

# Digitalni prozor u hrvatsku prošlost: digitalna humanistika i analitička funkcionalnost alata Google Ngram Viewer u povijesnoj analizi i nastavi povijesti

TADO JURIĆ

Hrvatsko katoličko sveučilište

Zagreb, Hrvatska

tado.juric@unicath.hr

Ovaj rad istražuje potencijal primjene velikih skupova podataka (*Big Data*) u povijesnoj analizi i nastavi, s posebnim naglaskom na analitičku funkcionalnost digitalnoga alata Google Ngram Viewer (GNV). Interdisciplinarnim/transdisciplinarnim pristupom rad demonstrira kako se GNV može koristiti za analizu, kontekstualizaciju i komparaciju međudržavnih odnosa Hrvatske i Austrije te kako njegova primjena može didaktički obogatiti nastavu povijesti. Ujedno rad nizom konkretnih primjera i metodoloških smjernica funkcionira kao priručnik za korištenje GNV-a i u istraživanju drugih polja humanističkih znanosti.

Analitički fokus usmjeren je na odabrane ključne povijesne događaje i ličnosti hrvatske povijesti od XVI. do XX. stoljeća, s obradom unutar njemačkoga jezičnoga korpusa. Središnje istraživačko pitanje jest može li se GNV, preko analize njemačkoga korpusa, koristiti kao pouzdan alat za kvantitativnu interpretaciju povijesnih i društvenih trendova relevantnih za hrvatski kontekst. Ciljevi rada uključuju kritičku analizu metodoloških prednosti i ograničenja GNV-a (napose u obrazovnom kontekstu) te usporedbu dobivenih kvantitativnih rezultata sa spoznajama historiografije u svrhu procjene njihove interpretativne valjanosti.

Osim političke povijesti, istraživanje ukazuje na širok spektar primjene GNV-a u analizama povezanim s književnošću, umjetnošću, jezikom i gospodarstvom. Time se afirmira šire shvaćanje digitalne humanistike, koje nadilazi konvencionalne okvire prisutne u hrvatskoj akademskoj zajednici. Zaključno se upozorava na nedostatke i ističe potreba za daljnjim istraživanjima.

Ključne riječi: *Big Data*; Google Ngram Viewer; Austrija; Hrvatska; digitalna humanistika; njemački korpus

## Analički dio rada

### *Uvod – Big Data i digitalna humanistika*

U posljednjem desetljeću digitalni alati sve se češće integriraju u metodologiju humanističkih znanosti, uključujući i povijesnu struku. No, njihova primjenjivost i metodološka vrijednost i dalje su predmet rasprave među znanstvenicima. Pitanje nije samo kako ih koristiti, nego na koji način njihova uporaba oblikuje epistemološki okvir povijesnoga istraživanja. S obzirom na taj izazov, ovaj rad polazi od teorijskoga problema primjene digitalnoga alata Google Ngram Viewer (GNV) u istraživanju povijesti, sagledavajući njegove potencijale, ograničenja i metodološke implikacije. U drugom dijelu rada fokus se premješta na funkcionalnu primjenu GNV-a u analizi povijesne upotrebe jezičnih pojmova, čime se demonstrira kako digitalni resursi mogu unaprijediti historiografsko promišljanje, ali i kako ovaj digitalni alat koristiti u nastavi povijesti.

Ubrzani razvoj digitalnih tehnologija uvelike je transformirao načine prikupljanja, obrade i interpretacije podataka unutar svih znanstvenih polja. Poseban zamah pritom u društvenim i humanističkim znanostima dobivaju tzv. veliki skupovi podataka (engl. *Big Data*), koji omogućuju kvantitativnu analizu opsežnih tekstualnih korpusa. Takvi su resursi prethodno bili nedostupni u svojem sadašnjem opsegu ili su bili metodološki zahtjevni za obradu tradicionalnim analitičkim postupcima.

Jedan od suvremenih pristupa u digitalnoj humanistici jest *rudarenje teksta* (engl. *text mining*)<sup>1</sup>, koje omogućuje analizu velikih količina tekstualnih podataka pomoću računalnih metoda. U tom je kontekstu GNV<sup>2</sup> koristan digitalni alat koji omogućuje kvantitativnu analizu (među ostalim i povijesnih) tekstova temeljenih na korpusu digitaliziranih knjiga. Iako ograničen na određene jezike, GNV omogućuje istraživačima pratiti učestalost pojmova tijekom stoljeća, čime se otvaraju mogućnosti za interpretaciju kulturnih i društvenih promjena. Ovaj pristup dio je širega područja „digitalne humanistike”, koja kombinira računalne alate i metode s tradicionalnim istraživačkim tehnikama.<sup>3</sup>

Korištenjem alata poput GNV-a moguće je kvantificirati promjene u upotrebi pojmova, koncepata i diskursa tijekom vremena, što može poslužiti kao

<sup>1</sup> Vidi opširnije u: BERRY, FAGERJORD, *Digital Humanities*.

<sup>2</sup> Kako alat funkcionira moguće je pronaći na Googleovu objašnjenju: <https://books.google.com/ngrams/info>.

<sup>3</sup> OZA, „Digital Humanities – An Introduction”, 16. – Ova novonastajuća interdisciplinarna znanost integrira tradicionalne humanističke znanosti kao što su književnost, povijest, filozofija, lingvistika, umjetnost i antropologija s računalno podržanim pristupima, analizom podataka, vizualizacijom i interaktivnim tehnologijama. Vidi opširnije: BURDICK et al., *Digital Humanities*.

indikator kulturnih promjena (kulturometrija).<sup>4</sup> Važno je naglasiti da se pod „kulturom” u ovom kontekstu ne misli isključivo na objavljene knjige, nego na širi tekstualni odraz društvenih vrijednosti, ideologija i kolektivnih narativa koji se manifestiraju u književnosti, publicistici, znanstvenim radovima i drugim vrstama pisanih izvora. Digitalizirani korpusi predstavljaju reprezentativni uzorak kulturne produkcije određenoga razdoblja, iako, naravno, ne obuhvaćaju cjelokupnu kulturnu dinamiku. Stoga, iako ne mjeri kulturu izravno, GNV omogućuje praćenje kulturnih indikatora preko jezičnih promjena, što je metodološki opravdano ako se koristi uz svijest o ograničenjima i uz dopunu kvalitativnim analizama.

Analizirajući učestalost pojmova vezanih uz određene događaje ili osobe, moguće je identificirati promjene u diskursu i javnom interesu tijekom vremena ili pak identificirati ključne događaje, promjene i trendove u određenim povijesnim razdobljima, koji katkad „promaknu” klasičnim istraživačkim metodama.<sup>5</sup> Na primjer, analiza učestalosti pojmova kao što su „Austrija”, „Mađarska” ili „Jugoslavija” može pružiti uvide u političke i društvene promjene koje su utjecale na Hrvatsku, kontekstualizirati povijesne događaje i komparirati ih s povijesnim trendovima u drugim zemljama. Pritom nastaju nove mogućnosti za analizu i razumijevanje povijesnih podataka, a katkad i potpuno nove perspektive.<sup>6</sup>

Iako postoje jasni dokazi o utjecaju društvenih promjena na učestalost pojavljivanja određenih pojmova u tekstualnim korpusima, odnosno da su društvo i jezik uski povezani,<sup>7</sup> važno je pritom naglasiti da se rezultati takvih analiza trebaju interpretirati uz svijest o ograničenjima korpusa, njegovoj reprezentativnosti, kao i o društvenim i medijskim kontekstima koji utječu na produkciju i distribuciju tekstova. Ovakav pristup ne zamjenjuje klasične metode (među ostalim i povijesne analize), nego ih dopunjava kvantitativnim

<sup>4</sup> Kulturometrija, kao interdisciplinarni pristup unutar digitalne humanistike, usmjere-na je na kvantitativnu analizu kulturnih fenomena pomoću digitalnih alata, vizualizacijskih tehnika i računalne obrade podataka. Ova metoda omogućuje istraživačima da analiziraju obrasce u velikim korpusima kulturnih podataka – od književnosti i umjetnosti do medijskih sadržaja i društvenih mreža – s ciljem otkrivanja diskurzivnih i društvenih trendova. Izvor: MANOVICH, „Cultural Analytics, Social Computing and Digital Humanities”.

<sup>5</sup> Važno je pritom naglasiti da se objavljene knjige ne mogu izravno poistovjetiti s javnim interesom, osobito u razdoblju od XIX. do XXI. stoljeća, kad su se pojavili drugi, masovni-ji mediji (novine, radio, televizija, internet) sa znatno širim dosegom i utjecajem. Međutim, knjige – osobito one koje su bile široko distribuirane, prevedene ili korištene u obrazovnim sustavima – mogu poslužiti kao indikatori kulturne produkcije elite ili dominantnih ideološ-kih okvira u određenom razdoblju. Stoga je metodološki opravdano tvrditi da kvantitativna analiza tekstova može otkriti promjene u diskursu.

<sup>6</sup> Korištenjem alata poput GNV-a moguće je generirati vizualne prikaze jezičnih trendova koji ilustriraju evoluciju pojmova tijekom vremena te potencijalno upućuju na promjene u dis-kursu, ideološkim okvirima i kulturnim narativima. Iako GNV ne mjeri izravno javni interes, on može poslužiti kao indikator prisutnosti i relevantnosti određenih pojmova u kulturnoj produkciji.

<sup>7</sup> MICHEL et al., „Quantitative Analysis of Culture”, 176-182.

uvidima, potičući interdisciplinarnost i razvoj digitalne pismenosti među istraživačima i učenicima ili studentima.

## Google Ngram Viewer kao digitalni alat u rudarenju teksta

GNV<sup>8</sup> je digitalni alat u rudarenju teksta koji omogućuje istraživačima analizirati učestalost riječi ili fraza u korpusu knjiga objavljenih tijekom različitih povijesnih razdoblja.<sup>9</sup> Ovaj alat predstavljen je javnosti 2010. te je otad besplatno dostupan svim istraživačima. Digitalizirani korpus koji se koristi u alatima poput GNV-a predstavlja reprezentativni uzorak objavljene književne i znanstvene produkcije, ali ne nužno cjelokupnu kulturnu ili društvenu dinamiku. Stoga je važno interpretirati podatke uz svijest o ograničenjima izvora, uključujući jezičnu, žanrovsku i geografsku selektivnost.<sup>10</sup> Pokrivaju široko razdoblje od 1500. do 2022. godine i sve svjetske jezike (engleski, francuski, njemački, talijanski, kineski, ruski, španjolski i hebrejski). GNV tako raspolaze s više od 15 milijuna digitaliziranih knjiga i više od 600 milijardi riječi (od 2022.), što ga čini najvećom besplatno dostupnom internetskom arhivom na svijetu, čija zbirka uključuje publikacije od znanstvenih radova, državnih službenih dokumenata do dnevnih novina.<sup>11</sup>

Da bi se kompenzirao utjecaj promjenjivoga broja objavljenih knjiga tijekom vremena, podaci se normaliziraju tako da se učestalost pojedinih izraza izražava kao omjer u odnosu na ukupni broj riječi (ili knjiga) objavljenih u promatranoj godini. Takav pristup omogućuje usporedivost podataka tijekom različitih razdoblja.<sup>12</sup>

Naime, bez normalizacije bilo bi nemoguće usporediti učestalost određenoga n-grama tijekom vremena jer broj knjiga objavljenih 1500. godine nije isti kao broj knjiga objavljenih 2025. godine. Nadalje, s obzirom na to da se broj objavljenih knjiga iz godine u godinu povećava, usporedba sirovih podataka o učestalosti određenih riječi, izraza ili tema može biti pogrešna. Na primjer, ako se neka riječ pojavljuje 500 puta 1950. godine i 1000 puta 2000. godine, to ne znači nužno da je dvostruko češća – možda je 2000. godine jednostavno objavljeno više knjiga. Normalizacija u ovom kontekstu znači da se broj pojavljivanja neke riječi ili izraza prilagođava ukupnom broju riječi ili knjiga objavljenih te godine. Time se dobiva relativna učestalost, izražena kao postotak ili omjer, koji omogućuje pravednu usporedbu tijekom vremena.

<sup>8</sup> GOOGLE NGRAM VIEWER, <https://books.google.com/ngrams/>.

<sup>9</sup> JURIC, „Using digital humanities for understanding COVID-19”.

<sup>10</sup> KOKALOVIC, „Digitalizacija knjižnične građe u specijalnim knjižnicama”.

<sup>11</sup> JURIC, *Big (Crisis) Data in Social Sciences and Humanities*, 38-43.

<sup>12</sup> CHUMTONG, KALDEWEY, „Beyond the Google Ngram Viewer”.

GNV stoga prikazuje postotak broja pojavljivanja, pri čemu se postotak izračunava od ukupnoga broja knjiga objavljenih u određenoj godini.<sup>13</sup> Klimom na točku na dijagramu prikazuje se postotak pojavljivanja za tu godinu.<sup>14</sup>

## Rudarenje teksta – metodologija

Osnovni postupak kojim se služi GNV u obradi podataka jest rudarenje teksta. Riječ je o metodi za dobivanje strukturiranih informacija iz nestrukturiranoga teksta, čime se omogućuje otkrivanje smislenih odnosa. GNV izračunava koliko se često određeni n-gram pojavljuje u odabranom korpusu za određenu godinu, relativno prema ukupnom broju n-grama.<sup>15</sup> Google navodi da je točnost automatske obrade teksta u novijim korpusima oko 95% za engleski jezik, no ta procjena odnosi se primarno na tehničke aspekte obrade (OCR, POS tagging), a ne na semantičku točnost ili pouzdanost izvora. S obzirom na komercijalnu prirodu platforme i nedostatak transparentnosti u metodologiji, podatke je potrebno interpretirati uz svijest o mogućim ograničenjima i pristranostima.

Rudarenje teksta nudi velik potencijal za akademska istraživanja, uključujući tri glavna područja: 1. razvoj novih hipoteza<sup>16</sup>; 2. sustavno pretraživanje literature; 3. testiranje hipoteza. Ova metoda ubrzava istraživačke procese i omogućuje istraživačima da više vremena posvete analizi podataka, čime ih oslobađa velikoga vremenskog opterećenja koje obično prati prikupljanje podataka.<sup>17</sup>

Međutim, važno je poznavati granice i GNV-a i metode rudarenja podataka i koristiti ih s velikom oprezom i rigoroznim testiranjem, što je uostalom slučaj sa svakom metodologijom. Naime, ako se rudarenje teksta ne provodi ispravno, metoda može dovesti do iskrivljenih ili pogrešnih rezultata. Povjesničari bi trebali biti oprezni i nikad ne smiju uzimati rezultate rudare-

<sup>13</sup> UNIVERSITY OF LONDON, „An introduction to text mining”.

<sup>14</sup> Podaci generirani određenim upitima zatim se mogu izvesti kao popis i obraditi pomoću alternativnih softverskih paketa (npr. „R”), posebno programima za proračunske tablice.

<sup>15</sup> U računalnoj lingvistici, n-gram je slijed od n elemenata iz određene tekstualne sekvence. Elementi mogu biti fonemi, slogovi, slova ili riječi. Google Ngram baza podataka podržava sekvence n-grama do pet elemenata. Na primjer, „ja” je 1-gram, a „ja sam” je 2-gram. To znači da kad istraživač traži riječ (unigram), dobiva postotak te riječi u odnosu na sve druge riječi pronađene u korpusu knjiga za određenu godinu. Ako istraživač unese više od jedne riječi ili fraze, svaka riječ ili fraza bit će predstavljena crtom u boji da bi se stvorio kontrast u odnosu na druge pojmove pretraživanja. Vidi opširnije u: KARCH, „How to Use the Ngram Viewer Tool in Google Books”.

<sup>16</sup> Analizom frekvencije pojmova, njihovih suodnosa i semantičkih veza istraživač može uočiti neočekivane povezanosti koje nisu bile predmet prethodnih istraživanja. Na primjer, u analizi tisuća medicinskih članaka rudarenje teksta može otkriti da se pojam „mikrobiom” sve češće pojavljuje uz „mentalno zdravlje”, što može potaknuti formulaciju nove hipoteze o njihovoj povezanosti.

