Nacrt kurikuluma grčkog jezika temeljen na frekvencijskoj analizi korpusa

Soldo, Petar

Master's thesis / Diplomski rad

2021

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, University of Zagreb, Faculty of Humanities and Social Sciences / Sveučilište u Zagrebu, Filozofski fakultet**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:131:080980

SVEUČILIŠTE U ZAGREBU
FILOZOFSKI FAKULTET
ODSJEK ZA KLASIČNU FILOLOGIJU
SMJER: GRČKI JEZIK, NASTAVNIČKI
Ak. god. 2020./2021.

Petar Soldo

Nacrt kurikuluma grčkog jezika temeljen na frekvencijskoj analizi korpusa

Diplomski rad

Mentor: dr. sc. Neven Jovanović, red. prof.

Zagreb, siječanj 2021.

Izjava o akademskoj čestitosti

Izjavljujem i svojim potpisom potvrđujem da je ovaj rad rezultat mog vlastitog rada koji se temelji na istraživanjima te objavljenoj i citiranoj literaturi. Izjavljujem da nijedan dio rada nije napisan na nedozvoljen način, odnosno da je prepisan iz necitiranog rada, te da nijedan dio rada ne krši bilo čija autorska prava. Također izjavljujem da nijedan dio rada nije korišten za bilo koji drugi rad u bilo kojoj drugoj visokoškolskoj, znanstvenoj ili obrazovnoj ustanovi.

_____

(potpis)

Na početku, nekoliko zahvala.

Zahvaljujem svojem mentoru, prof. Nevenu Jovanoviću, na pomoći prilikom osmišljavanja i izrade ovog rada. Njemu i svim ostalim nastavnicima Odsjeka za klasičnu filologiju hvala i za podršku i nemalu strpljivost tijekom cijelog studija.

Hvala i profesorima Klasične gimnazije, posebno Korneliji Pavlić i Mislavu Gjurašinu, koji su me svojim znanjem i „pedagoškim amorom" uvjerili u klasičnu filologiju.

Hvala svim prijateljima koji su na raznovrsne načine pomagali u okončanju mog studija i ostvarenju ovog rada (posebno Luki za strpljivu pomoć oko statistike i Heleni za hrvatsku i englesku lekturu).

Na kraju, najveća zahvala mojoj obitelji na pomoći, podršci i razumijevanju tijekom cijelog obrazovanja. Siguran sam da nije bilo lako.

Sadržaj

1. Uvod

Grčki je jezik, uz latinski, jedan od najdugovječnijih nastavnih predmeta u našem obrazovnom sustavu. Njegov trag u školama možemo slijediti od osnutka Klasične gimnazije u Zagrebu, 1604. godine. Premda se radi o vrlo dugoj tradiciji poučavanja (a možda i zbog toga), metodika nastave grčkog jezika nije doživljavala drastičnije promjene. Razlozi su tome vjerojatno mnogi i raznorodni, a zaslužuju vlastito istraživanje, svakako izvan stranica ovog rada. U siječnju 2019. godine objavljen je novi kurikulum predmeta Grčki jezik (Ministarstvo znanosti i obrazovanja, 2019.), u sklopu opće obrazovne reforme naziva *Škola za život*. Taj dokument definira, među ostalim, ciljeve, domene, ishode i okvirne sadržaje kao upute za izvođenje nastave grčkog jezika.

Jedna od glavnih promjena koju novi kurikulum donosi u odnosu na stari nastavni plan i program jest razumijevanje teksta kao središnji cilj učenja grčkog jezika. Time je učenje gramatičkih oblika usmjereno na sposobnost čitanja i prevođenja tekstova. Međutim, iako se novom kurikularnom reformom naglašava veća samostalnost i autonomija učitelja u provođenju nastave, kurikulum ipak propisuje koji se sadržaji moraju obraditi na određenoj razini učenja. Usporedimo li raspored gramatičkog sadržaja po razinama učenja novog kurikuluma i starog nastavnog plana i programa, vidjet ćemo da se oni značajno ne razlikuju.

U ovom će radu biti izložen alternativan slijed obradbe jezičnog gradiva (gramatike i vokabulara) koji će se temeljiti na frekvencijskoj analizi korpusa. Ideja je da se pokuša osmisliti takav kurikulum u kojem će se najprije obraditi oni gramatički oblici i riječi koji su u jeziku frekventniji, tj. učestaliji. Za konstrukciju takvog rasporeda gradiva treba najprije utvrditi frekvenciju riječi u grčkom jeziku, a za to će nam poslužiti frekvencijska analiza korpusa. Stoga će se dio ovog rada baviti i predstavljanjem metoda i rezultata istraživanja čiji je cilj pronalazak najučestalijih gramatičkih oblika i riječi u grčkom. Zbog informacija dostupnih u korpusu, a i zbog opsega ovog rada, u analizu gramatičkih oblika ući će samo frekvencija morfoloških pojava, a sintaksu ostavljamo za neku drugu priliku. Temelj su tog istraživanja principi korpusne lingvistike, a baza za provođenje istraživanja je korpus stvoren i označen na Katedri za grčki jezik Odsjeka za klasičnu filologiju Filozofskog fakulteta u Zagrebu.

