

Webové korpusy Aranea

Učebnica

pre učiteľov cudzích jazykov, prekladateľov, tlmočníkov,
filológov a študentov filologických odborov

Vladimír Benko

Anna Butašová

Mária Lalinská

Mária Paľová

Zuzana Puchovská

Arnaud Segretain

Mariana Zeleňáková

2019

Univerzita Komenského v Bratislave

© Ing. Vladimír Benko, PhD.

Univerzita Komenského v Bratislave, UNESCO Katedra plurilingválnej a multikultúrnej komunikácie

Doc. PhDr. Anna Butašová, CSc.

Univerzita Komenského v Bratislave, UNESCO Katedra plurilingválnej a multikultúrnej komunikácie

PaeDr. Mária Lalinská, PhD.

Univerzita Konštantína Filozofa v Nitre, Filozofická fakulta

PhDr. Mária Paľová, PhD.

Univerzita Pavla Jozefa Šafárika v Košiciach, Filozofická fakulta

Mgr. Zuzana Puchovská, PhD.

Univerzita Komenského v Bratislave, Filozofická fakulta

PaeDr. Arnaud Segretain

Univerzita Konštantína Filozofa v Nitre, Filozofická fakulta

Mgr. Mariana Zeleňáková, PhD.

Univerzita Pavla Jozefa Šafárika v Košiciach, Filozofická fakulta

2019

Recenzenti:

Mgr. Radka Mudrochová, PhD.

Mgr. Jana Levická, PhD.

Za odbornú stránku tejto publikácie zodpovedajú autori.

ISBN 978-80-223-4663-4

Obsah

Úvod	5
1. TEORETICKÁ ČASŤ	7
1.1 Jazykové korpusy	7
1.1.1 Čo je to vlastne korpus?	8
1.1.2 Korpus ako vzorka jazyka	8
1.1.3 Korpus ako nástroj na skúmanie jazyka	9
1.2 Využitie jazykových korpsov	12
1.2.1 Korpus a skúmanie jazyka	12
1.2.2 Korpus a vyučovanie/učenie sa cudzieho jazyka	13
1.2.3 Korpus a preklad	14
1.3 Korpusová lingvistika	14
1.3.1 Tvorba korpsov	16
1.3.2 Projekt Aranea	16
1.3.3 Iné projekty webových korpsov	17
1.4 Korpusový portál Aranea	18
1.4.1 Korpusový manažér NoSketch Engine	18
1.4.2 Vyhľadávanie slov a slovných spojení	21
1.4.3 Vyhľadávanie slovného tvaru	26
1.4.4 Vyhľadávanie slovného spojenia	28
1.4.5 Regulárne výrazy I	29
1.4.6 Bodka, plus a hviezdička	31
1.5 „Jednoduché hľadanie“	34
1.5.1 Kontextové vyhľadávanie	37
1.6 Kolokácie	39
1.6.1 Miery asociatívnosti a kolokačná štatistika	40
1.6.2 Kolokačná analýza	44
1.6.3 Kolokácie so slovným spojením	46
1.7 Anotácia korpsov	49
1.7.1 Súbor morfológických značiek (tagset)	50
1.7.2 Morfosyntaktická anotácia	51
1.8 Jazyk CQL	57
1.8.1 Jednoduchá syntax CQL	57
1.8.2 Regulárne výrazy II	60
1.8.3 Jednoduchá syntax CQL a regulárne výrazy	60
2. PRAKTICKÁ ČASŤ	68
2.1 Lexikológia	68
2.1.1 Skratky, skratkové slová a iniciálové slovo/akronym	68
2.1.2 Význam a použitie afixov	73
2.1.3 Porovnanie výskytu slov vzhľadom na lexikálne alebo pravopisné zvláštnosti	84
2.1.4 Spájateľnosť slova s iným slovom (vyhľadávanie kolokácií)	87
2.1.5 Výskyt zložených slov so spojovníkom	92
2.2 Morfológia	97
2.2.1 Rodové koncovky neživotných všeobecných podstatných mien	97
2.2.2 Homofónia rodových koncoviek neživotných všeobecných podstatných mien	104
2.2.3 Zmena významu neživotného všeobecného podstatného mena v závislosti od jeho gramatického rodu	112
2.2.4 Určitý a neurčitý člen pred vlastným podstatným menom	127
2.2.5 Pozícia prídavného mena v nominálnej syntagme	145