<sup>17</sup> ZIĘBA, „Google Books Ngram Viewer in socio cultural research”.

nja teksta „zdravo za gotovo”.<sup>18</sup> Međutim, budući da GNV omogućuje izravan pristup svakom digitaliziranom izvoru, sve je podatke moguće provjeriti (o ograničenjima metode vidi posljednje poglavlje).

Istraživač koji koristi GNV može mijenjati pretraživanje prema vremenskom okviru, tipu korpusa i nekoliko različitih jezika. Osim glagola i imenica, mogu se pretraživati i sve ostale vrste riječi. Za usporedbu više n-grama moguće je kombinirati ili odvojiti dva izraza te dijeliti ili množiti izraze da bi se usporedili n-grami različite učestalosti ili izolirale učestalosti jednoga n-grama u odnosu na drugi. Dodavanjem operatora „+” između n-grama istraživač može kombinirati više frekvencija u jednu (sve primjere vidi u sekciji rezultati). Dodavanjem operatora „-” između n-grama omogućuje se oduzimanje frekvencija s desne strane od frekvencija s lijeve strane, omogućujući time mjerenje povezanosti frekvencija.<sup>19</sup>

U sekciji rezultati ovoga rada prikazani su grafički prikazi našega istraživanja. Ordinata (os Y) prikazuje učestalost spominjanja kao postotak, a apscisa (os X) prikazuje godine analize. Dijagrami ilustriraju fluktuacije u učestalosti spominjanja pojmova tijekom vremena, s primjetnim vrhuncima i padovima, iz čega se mogu uočiti dinamika i trendovi. Različite crte u boji označavaju učestalost pojedinih pojmova i pokazuju kako se njihova važnost mijenjala tijekom stoljeća. Tako se može vidjeti koji su pojmovi bili dominantniji u određenim razdobljima. Posebnu pozornost treba posvetiti vrhuncima spominjanja ispitivanih pojmova, što upućuje na povećani interes i važnost u određenom korpusu odnosno društvu.<sup>20</sup>

Napominjemo da, iako je već brojnim testiranjima potvrđeno da je GNV koristan novi alat u istraživanju povijesti i socio-kulturnih fenomena<sup>21</sup>, neki istraživači<sup>22</sup> tvrde da podaci dobiveni pomoću njega nisu dovoljno pouzdani,<sup>23</sup> o čemu više u nastavku.

<sup>18</sup> *Isto.*

<sup>19</sup> Dodavanje operatora „/” između n-grama omogućuje izolaciju kretanja učestalosti relativno u odnosu na drugu. Dodavanjem operatora „\*” između n-grama učestalost lijevo množi se s učestalošću s odabranom vrijednošću, dakle s navedenim brojem. To omogućuje usporedbu dviju bitno različitih učestalosti. Dodavanje operatora „:” između n-grama koristi n-gram s lijeve strane u korpusu s desne strane i omogućuje usporedbu n-grama u različitim korpusima. Prikazivanje riječi u nekoliko gramatičkih kategorija može se postići dodavanjem koda „\_INF” kao sufiksa na korijen riječi. Unosom operatora „=>” moguće je prikazati odnos između riječi i njihov kontekst u rečenici. Za detaljne upute vidi: GOOGLE NGRAM VIEWER, „What does the Ngram Viewer do?”.

<sup>20</sup> Na relativnu učestalost određenih istraživanih pojmova (tablica 1) utječu različiti faktori, uključujući geopolitičke događaje, ekonomsku razmjenu, kulturne trendove i promjene u akademskim istraživanjima.

<sup>21</sup> ROJAS CASTRO, „Big Data in the Digital Humanities”, 62.

<sup>22</sup> Vidi opširnije: ZIĘBA, „Google Books Ngram Viewer in socio cultural research”.

<sup>23</sup> Vidi poglavlje „Ograničenja metode” u ovom radu.

## Pregled dosadašnjih istraživačkih spoznaja

U društvenim znanostima Kesebir i Kesebir<sup>24</sup> te Oishi i sur.<sup>25</sup> koristili su GNV kako bi dokazali da su moralni ideali i vrline znatno opali u američkoj javnoj raspravi. Oishi i sur. također su analizirali pojam sreće tijekom vremena i u različitim kulturama, Lucier<sup>26</sup> je istraživao odnose između znanosti i kapitalizma, a Kumar i Sahu<sup>27</sup> pratili su povijest marketinga. Ponajbolji pregled literature dostupan je kod Zięba.<sup>28</sup>

Michel i sur.<sup>29</sup> pokazali su da GNV omogućuje istraživačima kvantitativno proučavanje kulturnih trendova. Autori su istraživali kolektivno pamćenje i razvili dijagrame interesa za različite događaje između 1875. i 1975. da bi usporedili uspon i pad slave najpoznatijih osoba. Pritom su pokazali da ovaj pristup može pružiti uvide u različita područja kao što su leksikografija, kolektivno pamćenje, razvoj tehnologije, cenzura i povijesna epidemiologija. U slučaju gripe, vrhovi kulturnoga interesa pokazali su točnu podudarnost s poznatim povijesnim epidemijama.<sup>30</sup>

Newberry i sur.<sup>31</sup> koristili su GNV za analizu promjena engleskoga jezika od XII. do XXI. stoljeća. Greenfield je koristila GNV za testiranje svoje teorije o utjecaju individualizma na vrijednosti, ponašanje i psihologiju pojedinca. Acerbi i sur.<sup>32</sup> koristili su GNV za istraživanje prikaza emocija u knjigama u XX. stoljeću. Pokazali su da stresni i nasilni povijesni događaji ostavljaju tragove u izražavanju emocija u knjigama, tako da je moguće prepoznati „sretna” i „tužna” razdoblja povijesti, ovisno o prikazu i upotrebi riječi za određene emocije u knjigama. GNV je tako omogućio znanstvenicima da steknu dublje uvide u razna tematska područja poput razlika među spolovima, emocija, osobnosti, spoznaje, psihoterapije, moralnih vrijednosti, obrazovanja, prirode te razvoja individualizma i kolektivismu.<sup>33</sup>

Spomenuti napredak u analizi podataka iz uvoda u Hrvatskoj pak nije vidljiv. Upotreba *Big Data* pristupa i digitalne humanistike u hrvatskoj znanosti je rijetka<sup>34</sup>, a zasebnih radova o *Big Data* pristupu u istraživanju hrvatske povijesti zasad nema.

<sup>24</sup> P. KESEBIR, S. KESEBIR, „The Cultural Salience of Moral Character and Virtue”.

<sup>25</sup> OISHI et al., „Concepts of happiness across time and cultures”.

<sup>26</sup> LUCIER, „The Origins of Pure and Applied Science”.

<sup>27</sup> KUMAR, SAHU, „The Evolution of Marketing History”.

<sup>28</sup> ZIĘBA, „Google Books Ngram Viewer in socio cultural research”, 364-366.

<sup>29</sup> MICHEL et al., „Quantitative Analysis of Culture”, 176-182.

<sup>30</sup> MICHEL, „Supporting Online Material for Quantitative Analysis of Culture”.

<sup>31</sup> NEWBERRY et al., „Detecting evolutionary forces in language change”.

<sup>32</sup> ACERBI et al., „The Expression of Emotions in 20th Century Books”.

<sup>33</sup> JURIC, *Big (Crisis) Data in Social Sciences and Humanities*, 33.

<sup>34</sup> Iznimke su: JURIC, „Using digital humanities for understanding COVID-19” te JURIC, *Big (Crisis) Data in Social Sciences and Humanities*.

## Ograničenja metode

Korištenje Google NGrama za istraživačke svrhe izazvalo je kontroverzne rasprave.<sup>35</sup> Iako velik broj studija koje koriste GNV upućuje na pouzdanost metode, u nekoliko radova iznesene su metodološke sumnje.<sup>36</sup> Kritizirano je da se skup podataka temelji na netočnom optičkom prepoznavanju znakova (OCR), uključuje previše znanstvene literature i da su mnogi tekstovi pogrešno datirani i kategorizirani.

Glavne kritike odnose se na nedostatke OCR-a, osobito u vezi sa semantičkim pogreškama skeniranja koje mogu uzrokovati probleme s učestalošću riječi zbog sličnosti između slova *f* i *s* u engleskom korpusu (npr. *fail* i *sail*). Osim toga, neuredni metapodaci mogu dovesti do prikazivanja učestalosti riječi u pogrešnim kontekstima. Kritičari također ističu etička pitanja, koja su slična onima vezanim uz umjetnu inteligenciju.<sup>37</sup> Osobito naglašavaju koliko je važno preispitati pretpostavke, vrijednosti i predrasude koje su temelj ove nove istraživačke ere jer doba velikih podataka tek počinje.

GNV korpus također je bio predmet kritike u vezi s pitanjima njegove reprezentativnosti. Chumtong i Kaldewey u obranu metode ističu da vrijednost GNV-a kao istraživačkoga alata ne leži prvenstveno u njegovoj točnosti, nego u njegovu potencijalu za „brze i grube heurističke analize”.<sup>38</sup> Davis smatra da je sučelje GNV-a previše simplificirano jer mu nedostaju funkcije poput korištenja kolokacija pri pretraživanju, kao i smisleno korištenje većih dijelova rečenice.<sup>39</sup> Međutim, zato postoji kritički pristup unutar kojega je GNV samo inicijalni ili provedbeni dio postupka dobivanja zaključaka. Nikako jedini, na što će u ovom radu biti višekratno upozoreno.

Dotadna ograničenja GNV-a uključuju nedostatak uzimanja u obzir različitih konteksta u kojima se analizirane riječi koriste jer se točna značenja moraju izvući iz originalnih tekstova. Naime, povećana učestalost određenoga pojma ne znači nužno da njegova relevantnost raste, nego da se o njemu naširoko raspravlja. Nadalje, nisu sve knjige i tekstovi digitalizirani, stoga mogu postojati praznine u korpusu.

Iako GNV uz opciju „Pretraži Google Books” („Search in Google Books”) omogućuje prikaz isječaka knjiga iz kojih riječi potječu, ručno prikupljanje takvih podataka za milijune konteksta bilo bi nepraktično zbog vremena i truda koji su potrebni. Stoga, prema Zięba, GNV treba koristiti za jednostavne

<sup>35</sup> YOUNES, REIPS, „Guideline for improving the reliability of Google Ngram studies”.

<sup>36</sup> GOODING, „Mass digitization and the garbage dump”; PECHENICK, DANFORTH, DODDS, „Characterizing the Google Books corpus”; PETTIT, „Historical time in the age of big data”; KOPLINIG, „The impact of lacking metadata”; HILPERT, GRIES, „Assessing Frequency Changes in multistage Diachronic Corpora”.

<sup>37</sup> Vidi opširno u: HARARI, *Nexus: A Brief History of Information Networks*.

<sup>38</sup> CHUMTONG, KALDEWEY, „Beyond the Google Ngram Viewer”.

<sup>39</sup> DAVIS, „Making Google Books n-grams Useful”.

studije učestalosti riječi, a ne kao jedini alat za istraživanje složenih sociokulturnih transformacija.<sup>40</sup>

Lakoff se slaže da prisutnost većine riječi i promjene u njihovoj učestalosti možda ne otkrivaju izravno specifične vrijednosti istraživanoga fenomena, ali svakako mogu poslužiti kao indikatori.<sup>41</sup> Solovjev i sur.<sup>42</sup>, međutim, odbacuju spomenute kritike tvrdeći da se GNV korpus može smatrati reprezentativnim. Pritom naglašavaju da je to najveći korpus koji je ikad postojao. Ipak, i oni se slažu da je daljnji razvoj metodologije još u početnoj fazi i zahtijeva šire sudjelovanje akademske zajednice u njezinu usavršavanje.

Da bi se otklonili nedostaci, Younes i Reips<sup>43</sup> predlažu nekoliko metodoloških postupaka, uključujući unakrsnu validaciju istraživanjem različitih jezičnih korpusa, korištenje sinonima te primjenu standardizacijskih postupaka. Oni predlažu proučavanje svake riječi zajedno s tri odabrana sinonima iz relevantnih rječnika. Druga mogućnost za poboljšanje rezultata je preprocesiranje GNV-ovih „sirovih podataka”.

Unatoč nedostacima, GNV ipak neupitno nudi potencijal za poboljšanje našega razumijevanja kulturnih i jezičnih trendova tijekom vremena. Verzije iz 2012. i 2019. s više knjiga i poboljšanim OCR-om, knjižničnim i izdavačkim metapodacima dodatno su unaprijedile njegove mogućnosti.<sup>44</sup>

## Primjena u nastavi povijesti

U kontekstu nastave povijesti, digitalni alat GNV nudi više od puke ilustracije jezičnih frekvencija – njegova primjena doprinosi razvijanju temeljnih kompetencija učenika/studenata u suvremenom obrazovnom okruženju. Prvo, korištenje GNV-a omogućuje razvijanje digitalne pismenosti, uključujući sposobnost pretraživanja, interpretacije i kritičkoga vrednovanja digitalnih podataka. Drugo, alat potiče kritičko mišljenje pri evaluaciji povijesnih izvora: učenici nisu tek pasivni primatelji informacija, nego aktivno problematiziraju odnose između jezika, ideologije i povijesnoga konteksta. Treće, GNV se pokazuje korisnim u razumijevanju povijesne kontekstualizacije jezičnih izraza – analizom promjena u uporabi pojmova tijekom vremena učenici razvijaju osjetljivost za temporalne i diskurzivne pomake u povijesnoj naraciji.

Primjeri nastavne primjene uključuju zadatke poput analize učestalosti pojmova kao što su „revolucija”, „nacija”, „demokracija” ili „rat” u različitim povijesnim razdobljima, čime se potiče povezivanje jezičnih trendova s političkim, društvenim i kulturnim kontekstima. Učenici mogu samostalno formulirati istraživačka pitanja, uspoređivati pojmove među jezicima ili

<sup>40</sup> ZIĘBA, „Google Books Ngram Viewer in socio cultural research”.

<sup>41</sup> LAKOFF, „What Words Don't Tell Us”.

<sup>42</sup> SOLOVYEV, BOCHKAREV, AKHTYAMOVA, „Google Books Ngram”.

<sup>43</sup> YOUNES, REIPS, „Guideline for improving the reliability of Google Ngram studies”.

