
Značaj paralelnih korpusa za istraživanje frazemskih konstrukcija u nemačkom i srpskom jeziku

Naučni rad

DOI: 10.18485/judig.2025.1.ch21

Kristina Ilić¹,  0009-0002-7426-5530

Apstrakt

Paralelni korpusi su od velikog značaja za kontrastivna istraživanja frazemskih konstrukcija u različitim jezicima. Prema Dobrovoljskom (2011, str. 114), frazemske konstrukcije definisane su kao leksički delimično specifični parovi oblika i značenja, čija semantika proizlazi ne samo iz leksičkog značenja njihovih komponenti i načina njihove sintaktičke povezanosti, već i iz same konstrukcije kao celine. Frazemske konstrukcije su idiomatske kombinacije reči koje se sastoje od određenih leksičkih elemenata, tzv. sidra i određenog broja praznih mesta, tzv. slotova. Ove praznine popunjavaju se tzv. filerima, pri upotrebi konstrukcije (Dobrovol'skij 2011, str. 114). Konstrukciona gramatika je mlada grana lingvistike, a istraživanje frazemskih konstrukcija uz pomoć korpusa, naročito paralelnih korpusa u slučaju kontrastivnih istraživanja, sve više se razvija. U ovom trenutku može se reći da su paralelni korpusi nemačkog i srpskog jezika slabije zastupljeni, za razliku od paralelnih korpusa nemačkog i drugih većih svetskih jezika ili paralelnih korpusa drugih većih svetskih jezika i srpskog jezika. Cilj ovog rada je predstavljanje do sada dostupnih resursa u vidu paralelnih korpusa nemačkog i srpskog jezika i mogućih načina njihove upotrebe u kontrastivnom istraživanju frazemskih konstrukcija na konkretnim primerima, a potom i predlog optimalnog resursa za kontrastivno istraživanje frazemskih konstrukcija u nemačkom i srpskom jeziku.

¹ Filološki fakultet Univerziteta u Beogradu, kristina.ilic997@gmail.com

Ključne reči: Konstrukciona gramatika, korpusna lingvistika, frazemske konstrukcije, srpski jezik, nemački jezik

1. Uvod

Ovaj rad zasnovan je na sponi između konstrukcione gramatike i korpusne lingvistike i prikazuje praktičnu upotrebu paralelnih korpusa iz perspektive krajnjeg korisnika, koji sebi postavlja kao cilj istraživanje frazemskih konstrukcija.

Teorijsku osnovu rada sa jedne strane čine postulati konstrukcione gramatike i frazemske konstrukcije kao jedan od njenih predmeta istraživanja, dok sa druge strane rad predstavlja tipologiju i pruža opis različitih postojećih paralelnih korpusa, sa fokusom na paralelne korpusne srpskog i nemačkog jezika (*SrpNemKor*, *InterCorp* i *OpenSubtitles 2018 parallel – German*) i srpskog i italijanskog (*SerbItaCor3*).

Različiti nivoi složenosti frazemskih konstrukcija zahtevaju korišćenje upita različitog nivoa kompleksnosti za pretragu korpusa. U ovom radu će biti obrađena tri nivoa složenosti frazemskih konstrukcija i upita pomoću kojih se one pretražuju u korpusu.

Empirijsko istraživanje sastoji se od opisa različitih mogućnosti pretrage korpusa korišćenjem konkretnih primera frazemskih konstrukcija, u cilju predstavljanja postojećih i nedostajućih potrebnih funkcionalnosti korpusa, kako bi se dobili optimalni rezultati.

Cilj ovog rada je dati predlog optimalnog resursa za kontrastivno istraživanje frazemskih konstrukcija u nemačkom i srpskom jeziku, u pogledu sadržaja i funkcionalnosti.

2. Paralelni korpusi

Predmet istraživanja korpusne lingvistike predstavljaju korpusi različitog obima i sadržaja, koji su pretraživi uz pomoć korpusnih menadžera, različitih funkcionalnosti. Korpusni menadžeri predstavljaju kompjuterske programe koji omogućavaju izgradnju, uređivanje, obeležavanje tj. anotaciju i pretraživanje korpusa².

Paralelni korpusi predstavljaju jednu od vrsta korpusa. „Paralelni korpusi (poravnati korpusi) sastoje se od nekoliko semantički ekvivalentnih tekstova, koji su poravnati na određenom nivou — na nivou pasusa, rečenica, fraza ili reči.“ (Krstev & Vitas, 2011, str. 495). Svrha korpusa određuje: (1) izbor tekstova, (2) njihovu veličinu, (3) nivo poravnanja i (4) metod korišćen za poravnanje (Krstev & Vitas, 2011, str. 495–496). Autori ovog određenja

² <https://www.sketchengine.eu/glossary/corpus-manager/>, pristupljeno 07.11.2025.

svrhe korpusa, Krstev i Vitas, naglašavaju da poravnati korpus ne mora biti multilingvalan, već se može sastojati i iz dva ili više nezavisnih prevoda istog teksta na jednom jeziku (Krstev & Vitas, 2011, str. 495).

Blumel, citirajući McEnery & Xiao, prevodne korpuse (eng. *translation corpora*) deli na (1) paralelne korpuse, koji se sastoje od izvornih tekstova i njihovih prevoda na jedan ili više drugih jezika i oni mogu biti (a) jednosmerni (eng. *unidirectional*), (b) dvosmerni (eng. *bidirectional*) i (c) višesmerni (eng. *multidirectional*); zatim na (2) uporedne korpuse (eng. *comparable corpora*), koji se sastoje od sličnih tekstova i na (3) ekvivalentne korpuse, koji predstavljaju kombinaciju prva dva tipa (Blumel, 2014, str. 32, citirano prema McEnery & Xiao, 2008).

Uporedni korpusi se ne sastoje nužno od semantički ekvivalentnih tekstova, već od tekstova koji pripadaju istom domenu, istom vremenskom periodu itd.“ (Krstev & Vitas, 2011, str. 496).

