

Branko Tošović

12. 7. 2012

Upute za korišćenje Gralis-Korpus Branka Ćopića

1. Prvo treba da kopirate link

**[http://www-gewi.uni-
graz.at/cocoon/gralis/login?tid=6FA72B417919BD2DA13C37E26A79F3BB](http://www-gewi.uni-graz.at/cocoon/gralis/login?tid=6FA72B417919BD2DA13C37E26A79F3BB)**

i da ga unesete u internet:



2. Kada dobijete gornju sliku, neophodno je da upišete šifru:

pu65BL9ic

3. Zatim u desnom gornjem uglu



treba izabrati jezik interfejsa: njemački (DE), engleski (EN), sprski/hrvatski/bošnjački (BKS) ili ruski (RU), tako što ćete kliknuti na odgovarajuću latinsku skraćenicu.

Ako izaberete BKS, dobićete ovaku sliku:

Opis projekta

Gralis-Korpus je višejezički/višejezični paralelni korpus za istraživanje i učenje svih slovenskih/slavenskih jezika, posebno u odnosu na nj(u) u Gracu/Grazu (Austrija) u okviru saradnje/suradnje Instituta za slavistiku i Centra za modelovanje/modeliranje informacije Fakulteta humaniora i studija (CWB), razvijen u Institutu za mašinski obradu jezika u Stuttgartu/Stuttgартu. Rukovodilac/Rukovoditelj projekta je Branko Toš, programsku podršku/potporu na serveru Univerziteta/Sveučilišta u Gracu/Grazu upravlja Hubert Štigler (Stigler), dok Olga Lener (Lehner) ac

Korpus se sastoji od nekoliko potkorpusa pod(j)eđeljenih na makrogrupe (južnoslovenski/južnoslavenski, istočnoslovenski/istočnoslavenski, pojedine jezike i individualni korpusi – korpsi pisaca kao što je Ivo Andrić, Zoran Živković i dr.). Postoji takođe(r) mogućnost izbora dviju/duhajućih istočnoslovenske/istočnoslavenske. Posebnu skupinu čine slovenski/slavenski jezici upareni sa nj(u)emačkim jezikom.

Za sada postoji interfejs/interface na engleskom, nj(u)emačkom, bosanskom/bošnjačkom/hrvatskom/srpskom i ruskom jeziku. U narednom

Pošto se korpus nalazi u fazi izrade, slobodan dostup nije moguć. Svi oni koji su zainteresovani/zainteresirani da se njime koriste, mogu s

4. Ukoliko kliknete na skraćenicu za engleski jezik, rezultat će biti:

The screenshot shows the GRALIS korpus search interface. At the top, there are dropdown menus for 'Group' (set to 'All'), 'Corpus' (set to 'Gralis BKS Corpus'), 'Primary language' (set to 'Bosnian/Bosniac'), 'Author' (empty), 'Genre' (empty), and 'Functional style' (empty). To the right, there are settings for 'Aligned languages' (set to 'Croatian', 'Serbian'), 'Context size' (set to 'one sentence'), 'Page size' (set to '20', 'Corpus entries', 'as HTML'), and a checked 'Morphosyntactic Annotation' option. A map of Europe is visible in the background. Below the search form, a section titled 'Project description' provides information about the corpus, mentioning it is a polyglot parallel corpus for Slavic languages, created by the Department of Slavic Studies and the Center for Information-Modelling in the Humanities at the University of Graz. It notes the use of CWB software and the involvement of Branko Tošović, Hubert Stigler, and Olga Lehner. The interface supports English, German, Bosnian/Croatian/Serbian, and Russian.

