oblika	oseba	pogostost
--------	-------	-----------

EDNINA

delam	prva	25.021
delaš	druga	12.467
dela	tretja	84.450

DVOJINA

delava	prva	1.328
delata	druga	30
delata	tretja	4.940

MNOŽINA

delamo	prva	26.625
delate	druga	10.355
delajo	tretja	83.731

příklad tvarů „dela“ a „delava“

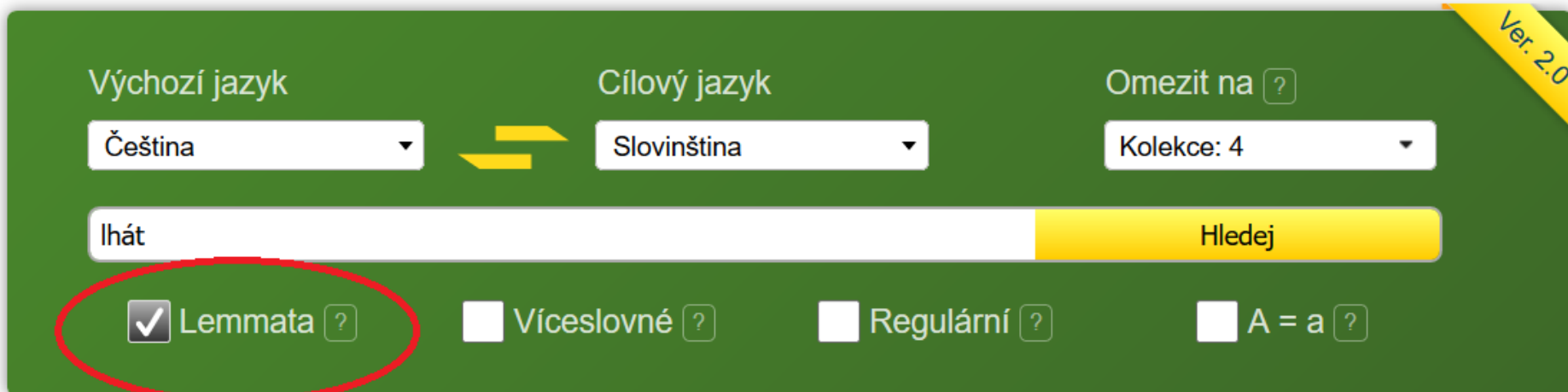
- Na konci hledání máme tedy tyto informace:

tvar v aplikaci TreQ	původní tvary	slovní druh
dela	del	podstatné jméno mužského rodu
	delati	nedokonavé sloveso
	delo / Delo	podstatné jméno středního rodu
	deti	dokonavé sloveso
delava	delati	nedokonavé sloveso

- Pokud musíte hledat význam termínů v slovinštině můžete používat slovinsko-český slovník: <https://glosbe.com/>

3. úkol: aplikace TreQ

- Teď opět hledáte tato samá slovesa v aplikaci TreQ, ale tentokrát vyberte formát výsledků „Lemmata“.



The screenshot shows the TreQ application interface with the following elements:

- Výchozí jazyk:** Čeština
- Cílový jazyk:** Slovinština
- Omezit na:** Kolekce: 4
- Search input:** lhát
- Search button:** Hledej
- Filters:**
 - Lemmata
 - Víceslovné
 - Regulární
 - A = a

A yellow banner in the top right corner indicates "Ver. 2.0".

- Napište si výsledek do souboru do druhé tabulky. Jaký rozdíl pozorujete?

4. úkol: Hledání v korpusu InterCorp

- Jděte opět na stránkách aplikace KonText
- Tentokrát klikněte na název korpusu, a pak na „všechny korpusy“

The screenshot shows the 'Hledat v korpusu' (Search in corpus) interface. The 'Korpus:' field contains 'syn2015', which is circled in red. A dropdown menu is open, showing 'Můj seznam' (My list) and 'Všechny korpusy' (All corpora), both circled in red. The 'Všechny korpusy' option is selected. Below the dropdown, there are two columns: 'Moje oblíbené korpusy' (My favorite corpora) and 'Vlajkové korpusy' (Flag corpora). The 'Vlajkové korpusy' column lists 'syn2015', 'ortofon v1', 'oral v1', and 'syn_v8'. At the bottom of the dropdown, there is a note: 'Stiskněte [Tab] pro vyhledávání ostatních korpusů. Použijte šipky pro pohyb mezi položkami.'