Uzimajući u obzir rezultate spomenutog istraživanja, predstavit ćemo, dakle, nacrt kurikuluma nastave grčkog jezika temeljen na frekvencijskoj analizi korpusa. Tako dobiven kurikulum trebao bi bolje pripremiti učenike za susret s izvornim grčkim tekstovima, što je jedna od

glavnih svrha učenja grčkog jezika te je, kako ćemo vidjeti, u skladu s glavnim ciljevima aktualnog kurikuluma, nastalog u sklopu nedavne kurikularne reforme.

Naš će kurikulum biti uspoređen s aktualnim, te ćemo locirati glavne metodičke teškoće koje bi frekvencijski pristup mogao donijeti i predložit ćemo, gdje to bude moguće, njihova potencijalna rješenja. Na samom ćemo kraju razmotriti daljnje mogućnosti koje donosi planiranje nastave potpomognuto analizom i obradbom korpusa.

Osim samim rezultatima, dio će ovog rada također biti posvećen nekim tehničkim i provedbenim problemima i izazovima koji su se pojavili prilikom obradbe korpusa. Iako su u fokusu rada rezultati istraživanja i kurikulum, pojedinosti vezane uz konkretne postupke u istraživanju važno je spomenuti iz nekoliko razloga. Osim što izlaganje takvih postupaka omogućava njihovu ponovljivost, olakšava i kritički osvrt na istraživanje ako se uoči neka greška. Osim toga, neki su specifikumi u strukturi podataka utjecali na način na koji su rezultati obrađeni i prikazani, a samim time su važni za naše shvaćanje i korištenje istih. Nadalje, koliko je autoru rada poznato, u nas ne postoje objavljeni radovi koji bi se specifično bavili ovakvom analizom ove vrste korpusa (tj. korpusa koji je sačinjen od lektirnih tekstova za početno učenje grčkog i koji je ručno anotiran) pa detaljnije dokumentiranje ovog istraživanja može doprinijeti javno dostupnom znanju. I na kraju, budući da je jedna od svrha ovog rada potaknuti druge da razmisle o ovakvom pristupu poučavanju jezika, autor se nada da će njegova iskustva moći bar u nekoj mjeri pomoći onima koji bi ovakva ili slična istraživanja provodili ubuduće.

2. Cilj i hipoteza

Cilj je ovog rada ispitati može li se nastava grčkog jezika organizirati tako da se jezični sadržaji obrađuju u skladu s intenzitetom kojim se pojavljuju u jeziku, odnosno tako da se najprije uče oni oblici koji su učestaliji. Ovdje se pod pojmom „jezični sadržaji" misli i na gramatičke (morfološke) oblike i na vokabular.

Pretpostavka je da raspored gradiva u trenutnom kurikulumu ne prati učestalost oblika u jeziku, već, nasljedujući klasične gramatike grčkog jezika (poput gramatike Musić i Majnarić, 1994.), oblike predstavlja redoslijedom koji prati svojevrsnu „tvorbenu logiku" i načelo jednostavnosti. To bi značilo da se oblici obrađuju tako da se najprije obrade svi oblici, primjerice, prezentske osnove, budući da se svi oni tvore dodavanjem različitih nastavaka na istu osnovu (tvorbena logika) ili da se najprije obrađuju imenice vokalskih deklinacija budući da je kod njih lakše uočiti podjelu na korijen i završetak (jednostavnost tvorbe). Ova dva pojma odgovarala bi otprilike onome što Poljak (1991.) navodi kao didaktičke principe sistematičnosti i postupnosti. Premda je takva podjela sigurno metodički opravdana, postavlja se pitanje predstavlja li takav raspored gradiva stvarnu situaciju u jeziku. Budući da se u aktualnom kurikulumu grčkog jezika naglašava kako je „gramatika u službi teksta, a ne tekst u službi gramatike" i navodi se kako je „temeljni cilj učenja Grčkog jezika razumijevanje teksta" (MZO, 2019.), svakako je vrijedno razmotriti plan nastave u kojem učenici čim ranije usvajaju (a samim time i dulje uvježbavaju) one oblike s kojima će se najviše susretati.