U pogledu sadržaja, „Paralelni korpusi kao kolekcije tekstova na dva ili više jezika najčešće sadrže administrativne, pravne ili novinske tekstove, te zapravo predstavljaju korpuse podjezika”, pri čemu su „višejezični paralelni korpusi književnih tekstova najčešće mnogo manji u odnosu na korpuse pravnih ili novinskih tekstova jer njihova priprema zahteva više vremena i rada, ali je i njihov broj poslednjih godina u porastu” (Andonovski, 2019, str. 29–30).

Kako navode Doval i Sančez Nieto, paralelni korpusi se istražuju najpre kao proizvodi, ali se može izučavati i proces njihove izrade, koji podrazumeva „kompilaciju, anotaciju, poravnanje, indeksiranje i vršenje upita nad korpusom”³ (Doval & Sánchez Nieto, 2019, str. 6–17).

Korpusi, pored toga što mogu predstavljati sam predmet istraživanja, kao u slučaju korpusne lingvistike, služe i kao alati koji se koriste za sprovođenje različitih lingvističkih istraživanja. Andonovski (2019), citirajući Dobrića, navodi različite oblasti lingvistike u kojima korpusi nalaze široku primenu, kao što su morfologija, sintaksa, semantika, diskurs, pragmatika, leksikologija, leksikografija (Andonovski, 2019, str. 10–11, prema Dobrić 2009, str. 362), pri čemu naglašava koje sve mogućnosti korpusi nude.

Kako navodi Andonovski, u zavisnosti od nivoa anotacije, korpusi omogućavaju istraživačima da uoče različite primere upotrebe jezika, da pozivanjem seta reči jednog jezika i njihovih gramatičkih oblika prilikom istraživanja sagledaju frekventnost pojavljivanja određene reči ili fraze postavljene kroz upite za pretragu, u kojim se sve oblicima i varijantama postavljena reč ili fraza pojavljuje, kao i u kakvoj je semantičkoj korelaciji

³ Svi prevodi direktnih citata u radu su od strane autora.

sa drugim rečima i frazama i njihovim oblicima u datom korpusu (Andonovski, 2019, str. 11).

Navedene funkcionalnosti pretrage korpusa pokazaće se kao izuzetno značajne za korpusna istraživanja frazemskih konstrukcija.

Paralelni i uporedni korpusi, prema Doval i Sančez Nieto (Doval & Sánchez Nieto, 2019, str. 4), nalaze primenu u (1) mašinskom prevodenju, i to u vidu podataka za trening prevodilačkih modela, prevodnih memorija ili leksičku i terminološku ekstrakciju, u višejezičnoj obradi prirodnih jezika (NLP) i u (2) lingvističkim istraživanjima u oblasti kontrastivne lingvistike i translatologije, prevodilačke prakse, leksikografije i oblasti podučavanja stranih jezika i prevodenja.

Analiza dvojezičnog korpusa, prema Modercu (2015), može se fokusirati na traduktološke i stilističke aspekte pojedinih prevoda ili više prevoda istovremeno, pratiti rad pojedinih prevodilaca ili porediti prevode istog dela od strane različitih prevodilaca (Moderc, 2015, str. 301).

Andonovski, citirajući Simeona, ističe značaj paralelnih korpusa za višejezičnu leksikografiju, i to u tome što „omogućavaju korisnicima da sagledaju uparene ekvivalente određenih segmenata teksta kao i širi kontekst koji ne postoji u klasičnim dvojezičnim rečnicima” (Andonovski, 2019, str. 24, prema Simeon 2002: 212).

Tvorci višejezičnog paralelnog korpusa *InterCorp*, Čermák & Rosen (Čermák & Rosen, 2012, str. 1), naglašavaju značaj paralelnih korpusa „posebno za jezike sa manjim brojem izvornih govornika, podržavajući i uporedna istraživanja i proučavanja jednog jezika iz perspektive drugih jezika”.

3. Postojeći paralelni korpusi srpskog i drugih jezika

S obzirom na činjenicu da se radi o jeziku sa malim brojem govornika tj. srpskom, a izrada korpusa predstavlja složen zadatak, paralelni korpusi srpskog i drugih svetskih jezika nisu brojni, kao ni opšti korpusi srpskog jezika, ali u poslednjim godinama beleži se njihov porast. Paralelni korpusi značajni za ovo istraživanje su:

- (1) *SrpNemKor*, dostupan na platformi *Biblisha*⁴, paralelni korpus nemačkog i srpskog jezika koji se sastoji od po 7 romana na oba jezika i njihovih prevoda, nastao u okviru doktorske disertacije *Mreža otvorenih podataka i jezički resursi u procesu izgradnje srpsko-nemačkog literarnog korpusa* (Andonovski, 2019);

⁴ <http://biblisha.jerteh.rs/>, pristupljeno 08.08.2025.

- (2) *InterCorp*, dostupan preko korpusnog menadžera *KonText*⁵, pri čemu je fokus na podkorpusu *DE-SR* iz centralnog dela korpusa, pod nazivom *Core* (Rosen, 2016);
- (3) *SerbItaCor3*, dostupan na platformi *NoSke*⁶, paralelan korpus italijanskog i srpskog jezika, sačinjen od 243 tekstova, odnosno 267 uparenih tekstova (Moderc et al., 2023);
- (4) *OpenSubtitles 2018 parallel – German*, koji sadrži prevode filmova i deo je kolekcije *OpenSubtitles 2018 parallel*⁷, sačinjene od 60 korpusa na 58 jezika, pohranjen na korpusnom menadžeru *Sketch Engine* (Lison & Tiedemann, 2016).

U Tabeli 1 dat je uporedni prikaz ovih korpusa, prema jezicima, izvorima, nivou poravnanja, podržanosti *CQL-a*⁸ i funkcionalnostima korpusnog menadžera.