<sup>44</sup> Vidi opširnije u: JURIC, *Big (Crisis) Data in Social Sciences and Humanities*.

korpusima te prezentirati zaključke u obliku digitalnih plakata, eseja ili rasprava. Takav pristup nastavi povijesti osnažuje učenike/studente za aktivno i istraživačko učenje bez obzira na razinu obrazovanja – od srednjoškolskih programa do visokoškolskih kolegija.<sup>45</sup>

*Tablica 1. Primjer nastavne primjene*

| <b>Aktivnost</b>                     | <b>Opis zadatka</b>  | <b>Ishodi</b>   |
|--------------------------------------|--|---|
| <b>Analiza pojma tijekom vremena</b> | Učenici/studenti odabiru jedan povijesno važan pojam (npr. „revolucija”, „imperij”, „nacija”) i pomoću GNV-a analiziraju njegovu učestalost u razdoblju 1800. – 2000.  | Razumijevanje dinamike pojmovne upotrebe u različitim povijesnim kontekstima.                 |
| <b>Usporedba jezika</b>              | Uspoređuje se frekvencija istoga pojma u engleskom, njemačkom i hrvatskom korpusu (ako je dostupan), uz raspravu o kulturnim i političkim implikacijama.   | Razvijanje interkulturalne analize i kritičkoga odnosa prema jeziku kao povijesnom fenomenu.  |
| <b>Izrada mikroistraživanja</b>      | Učenici/studenti formuliraju istraživačko pitanje i prezentiraju rezultate GNV analize u obliku eseja, prezentacije ili digitalnoga plakata (npr. „Kako se mijenjao diskurs o ‘demokraciji’ u XX. stoljeću?”). | Primjena digitalnoga alata u konstruiranju povijesne argumentacije i prezentaciji zaključaka. |
| <b>Kritička refleksija</b>           | Piše se refleksivni osvrt o prednostima i ograničenjima GNV-a kao historiografskoga alata.   | Razvijanje epistemološke svjesnosti i kritičkoga odnosa prema digitalnoj metodologiji.        |

## **Interpretativni dio rada**

### *Politički i kulturni odnosi između Hrvatske i zemalja njemačkoga govornog područja – testiranje digitalnoga alata GNV i diskusija*

Fokus rada je na percepciji hrvatskih povijesnih ličnosti i ključnih događaja u njemačkom korpusu. Analiza se temelji na njemačkom korpusu Google Books, koji obuhvaća digitalizirane knjige objavljene na njemačkom jeziku, a dostupne su preko Googleove platforme. Korpus uključuje širok spektar žanrova – od znanstvenih i historiografskih publikacija preko publicistike do

<sup>45</sup> Vidi: BANERJEE, „The Impact of Digital Tools on the Teaching of History”.

književnih djela – bez eksplicitne tematske selekcije. Korpus nije ograničen na historiografske naslove, nego reflektira opći izdavački profil knjižnoga tržišta u danom razdoblju.<sup>46</sup>

Takva heterogenost korpusa omogućuje uvid u šire diskurzivne obrasce, ali istovremeno zahtijeva oprez pri interpretaciji rezultata jer učestalost pojmovna može biti uvjetovana žanrovskim, stilskim ili ideološkim karakteristikama izvora. Stoga se rezultati GNV analize ne tumače kao izravni pokazatelj historiografske recepcije, nego kao indikativni signal javnoga interesa i kulturne prisutnosti određenih povijesnih ličnosti i događaja u njemačkom jezičnom prostoru. Pritom je teško razgraničiti izvore iz Austrije, Njemačke i Švicarske kao zemalja njemačkoga govornog područja.<sup>47</sup>

Navedeni digitalni alat testirat ćemo na primjeru bilateralnih odnosa između Hrvatske i Austrije. Pritom ćemo pokušati potvrditi hipotezu da se primjenom pristupa *Big Data* mogu dobiti širi uvidi u razvoj i dinamiku bilateralnih odnosa. Za širu kontekstualizaciju koristit ćemo komparaciju odnosa Austrija – Hrvatska i Austrija – Srbija.

Istraživačko pitanje glasi: može li se GNV, analizom njemačkoga korpusa, koristiti kao pouzdan alat za kvantitativnu interpretaciju povijesnih i društvenih trendova relevantnih za hrvatski kontekst? Cilj istraživanja je procijeniti metodološke prednosti i ograničenja korištenja GNV-a u humanističkim istraživanjima i nastavi povijesti (usporediti kvantitativne podatke s historiografskim izvorima da bi se provjerila interpretativna vrijednost alata). Hipoteze rada su: H1 – GNV može identificirati relevantne povijesne obrasce u njemačkom korpusu koji se odnose na hrvatski kontekst; H2 – unatoč jezičnim i korpusnim ograničenjima, GNV može poslužiti kao didaktički alat u nastavi povijesti.

Grafički prikazi i analiza pojmova pomoću GNV-a u sljedećim odjeljcima ilustrirat će kako se ovaj pristup može koristiti za dobivanje dubljih uvida u povijesne odnose Hrvatske i Austrije koristeći njemački korpus od 1600. do 2022. godine. Temeljna pretpostavka ovoga rada jest da analiza pomoću GNV-a može pružiti indikativne uvide u jezične obrasce koji se povezuju s javnim interesom unutar određenoga društvenog i kulturnoga konteksta. Iako takvi kvantitativni podaci nisu izravan ili potpun odraz stvarnih političkih i kulturnih događaja, oni mogu reflektirati diskurzivne promjene i temporalne pomake u javnom govoru, što ih čini relevantnim u interpretaciji povijesnih narativa.<sup>48</sup>

<sup>46</sup> MICHEL et al., „Quantitative Analysis of Culture”.

<sup>47</sup> Ovo je moguće samo „ručno”, provjeravajući svaki zasebni tekst opcijom „Pretraži Google Books”.

<sup>48</sup> Ograničenja ovakva pristupa odnose se prvenstveno na karakter korpusa, nedostatak kontekstualizacije samih pojmova te potrebu za oprezom pri interpretaciji korelacija između jezične frekvencije i društvene stvarnosti.

Tablica 2. Izbor istraživanih pojmova na njemačkom jeziku (i prijevod)

|  |   |
|--|---|
| <b>Percepcija Hrvatske i regije Jugoistočne Europe u zemljama njemačkoga govornog područja</b>   | <b>Politika i politička povijest: odnosi Austrija – Hrvatska</b>  |
| <ul style="list-style-type: none"> <li>- „Südosteuropa“, „Balkan“, „Westbalkan“ (Balkan, Jugoistočna Europa, Zapadni Balkan)</li> <li>- Gradovi: „Zagreb“, „Agram“</li> <li>- Odnos/ovisnost pojmova u istom kontekstu i usporedba: „Balkan=&gt;Österreich, Südosteuropa=&gt;Österreich (Austrija)“</li> <li>- Učestalost spominjanja pridjeva „jugoslawische“ (jugoslavenski*) i riječi koje slijede (naredba *)<sup>49</sup></li> <li>- „Österreichisch-Ungarische Monarchie und Serbien“, „Österreichisch-Ungarische Monarchie und Kroatien“ (AU i Srbija; AU i Hrvatska)</li> <li>- Koja riječ najčešće slijedi iza pojma „Kroaten sind“ (Hrvati su) (naredba =&gt;*)</li> </ul> | <ul style="list-style-type: none"> <li>- „Königreich Kroatien“, „Königreich Serbien“ (Kraljevina Hrvatska, Kraljevina Srbija)</li> <li>- „Österreich und Serbien, Österreich und Kroatien“, „Österreich und Bosnien“</li> <li>- „Kroatien“, „Serbien“ (Hrvatska, Srbija) (naredba =&gt;*)</li> <li>- „Kaiser=&gt;Kroaten, Kaiser=&gt;Serben“ (car=&gt;Hrvati; car=&gt;Srbi) (naredba =&gt;)</li> <li>- „Türken=&gt;Kroaten, Türken=&gt;Serben“ (Turci/Osmanlije=&gt;Hrvati/Srbi)</li> <li>„Krieg in Kroatien“ und „Krieg in Bosnien“ (rat u Hrvatskoj; rat u BiH) (naredba „zagrada“)</li> <li>- Zweiter Weltkrieg: „Ustaša, Ustascha, Četnik, Tschetnik, Partisan“ (Drugi svjetski rat)</li> <li>- „Politik=&gt;Kroaten, Politik=&gt;Serben“ (politika)</li> </ul> |
| <b>Kultura, umjetnost, književnost, glazba, gospodarstvo, jezik</b>  | <b>Ličnosti, prezimena, genealogija</b>   |
| <ul style="list-style-type: none"> <li>- „(Kultur in Kroatien), (Kultur in Serbien), (Kunst in Kroatien), (Kunst in Serbien) (Wirtschaft in Kroatien)“ (kultura, umjetnost, književnost, glazba, gospodarstvo)</li> <li>- „Kroatisch“, „Serbisch“ (hrvatski, srpski jezik)</li> <li>- germanizmi u hrvatskom jeziku</li> </ul>   | <ul style="list-style-type: none"> <li>- (pretraga najpoznatija hrvatska prezimena)</li> <li>- (pretraga najpoznatiji političari u XIX. i XX. stoljeću u Hrvatskoj i Srbiji i njihova prepoznatljivost u njemačkom korpusu)</li> <li>- Analiza sentimentata: „Stjepan Radić, Nikola Pašić, Ante Starčević...“</li> <li>- proširena analiza: „Pretraži Google Books“ (direktna pretraga u Google Books)</li> </ul>   |

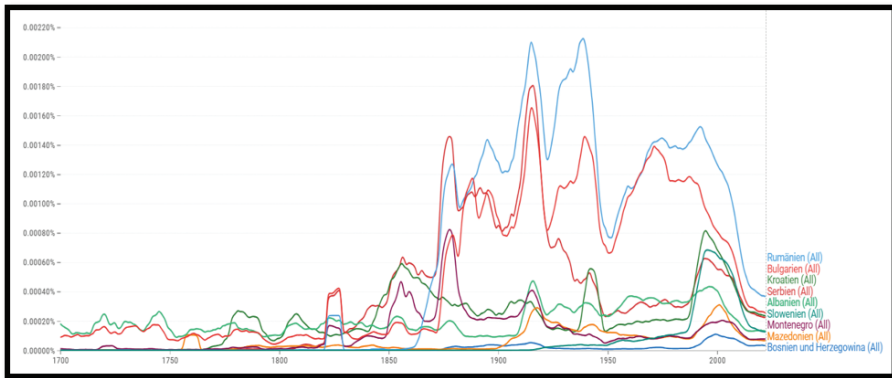
Pri testiranju metode najprije ćemo analizirati percepciju regije Jugoistočne Europe u zemljama njemačkoga govornog područja, uslijedit će primjeri iz 1. političke povijesti (relacije Austrija – Hrvatska), za čim slijedi analiza iz područja 2. književnosti, umjetnosti i lingvistike te 3. primjeri iz povijesnih migracija i genealogije te analiza povijesnih ličnosti.<sup>50</sup>

<sup>49</sup> Zbog nedostatka prostora ograničili smo se na ovaj primjer. Daljnji bi radovi zbog potpunosti istraživanja mogli usporediti učestalost pridjeva „jugoslavenski“ s „hrvatski“, „srpski“, „bosanski“, „bošnjački“.

<sup>50</sup> Treba napomenuti da se često radi o vrlo složenim analizama konteksta i pojmovlja te se istraživač treba potruditi biti precizan u opisu načina formuliranja upita i objašnjenju što se i zašto pita.

## Percepcija regije Jugoistočne Europe u zemljama njemačkoga govornog područja

Fokus u ovom odjeljku bit će na pitanju koliki je interes za regiju Jugoistočne Europe i posebice Hrvatsku te opada li taj interes ili raste? Pritom treba imati na umu da je percepcija Jugoistočne Europe u zemljama njemačkoga govornog područja često oblikovana stereotipima i klišejima<sup>51</sup> koji proizlaze iz povijesnih veza i političkih događaja, ali i suvremenih političkih diskursa i medijskoga izvještavanja koji ističu samo određene aspekte.<sup>52</sup> *Big Data* pristup može biti od koristi u ispravljajući određenih stereotipa koji su se uvrijedili u njemačkom društvu (primjerice da je Hrvatska dio Balkana).<sup>53</sup>



Slika 1. Učestalost spominjanja država Jugoistočne Europe u njemačkom korpusu „Bosna i Hercegovina”, „Hrvatska”, „Srbija”, „Crna Gora”, „Slovenija”, „Albanija”, „Bugarska”, „Rumunjska” (1700. – 2022.)<sup>54</sup>

Podaci GNV-a pokazuju da je bilo nekoliko razdoblja u kojima su se zemlje Jugoistočne Europe tijekom povijesti češće spominjale.

Češće spominjanje zemalja poput „Rumunjske”, „Bugarske”, „Srbije” i „Hrvatske” u XIX. stoljeću može se povezati s oslobodilačkim pokretima i sve većim nacionalnim težnjama u Jugoistočnoj Europi. Balkanski ratovi (1912. – 1913.) i rastući otpor protiv Osmanskoga Carstva igrali su ključnu ulogu i našli svoj odraz u njemačkoj literaturi, odnosno njemačkom korpusu toga vremena.<sup>55</sup>

<sup>51</sup> Vidi opširnije u: CLEWING, SCHMITT, *Geschichte Südosteuropas*.

<sup>52</sup> BRUNNBAUER, „Europa und der Balkan: Fremd- und Selbstzuschreibungen”.

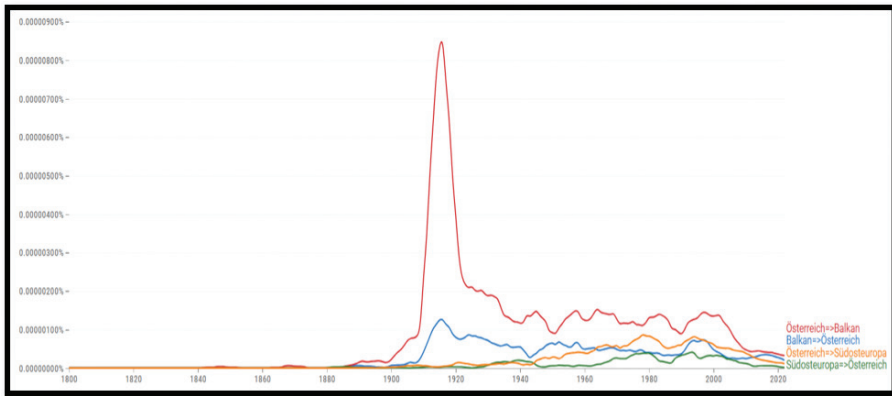
<sup>53</sup> Radi se o geopolitičkom pojmu vrlo rastezljiva opsega čije je definiranje nerijetko obojeno negativnom konotacijom. No, istovremeno je pojam sveprisutan i polivalentan u svojem značenju u njemačkoj literaturi.

<sup>54</sup> Njem. *Bosnien und Herzegowina, Kroatien, Serbien, Montenegro, Slowenien, Albanien, Bulgarien, Rumänien*.

<sup>55</sup> Vidi: CLEWING, SCHMITT, *Geschichte Südosteuropas*.

Znan porast spominjanja početkom XX. stoljeća pokazuje dalekosežne geopolitičke promjene uzrokovane Prvim svjetskim ratom i posljedičnim preuređenjem političkih granica u Europi. Zemlje poput „Srbije” i „Bosne i Hercegovine” češće se spominju u tom razdoblju, što upućuje na njihovu središnju ulogu u tim sukobima. To je i prvi dokaz pouzdanosti metode, odnosno dokaz da se uvidi GNV-a preklapaju s poznatim historiografskim podacima.

Može se primijetiti da pojmovi „Balkan”, „Jugoistočna Europa” i „Zapadni Balkan” u usporedbi dužega razdoblja opadaju<sup>56</sup>, što implicira smanjenje interesa za regiju u njemačkom govornom području ili pak promjenu fokusa s regionalnoga na nacionalni/etnički koncept.<sup>57</sup>



Slika 2. Odnos/ovisnost pojmova u istom kontekstu i usporedba „Balkan=>Österreich (Austrija), Österreich=>Balkan, Südosteuropa (Jugoistočna Europa)=>Österreich, Österreich=>Südosteuropa” (naredba =>) (1800. do 2022.)

U ovom primjeru treba napomenuti da se radi o naredbi „=>” nakon analizirane riječi, što omogućuje analizu odnosa/zavisnosti pojmova u istom kontekstu. Naredba omogućuje izdvajanje pojmova koji su korišteni u istim tekstovima, odnosno kontekstne veze.

Tijekom XVIII. stoljeća odnosi između pojmova poput „Balkan” i „Austrija”, kao i „Jugoistočna Europa” i „Austrija”, rijetko su se pojavljivali u njemačkom korpusu, što na prvi pogled može sugerirati slabiji interes za jugoi-

<sup>56</sup> Za izvatke iz članaka, knjiga i publikacija vidi: [https://books.google.com/ngrams/graph?content=%E2%80%9EBosnien+und+Herzegowina%E2%80%9C-2C+%E2%80%9EKroatien%E2%80%9D%2C+%E2%80%9ESerbien%E2%80%9C-2C+%E2%80%9EMontenegro%E2%80%9C%2C+%E2%80%9ESlowenien%E2%80%9C-2C+%E2%80%9EAlbanien%E2%80%9C%2C+Bulgarien%2C+Rum%3%A4nien&year\\_start=1700&year\\_end=2022&corpus=de&smoothing=3](https://books.google.com/ngrams/graph?content=%E2%80%9EBosnien+und+Herzegowina%E2%80%9C-2C+%E2%80%9EKroatien%E2%80%9D%2C+%E2%80%9ESerbien%E2%80%9C-2C+%E2%80%9EMontenegro%E2%80%9C%2C+%E2%80%9ESlowenien%E2%80%9C-2C+%E2%80%9EAlbanien%E2%80%9C%2C+Bulgarien%2C+Rum%3%A4nien&year_start=1700&year_end=2022&corpus=de&smoothing=3).