Osim gramatičkih oblika, u sklopu ovog istraživanja ispitat će se i koje su riječi najfrekventnije te predložiti osnovni vokabular (engl. *core vocabulary*) koji bi se mogao koristiti kao kriterij za odabir riječi koje će učenici morati znati i koje će im omogućiti čim lakše snalaženje u tekstu. Taj bi popis mogao služiti kao osnova za stvaranje nastavnog sadržaja (vježbi, tekstova, popisa riječi) i za provjeru usvojenosti vokabulara.

3. Metodologija

Budući da je cilj rada usporediti trenutni kurikulum s učestalošću riječi u jeziku, treba s jedne strane proučiti aktualni kurikulum, a s druge provesti frekvencijsku analizu na odabranom jezičnom uzorku.

Analiza kurikuluma provedena je proučavanjem i pregledavanjem službenih dokumenata koje je objavilo Ministarstvo znanosti i obrazovanja, a koji definiraju kurikulumske sadržaje za nastavu Grčkog jezika u osnovnim i srednjim školama. To su, napose, dokumenti *Kurikulum nastavnog predmeta Grčki jezik za osnovne škole i gimnazije* (MZO, 2019.) i *Godišnji izvedbeni kurikulum za šk. g. 2020./2021.* (MZO, 2020.). U njima možemo provjeriti koji se sadržaji i kojim redoslijedom obrađuju u trenutnoj nastavi Grčkog jezika. Uz kurikulume, u obzir su uzeti i aktualni srednjoškolski udžbenici za Grčki jezik.

Utvrđivanje učestalosti oblika postignuto je frekvencijskom analizom korpusa tekstova koji sačinjavaju *Čitanku za Grčku morfologiju 1* (Bratičević, Čengić, Jovanović, Rezar, Šoštarić i Zubović, 2019.), a koja je prvi put objavljena 2019. kao nastavni materijal za prvi semestar studija Grčkog jezika i književnosti na Filozofskom fakultetu Sveučilišta u Zagrebu. Tekstove spomenute čitanke morfološki su označili studenti grčkog jezika koristeći se okruženjem za anotaciju tekstova *Arethusa*[1]. Time je dobiven ručno analiziran korpus tekstova koji je objavljen kao repozitorij na platformi *GitHub* u obliku XML dokumenata.

Podatci su iz korpusa izvučeni koristeći upitni (engl. *query*) programski jezik *XQuery*[2], koji je namijenjen pretraživanju i stvaranju upita za XML datoteke, a obradba podataka provedena je uz pomoć skripti napisanih u programskom jeziku *Python*[3]. Za statističku obradbu i kao referenca za prikaz podataka korišten je priručnik *Statistics in Corpus Linguistics* (Brezina, 2018.).

Za provjeru valjanosti istraživanja korištena su i dva slična rada, od koji se jedan bavi samo kreiranjem osnovnog vokabulara za grčki jezik (Major, 2008.), a drugi samo učestalošću oblika u grčkom jeziku (Mahoney, 2004.). Za osmišljavanje i korekciju istraživanja bio je važan i rad Leecha (2001.) u kojem se raspravlja općenito o korištenju frekvencijskih korpusa u poučavanju stranih jezika.

[1] Za više o *Arethusi*, vidi https://www.perseids.org/tools/arethusa/app/#/ (pristupljeno 28. rujna 2020.)
[2] Za više o *XQueryju*, vidi https://www.w3schools.com/xml/xquery_intro.asp (pristupljeno 28. rujna 2020.)
[3] Za više o *Pythonu*, vidi https://www.w3schools.com/python/python_intro.asp (pristupljeno 28. rujna 2020.)

4. Trenutni kurikulum

Prije no što krenemo s istraživanjem, pokušat ćemo definirati riječ kurikulum. Cindrić, Miljković i Strugar (2010.) donose nekoliko različitih načina da se odredi pojam „kurikulum“ (lat. *curriculum*), ali svima im je zajedničko da kurikulum označuje organizaciju i planiranje nastave te da su elementi koji ga čine ciljevi, zadatci i sadržaji učenja, uvjeti, organizacija i tehnologija učenja i poučavanja te vrednovanje učenikovih postignuća. Nas će u ovom radu najviše zanimati sadržaji i njihov raspored u kurikulumu.

Valja najprije pogledati kojim redom aktualni kurikulum obrađuje pojedine oblike u grčkom jeziku. Kako bismo ovo ispitali, zavirit ćemo u dva dokumenta, odnosno u dva kurikuluma. Prvi od njih je *Kurikulum nastavnog predmeta Grčki jezik za osnovne škole i gimnazije* (MZO, 2019.), odnosno predmetni kurikulum za Grčki jezik. U njemu možemo pronaći općenito prezentiran plan izvođenja nastave. Sastoji se od sljedećih dijelova: svrhe i opisa nastavnog predmeta, ciljeva učenja i poučavanja i domena od koji je predmet sačinjen. Velik je dio kurikuluma posvećen i razradi odgojno-obrazovnih ishoda, sadržaja i razina usvojenosti po razredima i domenama. Uz to, dokument opisuje povezanost s drugim predmetima i međupredmetnim temama, ukratko govori o učenju i poučavanju predmeta te opisuje vrednovanje usvojenosti ishoda.