Tabela 1: Korpusi obuhvaćeni istraživanjem

korpus (platforma) i autor	jezici	izvori	poravnanje	CQL	najvažnije funkcionalnosti
<i>SrpNemKor</i> (Biblisha) Jelena Andonovski	srpski i nemački	7 srpskih romana 7 nemačkih romana (različitih)	nivo rečenice	ne	smer srpski → nemački
<i>SerbItaCor3</i> (NoSke) Saša Moderc	srpski i italijanski	267 prevoda ukupno (ima i više prevoda jednog književnog dela)	nivo rečenice	da	parallel concordance, concordance, word list, keywords, frequency text type analysis
<i>InterCorp</i> (KonText) Alexandr Rosen	nemački i srpski	core (književna dela, stručne knjige) collections (<i>Syndicate</i> , <i>Acquis</i> , <i>Europarl</i> , <i>Open Subtitles</i> , Biblija)	nivo rečenice prema češkom	da	parallel concordance, word list, keywords, frequency, collocations,
<i>OpenSubtitles 2018 parallel – German</i> (Sketch Engine) Jörg Tiedemann	engleski ↔ nemački ↔ srpski	titlovi za film sa www.OpenSubtitles.org	nivo rečenice prema engleskom	da	parallel concordance concordance word list keywords frequencyword sketch, thesaurus, n-grams,

⁵ <https://intercorp.korpus.cz/>, pristupljeno 08.08.2025.

⁶ <https://noske.jerteh.rs/>, pristupljeno 08.08.2025.

⁷ <https://www.sketchengine.eu/opensubtitles-parallel-corpora/>, pristupljeno 08.08.2025.

⁸ *Korpusni jezik za upite (CQL)* je poseban kod ili jezik za upite koji se koristi [...] za pretragu složenih gramatičkih ili leksičkih obrazaca ili za korišćenje kriterijuma pretrage koji se ne mogu podesiti pomoću standardnog korisničkog interfejsa.

<https://www.sketchengine.eu/documentation/corpus-querying/>, pristupljeno 08.08.2025.

					text type analysis
--	--	--	--	--	--------------------

4. Frazemske konstrukcije

„Frazemske konstrukcije se mogu definisati kao konstrukcije koje kao celina imaju leksičko značenje, pri čemu su određene pozicije u njihovoj sintaksičkoj strukturi popunjene leksičkim elementima, dok druge predstavljaju slotove (eng. Slots), koji moraju biti popunjeni“ (Dobrovol'skij, 2011, str. 114).

Prema Dobrovol'skom, frazemske konstrukcije su jedan od podtipova frazema. Ovaj autor deli frazeme na (1) idiome, koji pokazuju „izrazite nepravilnosti u odnosu forme i značenja“, kao npr. *ins Graß beißen* (srp. *umreti; otegnuti papke*); (2) kolokacije, koje se sastoje od osnovne konstituente sa punim značenjem i vezane komponente sa oslabljenim ili metaforičkim značenjem, u koje spadaju i funkcionalni glagolski spojevi, kao npr. *eine Entscheidung treffen* (srp. *doneti odluku*); (3) situacione klišeje tj. rutinske formulacije, npr. *Guten Tag!* (srp. *Dobar dan!*); (4) izreke, npr. *Morgenstund hat Gold im Mund.* (srp. *Ko rano rani, dve sreće grabi.*); (5) gramatičke frazeme, npr. *geschweige denn* (srp. *a kamoli*) i (6) frazemske konstrukcije tj. frazeošablone, kao na primer *So ein Glück!*, sa apstraktnom shemom/šablonom (nem. *Pattern*) [*so EIN N*]⁹ (Dobrovol'skij, 2011, str. 111).

To su stabilne višečlane jedinice, koje se sastoje od popunjenih delova, tzv. sidra (nem. *Anker*) i slobodnih slotova (nem. *Slots*), koji se prilikom upotrebe popunjavaju leksemama ili kombinacijama leksema koje se nazivaju fileri (nem. *Filler* ili *Slotfüller*) (Pavlova, 2022, str. 594).

Prema Ilić i Šlund (2024), „do ekvivalenata FK [frazemskih konstrukcija] može se doći na nekoliko različitih načina, u zavisnosti od dostupnosti resursa: (1) pretragom u paralelnim korpusima nemačkog i ciljnog jezika; (2) u prevodima književnih dela sa nemačkog govornog područja na ciljni jezik, ručno ili uz pomoć paralelnih korpusa književnih dela; ili (3) samostalnim prevodom“ (Ilić & Šlund, 2024, str. 291). U ovom radu ćemo se fokusirati na prva dva slučaja, jer je fokus rada na upotrebi korpusa.

Prema složenosti upita za pretragu u jednojezičnom korpusu, u svojoj prezentaciji Krotova (2024) razlikuje tri nivoa složenosti frazemskih konstrukcija. Na *Niveau 1* nalaze se frazemske konstrukcije sa malo ili bez ograničenja kod filera, malo ili bez varijacije kod sidra, te za istraživanje nije nužno koristiti anotirane korpusne, već su dovoljni pretraga na Guglu,

⁹ N označava da na tom mestu stoji imenica.

onlajn rečnici i *Google Books Ngram-Viewer*¹⁰. Primeri takvih frazemskih konstrukcija su *es sei denn X* (srp. *sem/osim ako ne*) i *ob X oder nicht* (srp. *bio/bila/bilo X ili ne*). Za frazemske konstrukcije na Nivou 2 preporučljivo je koristiti korpusne alate sa anotiranim korpusima koji sadrže anotaciju vrsta reči, tzv. *POS-Tagging*, i mogućnost navođenja željenog razmaka između elemenata u okviru upita¹¹, kako bi upit bio precizan. Frazemska konstrukcija kod koje je takav postupak potreban je, na primer, *du und dein X/ du mit deinem X* (srp. *ti i tvoj/tvoja/tvoje/ tvoji X*). Za istraživanje frazemskih konstrukcija na Nivou 3 potrebno je da korpusni alati imaju mogućnost pretrage ponavljanja jednog elementa i da omogućavaju generisanje frekvencijskih lista za filere, kao što je slučaj kod korpusnog menadžera *Sketch Engine*. Alternativa bi bili regularni izrazi ili programski jezik *Python* za generisanje frekvencijskih lista. Ovom nivou, na primer, pripada frazemska konstrukcija *Kind ist Kind* (srp. *dete je dete*) sa apstraktnom shemom *X' ist/sind [immer] X' (X' je/su X')* (Krotova, 2024). Ove klasifikacije ćemo se pridržavati i u ovom radu.