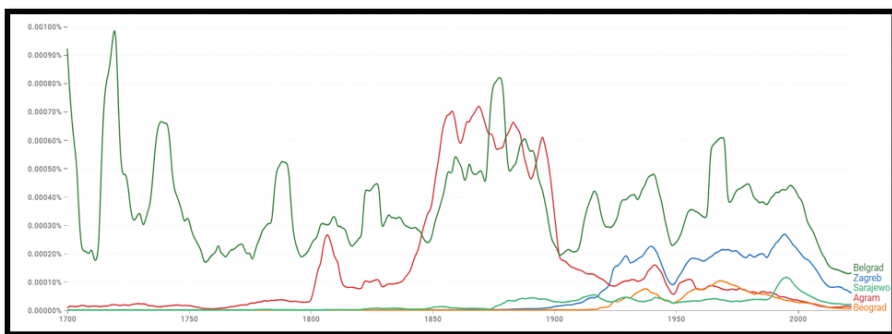
<sup>57</sup> To bi valjalo dodatno „ispitati” npr. poredbom učestalosti pojave regionalni vs nacionalni pojmovi. U vezi s terminologijom i njezinom uporabom od XIX. stoljeća nadalje vidi: JURIC, BAKOTA, RUKAVINA, „What is the Balkans?”, 75-82.

stočne periferije Habsburškoga Carstva. Međutim, takva interpretacija bila bi metodološki pojednostavnjena. Naime, pojmovi „Balkan” i „Jugoistočna Europa” etabliraju se u njemačkom jeziku tek u XIX. stoljeću, što znači da njihova učestalost u ranijem razdoblju ne može služiti kao vjerodostojan pokazatelj stvarne razine interesa ili prisutnosti tih regija u kulturnom imaginariju. U tom kontekstu terminološka evolucija sama po sebi postaje predmet historiografske analize s obzirom na to da reflektira šire promjene u geografskim konceptima, političkoj nomenklaturi i znanstvenom diskursu.

U XIX. stoljeću uočava se znatan porast učestalosti povezanih pojmova „Austrija=>Balkan”. To se razdoblje podudara s habsburško-osmanskim ratovima i kasnijom ekspanzijom Habsburgovaca. I ovaj uvid odgovara općeprihvaćenim historiografskim spoznajama jer je Habsburška Monarhija uslijed ekspanzionističke politike igrala ključnu ulogu u političkom preuređenju Balkana nakon povlačenja Osmanskoga Carstva.

U XX. stoljeću korištenje relacije „Austrija=>Jugoistočna Europa” upućuje na sve veće zanimanje za ovu regiju. Izvaci iz knjiga koje se pojavljuju u korpusu u razdoblju porasta relacije (povijesne studije, političke analize i publicistički tekstovi) koji tematiziraju austrijsku ulogu na prostoru jugoistočne Europe to potkrepljuju.<sup>58</sup>

Povećanje učestalosti relacija početkom XX. stoljeća odražava političke preokrete, posebno balkanske ratove, Prvi svjetski rat i raspad Habsburške Monarhije. Vrhunci učestalosti spominjanja pojmova u njemačkom korpusu snažno koreliraju s važnim povijesnim događajima i geopolitičkim promjenama. Jasno se vidi da je pojam „Balkan” bio mnogo češće korišten od pojma „Jugoistočna Europa”. Daljnji porast može se uočiti 1990-ih, što je povezano s raspadom Jugoslavije i Domovinskim ratom, ali i širim tranzicijskim procesima na tzv. Zapadnom Balkanu i u Jugoistočnoj Europi.



Slika 3. Učestalost spominjanja najpoznatijih hrvatskih i srpskih gradova „Zagreb”, „Agram”, „Beograd” („Belgrad”) od 1700. do 2022.

<sup>58</sup> Vidi: [https://books.google.com/ngrams/graph?content=%C3%96sterreich%2C+Balkan&year\\_start=1700&year\\_end=2022&corpus=de&smoothing=3](https://books.google.com/ngrams/graph?content=%C3%96sterreich%2C+Balkan&year_start=1700&year_end=2022&corpus=de&smoothing=3).

„Beograd” (žuto) i „Belgrad” (zeleno): Spominjanja obaju pojmova pokazuju znatan porast tijekom važnih povijesnih razdoblja poput balkanskih ratova, Prvoga i Drugoga svjetskog rata te Domovinskoga rata 1990-ih. Češće spominjanje Beograda odražava središnju ulogu Beograda u regiji, ali i kao generatora političke nesigurnosti.

„Zagreb” (plavo) i „Agram”<sup>59</sup> (crveno): Spominjanje Zagreba osobito raste u XIX. i XX. stoljeću, što je povezano s njegovom rastućom važnošću kao glavnoga grada Hrvatske i njegovom ulogom u kontekstu Habsburške Monarhije.

„Sarajevo” (svijetloplavo): Spominjanje grada Sarajeva pokazuje izražene vrhunce tijekom važnih povijesnih događaja poput ubojstva nadvojvode Franje Ferdinanda 1914. te opsade Sarajeva tijekom rata u Bosni 1990-ih. Ta spominjanja ponovno odražavaju da GNV pravilno bilježi geopolitičku važnost istraživanih pojmova.

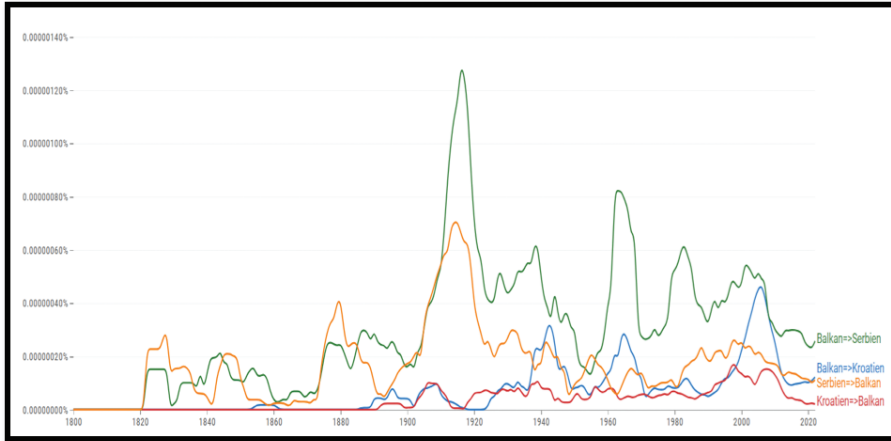
Korištenjem GNV-a može se analizirati i suradnja između gradova. U tu svrhu kao naredbu pretrage koristili smo veznik „i” („Graz i Zagreb”, „Beč i Zagreb”).

„Graz i Zagreb”: Spominjanja dvaju gradova u istom kontekstu bitno rastu krajem XIX. i početkom XX. stoljeća, što upućuje na intenzivnije kulturne, ekonomske i političke veze između tih gradova. To je zanimljivo jer je uobičajena i očekivana veza Zagreb-Beč. Analiza pokazuje da su tijekom razdoblja Habsburške Monarhije postojale bliske veze između Graza i Hrvatske, što se očituje u čestim spominjanjima tih pojmova. Vrhunci spominjanja mogu upućivati na migracije te obrazovne i ekonomske aktivnosti Hrvata u Grazu. To je potvrđeno daljnjim istraživanjem opcijom „Pretraži Google Books”. Međutim, na umu treba imati i da krajem XIX. stoljeća, kad započinje postupno rasformiranje Vojne krajine kao zasebnoga administrativno-vojnog entiteta unutar Habsburške Monarhije, Ratno vijeće u Grazu ostaje jedno od ključnih regionalnih vojno-upravnih tijela. U tom razdoblju dolazi do znatnih strukturnih promjena u upravljanju graničnim područjima, pri čemu se vojna administracija sve više podređuje civilnim strukturama vlasti. Smještaj Ratnoga vijeća u Grazu, s obzirom na njegovu zemljopisnu blizinu i stratešku važnost, reflektira reorganizaciju vojnoga sustava na prijelazu iz klasičnoga pograničnog modela u modernije oblike upravljanja unutar Austro-Ugarske. Ovaj primjer pokazuje da GNV može doprinijeti povijesti odnosa između određenoga grada i određenoga naroda te je pogodan za korištenje u nastavi.

„Beč i Zagreb (te Agram)”: Iako su u doba Habsburške Monarhije postojale bliske veze između Hrvatske i Austrije, to se, na naše iznenađenje, ne odražava u spominjanju ovoga para gradova (naime, ipak je Beč imao neusporedivo veću ulogu od Graza). Stoga bi svakako bilo zanimljivo dodatno istražiti zašto veze između tih gradova nisu bile u fokusu njemačke politike, literature i znanosti. Ovo je primjer kako GNV može poslužiti za oblikovanje novih istraživačkih pitanja.

---

<sup>59</sup> Agram je povijesni njemački naziv za Zagreb.



Slika 4. Odnos/zavisnost pojmova u istom kontekstu i komparacija „Balkan=>Hrvatska, Hrvatska=>Balkan, Balkan=>Srbija, Srbija=>Balkan” (naredba =>) (1800. do 2022.)

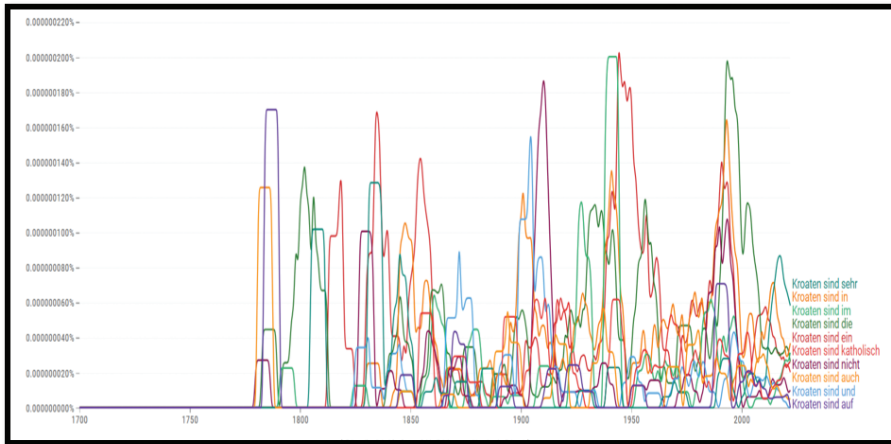
Ovaj primjer može poslužiti kao istraživačko pitanje koja se zemlja više percipira kao „balkanska” (Hrvatska ili Srbija). Primjeri pokazuju da je neupitno riječ o Srbiji.<sup>60</sup>

Analiza učestalosti relacije „Balkan=>Srbija” pokazuje znatan porast u XIX. stoljeću, osobito tijekom ratova za oslobođenje i rastuće težnje Srbije za autonomijom od Osmanskoga Carstva. Srpski ustanci (1804. – 1813. i 1815. – 1817.) protiv osmanske vlasti i uloga Srbije u balkanskim ratovima (1912. – 1913.) dovode do osobito povećanog interesa na njemačkom govornom području.

„Balkan=>Hrvatska”: I ovdje učestalost raste u XIX. stoljeću (ipak znatno manje nego u slučaju Srbije), što odražava politička zbivanja u Hrvatskoj, uključujući hrvatski otpor prema habsburškoj vladavini i hrvatsko-ugarske pregovore (1868.).

Daljnji znatan porast relacije „Balkan=>Srbija” početkom XX. stoljeća snažno korelira s ulogom Srbije u Prvom svjetskom ratu i kasnijim osnivanjem Kraljevine Srba, Hrvata i Slovenaca. I u slučaju Hrvatske učestalost raste u tom razdoblju, što odražava rastuću političku važnost Hrvatske u kontekstu nove južnoslavenske države. Relacije i ovdje pokazuju ponovno povećanje 1990-ih, što je povezano s raspadom Jugoslavije i posljedičnim ratom.

<sup>60</sup> Balkan se često definira iz perspektive zapadnih centara moći (Berlin, London), a Srbija se u tim narativima neupitno pozicionira kao „balkanska” država, dok se za Hrvatsku često koristi kvalifikator „granična”. Istraživanja pokazuju da se Srbija percipira kao agresivna, ratoborna i nacionalistička, što su stereotipi često povezani s „balkanskim” identitetom. Srpska politička povijest obilježena je izvaninstitucionalnim prevratima, autokracijom i mitologizacijom, što se uklapa u zapadni stereotip „balkanske nestabilnosti”.



Slika 5. Koja riječ najčešće slijedi nakon fraze „Hrvati su” (naredba =>\*) (1700. do 2022.)

U ovom primjeru (slika 5) vidimo da grafikon može izgledati potpuno kaotično te je u tom slučaju nužno koristiti opciju „Pretraži Google Books”, koja omogućuje sistematiziranje.

Ovaj primjer pokazuje da GNV nudi velik potencijal za analizu percepcije naroda u međunarodnom okruženju – mogućnost analize stereotipa i kulturne i povijesne percepcije naroda Jugoistočne Europe na njemačkom govornom području. GNV grafikon s opcijom „=>\*” prikazuje različite riječi koje se pojavljuju nakon fraze „Hrvati su”. Neke od najčešćih su „ponosni”, „hrabri”, „nacionalistički”, „borbeni”, „tradicionalni”. Opcija „Pretraži Google Books” pokazuje da se pojmovi mijenjaju ovisno o povijesnom kontekstu: za vrijeme ratova ili političkih sukoba, primjerice, dominiraju pojmovi poput „borbeni” ili „nacionalistički”.

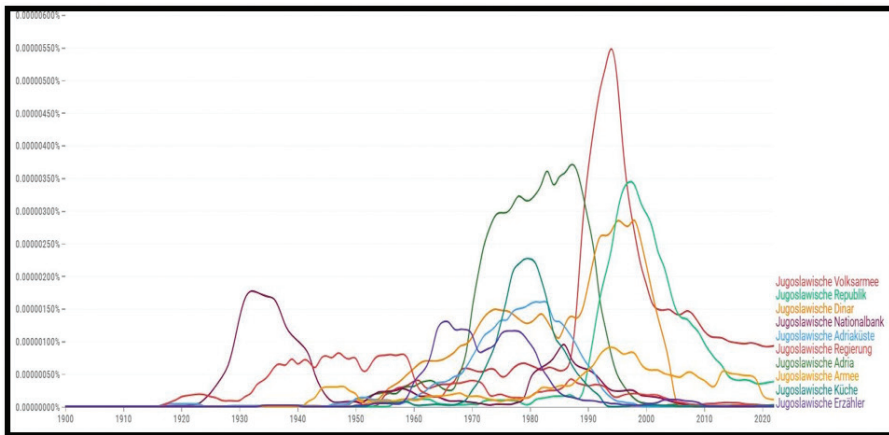
Od 1700. do 1800. godine spominjanja su relativno rijetka, što bi se pogrešno moglo interpretirati na način da ti nalazi pokazuju da Hrvati nisu bili u fokusu javnosti njemačkoga govornog područja. Naime, ovdje je nužno na umu imati širenje hrvatskoga imena i učvršćivanje nacionalne zajednice koja se formira tek sredinom XIX. stoljeća. Stoga je to ujedno jedan od primjera kako ovaj pristup može navesti na pogrešne zaključke ako se ne koristi uz kritičku prosudbu.