4. úkol: Hledání v korpusu InterCorp

- Vyberte korpus „InterCorp“, a pak „InterCorp v12 - Czech“

Můj seznam | [Všechny korpusy](#)

řada SYN řada ORAL **InterCorp** synchronní diachronní mluvený psaný webový

současná verze starší verze čeština cizojazyčný paralelní srovnatelný autorský

reprezentativní akviziční specializovaný

Přidrže klávesu CTRL/Command, chcete-li vybrat více štítků

Část názvu nebo popisu

Stiskněte [Tab] pro zobrazení vašich oblíbených korpusů

Můj

řada SYN řada ORAL **InterCorp** synchronní

současná verze starší verze čeština

reprezentativní akviziční specializovaný

Část názvu nebo popisu

InterCorp v12 - Arabic 43k

InterCorp v12 - Belarusian 7M

InterCorp v12 - Bulgarian 36M

InterCorp v12 - Catalan 10M

InterCorp v12 - Czech 254M

InterCorp v12 - Danish 71M

4. úkol: Hledání v korpusu Intercorp

- Pak klikněte na „Přidat korpus“ a „InterCorp v12 – Slovenian“

The screenshot shows the Intercorp search interface. At the top, the 'Korpus:' field is set to 'InterCorp v12 - Czech'. Below it, the 'Typ dotazu:' is set to 'Základní'. The 'Dotaz:' field is empty. A dropdown menu is open under the 'Zarovnané korpusy' section, listing various corpora. The option 'InterCorp v12 - Slovenian' is highlighted in blue. A red circle highlights the dropdown arrow icon at the top right of the menu. Another red circle highlights the 'InterCorp v12 - Slovenian' option. The 'Hledat' button is visible at the bottom left.

Korpus: ☆

Typ dotazu: ?

[Klávesnice](#) | [Předchozí dotazy](#)

Dotaz:

V dotazu CQL můžete vložit další řádek stisknutím klávesy Shift+ENTER ▶

Zarovnané korpusy

- ▶ **Specifikovat kontext**
- ▶ **Omezit hledání**

i Korpus InterCorp obsahuje kromě běžných zdrojů (Acquis Communautaire, EuroParl apod.) také množství dalších zdrojů. Pro zúžení hledání v korpusu po rozkliknutí této zprávy můžete definovat vlastní subkorpus. V obou případech se názvy relevantních textů se zobrazí po vyhledání.

Dropdown menu options:

- Přidat korpus --
- InterCorp v12 - Czech
- InterCorp v12 - Norwegian
- InterCorp v12 - Polish
- InterCorp v12 - Portuguese
- InterCorp v12 - Romani
- InterCorp v12 - Romanian
- InterCorp v12 - Russian
- InterCorp v12 - Slovak
- InterCorp v12 - Slovenian**
- InterCorp v12 - Albanian
- InterCorp v12 - Serbian

Co je InterCorp?

- InterCorp je paralelní korpus, pomocí kterého lze porovnávat překlady jednoho samého textu v různých jazycích.
- Během práce s takovým korpusem lze očekávat, že v daném jazyce se u nějakých vět překlad hodně liší od textu v druhém jazyce.

4. úkol: Hledání v korpusu Intercorp

1. Podobně jako u předchozí lekce vyberte „CQL“ jako „typ dotazu“
2. Hledejte tato samá slovesa pomocí dotazu odpovídající obligatorním funktorům valenčního rámce (můžete nejdříve zkusit s dotazem, který jste už vymysleli v rámci předchozí lekce, případně ho trochu modifikovat pokud bude potřebné).
3. Pro každý výsledek v češtině, který odpovídá tomuto rámci, dívejte se na slovinský ekvivalent. Pokud je tento ekvivalent jasným překladem české věty, kopírujte tyto dvě ekvivalence do souboru.
4. Dělejte to pro každé sloveso tak, abyste měli vždycky 5 vět v slovinštině.

5. úkol: Tvoření valenčních rámců

6. Na základě těchto slovinských vět tvořte valenční rámec pro použitá slovesa, který obsahuje alespoň obligatorní funktoři.

7. Jestli tyto slovinské věty nepoužívají pokaždé to samé sloveso, soustředte se na sloveso, který je nejvíce používán.

8. Použijte aplikaci Sloleks pokud potřebujete zjistit pád nějakých slov.

Shrnutí úkolů

Ve finální verzi souboru máte:

1. Výsledky z 1. hledání v aplikaci TreQ
2. Do tabulky, spolu s těmi výsledky, doplňující údaje pomocí aplikace Sloleks
3. Druhou tabulku s výsledky 2. hledání v aplikaci TreQ a doplňující údaje pomocí aplikace Sloleks
4. 5 vět v češtině a v slovinštině pro každé sloveso z korpusu InterCorp
5. Valenční rámec pro každé slovinské sloveso

Ještě pár otázek

- Na základě této práce se slovinskými slovesy, i když nemáte mnoho dat, máte pocit, že valence slovinských sloves se hodně liší od valence českých sloves?
- Už jste použili předtím paralelní korpus? K čemu si myslíte, že by ještě mohli sloužit?

- Tento úkol pošlete ve formátu mojeJméno-cvrtalekce.doc(/.odt...) na adrese gregoire.labbe@ff.cuni.cz nejpozději do 22.05.2020 včetně.
- Neváhejte se ptát v případě, že máte otázku!