5. Frazemske konstrukcije i pretraga u paralelnom korpusu: rezultati istraživanja

Rezultati istraživanja pokazuju da se frazemske konstrukcije u nemačkom jeziku mogu efikasno identifikovati i analizirati u paralelnim korpusima, ali da način njihove pretrage zavisi od formalne strukture same konstrukcije. Stoga je važno da korpusni menadžeri poseduju određene tehničke funkcionalnosti. Konstrukcije su klasifikovane na osnovu stepena složenosti, uzimajući u obzir tehničke funkcionalnosti i složenost upita potrebnih za pretragu u korpusu.

Nivo 1. Za analizu konstrukcija koje imaju jednostavno sidro i jedan slot sa velikim brojem mogućih filera, poput *es sei denn X* (srp. *osim ako ne X*) ili *durch und durch X* (srp. *skroz na skroz X*), odnosno *X durch und durch* (srp. *X skroz na skroz*), moguće je koristiti jednostavne metode pretrage u paralelnim korpusima. Za njihovu pretragu nije potrebna POS anotacija i dovoljno je u polje za pretragu uneti sidro *durch und durch*. U korpusu paralelnih tekstova se vidi da li je i u drugom jeziku upotrebljena frazemska konstrukcija.

S obzirom na to da je za *SrpNemKor* na platformi *Biblisha* moguća samo pretraga u smeru srpski → nemački, moguće je doći do prevodnih ekvivalenata za konstrukciju *durch und durch X* (srp. *skroz na skroz X*) tako

¹⁰ <https://books.google.com/ngrams/>, pristupljeno 15.07.2025.

¹¹ Na primer, navođenje {0,2} posle elementa frazemske konstrukcije u okviru upita znači da se naredni element nalazi odmah pored, nakon jednog tokena ili nakon dva tokena.

što se postavi upit koji sadrži njen mogući prevod. Pogodak je pronađen uz pomoć jednog od sinonima izraza *skroz na skroz*, odnosno uz pomoć reči *potpuno*, kao što se vidi na Slici 1, što znači da se prevodilac nije opredelio za frazemsku konstrukciju u srpskom jeziku, već je posegnuo za prilogom sa sinonimnim značenjem.

Das Parfum / Patrick Süskind = Parfum / Petnik Zorand, ID: 11.1.097 Metadata	n2174 Sie mussten durch und durch getränkt sein von seinem Geruch.	n2174 Mora da su potpuno natopljene njegovim mirisom.
---	---	--

Slika 1: Pretraga u korpusu SrpNemKor na osnovu reči „potpuno“

Unošenjem sidra u polje za pretragu može se vršiti pretraga i u korpusima koji se nalaze na platformama koje se baziraju na korpusnim menadžerima *NoSketchEngine* i *SketchEngine*, odnosno u korpusu *SerbItaCor3*, kao što se vidi na Slici 2 i u korpusu *OpenSubtitles 2018 parallel – German*, prikazanom na Slici 3¹². Moguće je analizirati frekvenciju filera ili slotova koji se nalaze u pretraživanoj konstrukciji. Ukoliko je potrebno analizirati frekvenciju u ciljnom jeziku, to je moguće samo unošenjem ekvivalentne konstrukcije i u ciljnom jeziku u pretragu¹³. U tom slučaju dobijamo samo rezultate u kojima je konstrukcija prevedena odgovarajućom konstrukcijom. Nije moguće dobiti frekvenciju različitih ekvivalenata konstrukcije u ciljnom jeziku, što bi predstavljalo značajan podatak.



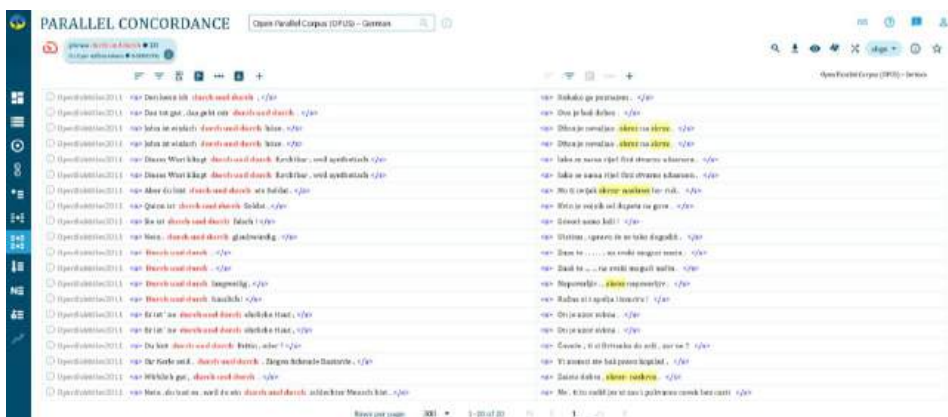
Slika 2: Pretraga konstrukcije „skroz naskroz X/X skroz naskroz“ u korpusu SerbItaCor3.

Na isti način vrši se i pretraga u korpusu *InterCorp* preko platforme *KonText*. U rezultatu, obeležavanjem bilo koje reči u izvornom tekstu, ne samo reči obuhvaćenih pretragom, automatski se istom bojom označi ekvivalent te reči u ciljnom tekstu, ukoliko ga ima, kao što se vidi na Slici

¹² Korpus *OpenSubtitles 2018 parallel*, kako se vidi na Slici 3, ima sledeći nedostatak: paralelni tekstovi ne predstavljaju nužno izvorni tekst i njegov prevod, već je kod oba teksta često u pitanju prevod sa engleskog. To otežava lingvistička i translatološka istraživanja.

¹³ <https://www.sketchengine.eu/guide/parallel-concordance-searching-translations/#toggle-id-2>, pristupljeno 07.08.2025.

4. Direktnog ekvivalenta pojedinačne reči nema ukoliko je prevodilac izabrao neko drugo rešenje. Dolazi i do grešaka u obeležavanju. Na platformi *Sketch Engine* nije moguće kliknuti na pojedinačne reči i dobiti obeležene ekvivalente.