5. Sada treba da u gornjem uglu izaberete Gralis-potkorpus, u datom slučaju Gralis-Korpus Branka Čopića (skraćeno Gralis Čopić Korpus):

The screenshot shows the 'Korpusi' dropdown menu open, displaying a list of available corpora. The menu is titled 'Korpusi:' and includes options like 'Primarni jezik:', 'Autor:', 'Žanr:', and 'Funkcionalni stil:'. The dropdown list contains the following items:

- Čopićev Gralis-Korpus
- Copicev Gralis-Korpus
- Gralis BKS-Korpus
- Gralis BKS-Korpus (N)
- Gralis Ist-Korpus
- Gralis Ist-Zap-Korpus
- Gralis Južn-Korpus
- Gralis Južn-Zap-Korpus
- Gralis Mak-Korpus
- Gralis Ru-N(j)em-Korpus
- Gralis Slav-Korpus
- Gralis Zapad-Korpus

6. Kao rezultat dobijete sljedeću sliku:

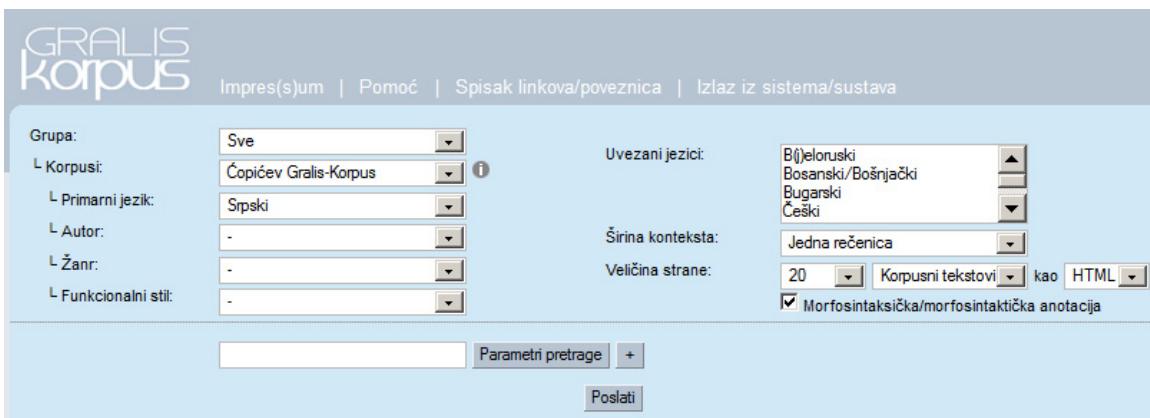
The screenshot shows the 'Primarni jezik:' dropdown menu open, displaying a list of primary languages. The menu is titled 'Primarni jezik:' and includes options like 'Autor:', 'Žanr:', and 'Funkcionalni stil:'. The dropdown list contains the following items:

- Bosanski/Bošnjački
-
-
-

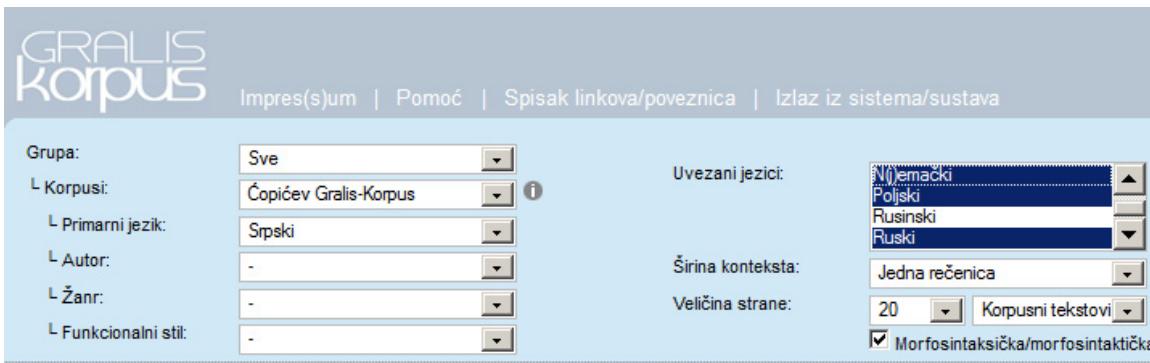
Slijedi izbor primarnog jezika. U našem slučaju to je srpski.



