

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>**

i da ga unesete u internet:



GRALIS is a multilingual parallel corpus for investigations on Slavic languages and German. The use of the corpus is restricted for research and education purposes only. To receive the permission to use the GRALIS-Corpus please contact branko.tosovic@uni-graz.at

Please log in using the correct password

..... Login

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 ovakvu sliku:

**Opis projekta**

Gralis-Korpus je višejezički/višejezični paralelni korpus za istraživanje i učenje svih slovenskih/slavenskih jezika, posebno u odnosu na n(j) u Gracu/Grazu (Austrija) u okviru saradnje/suradnje Instituta za slavistiku i Centra za modelovanje/modeliranje informacije Fakulteta humanar korpusa služi CVB, razvijen u Institutu za mašinsku obradu jezika u Štutgartu/Stuttgartu. 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(ij)eljenih na makrogrupe (južnoslovenski/južnoslavenski, istočnoslovenski/istočnoslavenski, pojedine jezike i individualni korpusi – korpusi pisaca kao što je Ivo Andrić, Zoran Živković i dr.). Postoji takođe(r) mogućnost izbora dviju/dv istočnoslovenske/istočnoslavenske. Posebnu skupinu čine slovenski/slavenski jezici upareni sa n(j)emačkim jezikom.

Za sada postoji interfejs/interface na engleskom, n(j)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:

GRALIS Korpust

Imprint | Help | Links | Logout

Department of Slavic Studies | GAMS | GRALIS

UNI GRAZ

Group: All

Aligned languages: Croatian, Serbian

Context size: one sentence

Page size: 20 Corpus entries as HTML

Morphosyntactic Annotation

Search options

Submit

### Project description

The Gralis corpus is a polyglot parallel corpus for the study and learning of all Slavic languages, with special regard to German, created in a cooperative effort by the Department of Slavic Studies and the Center for Information-Modelling in the Humanities at the University of Graz. The Client software CVB was developed by the Institute for Natural Language Processing in Stuttgart. Branko Tošović serves as the project leader, project coordination is observed by Arno Wonsch. Hubert Stigler is in charge of software development and server processing, while Olga Lehner is responsible for the Corpus administration.

The corpus consists of several sub-corpora which are divided into macro-groups (the Southern, Eastern and Western Slavic languages) and micro-groups (corpora for single languages and individual corpora – corpora of the works of writers like Ivo Andrić, Zoran Živković and others). Optionally, users can choose between two macro-groups (e.g. Southern and Eastern Slavic). A special group contains Slavic languages parallelised with German.

The interface currently supports English, German, Bosnian/Croatian/Serbian and Russian language, while support for all Slavic languages is under development.

Due to the corpus still being in the development stage, free access is not possible. People interested in the use of the corpus can request access at: [branko.tosovic@uni-graz.at](mailto:branko.tosovic@uni-graz.at).

5. Sada treba da u gornjem uglu izaberete Gralis-potkorporus, u datom slučaju Gralis-Korporus Branka Ćopića (skraćeno Gralis Ćopić Korporus):

GRALIS Korpust

Impres(s)um | Pomoć | Spisak

Grupa: Sve

↳ Korpust: Ćopićev Gralis-Korporus

↳ Primarni jezik: Ćopićev Gralis-Korporus

↳ Autor: Gralis BKS-Korporus

↳ Žanr: Gralis Ist-Korporus

↳ Funkcionalni stil: Gralis Ist-Zap-Korporus

Gralis Južn-Korporus

Gralis Južn-Zap-Korporus

Gralis Mak-Korporus

Gralis Ru-Nijem-Korporus

Gralis Slav-Korporus

Gralis Zapad-Korporus

6. Kao rezultat dobićete sljedeću sliku:

GRALIS Korpust

Impres(s)um | Pomoć | Spisak

Grupa: Sve

↳ Korpust: Ćopićev Gralis-Korporus

↳ Primarni jezik: Bosanski/Bošnjački

↳ Autor: -

↳ Žanr: -

↳ Funkcionalni stil: -

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

GRALIS Korpus  
Impres(s)um | Pomoć | Spisa

Grupa: Sve

L Korpusi: Čopićev Gralis-Korpus

L Primarni jezik: Bosanski/Bošnjački

L Autor: Bosanski/Bošnjački

L Žanr: Hrvatski

L Funkcionalni stil: Srpski

Kada kliknete na srpski jezik, otvoriće se prozor:

GRALIS Korpus  
Impres(s)um | Pomoć | Spisak linkova/poveznica | Izlaz iz sistema/sustava

Grupa: Sve

L Korpusi: Čopićev Gralis-Korpus

L Primarni jezik: Srpski

L Autor: -

L Žanr: -

L Funkcionalni stil: -

Uvezani jezici: Bjeloruski, Bosanski/Bošnjački, Bugarski, Češki

Širina konteksta: Jedna rečenica

Veličina strane: 20

Korpusni tekstovi kao HTML

Morfosintaksička/morfosintaktička anotacija

Parametri pretrage +

Poslati

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.

GRALIS Korpus  
Impres(s)um | Pomoć | Spisak linkova/poveznica | Izlaz iz sistema/sustava

Grupa: Sve

L Korpusi: Čopićev Gralis-Korpus

L Primarni jezik: Srpski

L Autor: -

L Žanr: -

L Funkcionalni stil: -

Uvezani jezici: Njemački, Poljski, Rusinski, Ruski

Širina konteksta: Jedna rečenica

Veličina strane: 20

Korpusni tekstovi

Morfosintaksička/morfosintaktička

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

Veličina strane: 20

Sve

10

20

30

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.

Ćopićev Gralis-Korpus	
ukupno	490257
N(j)emački	7368
Ruski	13478
Srpski	469411

---

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.\***

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	najgore
11	najlepše
11	najmlađi
10	najmanju
9	najbliži
6	.....

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

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

Grupa: Sve  
 Korpusi: Čopićev 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

Korpusi: Čopićev Gralis-Korpus; Primarni jezik: Srpski;  
 Parametar/Parametri pretrage: naj\* (Velika i mala slova: Ne; Pretraga po lemmama: Ne;)

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

**Artiljerac Marko Medić**  
 Srpski Nije druge nego tući se ovdje , pa ja nadbiti ove naše , ja opet nekako . . . Možda se Nijemac ponovo pripovrati , a možda . . . ako te najzad uhvate , pritrpi se malo dok opale u te , nije ni to najgore , a poslije ćemo već vidjeti .

**Artiljerac Marko Medić**  
 Srpski Nije druge nego tući se ovdje , pa ja nadbiti ove naše , ja opet nekako . . . Možda se Nijemac ponovo pripovrati , a možda . . . ako te najzad uhvate , pritrpi se malo dok opale u te , nije ni to najgore , a poslije ćemo već vidjeti .

**Bitka s aždajom**  
 Srpski Evo aždaje ! - tek bi iznenada povikalo najstarije 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 , nagnuli da bježe i da se kriju iza udaljenih gomila kamenja .

**Bitka s aždajom**  
 Srpski Pazite , spremite se dobro , sad će ona izći ! - opet bi viknulo najstarije čobanče , pirgav i nestašan dječak , žutih mačjih očiju , a svi ostali čobančići , pogleli iza gomila kamenja . Iznirali su njemu " aradu " svijetlih različitih očiju i unta su ustrvali anonami etarisa dnevno

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:

B(j)eloruski  
 Bosanski/Bošnjački  
 Bugarski  
 Češki

Jedna rečenica

Sve Korpusni tekstovi kao HTML

Morfosintaksička/morfosintaktička anotacija HTML  
 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 i mo te e)
%c	Nerazlikovanje velikih i malih slova	na %c