Slika 3: Pretraga konstrukcije „durch und durch X/X durch und durch“ u korpusu OpenSubtitles



Slika 4: Pretraga konstrukcije „durch und durch X/X durch und durch“ u korpusu InterCorp

Primeri pogrešnog ili izostalog obeležavanja ekvivalenata u korpusu *InterCorp* prikazani su na Slici 5, na primeru frazemske konstrukcije *es sei denn X* (srp. *osim ako ne X*).

Nešto složenija za pretragu je konstrukcija *ob X oder nicht*, s obzirom na to da se filer nalazi između dva dela sidra. Stoga bi trebalo upotrebiti *CQL* upit *[word="ob"] [word="oder"] [word="nicht"]* kako bi se dalo uputstvo da je mesto slotu između dva dela sidra, odnosno posle reči *ob*.

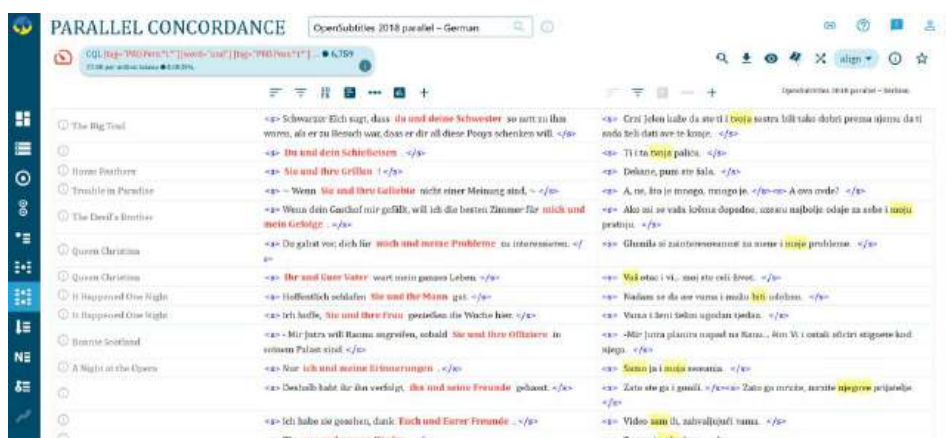


Slika 5: Primeri pogrešnog ili izostalog obeležavanja ekvivalenata u korpusu InterCorp

Nivo 2. Kod konstrukcija koje uključuju lične i posesivne zamenice u sidru, pretraga postaje znatno složenija. Takve su konstrukcije *du und dein X* i njene varijante *du [schon wieder/immer] mit deinem X* i *du mit deinem X [schon wieder?/!]*, čije apstraktne sheme glase $\text{Pron}_{\text{pers}} \text{ nom}/N'_{\text{nom}}$ und dein(e)/sein(e)/ihr(e)/euer(eure)... N'_{nom} , odnosno $\text{Pron}_{\text{pers}} \text{ nom}/N'_{\text{nom}}$ [schon wieder/immer] mit deinem/seinem/deiner/seiner/seinen/ihrer... N'_{nom} [schon wieder?/!], čiji bi srpski ekvivalent bio *Ti i tvoj/tvoja/tvoje* N_{nom} . Zamenice i glagol mogu stajati u bilo kojem licu. Kod ovakvih konstrukcija, neophodno je koristiti anotirane korpuse sa POS oznakama.

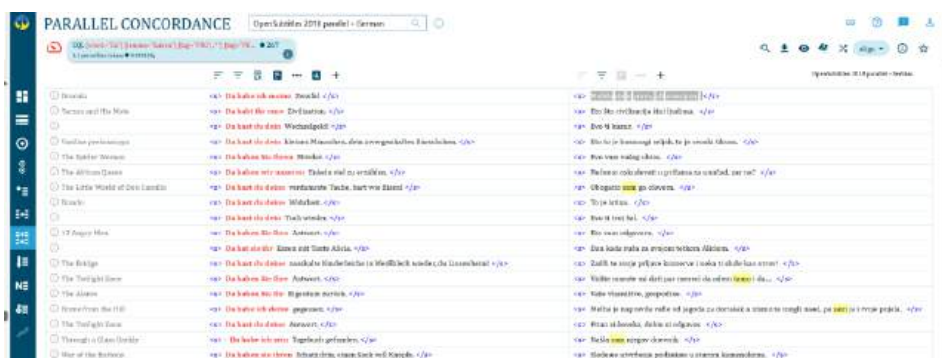
Kada se istražuje konstrukcija *du und dein X* u korpusnim menadžerima koji podržavaju CQL, ukoliko se postavi opšti upit `[tag="PRO\..*" | tag="N.Name.Nom.*" | tag="N.Reg.Nom.*"] [word="und"] [tag="PRO.Poss.*"] [tag="N.*"]`, koji bi na prvi pogled obuhvatio sva moguća lica i dao najbolje rezultate, dobije se veliki broj rezultata koji nisu frazemska konstrukcija, već obuhvataju različita lica lične i posesivne zamenice ili nemaju značenje konstrukcije, već predstavljaju složeni subjekat, kao npr. *Nora und ihr Mann kommen heute Abend.* (srp. *Nora i njen suprug dolaze večeras*).

Kako bi se suzila pretraga, potrebno je u okviru CQL upita precizirati da je kod lične zamenice i kod prisvojne zamenice u pitanju isto lice: `[tag="PRO.Pers.*1*"] [word="und"] [tag="PRO.Poss.*1*"] [tag="N.*"]`, pri čemu se dobiju rezultati prikazani na Slici 6. Ovom pretragom su obuhvaćena sva lica, ali su zamenice u pojedinačnim primerima u istom licu. Zatim je potrebno ručno izdvojiti primere u kojima se javlja tražena konstrukcija, kao što su *Du und dein Schießseisen* i ekvivalent *Ti i tvoja palica* sa Slike 6. Ova pretraga ne daje rezultate u korpusu *InterCorp*.