Od XIX. stoljeća bilježi se zamjetan porast učestalosti pojmova povezanih s hrvatskim nacionalnim identitetom u njemačkom jezičnom korpusu. Takva tendencija može se interpretirati kao odraz rastuće političke artikulacije koncepta hrvatske nacije, odnosno sve učestalije prisutnosti toga koncepta u javnom diskursu. Ovaj primjer mogao bi istraživača navesti da govori o porastu „povijesne važnosti” hrvatskoga naroda. Međutim, taj je pojam normativno neodređen, pa je analitički prikladnije promatrati njegovu semantičku dinamiku kroz prizmu političkoga značenja i diskurzivne prepoznatljivosti unutar

konkretnoga društveno-povijesnoga konteksta. Stoga je ovo primjer kako se nalazima može lako manipulirati ako se ne zadrži nepristranost.

U ovom primjeru valja istaknuti i nalaz „Hrvati su katolici”, koji upućuje na učestalu tematsku vezu između etničkoga i religijskoga identiteta, osobito u razdoblju Austro-Ugarske, kad je vjerska pripadnost igrala ključnu ulogu u administrativnoj klasifikaciji naroda. GNV tako omogućuje kvalitativnu interpretaciju kvantitativnih podataka, gdje se pojavnost sličnih konstrukcija koristi kao indikator kulturne reprezentacije i stereotipizacije u javnom diskursu. Ovaj pristup pomaže u razumijevanju kako se nacije identificiraju i kategoriziraju u pojedinom jeziku odnosno kulturi te u kojim se kontekstima religijska oznaka koristi kao ključni identitetski marker.<sup>61</sup>

U nastavi povijesti ovakvi nalazi mogu poslužiti kao osnova za: 1. raspravu o konstrukciji nacionalnoga identiteta kroz jezik i ideologiju; 2. usporedbu s drugim narodima i njihovim religijskim povezivanjem (npr. „Poljaci su katolici”, „Srbi su pravoslavci”); 3. analizu medijskih, političkih ili znanstvenih diskursa koji utječu na oblikovanje percepcije etničkih skupina.



Slika 6. Učestalost spominjanja pridjeva „jugoslavenski” u njemačkom korpusu i rezultati povezanih riječi (naredba \*) (1700. do 2022.)

U ovom primjeru pokazat ćemo mogućnost analize s unosom parametra „\*” nakon unesenoga pojma. Ova naredba omogućuje prikaz svih najčešćih riječi koje su povezane s pridjevom „jugoslavenski”. Najfrekventnije su sintagme „Jugoslavenska narodna armija” i „Jugoslavenska republika”. Treba naglasiti da pojam „Jugoslavenska republika” u formalnom smislu nikad nije

<sup>61</sup> Iako se ne može tvrditi da frekvencijski podaci iz GNV-a izravno odražavaju kompleksne društvene stavove, njihova prisutnost upućuje na kulturološku percepciju u javnom diskursu analiziranoga razdoblja. U tom smislu, jezični obrasci postaju indikatori dominantnih narativa i stereotipa, pri čemu GNV funkcionalno služi kao alat za identifikaciju tematskih središta u kulturnim reprezentacijama.

postojao. Unatoč tome, njegova pojavnost u knjižnom korpusu sugerira postojanje kolokvijalne, popularne i netočne upotrebe u javnom diskursu.<sup>62</sup> Učestalost tih izraza najizraženija je sredinom XX. stoljeća (osnivanje Socijalističke Federativne Republike Jugoslavije), kao i početkom 1990-ih (raspad SFRJ).

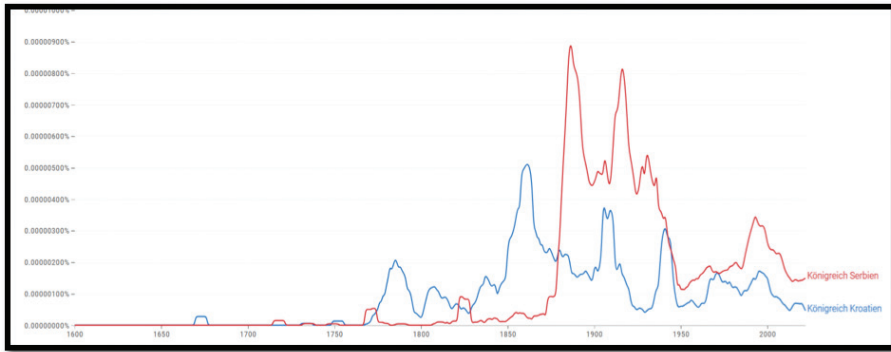
Osim političkih događaja, susljedno se može pratiti i bilježenje razvoja turizma od 1970-ih, što pokazuju izrazi „jugoslavenska obala” i „jugoslavensko Jadransko more”.<sup>63</sup> Istovremeno ovi primjeri pokazuju da se hrvatsko ime u njemačkom korpusu „utopilo” u pojam Jugoslavije, odnosno da je diferenciranje Hrvatske od Jugoslavije rijetko.

*Politika i politička povijest: diskurzivna prisutnost Austrije u kontekstu Jugoistočne Europe u njemačkom korpusu (relacija Austrija – Hrvatska i Austrija – Srbija)*

Na početku ovoga poglavlja, kad ćemo češće spominjati pojam „Austrija”, važno je istaknuti da se analiza temelji na njemačkom jezičnom korpusu, koji obuhvaća publikacije objavljene na njemačkom jeziku neovisno o zemljopisnom porijeklu autora ili izdavača. Budući da korpus nije teritorijalno segmentiran (npr. na austrijski, pruski ili bavarski potkorpus), pojam „Austrija” promatra se kao jezični entitet unutar njemačkoga korpusa, a ne kao izdvojena nacionalna ili regionalna komponenta. Stoga se učestalost relacija poput „Austrija – Hrvatska” i „Austrija – Srbija” odnosi na pojavljivanje tih pojmova u njemačkom jeziku, pri čemu se ne pretpostavlja da su svi izvori austrijskoga porijekla. Ipak, s obzirom na povijesnu povezanost Austrije s Jugoistočnom Europom, može se očekivati da je diskurzivna prisutnost Austrije u kontekstu Hrvatske i Srbije izraženija nego kod drugih njemačkih regija, što se reflektira u jezičnim obrascima.

<sup>62</sup> Vidi: [https://books.google.com/ngrams/graph?content=Jugoslawische+\\*&year\\_start=1900&year\\_end=2022&corpus=de&smoothing=3](https://books.google.com/ngrams/graph?content=Jugoslawische+*&year_start=1900&year_end=2022&corpus=de&smoothing=3).

<sup>63</sup> Već krajem XIX. i početkom XX. stoljeća hrvatska obala – osobito područja poput Opatije i Crikvenice – bilježi intenzivan razvoj turističke infrastrukture, pri čemu češki turisti igraju ključnu ulogu kao pokretači jadranskoga turizma. Opatija je 1889. proglašena prvim morskim klimatskim lječilištem Austro-Ugarske, a Crikvenica započinje svoj turistički razvoj još 1888. godine. Ti primjeri pokazuju da je Hrvatska kao odredište bila prepoznata i imenovana u međunarodnom kontekstu, osobito među srednjoeuropskim gostima, prije nego što je jugoslavenski okvir postao dominantan u korpusima kao što je Google Books. Stoga se može zaključiti da je diferencijacija Hrvatske i Jugoslavije postojala u ranijim razdobljima, ali da se u kasnijem diskursu – osobito u njemačkom jeziku – hrvatski identitet diskurzivno preklapio s jugoslavenskim, što je vidljivo u frekvencijskim podacima.



Slika 7. Učestalost spominjanja pojmova „Kraljevina Hrvatska”, „Kraljevina Srbija”<sup>64</sup> (1600. do 2022.)

Slika 7 prikazuje frekvencijsku usporedbu izraza „Kraljevina Hrvatska” i „Kraljevina Srbija” u njemačkom korpusu Google Books, čime se otvara prostor za istraživačko pitanje o diskurzivnoj prisutnosti dviju zemalja kao političkih entiteta u njemačkom jezičnom prostoru. Takva analiza omogućuje indirektnu rekonstrukciju percepcije bilateralnih odnosa, pri čemu se učestalost pojmova promatra kao indikator javnoga interesa, političke vidljivosti i kulturne reprezentacije.

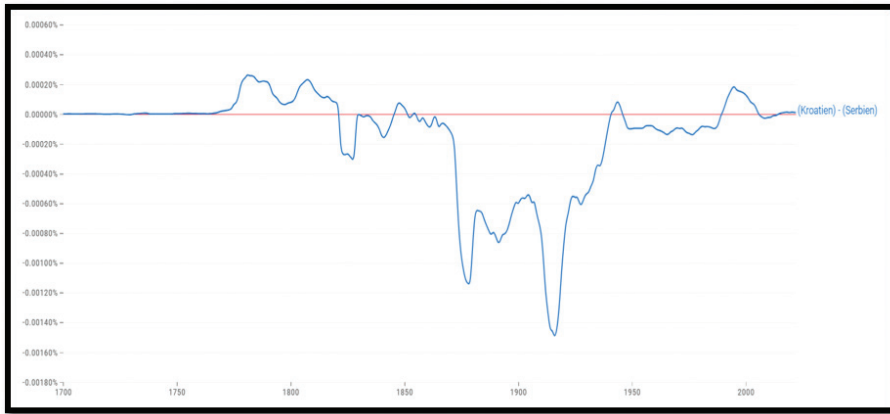
Frekvencijski vrhunci izraza „Kraljevina Srbija” zapaženi su sredinom XIX. stoljeća, što se može povezati sa srpskim ustancima, dinastičkim promjenama i postupnim oblikovanjem srpske državnosti. S druge strane, pojam „Kraljevina Hrvatska”<sup>65</sup> bilježi povećanu pojavnost oko 1850-ih, u kontekstu nacionalnoga preporoda, ilirskoga pokreta, i političkih reformi unutar Habsburške Monarhije.

Ovi podaci upućuju na korelaciju između jezične frekvencije i povijesnih zbivanja, pri čemu se diskurzivna prisutnost Srbije u njemačkom korpusu pokazuje kvantitativno izraženijom od one Hrvatske. Takva razlika može se interpretirati kao rezultat geopolitičkih interesa, diplomatskih dinamika i kulturnih narativa koji su oblikovali njemačku percepciju Jugoistočne Europe. Dodatna potvrda te tendencije vidljiva je na slici 8, gdje se nastavlja analiza političkih odnosa kroz jezične obrasce. Analiza pokazuje da se Srbija u političkim odnosima znatno češće spominje od Hrvatske.<sup>66</sup>

<sup>64</sup> Njem. *Königreich Kroatien, Königreich Serbien*.

<sup>65</sup> Dodatna istraživanja mogu provjeriti učestalost pojave termina „Trojedna kraljevina” te „Kraljevstvo Hrvatske, Slavonije i Dalmacije”.

<sup>66</sup> Napomena: slični rezultati dobivaju se i kod pretraga bez dodatka „kraljevina” (*Königreich*).



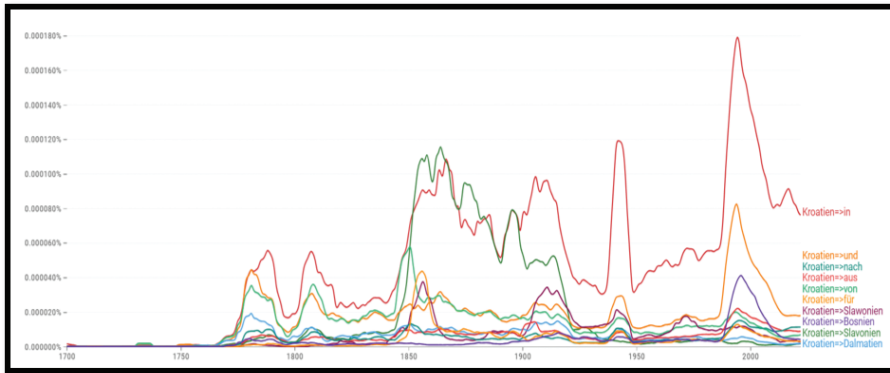
Slika 8. Unos „Hrvatska - Srbija” (naredba -) (1700. do 2022.)

Slika 8 pokazuje da unos „-” omogućuje prikaz razdoblja u kojima se Hrvatska spominjala manje od Srbije. Češće spominjanje pojma „Srbija” u XIX. i XX. stoljeću sugerira da je Srbija igrala važniju ulogu od Hrvatske u političkim i kulturnim diskursima na njemačkom govornom području.<sup>67</sup> To također može upućivati na geopolitičku važnost Srbije u to vrijeme, ali je svakako uvjetovano i sukobom Austrije i Njemačke sa Srbijom, za razliku od Hrvatske, koja je percipirana sastavnim dijelom Habsburške i poslije Austro-Ugarske Monarhije.

Ovo je ujedno primjer kako se metoda može koristiti u nastavi povijesti. Primjer može poslužiti kao polazište za kritičku raspravu o 1. geopolitičkom značenju Srbije u odnosu na Hrvatsku u kontekstu njemačko-austrijskih interesa te 2. diskurzivnim razlikama u percepciji Jugoistočne Europe u različitim povijesnim razdobljima. Takva analiza omogućuje nastavnicima da povežu jezične obrasce s međunarodnim političkim kontekstom, otvore raspravu o kolonijalnim, imperijalnim i geostrateškim perspektivama njemačko-austrijskoga diskursa i postave pitanje kako se povijest „piše” ili konstruira kroz jezik.<sup>68</sup>

<sup>67</sup> Vidi izvratke iz knjiga: [https://books.google.com/ngrams/graph?content=Serbien%2C-Kroatien&year\\_start=1800&year\\_end=2022&corpus=de&smoothing=3](https://books.google.com/ngrams/graph?content=Serbien%2C-Kroatien&year_start=1800&year_end=2022&corpus=de&smoothing=3).

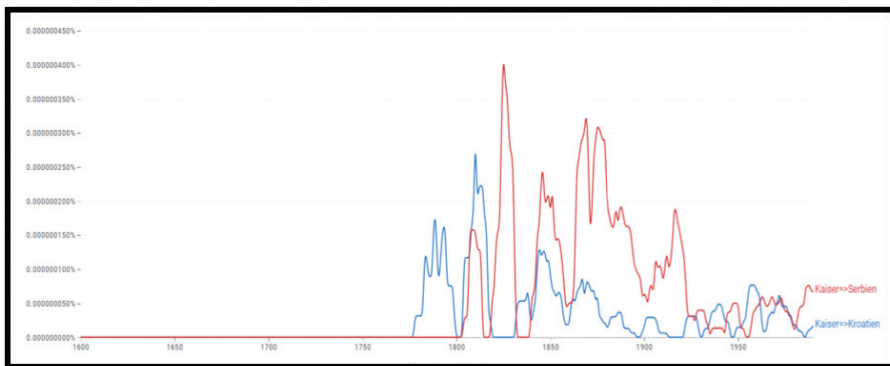
<sup>68</sup> Mogući nastavni zadatak: učenici/studenti analiziraju frekvencijsku uporabu pojmova „Kraljevina Srbija” i „Kraljevina Hrvatska” u njemačkom korpusu pomoću GNV-a te raspravljaju: Zašto se određeni pojmovi učestalije pojavljuju? Koje povijesne događaje možemo povezati s tim učestalostima? Kako politički kontekst utječe na jezične reprezentacije? Ovim zadatkom učenici razvijaju kritičku pismenost i analitičko mišljenje, vještine rada s digitalnim izvorima i osjetljivost na jezično oblikovanje povijesne stvarnosti.



Slika 9. Koja riječ najčešće slijedi iza pojma „Hrvatska” (naredba =>\*) (1700. do 2022.)

Ovdje koristimo naredbu „=>\*”, koja omogućuje uvid u to koja riječ najčešće slijedi nakon pojma „Hrvatska”, odnosno koje su najčešće kontekstne veze. Može se primijetiti da se pojam „Bosna i Hercegovina” u analiziranom razdoblju često spominje u vezi s Hrvatskom, što upućuje na bliske odnose između te dvije zemlje. Pojmovi poput „Slavonija” i „Dalmacija” su pak očekivani jer su bili sastavni dio naziva Kraljevine tijekom različitih povijesnih perioda.