Kada kliknete na srpski jezik, otvorice se prozor:



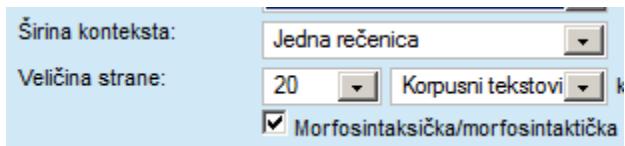
7. Ako želite da, pored srpskog jezika, imate još jedan, dva pa i više uparenih jezika, treba da u spisku sa desne strane izaberete odgovarajući jezik / odgovarajuće jezike, pri čemu ih treba istovremeno markirati pomoću Ctrl (Control) i desnog dugmeta miša.



8. Nakon toga treba da se izabere broj želenih rezultata pretraživanja (1. opcija: da se dobiju svi rezultati, 2. da se dobijete 10, 20 ili 30). Preporučujem da za početak izaberete opciju SVE.



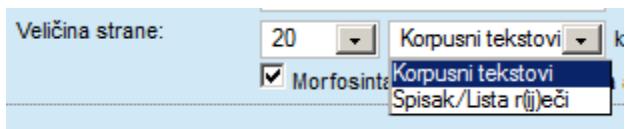
8. Poslije toga možete da se opredijelite za rezultat u obliku rečenica (Korpusni tekstovi) ili spiska pojavnica (Spisak/Lista riječi).



Veliki dio korpusa je morfosintaksički obilježen pa ako stoji:

Morfosintaksička/morfosintaktička

to znači da su anotirane podvučene riječi u dobijenim rečenicama. Prije nego što želite da rezultate pretraživanja prekopirate i unesete u word-tekst preporučujem da ovu opciju isključite.



9. Ako želite da saznate koji se tekstovi Branka Ćopića nalaze u Gralis-Korpusu, treba da kliknete na latinsko slovo „i“ s desne strane pa ćete dobiti ovakav spisak.

1	Srpski	Artiljerac Marko Medić
2	Srpski	Bitka s aždajom
3	Srpski	Bitka s đavolom
4	Srpski	Bog i batina
5	Srpski	Bunar bez vode

12. Najzad, možete da tražite riječi, oblike, morfeme, afikse, sintagme i sl. Za početak preporučujem da probate da dobijete sve pojavnice sa **naj-**, za šta je potrebno da napišete tačku i zvjezdicu poslije ove morfeme:

naj.*

The screenshot shows the GRALIS Korpus search interface. In the search bar at the bottom left, the query **naj.*** is entered. To the right of the search bar are two buttons: **Parametri pretrage** and **+**. Below these buttons is a **Poslati** button. On the left side of the interface, there are several dropdown menus for filtering search results based on group, corpus, primary language, author, genre, and functional style. On the right side, there are dropdown menus for linked contexts, sentence width, and output format. A checked checkbox labeled "Morfosintaktička/morfosintaktička anotacija" is visible. The top right corner features links to other languages: de, en, bks, ru, and the UNI GRAZ logo.

13. Ako ste izabrali opciju za dobijanje spiska, rezultat će ovako izgledati:

Spisak/Lista r(ij)eči

Korpusi: Čopićev Gralis-Korpus; Primarni jezik: Srpski;

Parametar/Parametri pretrage:

naj.* (Velika i mala slova: Ne; Pretraga po lemmama: Ne;)

Pronađeno je / Pronađena su 333 različita oblika / različitih oblika sa 1102 pojavnice/pojavnica.

142 **najzad**

85 **najprije**

67 **najbolje**

55 **najviše**

52 **Najzad**

36 **najmanje**

32 **Najprije**

14 **Najbolje**

14 **najveći**

13 **najbolji**

12 **najbliže**

11 **najgorje**

11 **najljepše**

11 **najmlađi**

10 **najmanju**

9 **najbliži**

...