Slika 6: Rezultati pretrage [tag="PRO.Pers.*1*"] [word="und"]
[tag="PRO.Poss.*1*"] [tag="N.*"]

Takva je i konstrukcija *Da hast du [dein/deinen/deine] N_{Akk} [!]* (srp. *Evo ti tvoj/tvoja N_{nom}*), sa apstraktnom shemom *Da habe/hast/hat/habt/haben Pron_{pers} nom [Pron_{poss} akk] N_{Akk} [!]*. Ona se u korpusu može pretraživati uz pomoć CQL upita [word="Da"] [lemma="haben"] [tag="PRO\..*"] [tag="PRO.Poss.*"], pri čemu veliki deo rezultata sadrži tu konstrukciju, kao što se vidi na Slici 7.



Slika 7: Rezultati pretrage [word="Da"] [lemma="haben"]
[tag="PRO\..*"] [tag="PRO.Poss.*"]

Zbog razlike u tagovima u različitim korpusima, potrebno je prilagoditi upit za pretragu u korpusu *InterCorp*, koji onda glasi [word="Da"] [lemma="haben"] [tag="P.*"] [tag="P.*"], a rezultati su prikazani na Slici 8. Među ovim rezultatima, frazemske konstrukcije su manje zastupljene nego u slučaju korpusa *OpenSubtitles*, s obzirom na to

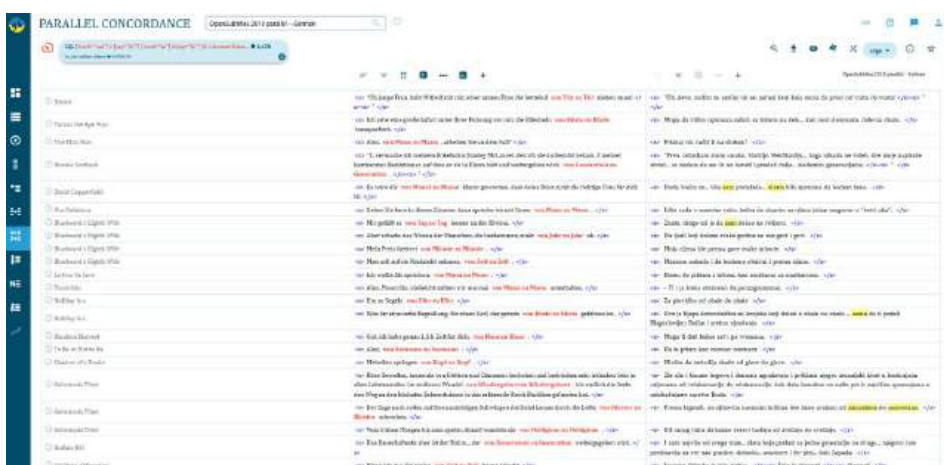
da je korpus ograničen na 13 romana koji su prevedeni sa nemačkog na srpski. Osim toga, ova konstrukcija je karakteristična za govorni jezik, pa je zastupljenija u filmskim titlovima.

The screenshot shows the konText search interface. The search query is `[word='Da'] [lemma='haben'] [tag='P.*'] [tag='P.*']`. The results are displayed in a table with columns for the search results and the corresponding German text. The results are sorted by frequency, with the most frequent results at the top.

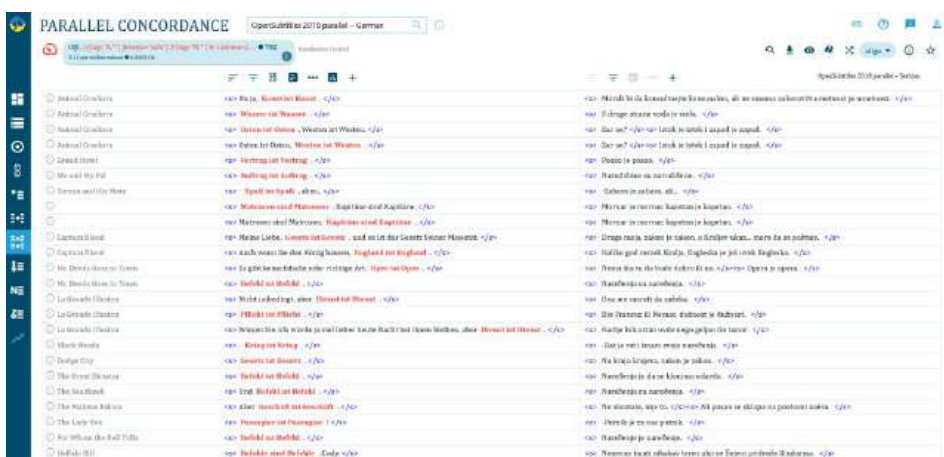
Rank	Lemma	Frequency	Text
1	Da	100	Da ich ein Kind bin, das Träumen...
2	Da	80	Da ich ein Kind bin, das Träumen...
3	Da	60	Da ich ein Kind bin, das Träumen...
4	Da	40	Da ich ein Kind bin, das Träumen...
5	Da	30	Da ich ein Kind bin, das Träumen...
6	Da	20	Da ich ein Kind bin, das Träumen...
7	Da	10	Da ich ein Kind bin, das Träumen...
8	Da	10	Da ich ein Kind bin, das Träumen...
9	Da	10	Da ich ein Kind bin, das Träumen...
10	Da	10	Da ich ein Kind bin, das Träumen...
11	Da	10	Da ich ein Kind bin, das Träumen...
12	Da	10	Da ich ein Kind bin, das Träumen...
13	Da	10	Da ich ein Kind bin, das Träumen...
14	Da	10	Da ich ein Kind bin, das Träumen...
15	Da	10	Da ich ein Kind bin, das Träumen...
16	Da	10	Da ich ein Kind bin, das Träumen...
17	Da	10	Da ich ein Kind bin, das Träumen...
18	Da	10	Da ich ein Kind bin, das Träumen...
19	Da	10	Da ich ein Kind bin, das Träumen...
20	Da	10	Da ich ein Kind bin, das Träumen...