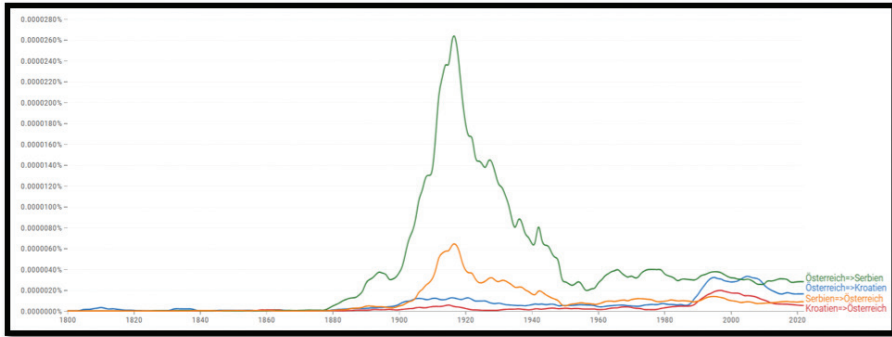
Česta uporaba prijedloga „u” upućuje na to da se Hrvatska često spominje u geografskim ili političkim kontekstima (Hrvatska u Jugoslaviji, Hrvatska u Monarhiji itd.). Riječi „protiv” i „za” daju indicije da postoje brojne rasprave u njemačkom korpusu u kojima pojam „Hrvatska” služi kao predmet argumenta ili debata.



Slika 10. Odnos/zavisnost pojmova u istom kontekstu i komparacija „car=>Hrvati; car=>Srbi” (naredba =>) (1600. do 1990.)

Relacija „car=>Hrvati” (njem. *Kaiser, Kroaten*) pokazuje kontinuirane vrhunce u učestalosti spominjanja od XVI. stoljeća do raspada Monarhije 1918. godine, što upućuje na dugotrajnu povijesnu i političku važnost toga odnosa

u njemačkom diskurzivnom prostoru.<sup>69</sup> Odnos „car=>Srbi”<sup>70</sup> osobito je pak naglašen u razdobljima političkih sukoba te otkriva da je odnos između Srba i austrijskoga cara bio obilježen vojnim sukobima, političkim napetostima i teritorijalnim sporovima.<sup>71</sup>



Slika 11. Odnos/zavisnost pojmova u istom kontekstu i komparacija „Austrija=>Hrvatska, Hrvatska=>Austrija, Austrija=>Srbija, Srbija=>Austrija” (naredba =>) (1800. do 2022.)

Učestalost spominjanja analiziranoga konteksta raste krajem XIX. stoljeća, što bi moglo biti povezano s germanizacijom i teritorijalnim težnjama Monarhije. Također se pojavljuju značajni vrhunci u prvoj polovini XX. stoljeća, osobito tijekom oba svjetska rata. Ponovno se može primijetiti da je relacija/zavisnost pojmova „Austrija=>Srbija” izraženija, što sugerira da je Austrija pokazivala veći (politički) interes za Srbiju u odnosu na Hrvatsku.

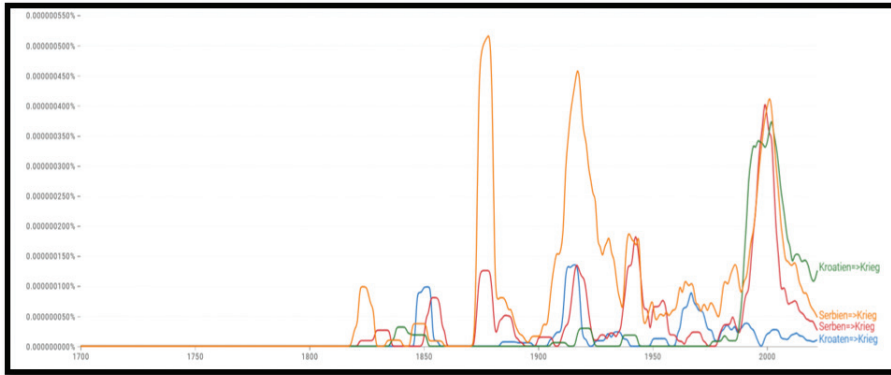
Ovdje treba imati na umu da se takva razlika u frekvenciji može tumačiti i kao refleksija geopolitičkoga položaja Srbije izvan Habsburške Monarhije, pri čemu je austrijski interes proizašao iz imperijalnih ambicija širenja utjecaja dinastije Habsburg prema jugoistočnom Balkanu. S druge strane, Hrvatska se već nalazila unutar Monarhije – i to pretežito u ugarskom administrativnom okviru, s iznimkom Dalmacije, koja je bila pod austrijskom upravom – čime je politička relacija bila stabilnija i formalizirana, a samim time manje prisut-

<sup>69</sup> Vidi izvatke iz knjiga: [https://books.google.com/ngrams/graph?content=Kaiser+%3D%3E+Kroaten&year\\_start=1600&year\\_end=2022&corpus=de&smoothing=3](https://books.google.com/ngrams/graph?content=Kaiser+%3D%3E+Kroaten&year_start=1600&year_end=2022&corpus=de&smoothing=3).

<sup>70</sup> Za daljnju analizu bilo bi korisno istražiti frekvenzijsku i semantičku pojavnost konstrukcija poput „kralj – Hrvati” i „kralj – Srbi”. S obzirom na činjenicu da je austrijski car istovremeno nosio titulu kralja Hrvatske i Ugarske, a poslije su i Karađorđevići formalno postali hrvatski kraljevi u okviru Kraljevstva Srba, Hrvata i Slovenaca, takva analiza može rasvijetliti slojevitost političkoga suvereniteta i diskurzivnu varijabilnost pojma „kralj” u odnosu na različite nacionalne zajednice. Te konstrukcije također otvaraju pitanje diskurzivnoga razgraničenja i (ne) ujednačenosti u prikazu monarhijskih odnosa prema Hrvatskoj i Srbiji. Frekvenzijska analiza pojmova može upozoriti na razlike u političkom tumačenju, legitimacijskom jeziku i simboličkom pozicioniranju unutar njemačkoga govornog prostora. Takvi nalazi bili bi dragocjeni u kontekstu istraživanja percepcije vladarske vlasti, nacionalne pripadnosti i kulture sjećanja.

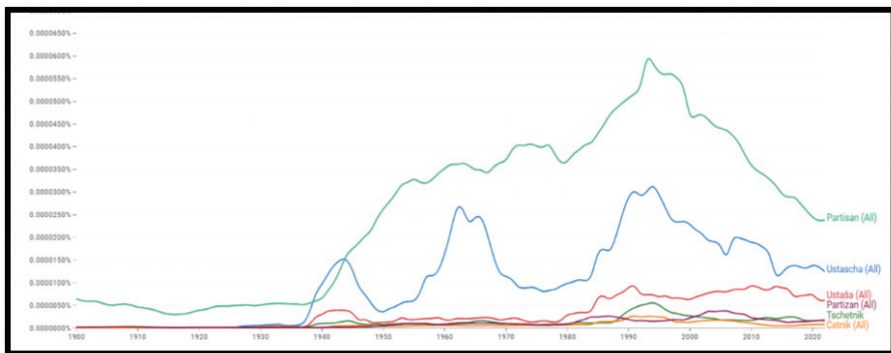
<sup>71</sup> Vidi opširno u: CLEWING, SCHMITT, *Geschichte Südosteuropas*.

na u vanjskopolitičkom diskursu. U tom smislu, učestalost pojma „Srbija” u relaciji s „Austrija” u korpusu ne mora nužno označavati veći interes za samu zemlju, nego može upućivati na dinamičniju prirodu bilateralnih odnosa, uključujući diplomatska nastojanja, konflikte i strateške projekcije.



Slika 12. Komparacija spominjanja Hrvata i Hrvatske te Srba i Srbije zajedno s pojmom „rat” (Krieg) („Hrvatska=>rat, Srbija=>rat”) (Unos =>) (1700. do 2022.)

Porast učestalosti pojmova „Srbi” i „Srbija” u kontekstu riječi „Krieg” u njemačkom korpusu tijekom određenih razdoblja (npr. oko 1914., 1941., 1990.) može se interpretirati kao odraz povećanoga interesa i za Srbiju i Hrvatsku u ratnim kontekstima, što korelira s poznatim povijesnim događajima poput balkanskih ratova, Prvoga i Drugoga svjetskog rata i sukoba 1990-ih.<sup>72</sup> Dugoročni učinci rata, povezanost s ratom u Bosni te izazovi političke i društvene obnove, suočavanja s prošlošću i ratnim zločinima također su snažno zastupljene teme u njemačkom korpusu, što je vidljivo iz daljnje analize „Pretraži Google Books”.



Slika 13. Analiza sentimentata (i učestalost spominjanja) pojmova „Ustaša, Ustascha, Četnik, Tschetnik, Partisan” (1900. do 2022.)

<sup>72</sup> Vidi opširno u: SUNDHAUSSEN, „Postjugoslawische Kriege (1991-95, 1998/99)”.

Vrhunci spominjanja navedenih pojmova koreliraju s važnim povijesnim događajima i rastućim međunarodnom interesom za povijest Jugoistočne Europe tijekom Drugoga svjetskog, ali i Domovinskoga rata. Naime, uočava se da je kraj XX. stoljeća donio ponovno suočavanje s poviješću i političkim ideologijama regije.

Navedeni primjer prikazuje kako se GNV može koristiti za analizu sentimentata, odnosno naklonost određenim pojmovima, iz čega se onda mogu izvući zaključci i o široj slici. Primjerice, pojam „partizan” u njemačkom korpusu češće se vidi u pozitivnom svjetlu,<sup>73</sup> što pokazuje opcija „Pretraži Google Books”, dok pojmovi „ustaša” i „četnik” imaju često negativne konotacije te su naglašene uloge tih skupina u sukobima i zločinima tijekom Drugoga svjetskog rata. Ovaj primjer otvara nova istraživačka pitanja, primjerice zašto partizanski zločini nisu dovoljno zastupljeni u njemačkom korpusu.

Primjer nastavnoga zadatka bio bi da učenici pomoću opcije „Pretraži Google Books” identificiraju primjere iz kojih se može razabrati konotacija pojmova „ustaša”, „četnik” i „partizan” (pozitivna/negativna/neutralna). Analitička pitanja mogla bi biti: Zašto se pojam „partizan” pojavljuje u pozitivnom kontekstu? Koji su politički ili ideološki faktori utjecali na negativne konotacije pojmova „ustaša” i „četnik”? Zašto su partizanski zločini slabije zastupljeni u njemačkom korpusu i kako to utječe na povijesnu ravnotežu prikaza?<sup>74</sup>

I u interdisciplinarnom istraživanju, na razmeđu politologije i povijesti, GNV može biti od velike koristi – primjerice u analizi povijesnoga značenja i percepcije određenih političara. Naime, ovom metodom moguće je utvrditi međunarodni javni interes za određene povijesne ličnosti tijekom određenoga razdoblja, njihovu ulogu u bilateralnim i međunarodnim odnosima te percepciju u različitim kulturnim prostorima ili jezičnim regijama. Njome se može odrediti i analiza sentimenta prema određenim političarima te vidjeti koje su kulturne regije bile sklonije kojem političaru. Primjerice, rezultati učestalosti spominjanja Franje Tuđmana i Slobodana Miloševića potpuno su drugačiji u njemačkom<sup>75</sup> nego u engleskom<sup>76</sup> korpusu. Ovakav tip analize pruža uvide u to jesu li analizirane povijesne ličnosti bile pozitivno, neutralno ili negativno

<sup>73</sup> Ovdje treba imati na umu da se izraz „partizan” ne veže isključivo za Narodnooslobodilačku borbu, nego općenito gerilsku borbu – često u antifašističkom kontekstu (postoje tako i grčki partizani i francuski partizani).

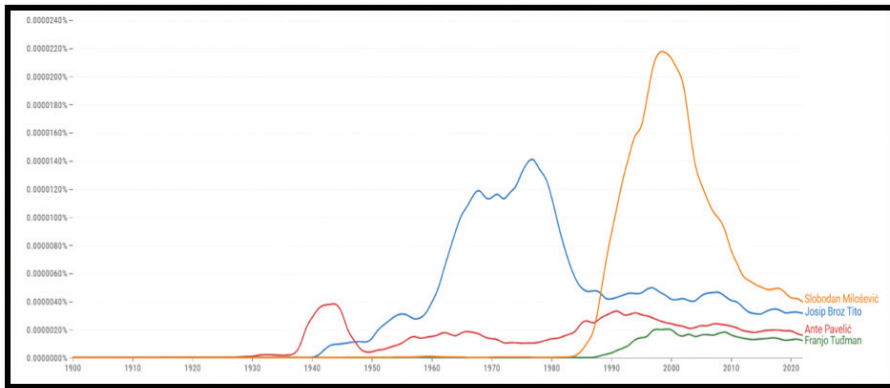
<sup>74</sup> Ishodi učenja: 1. razviti sposobnost interpretacije povijesnoga sadržaja u digitalnom kontekstu; 2. povezati jezične podatke s historiografskim narativima i ideološkim utjecajima; 3. osmisliti istraživačko pitanje temeljeno na digitalnoj analizi; 4. raspravljati o etičkim dimenzijama historiografskoga prikaza sukoba.

<sup>75</sup> Vidi GNV: [https://books.google.com/ngrams/graph?content=Franjo+Tu%C4%91man%2C+Slobodan+Milo%C5%A1evi%C4%87&year\\_start=1800&year\\_end=2022&case\\_insensitive=true&corpus=de&smoothing=3](https://books.google.com/ngrams/graph?content=Franjo+Tu%C4%91man%2C+Slobodan+Milo%C5%A1evi%C4%87&year_start=1800&year_end=2022&case_insensitive=true&corpus=de&smoothing=3).

<sup>76</sup> Vidi GNV: [https://books.google.com/ngrams/graph?content=Franjo+Tu%C4%91man%2C+Slobodan+Milo%C5%A1evi%C4%87&year\\_start=1800&year\\_end=2022&case\\_insensitive=true&corpus=en&smoothing=3](https://books.google.com/ngrams/graph?content=Franjo+Tu%C4%91man%2C+Slobodan+Milo%C5%A1evi%C4%87&year_start=1800&year_end=2022&case_insensitive=true&corpus=en&smoothing=3).

prikazivane, a dodatne naredbe u GNV-u (vidi metodologiju) pomažu razumjeti šire tematske kontekste. Smatramo da je to ponovno dobar primjer za nastavu povijesti.

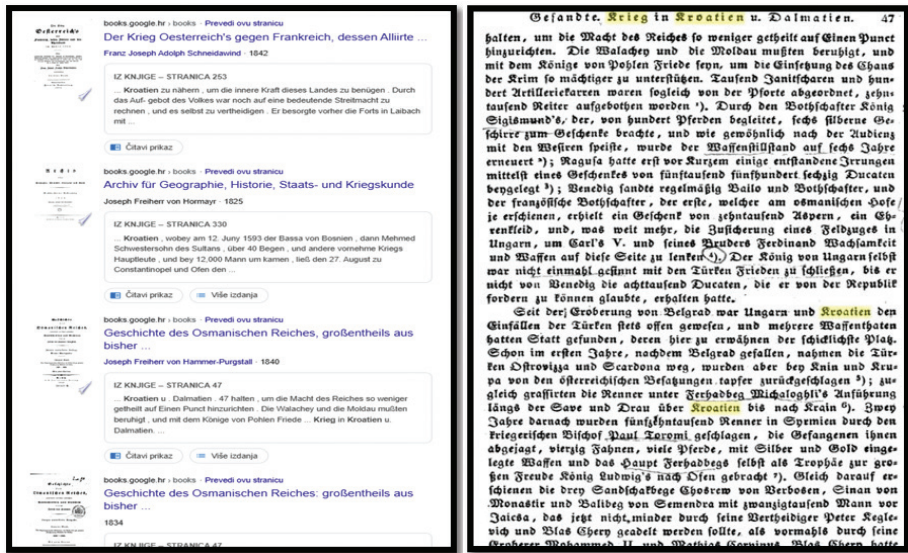
Daljnja istraživačka pitanja mogu pomoću ove metode provjeriti opću percepciju javnosti prema Hrvatskoj primjerice u Francuskoj i Španjolskoj. Daljnje znanstveno istraživanje uz pomoć opcije „Pretraži Google Books” može pokazati jesu li spominjanja tih ličnosti u različitim razdobljima politički instrumentalizirana i u kojim se kulturnim prostorima to češće događalo.