Ukoliko ste pak odabrali opciju za dobijanje rečenica, imaćete sljedeći rezultat:

korporus Impres(s)um | Pomoć | Spisak linkova/poveznica | Izlaz iz sistema/sustava Institut za slavistiku | GAMS | GRALIS **UNI GRAZ**

Grupa: Sve
Korpsi: Čopičev Gralis-Korpus
Primarni jezik: Srpski
Autor: -
Žanr: -
Funkcionalni stil: -

Uvezani jezici: B(j)eloruski, Bosanski/Bošnjački, Bugarski, Češki
Širina konteksta: Jedna rečenica
Veličina strane: Sve, Korpusni tekstovi, kao HTML
Morfosintaksička/morfosintaktička anotacija

naj.* Parametri pretrage + Poslati

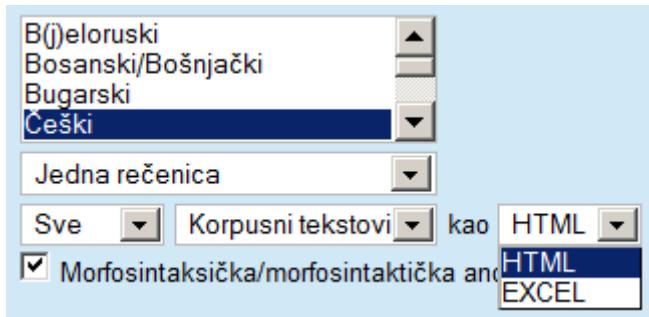
Rezultati pretrage
Korpsi: Čopičev Gralis-Korpus; Primarni jezik: Srpski;
Parametar/Parametri pretrage:
naj.* (Velika i mala slova: Ne; Pretraga po lemmama: Ne;)

1102 unos/unosa u 72 izvoru/izvora
Unos/Unosi 1 sve do 1102

Artijerac Marko Medić
Srpski Nije druge nego tući se ovdje , pa jai nadbiti ove naše , jai opet nekako . . . Možda se Nijemac ponovo pripovrati , a možda . . . ako te najzad uhvate , pritripi se malo dok opale u te , nije ni to najgore , a poslije čemo već vidjeti .
Artijerac Marko Medić
Srpski Nije druge nego tući se ovdje , pa jai nadbiti ove naše , jai opet nekako . . . Možda se Nijemac ponovo pripovrati , a možda . . . ako te najzad uhvate , pritripi se malo dok opale u te , nije ni to najgore , a poslije čemo već vidjeti .
Bitka s aždajom
Srpski Evo aždaje ! - tek bi iznenada povikalo najstarje čobanče , kad bi " grad " bio već završen , a onda bi svi , jezom od koje je zamiralo srce , u isti mah i prestrašeni i golicavo srećni ,agnuli da bježe i da se kriju iza udaljenih gomila kamenja .
Bitka s aždajom
Srpski Pazite , spremite se dobro , sad će ona izići ! - opet bi viknulo najstarje čobanče , pigrav i nestašan dječak , žutih mačjih očiju , a svi ostali čobančići , polegli iza nornice kamenja , bukvili su ovima " oratu " , cijelih roženih rđavih i kruških vlasova , nornici stanuju .

Kao što sam napomenuo, podvučene riječi imaju morfosintaksičku-anotaciju. Da biste u nju ušli, morate imati posebnu šifru.

14. Rezultate pretrage možete da automatski dobijete u Excelu, koji će se sam otvoriti. Za to je potrebno da izaberete opciju Excel:



15. Složeniji načini pretrage objašnjeni su u rubrici Pomoć:

**GRALIS
korpus**

Impres(s)um | Pomoć | Spisak

Grupa:	Sve
└ Korpusi:	Ćopićev Gralis-Korpus
└ Primarni jezik:	Srpski
└ Autor:	-
└ Žanr:	-
└ Funkcionalni stil:	-

naj.*

Pomoć

Znak	Opis	Prim(j)er
.	Bilo koji znak	narod.
*	Ponavljanje znaka	po*
.*	Naći znak u bilo kojoj konkordansi.	po.*
+	Najmanje jedan znak (jedno slovo)	po+
?	Izostavljanje	po?
&	Sastavljanje (konjunkcija)	po & od
!	Negiranje	
	Rastavljanje (disjunkcija)	radi(ti m š j mo te e)
%c	Nerazlikovanje velikih i malih slova	na %c