Slika 8: Rezultati pretrage `[word="Da"] [lemma="haben"] [tag="P.*"] [tag="P.*"]`

Nivo 3. Treći nivo složenosti obuhvata konstrukcije koje sadrže rekurentne elemente, odnosno kod kojih se ista reč, najčešće imenica, javlja kao filer u oba slotu. Tipični primeri uključuju konstrukcije poput *von Tag zu Tag* (srp. *iz dana u dan*) i *Kind ist Kind* (srp. *dete je dete*), odnosno apstraktne sheme *von X' zu X'* i *X' ist X'*. Za pretragu ovih konstrukcija u korpusnim menadžerima *Sketch Engine* i *NoSketch Engine* može se koristiti upit `[word="von"] 1:[tag="N.*"] [word="zu"] 3:[tag="N.*"] & 1.lemma=3.lemma`, pri čemu se dobiju rezultati sa Slike 9, odnosno `1:[tag="N.*"] [lemma="sein"] 2:[tag="N.*"] & 1.lemma=2.lemma`, sa rezultatima sa Slike 10, pri čemu je izuzetno važno u upitu naznačiti koje reči su iste.

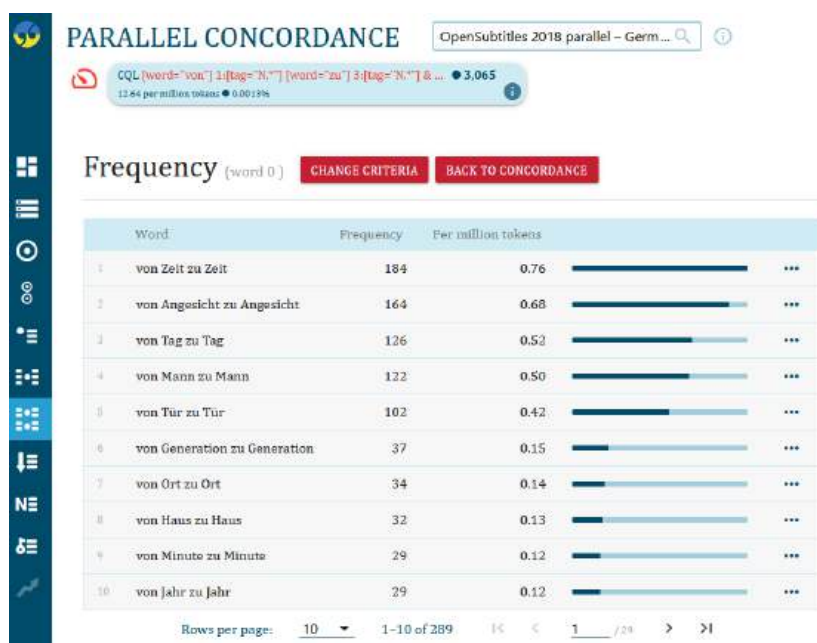


Slika 9: 1:[tag="N.*"] [word="zu"] 3:[tag="N.*"] & 1.lemma=3.lemma



Slika 10: 1:[tag="N.*"] [lemma="sein"] 2:[tag="N.*"] & 1.lemma=2.lemma

Frekvencijske liste u alatima poput *Sketch Engine* dodatno pomažu u identifikaciji najčešćih leksičkih jedinica koje se javljaju kao fileri. Metodološki, one zahtevaju kombinaciju korpusnih alata sa kvantitativnim analizama. Takva lista prikazana je na Slici 11.



Slika 11: Frekvencijska lista filera za konstrukciju von X' zu X'

U svim navedenim slučajevima, jasno je da izbor metode pretrage korpusa direktno utiče na uspešnost analize. Jednostavne konstrukcije mogu se pratiti čak i bez naprednih alata, dok fleksibilnije i kompleksnije strukture zahtevaju preciznu anotaciju i naprednu pretragu.

6. Zaključak

Rezultat ovog istraživanja predstavlja obrazloženi predlog funkcionalnosti koje treba da ima idealni korpusni alat za istraživanje frazemskih konstrukcija.

Na osnovu prikazanih karakteristika postojećih korpusa i opisa funkcionalnosti koje su potrebne za pretragu i istraživanje frazemskih konstrukcija, zaključuje se da bi idealan paralelni korpus nemačkog i srpskog jezika sadržao (1) romane na oba jezika paralelizovane na nivou rečenice, (2) POS anotaciju, (3) više različitih prevoda jednog književnog dela, (4) paralelizovane titlove, pri čemu je izvorni jezik nemački, a ciljani jezik srpski i obrnuto i (5) dodatno, ukoliko je moguće: ekvivalentni korpus koji bi sadržao i uporedni korpus sa novinskim tekstovima o događajima na oba govorna područja.

Idealan korpusni menadžer bi omogućavao prvo mogućnost pretrage u oba smera, drugo mogućnost pretrage uz pomoć CQL-a, treće

mogućnost selekcije bilo koje reči u konkordansi i automatsko obeležavanje njenog ekvivalenta u ciljnom tekstu i četvrto mogućnost generisanja frekvencijskih lista reči iz upita i rezultata.

Pretrage u paralelnim korpusima dragocene su za kontrastivne analize frazemskih konstrukcija i kontrastivna istraživanja uopšte. To je naročito značajno za jezike u čijim istraživanjima konstrukciona gramatika nije još razvijena, pa se ovakvim kontrastivnim istraživanjima doprinosi i razvoju nauke o ciljnom jeziku. Stoga je važno razvijati i unapređivati paralelne korpus srpskog i nemačkog, kao i srpskog i drugih jezika.

Zahvalnica

Ovo istraživanje sprovedeno je u okviru COST akcije PhraConRep Višejezični repozitorijum frazemskih konstrukcija na jezicima centralne i istočne Evrope („A Multilingual Repository of Phraseme Constructions in Central and Eastern European Languages”), koja ima za cilj da kontrastivno istraži frazemske konstrukcije nemačkog i srpskog jezika, pri čemu upotreba različitih alata i paralelnih korpusa, predstavljenih u ovom istraživanju, može odigrati važnu ulogu.