Slika 14. Usporedba učestalosti spominjanja najpoznatijih političara u Hrvatskoj i Srbiji u XX. stoljeću „Josip Broz Tito, Ante Pavelić, Franjo Tuđman, Slobodan Milošević” (1900. – 2022.)

Ovaj primjer usporedbe najpoznatijih političara u Hrvatskoj i Srbiji u XX. stoljeću koristan je za nastavu jer omogućuje učenicima da: 1. usporede diskurzivnu prisutnost političara u različitim jezičnim i kulturnim korpusima; 2. razviju kritičko mišljenje o tome kako se povijesne ličnosti „instrumentaliziraju” u javnom prostoru; 3. istraže kontekstne razlike: npr. zašto je Tuđman češće spominjan u anglo-američkom prostoru, a Milošević u njemačkom?; 4. koriste opciju „Pretraži Google Books” za kvalitativnu analizu – pronalaze nje konkretne knjige i odlomaka koji objašnjavaju vrhunce u grafovima te istraživanje jesu li spominjanja tih ličnosti u različitim razdobljima politički instrumentalizirana i u kojim se kulturnim prostorima to češće događalo.

Dok GNV grafikoni pružaju pregled učestalosti spominjanja, opcija „Pretraži Google Books” pruža dublje uvide u istraživane teme, omogućujući pregled pojedinačnih skeniranih knjiga, ali i povijesnih dokumenata, te tako identificiranje specifičnih povijesnih konteksta i diskursa koji nisu odmah vidljivi u grafičkom prikazu (vidi sliku 15).



Slika 15. GNV opcija „Pretraži Google Books“ – napredna analiza (izvori za daljnje istraživanje)

GNV u svrhu napredne analize nudi poveznice na digitalizirane tekstove, pri čemu su istraživani pojmovi na svakoj pojedinačnoj stranici označeni žutom bojom (vidi sliku 15), što uvelike olakšava pregled, ali i daljnju kontekstualizaciju. Time se ostvaruje velika ušteda vremena, ali to nikako ne isključuje pretraživanje originalnih tekstova u knjižnicama ili arhivima.

### Književnost, umjetnost, jezik i gospodarstvo

Primjena GNV-a otvara niz mogućnosti i u drugim područjima humanističkih i društvenih znanosti. Analiza omogućuje vrijedne uvide u razvoj književnih, umjetničkih, jezičnih i gospodarskih fenomena, osobito kad se promatraju dugoročni trendovi i semantičke promjene. Oni su nerijetko čak i sadržajni za književnost i kulturu nego za političku povijest.

Može se, primjerice, istraživati učestalost književnih pojmova i autora u različitim epohama ili popularnost različitih književnih žanrova u pojedinim zemljama te učiniti komparaciju. U području umjetnosti mogu se analizirati umjetnički pravci kao što su renesansa, barok, impresionizam i njihova važnost tijekom vremena (ili pak spominjanje umjetnika). Time se mogu komparirati europski kulturni utjecaji na pojedine zemlje Jugoistočne Europe, što može imati brojne primjene u međunarodnim odnosima. Primjerice, u kontekstu EU integracija može se analizirati koliko su kulturni utjecaji odredili

sposobnost određenoga društva da prihvati „europske vrijednosti” i norme.<sup>77</sup>

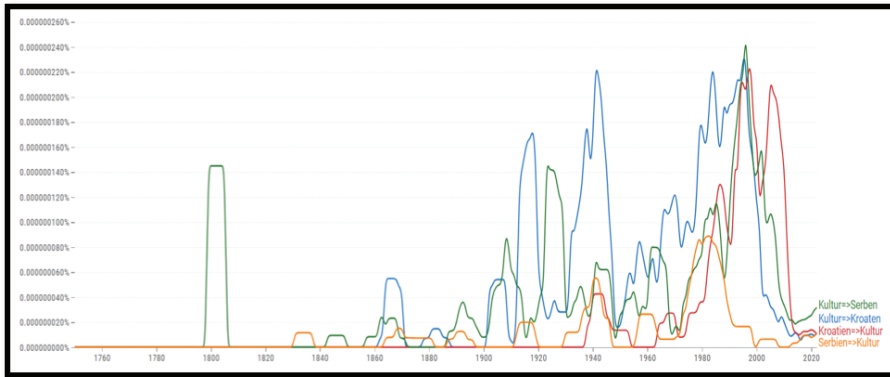
Ova metoda također je korisna za proučavanje jezičnih promjena: mogu se pratiti promjene u korištenju određenih riječi i fraza tijekom desetljeća ili utjecaj stranih jezika. Za Hrvatsku je posebno zanimljiv utjecaj germanizama, što je ponovno primjer kako GNV koristiti u nastavi.<sup>78</sup>

U području gospodarstva, primjerice, može se istraživati i uspoređivati učestalost pojmova poput kapitalizma, socijalizma i industrijalizacije u pojedinim zemljama, ili pak ekonomske teorije i ekonomisti poput Karla Marxa, Adama Smitha itd., iz čega se mogu izvući brojne implikacije i uočiti sklonosti određenoga društva u nekom specifičnom povijesnom vremenu. Primjer nastavnoga sata izgledao bi ovako: učenici se okupljaju u skupinama i otvaraju GNV na svojem mobitelu. Svaka skupina dobiva zadatak istražiti učestalost pojmova na engleskom jeziku *capitalism, socialism, industrialization, Karl Marx i Adam Smith* od 1800. do 2000. godine. Dok pregledavaju grafove, bilježe ključne promjene u učestalosti – nagli porast interesa za pojam „socijalizam” tijekom hladnoga rata, dominaciju „kapitalizma” nakon 1990., ili kontinuiranu prisutnost „industrijalizacije” u XIX. stoljeću. Druga je mogućnost usporedba u njemačkom, engleskom i talijanskom korpusu (svaka grupa dobiva drugi korpus).

Učenici zatim povezuju uočene trendove s povijesnim razdobljima: industrijska revolucija, ideološke borbe XX. stoljeća, tranzicija iz socijalizma u kapitalizam u postsocijalističkim zemljama. Raspravljaju o tome kako se ideološke sklonosti društava odražavaju u jeziku i objavljenim tekstovima te kako se interes za pojedine ekonomiste mijenjao tijekom vremena. Na kraju raspravljaju o ulozi digitalnih alata poput GNV-a u razumijevanju prošlosti. Uviđaju da broj pojavljivanja riječi ne govori sve – potrebno je kontekstno tumačenje, kritički pristup i povezivanje s društvenim i političkim okolnostima. Time razvijaju ne samo digitalnu pismenost nego i sposobnost povijesne interpretacije.

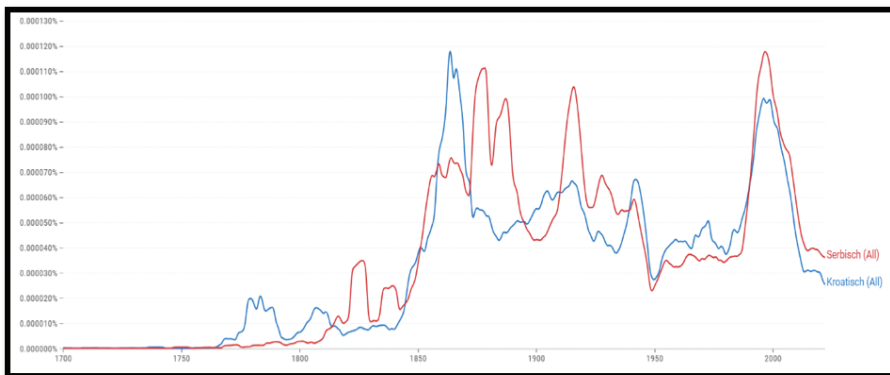
<sup>77</sup> Mogu se komparirati europski kulturni utjecaji na pojedine zemlje jugoistočne Europe, što može imati brojne primjene u međunarodnim odnosima (primjerice u kontekstu EU integracija). Vidi opširno u: JURIC, *Westbalkan-Erweiterung der EU*.

<sup>78</sup> Google Ngram Viewer trenutno ne podržava pretraživanje korpusa na hrvatskom jeziku – njegov korpus temelji se uglavnom na knjigama na engleskom, francuskom, njemačkom, španjolskom, ruskom, talijanskom i hebrejskom jeziku. No to ne znači da se ideja ne može prenijeti u nastavu stranih jezika.



Slika 16. Odnos/zavisnost pojmova u istom kontekstu i komparacija „kultura=>Hrvati, Hrvatska=>kultura, kultura=>Srbi, Srbija=>kultura” (operator =>) (1700. do 2022.)

Oba pojma „kultura” i „umjetnost”, u Hrvatskoj i Srbiji pokazuju znatne poraste u vremenima političkih i društvenih previranja. Posebno je zanimljivo uočiti da se „kultura” češće u kontekstima povezuje s pojmom Hrvatska u usporedbi sa Srbijom, što implicira da su kulturne veze njemačkoga govornog područja s Hrvatskom bile intenzivnije nego sa Srbijom.<sup>79</sup> Treba podsjetiti da smo već pokazali da je suprotan slučaj u usporedbi s političkim vezama. Stoga je ovo ponovno primjer kako se pomoću GNV-a mogu postaviti nove hipoteze, uz napomenu da bi hipoteze trebalo dublje analizirati istraživanjem pojedinačnih skeniranih knjiga i dokumenata (opcija „Pretraži Google Books”).



Slika 17. Učestalost spominjanja jezika „hrvatski”, „srpski” (1700. do 2022.)

<sup>79</sup> Analiza njemačkoga korpusa pokazuje da se pojam „kultura” češće pojavljuje uz „Hrvatsku” nego uz „Srbiju”. To ne implicira nužno intenzivnije kulturne veze, nego može ukazivati na različite obrasce percepcije i reprezentacije Hrvatske i Srbije u njemačkom diskurzivnom prostoru. Takva razlika može biti rezultat turističkih, medijskih, povijesnih ili kulturno-političkih čimbenika koji su oblikovali njemačku recepciju Hrvatske.

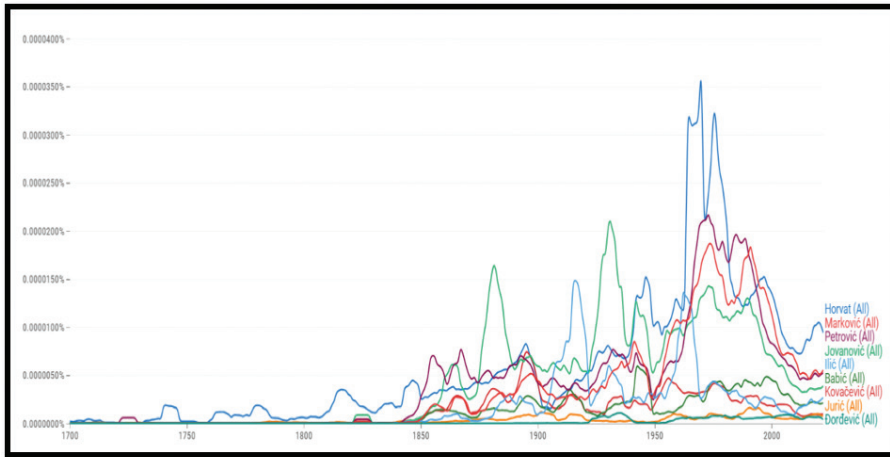
Učestalost spominjanja jezika „hrvatski” i „srpski” u njemačkom korpusu pokazuje slične uzorke u XIX. i XX. stoljeću, pri čemu je spominjanje srpskoga jezika nešto češće. Na ovom primjeru mogla bi se dalje provjeravati hipoteza je li to uslijed veće geografske rasprostranjenosti srpskoga jezika u regiji Jugoistočne Europe ili zbog „srbizacije”. Naime, jugoslavenska vlada poticala je korištenje srpskoga jezika u upravi, obrazovanju i medijima. I pojava pojma „srpskohrvatski”, koji je trebao predstavljati zajednički jezik Srba i Hrvata, poticala je taj proces. Valja naglasiti i da GNV točno uočava pojačan interes za hrvatski jezik tijekom hrvatskoga narodnog preporoda. Ovaj primjer stoga pokazuje da bi GNV mogao biti od velike važnosti i za jezikoslovce (ne nužno samo za njih)<sup>80</sup>. Naime, iako je fenomen „srbizacije” hrvatskoga jezika poznata teza, potpuno je druga perspektiva kad se ona vizualno prikaže (što ova metoda omogućuje).

### *Genealogija i povijesne ličnosti*

GNV može biti od velike koristi i u istraživanju genealogije jer omogućuje analizu učestalosti spominjanja određenih prezimena, ali i imena, unutar jedne obitelji tijekom vremena, čime se može dublje razumjeti prošlost pojedine obitelji, njezine migracije i povezanost između različitih generacija i mjesta.<sup>81</sup> Takvom analizom mogu se povezati brojni povijesni događaji i pratiti kako su se odražavali na određeno društvo ili određenu obitelj i pretke. Primjerice, kombinacija pretraživanja prezimena s određenim mjestima može pomoći razumjeti gdje su živjeli nečiji preci i kako su migrirali tijekom povijesti. Nadalje, GNV omogućuje analizu specifičnih povijesnih dokumenata, poput matičnih knjiga, osmrtnica i povijesnih adresara, koji pokazuju geografsku raspodjelu prezimena i daju dodatne uvide u migracijske obrasce određene obitelji. Na kraju, GNV može pokazati i kako su se osobna imena mijenjala u određenom narodu ili koja su kulturna značenja imala tijekom vremena.

<sup>80</sup> Zanimljivo bi bilo ispitati ono što trenutno u Austriji nazivaju BKS (*Bosnisch-Kroatisch-Serbisch*).

<sup>81</sup> GNV ne omogućuje prostornu (regionalnu) segmentaciju korpusa, ali unutar opcije „Pretraži Google Books” nudi uvid prema geografskom podrijetlu publikacije (npr. mjesto izdavanja, zemlja autora ili izdavača).



Slika 18. Povijesne migracije: učestalost spominjanja najpoznatijih hrvatskih<sup>82</sup> i srpskih prezimena u njemačkom govornom području „Horvat, Babić, Jurić, Marković, Ilić, Kovačević, Jovanović” (1700. do 2022.)

Slika 18 pokazuje kako GNV može dati doprinos u genealoškom istraživanju i istraživanju povijesnih migracija. Od 1700. do 1800. spominjanja najčešćih hrvatskih i srpskih prezimena „Horvat, Babić, Jurić, Marković, Ilić, Kovačević, Jovanović” relativno su rijetka, što upućuje na to da su veze i kretanja između zemalja njemačkoga govornog područja i hrvatskih odnosno srpskih regija manje učestali. Spominjanja se povećavaju od XIX. stoljeća te još intenzivnije u XX. stoljeću, što govori i o povećanoj migraciji. Međutim, ovdje treba imati na umu da brojne hrvatske „kmetske” obitelji, kao i srpske pod Osmanlijama, dugi niz stoljeća nisu ni imale prezime.<sup>83</sup> Prezimena su često bila rezervirana za plemićke obitelji ili su se počela koristiti tek kasnije, s porastom administrativnih potreba i uređenja društva.<sup>84</sup> Ova činjenica može utjecati na rezultate analize povijesnih dokumenata i učestalosti spominjanja prezimena tijekom povijesti te je ponovno pokazatelj zašto ovu metodu treba koristiti uz kritičko razmišljanje.