2.3 Využitie korpusov Aranea pri právnom preklade: Slovesný vid a jeho vyjadrenie v slovenskej a francúzskej časti korpusu Aranea	160
2.3.1 Teoretické hľadisko	160
2.3.2 Praktické hľadisko.....	164
Bibliografia	179

Úvod

Projekt „učebnica“ vznikol trochu ako „spojená nádoba“ s projektom viacjazyčných webových korpusov *Aranea*, v priebehu ktorého sme si postupne začali uvedomovať potrebu určitého návodu na prácu s týmito korpusmi. Táto potreba bola v našich predstavách veľmi jasná, bolo odhodlanie, len chýbala prax, ktorá by bola dodala týmto ideám tú náležitú silu pustiť sa do písania učebnice. Navyše učebnice, ktorá bude slúžiť, v to dúfame, nielen študentom jazykových odborov, ale aj učiteľom, prekladateľom, tlmočníkom a všetkým tým, ktorí k tejto modernej disciplíne, korpusovej lingvistike, majú vzťah a našli si k nej cestu.

So stále narastajúcim počtom a rozsahom korpusov *Aranea* a stále väčším počtom študentov, ktorí prejavujú záujem o prácu s korpusmi, myšlienka učebnice naberá na váhe a treba prejsť od idey k realizácii. Grant KEGA 022UK-4/2016 nám pomohol vytvoriť tím vysokoškolských učiteľov z viacerých fakúlt, prevažne romanistov, a projekt sa začal meniť na realitu.

Výsledkom našich úvah a práce je táto vysokoškolská učebnica, ktorá v sebe zahŕňa viacero aspektov práce s korpusom. V prvej teoretickej časti stručne definujeme a opisujeme korpusy vo všeobecnosti, zaoberáme sa tiež korpusovou lingvistikou, najmä však teoreticky a prakticky opisujeme základy práce s korpusmi *Aranea*. V ďalších kapitolách učebnice, zameraných na praktické využitie korpusov *Aranea*, má čitateľ možnosť sledovať krok po kroku prácu s korpusom, a to v oblasti lexikológie, morfológie a v oblasti právneho prekladu. Na tomto mieste je treba upresniť, že všetky praktické úlohy s korpusom *Aranea* sme realizovali na francúzskych korpusoch *Araneum Francogallicum*, a to z toho dôvodu, že prevažná časť autorského kolektívu sú odborníci vo francúzskej lingvistike, didaktike či preklade a tlmočení. Myslíme si však, že práca s jazykovými javmi vo francúzštine môže slúžiť ako podnet pri využívaní korpusov iných cudzích jazykov, ktoré projekt *Aranea* ponúka. V kapitolách o lexikálnych a morfológických javoch sa objavujú aj tzv. *didaktické aplikácie pre francúzsky jazyk*. Našou snahou bolo ukázať, že jazykové korpusy *Aranea* majú reálne využitie v praxi, pri výučbe cudzích jazykov, a to napríklad pri koncipovaní didaktických situácií či pri tvorbe cvičení alebo ako ilustračný materiál pri výklade niektorého jazykového javu. Pri jednotlivých úlohách s korpusom, samozrejme, nechýbajú cvičenia, ktoré čitateľovi pomôžu prácu s korpusom rozvíjať a upevňovať nadobudnuté zručnosti.

V úvode chceme ešte poukázať na to, že *Aranea* ako projekt je náplňou činnosti UNESCO Katedry plurilingválnej a multikultúrnej komunikácie Univerzity Komenského v Bratislave. Táto katedra má prioritne ako cieľ podporovať plurilingvizmus a cez jazyk poukázať na bohatstvo v strete rozličných kultúr.

Milí čitatelia, učebnica, ktorá sa Vám dostáva do rúk, má ambíciu byť Vaším sprievodcom, nielen návodom, byť akýmsi priateľom „na telefóne“.

Uvítame všetky Vaše pripomienky a návrhy na jej zlepšenie.

Za kolektív autorov Anna Butašová.