Literatura

- [1] Andonovski, J. S. (2019). *Mreža otvorenih podataka i jezički resursi u procesu izgradnje srpsko-nemačkog literarnog korpusa*. [Doktorska disertacija]. Beograd: Univerzitet u Beogradu, Filološki fakultet.
- [2] Blumel, B. (2014). *Learning in parallel: Using parallel corpora to enhance written language acquisition at the beginning level*. Dimension, 31–48.
- [3] Dobrovol'skij, D. (2011). Phraseologie und Konstruktionsgrammatik. In A. Lasch & A. Ziem (Hrsg.), *Konstruktionsgrammatik III. Aktuelle Fragen und Lösungsansätze* (S. 110–130). Stauffenburg.
- [4] Čermák, F. & Rosen, A. (2012). The Case of InterCorp, a multilingual parallel corpus. *International Journal of Corpus Linguistics*, 411–427.
- [5] Doval, I. & Sánchez Nieto, M. (2019). Parallel corpora in focus. An account of current achievements and challenges. U I. Doval, & M. Sánchez Nieto (ur.), *Parallel Corpora for Contrastive and Translation Studies. New resources and applications*, (str. 1–18). Amsterdam / Philadelphia: John Benjamins Publishing Company.
- [6] Krotova, E. (2024). Projektrelevante Webkorpora & Analysesoftware [Prezentacija]. *Training "Datenbearbeitung und Übersetzung im Repositorium auf GitHub" Osijek, Croatia. 4.–6. September 2024*. Preuzeto sa https://github.com/PhKW/PhKWB/blob/main/Workshops/Workshop_20240905.pdf

- [7] Krstev, C. & Vitas, D. (2011). An aligned English-Serbian corpus. U N. Tomović & J. Vujić (ur.), *ELLSIIR Proceedings (English Language and Literature Studies: Image, Identity, Reality)*, Volume I, 495–508. Belgrade: Faculty of Philology, University of Belgrade.
- [8] Pavlova, A. (2022). Mehrsprachige Datenbank der Phrasem-Konstruktionen. U A. Klosa-Kückelhaus, S. Engelberg, C. Möhrs, & P. Storjohann (ur.), *Dictionaries and society: Proceedings of the XX EURALEX International Congress* (str. 594–604). Mannheim: IDS.
- [9] Stanković, R., Krstev, C., Lazić, B. & Vorkapić, D. (2015). A bilingual digital library for academic and entrepreneurial knowledge management. *10th International Forum on Knowledge Asset Dynamics — IFKAD 2015*. Bari, Italy.
- [10] Илић К. & Шлунд, К., (2024). Увиди у изразу вишејезичног репозиторијума фраземских конструкција. У С. Марјановић (ур.) *Модерни речници у функцији просечног корисника: стари проблеми, савремени правци и нови изазови*, Лексикографски сусрети, Вол. 1 (str. 285–296). Београд: Филолошки факултет Универзитета у Београду.

Korpusi

- [1] Andonovski, J. S. (2019). *Mreža otvorenih podataka i jezički resursi u procesu izgradnje srpsko-nemačkog literarnog korpusa*. [Doktorska disertacija]. Beograd: Univerzitet u Beogradu, Filološki fakultet.
- [2] Lison, P. & Tiedemann, J. (2016). OpenSubtitles2016: Extracting large parallel corpora from movie and TV subtitles. U Calzolari, N., Choukri, K., Declerck, T., Goggi, S., Grobelnik, M., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J., & Piperidis, S. *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC-2016)*(923–929). Portorož: European Language Resources Association (ELRA).
- [3] Moderc, S., Stanković, R., Tomašević, A. & Škorić, M. (2023). An Italian-Serbian sentence aligned parallel literary corpus. *Review of the National Center for Digitization*, 43, 78–91. <https://www.ncd.matf.bg.ac.rs/issues/43/09.pdf>
- [4] Rosen, A. (2016). InterCorp – A look behind the façade of a parallel corpus. U E. Gruszczyńska (ur.), *Polskojęzyczne korpusy równoległe: Polish-language parallel corpora* (str. 21–40). I Warszawa: instytut Lingwistyki Stosowanej.

The Importance of Parallel Corpora for the Research of Phraseme Constructions in the German and Serbian languages

Kristina Ilić

Summary

Parallel corpora are of great importance for contrastive studies of phraseme constructions in different languages. According to Dobrovol'skij (2011: 114), phraseme constructions are defined as lexically partially specific pairs of forms and meanings (= constructions), whose semantics derives not only from the lexical meaning of their components and the manner of their syntactic connection but also from the construction as a whole. Phraseme constructions are idiomatic combinations of words that consist of certain lexical elements, so-called anchors, and a certain number of empty spaces, the so-called slots. These slots are filled with the so-called fillers when using the structure (Dobrovol'skij, 2011, p. 114). Constructional grammar is a young branch of linguistics, and the research of phraseme constructions with the help of corpora, especially parallel corpora in the case of contrastive research, is developing increasingly. This is particularly the case when the purpose of the corpus search is to identify translation equivalents of source language constructions in parallel texts of the target language, as these equivalents are often constructions themselves, thereby contributing to the investigation of phraseme constructions in the target language as well. At the moment, it can be said that the parallel corpora of the German and Serbian languages are less represented, in contrast to the parallel corpora of German and other major world languages, such as *COMBIDIGILEX* or *EuReCo*, or the parallel corpora of other major world languages and the Serbian language, such as *SrpEngKor*, *SrpFranKor*, *Evroteka* or *ParCoLab*. This paper aims to present the currently available resources in the form of parallel corpora of the German and Serbian languages, such as *SrpNemKor* within the tool *Bibliša*, corpora that are part of *Sketch Engine* or the corpus *InterCorp* within the Czech *National Corpus*. In addition, it will show the possible ways of their usage in the contrastive research of phraseme constructions on concrete examples, point out what else is needed for such research, and suggest the layout, scope, search possibilities, etc. of an optimal resource. This research is carried out as part of the COST action PhraConRep "A Multilingual Repository of Phraseme Constructions in Central and Eastern European Languages", which aims, among other things, to investigate the

phraseme constructions of German and of the Serbian language contrastively, whereby the use of different tools and parallel corpora, presented in this research, can play a crucial role in reaching that goal.

Key words: Construction grammar, corpus linguistics, phraseme constructions, Serbian language, German language