<sup>82</sup> Vidi primjerice: ŠIMUNOVIĆ, *Hrvatska u prezimenima*.

<sup>83</sup> Kmetske i podaničke obitelji u hrvatskim i srpskim krajevima često nisu imale prezimena sve do kasnoga XVIII. ili čak XIX. stoljeća. Prezimena su bila rezervirana za plemstvo, građanstvo i administrativne elite, a seljaci i podanici često su bili označeni samo osobnim imenom i patronimom. U osmanskom kontekstu prezimena su se često formirala iz funkcija, zanata ili nadimaka, a ne kao nasljedna identifikacijska oznaka. Uvođenje matica rođenih, vjenčanih i umrlih nakon Tridentskoga koncila (XVI. stoljeće) u katoličkim krajevima te defterski popisi u Osmanskom Carstvu potaknuli su potrebu za stabilnim identifikatorima. Na umu treba imati i da su se prezimena širila zbog poreznih, vojnih i pravnih potreba, a ne nužno zbog kulturne tradicije.

<sup>84</sup> Vidi primjerice: KEKEZ, „Urbarijalni nameti i katastarski popis kmetskih selišta na vlastelinstvu Brlog u 18. stoljeću”.

Prezime „Horvat” pokazuje najvišu učestalost spominjanja, što točno upućuje na to da je ovo prezime posebno učestalo u hrvatskoj populaciji, stoga se češće pojavljuje i u njemačkoj literaturi i izvorima.<sup>85</sup> GNV je u ovom primjeru također pokazao da opcijom „Pretraži Google Books” točno upućuje na dodatnu literaturu i izvore, kao što je primjerice knjiga Ivana von Bojničića *Der Adel von Kroatien und Slavonien* (Nürnberg, 1905.).

Za kraj, a u nedostatku više prostora, potrebno je još spomenuti da je GNV od koristi i u analizi brojnih drugih tematskih skupova. Primjerice, analiza iz područja zemljopisa i razvoja turizma pokazuje velik potencijal (pri čemu se mogu analizirati hrvatske znamenitosti, dvorci, prirodne ljepote i sl. te njihova prepoznatljivost na njemačkom i drugim govornim područjima).

## Zaključak

U ovom radu testirali smo digitalni alat GNV kroz različite analitičke prizme – geopolitiku, kulturne utjecaje, ratove, migracije i kulturnu produkciju. Fokusrajući se na odnose između Austrije i Jugoistočne Europe, s posebnim naglaskom na Hrvatsku, pokazali smo da GNV ne samo otkriva kvantitativne promjene u diskursu nego i pruža dublje uvide u dinamiku javnoga interesa tijekom više stoljeća. Istovremeno članak služi kao priručnik za rad pomoću navedene metode te pokazuje kako ju koristiti u nastavi.

Pokazali smo da GNV analiza, u vezi s predmetom našega istraživanja (njemački korpus), pouzdano odražava javni interes određenoga društva i kulture te autentične političke ili kulturne događaje. Demonstrirali smo kako se uz pomoć GNV-a mogu identificirati obrasci i trendovi u spominjanju ključnih povijesnih ličnosti, pojmova i pojava koje nadilaze mogućnosti klasičnih metoda. Većina primjera osmišljena je kao praktični vodič za primjenu GNV-a u nastavi povijesti (ne nužno), čime GNV postaje i relevantan pedagoški alat. Osim za političku povijest, korištenjem GNV-a također se mogu dobiti vrlo zanimljivi uvidi u području književnosti, umjetnosti, jezika i gospodarstva.

Međutim, nužno je prepoznati i metodološka ograničenja ovoga digitalnog alata. Iako su brojna istraživanja potvrdila njegovu osnovnu pouzdanost, u literaturi su zabilježene primjedbe na nedostatke vezane uz selekciju korpusa, kvalitetu OCR obrade te interpretaciju statističkih rezultata. Budući da je analiza velikih tekstualnih skupova i dalje u ranoj fazi razvoja, potrebno je uključiti širu akademsku zajednicu u standardizaciju postupaka, kritičko preispitivanje pretpostavki te daljnje usavršavanje metodologije.

---

<sup>85</sup> Vidi primjerice: ŠIMUNOVIĆ, *Hrvatska u prezimenima*.

## Literatura

ACERBI, Alberto; LAMPOS, Vassailo; GARNETT, Philipp; BENTLEY, Alexander. „The Expression of Emotions in 20th Century Books”. *PLOS One* 8 (2013), br. 3. <https://doi.org/10.1371/journal.pone.0059030>.

BANERJEE, Debatri. „The Impact of Digital Tools on the Teaching of History: A Comparative Study”. *International Journal for Multidisciplinary Research (IJFMR)* 6 (2024), br. 3: 1-5.

BERRY, David M.; FAGERJORD, Andres. *Digital Humanities: Knowledge and Critique in a Digital Age*. UK: Polity, 2017.

BOJNIČIĆ, Ivan von. *Der Adel von Kroatien und Slavonien*. Nürnberg, 1905.; faksimilni pretisak izdanja: J. Siebmacher, Wappenbuch, Bd. 3, Abt. 13, Nürnberg: Verlag von Bauer und Raspe, 1899.

BRUNNBAUER, Ulf. „Europa und der Balkan: Fremd- und Selbstzuschreibungen”. Pristup ostvaren 18. 9. 2024. [https://www.kas.de/c/document\\_library/get\\_file?uuid=e160c5a2-9020-a457-2f2c-78ee765bb069&groupId=252038](https://www.kas.de/c/document_library/get_file?uuid=e160c5a2-9020-a457-2f2c-78ee765bb069&groupId=252038) (14.11.2024).

BURDICK, Anne; DRUCKER, Johanna; LUNENFELD, Peter; PRESNER, Todd; JEFFREY, Schnapp. *Digital Humanities*. The MIT Press, 2012.

CHUMTONG, Jason; KALDEWEY, David. „Beyond the Google Ngram Viewer: bibliographic databases and journal archives as tools for the quantitative analysis of scientific and meta-scientific concepts”. *FIW Working paper*, No. 08, Forum Internationale Wissenschaft. Bonn, 2017.

CLEWING, Konrad; SCHMITT, Oliver Jens. *Geschichte Südosteuropas. Vom frühen Mittelalter bis zur Gegenwart*. Regensburg: Pustet Verlag, 2005.

DAVIS, Mark. „Making Google Books n-grams Useful for a Wide Range of Research on Language Change”. *International Journal of Corpus Linguistics* 19 (2014), br. 3: 401-416.

GOODING, Paul. „Mass digitization and the garbage dump: The conflicting needs of quantitative and qualitative methods”. *Literary and Linguistic Computing* 28 (2012), br. 3: 425-431.

GOOGLE NGRAM VIEWER. „What does the Ngram Viewer do?”. Pristup ostvaren 23. 10. 2024. <https://books.google.com/ngrams/info#>.

GOOGLE NGRAM VIEWER. Pristup ostvaren 11. 9. 2024. <https://books.google.com/ngrams/>.

HARARI, Yuval Noah. *Nexus: A Brief History of Information Networks from the Stone Age to AI*. New York: Fern Press, 2024.

HILPERT, Martin; GRIES, Stefan. „Assessing Frequency Changes in multistage Diachronic Corpora: Applications for Historical Corpus Linguistics and the Study of Language Acquisition”. *Literary and Linguistic Computing* 24 (2009), br. 4: 385-401.

JURIĆ, Tado. *Big (Crisis) Data in Social Sciences and Humanities: Predicting Crises*. Hamburg: Verlag Dr. Kovač, 2023.

JURIĆ, Tado. „Using digital humanities for understanding COVID-19: lessons from digital history about earlier coronavirus pandemic”. *Athens Journal of Technology & Engineering* 9 (2022), br. 2: 119-147.

JURIĆ, Tado. *Westbalkan-Erweiterung der EU. Europäisierungsprozess in Bosnien und Herzegowina, Serbien und Kroatien – ein Vergleich*. Hamburg: Verlag Dr. Kovač, 2012.

JURIĆ, Tado; BAKOTA, Anja; RUKAVINA, Bruno. „What is the Balkans?: Perception of the Balkans among Students in Croatia”. U: *ICONSR 2021. Proceedings & Abstracts Book*, ur. Kandemir Hamza, Özdaşlı Kürşat i Kala Ergin. Isparta: Association of Kutbilge Academicians, 2021, 75-98.

KARCH, Marziah. „How to Use the Ngram Viewer Tool in Google Books”. 2021. Pristup ostvaren 23. 10. 2024. <https://www.lifewire.com/google-books-ngram-viewer-1616701>.

KEKEZ, Hrvoje. „Urbarijalni nameti i katastarski popis kmetskih selišta na vlastelinstvu Brlog u 18. stoljeću”. *Povijesni prilozi* 25 (2006), br. 30: 79-102.

KESEBIR, Pelin; KESEBIR, Selin. „The Cultural Salience of Moral Character and Virtue Declined in Twentieth Century America”. *Journal of Positive Psychology* 7 (2012), br. 6: 471-480.

KOKALOVIĆ, Petra. „Digitalizacija knjižnične građe u specijalnim knjižnicama”. Završni rad, Sveučilište Josipa Jurja Strossmayera, 2023.

KOPLINIG, Alexander. „The impact of lacking metadata for the measurement of cultural and linguistic change using the Google Ngram data sets – Reconstructing the composition of the German corpus in times of WWII”. *Digital Scholarship in the Humanities* 32 (2015), br. 1: 169-188. <https://doi.org/10.1093/llc/fqv037>.

KUMAR, Nitu; SAHU, Manish. „The Evolution of Marketing History: a Peek Through Google Ngram Viewer”. *Asian Journal of Management Research* 1 (2010): 415-426.

LAKOFF, Robin. „What Words Don't Tell Us”. 2013. Pristup ostvaren 23. 11. 2024. <http://blogs.berkeley.edu/author/rlakoff/>.

LUCIER, Paul. „The Origins of Pure and Applied Science in Gilded Age America”. *ISIS* 103 (2012), br. 3: 527-536.

MANOVICH, Lev. „Cultural Analytics, Social Computing and Digital Humanities”. U: *The Datafied Society: Studying Culture through Data*, ur. Mirko Tobias Schäfer i Karin van Es. Amsterdam: Amsterdam University Press, 2017. <https://doi.org/10.25969/mediarep/12514>.

MICHEL, Jean-Baptiste. „Supporting Online Material for Quantitative Analysis of Culture Using Millions of Digitized Books”. 2012. Pristup ostvaren 23. 10. 2024. [www.sciencemag.org/cgi/content/full/science.1199644/DC1](http://www.sciencemag.org/cgi/content/full/science.1199644/DC1).

MICHEL, Jean-Baptiste; SHEN, Yuan Kui; PRESSER AIDEN, Aviva; VERRIS, Adrian; GRAY, Matthew K.; The Google Books Team; PICKETT, Joseph P.; HOIBERG, Dale; CLANCY, Dan; NORVIG, Peter; ORWANT, Jon; PIN-

KER, Steven; NOWAK, Martin A.; LIEBERMAN AIDEN, Erez. „Quantitative Analysis of Culture Using Millions of Digitized Books”. *Science* 331 (2011), br. 6014: 176-182. DOI: 10.1126/science.1199644.

NEWBERRY, Mitchell; AHERN, Christopher; CLARK, Robin; PLOTKIN, Joshua. „Detecting evolutionary forces in language change”. *Nature* 551 (2017), br. 7679: 223-226. DOI: 10.1038/nature24455. E-pub 2017 Nov 1. PMID: 29088703.

OISHI, Shigehiro; GRAHAM, Jesse; KESEBIR, Selin; GALINHA, Iolanda C. „Concepts of happiness across time and cultures”. *Personality and Social Psychology Bulletin* 39 (2013), br. 5: 559-577.

OZA, Preeti. „Digital Humanities – An Introduction”. U: *Multidimensionality of the Concept & Function of Digital Publisher*, ur. Pradeep Japee Gurdutta i Oza Preeti. Apple Books, 2020.

PECHENICK, Eitan Adam; DANFORTH, Christopher M.; DODDS, Peter Sheridan. „Characterizing the Google Books corpus: Strong limits to inferences of socio-cultural and linguistic evolution”. *PLOS One* (2015). <https://doi.org/10.1371/journal.pone.0137041>.

PETTIT, Michael. „Historical time in the age of big data: Cultural psychology, historical change, and the Google Books Ngram Viewer”. *History of Psychology* 19 (2016), br. 2: 141-153. DOI: 10.1037/hop0000023.

ROJAS CASTRO, Antonio. „Big Data in the Digital Humanities. New Conversations in the Global Academic Context”. *AC/E Digital Culture 2017*. (2017): 62-71.

SOLOVYEV, Vallery; BOCHKAREV, Vladimir; AKHTYAMOVA, Svetlana. „Google Books Ngram: Problems of Representativeness and Data Reliability”. U: *Data Analytics and Management in Data Intensive Domains*, ur. A. Elizarov, B. Novikov i S. Stupnikov. Communications in Computer and Information Science, vol. 1223. Springer, Cham. [https://doi.org/10.1007/978-3-030-51913-1\\_10](https://doi.org/10.1007/978-3-030-51913-1_10).

SUNDHAUSSEN, Holm. „Postjugoslawische Kriege (1991-95, 1998/99)”. U: *Lexikon zur Geschichte Südosteuropas*, ur. Konrad Clewing i Holm Sundhaussen. Wien: Böhlau, 2016.

ŠIMUNOVIĆ, Petar. *Hrvatska u prezimenima*. Zagreb: Golden marketing – Tehnička knjiga, 2008.

UNIVERSITY OF LONDON. „An introduction to text mining”. Pristup ostvaren 18. 9. 2024. <https://port.sas.ac.uk/mod/book/view.php?id=554&chapterid=331>.

YOUNES, Nadja; REIPS, Ulf Dietrich. „Guideline for improving the reliability of Google Ngram studies: Evidence from religious terms”. *PLOS One* 14 (2019), br. 3. <https://doi.org/10.1371/journal.pone.0213554>.

ZIEȚA, Anne. „Google Books Ngram Viewer in socio cultural research”. *Research in Language* 16 (2018), br. 3: 357-375. DOI: 10.2478/rela-2018-0015.

## SUMMARY

**A Digital Window into Croatia's Past: The Analytical Functionality of Google Ngram Viewer and Digital Humanities in Historical Analysis and History Teaching**

In this article, we examined the digital tool Google Ngram Viewer (GNV) through multiple analytical lenses – including geopolitics, cultural influences, wars, migration, and cultural production. By focusing on the relationship between Austria and Southeast Europe, with particular attention to Croatia, we demonstrated that GNV not only reveals quantitative shifts in discourse but also offers deeper insights into the dynamics of public interest across several centuries. At the same time, the paper functions as a practical guide for working with the method and illustrates its pedagogical applicability in history education.

Our analysis shows that GNV, when applied to the German-language corpus at the center of this research, reliably reflects patterns of public attention and mirrors authentic political or cultural events. We illustrated how the tool can uncover patterns and trends in the references to prominent historical figures, concepts, and phenomena – insights that often exceed the reach of traditional research methodologies. Most examples in the study have been designed as applied exercises, making GNV a relevant educational tool for teaching history (though not exclusively so). Beyond political history, GNV also enables engaging exploration in fields such as literature, the arts, linguistics, and economics.

However, it is essential to acknowledge the methodological constraints of this digital instrument. While numerous studies have confirmed its general reliability, critical voices in the literature have highlighted limitations related to corpus selection, the quality of OCR processing, and the interpretative scope of statistical outputs. As large-scale text analysis remains an emerging area, the standardization of procedures, reassessment of underlying assumptions, and further development of methodology all call for broader participation from the academic community.

Keywords: Big Data; Google Ngram Viewer; Austria; Croatia; Digital Humanities