Dijelovi kurikuluma koji nas u ovom trenu najviše zanimaju su opisi domena te ishodi, sadržaji i razine usvojenosti po razredima. Domenama se opisuju „područja kojima se učenik kreće proučavajući, učeći i sve se više približavajući konačnim ciljevima“, a definirane su ukupno tri domene: Jezična pismenost, Iskustvo teksta i komunikacija te Civilizacija i baština (MZO, 2019.). Nas će u ovom radu u prvom redu zanimati domene Jezična pismenost te Iskustvo teksta i komunikacija. Za provjeru popisa jezičnih sadržaja pogledat ćemo koji su jezični sadržaji predviđeni za obradbu na pojedinoj razini učenja. O redoslijedu kojim se gradivo iz morfologije obrađuje, bit će riječi kasnije. Kurikulum predviđa tri programa: osnovne škole, klasične gimnazije (nastavljači) i klasične gimnazije (početnici). Budući da se nastavljački program klasične gimnazije nadovezuje na program za osnovne škole, pogledajmo ih zajedno.[4]

[4] Sadržaji za 3. i 4. razred nisu navedeni jer radi o „sistematizaciji sadržaja realiziranih prethodnih godina“.

RAZRED	SADRŽAJI ZA OSTVARIVANJE ODGOJNO-OBRAZOVNIH ISHODA
7. OŠ	I. i II. deklinacija, prezentska osnova (indikativ, imperfekt, infinitiv, imperativ u aktivu i mediopasivu glagoli na –ω i εἶναι) te osobne i posvojne zamjenice i αὐτός.
8. OŠ	III. deklinacija, particip prezentske osnove, oblici prezentske osnove /indikativ, imperfekt, infinitiv, imperativ u aktivu i mediopasivu/ stegnutih glagola i glagola na –μι.
1. SŠ	Ostale zamjenice, komparacija, futurska osnova / indikativ, infinitiv, particip / i aoristna osnova / indikativ, infinitiv, imperativ, particip / aktivna, medijalna i pasivna, nepravilna komparacija; daktilski heksametar[5].
2. SŠ	Složeniji morfološki oblici, participi i infinitivi svih vremena uz stalnu primjenu na izvornome tekstu

Tablica 1. Sadržaji za ostvarivanje odgojno-obrazovnih ishoda u domeni Jezična pismenost raspoređeni po godinama učenja za nastavljački program

Možemo pogledati i kako izgledaju jezični sadržaji za učenike koji grčki tek počinju učiti u klasičnoj gimnaziji. Primijetit ćemo da se sadržaji obrađuju sličnim redom, s malo promijenjenom raspodjelom po razredima.

RAZRED	SADRŽAJI ZA OSTVARIVANJE ODGOJNO-OBRAZOVNIH ISHODA
1. SŠ	I., II. i III. deklinacija, prezentska osnova (bez konj. i opt.) u aktivu i mediopasivu.
2. SŠ	Zamjenice i oblici futurske, aoristne i pasivne osnove (bez konj. i opt.).
3. SŠ	Elegijski distih, alkejska i sapfička strofa, komparacija pridjeva, brojevi.
4. SŠ	Sistematizacija sadržaja realiziranih prethodnih godina.

Tablica 2. Sadržaji za ostvarivanje odgojno-obrazovnih ishoda u domeni Jezična pismenost raspoređeni po godinama učenja za početnički program

Iz ovoga bismo već mogli zaključiti koji je redoslijed obradbe pojedinih oblika, međutim, bilo bi dobro kada bismo imali uvid u konkretniju realizaciju sadržaja. Na sreću, općeniti se predmetni kurikulum dodatno razrađuje u dokumentu koji se naziva Godišnjim izvedbenim

[5] Premda bi se moglo činiti neobično što se daktilski heksametar, kao vrsta stiha, nalazi u domeni Jezična pismenost, situacija zaista jest takva u aktualnom *Kurikulumu*. Razlog tomu može biti nedostatak prikladnije domene ili potreba za poznavanjem fonoloških i prozodijskih pravila grčkog jezika.

kurikulumom, a koji je za šk. god. 2020./2021. objavilo Ministarstvo znanosti i obrazovanja.[6] U njemu se sadržaji predviđeni kurikulumom raspisuju po tjednima nastavne godine u kojima ih se preporučuje obraditi. Time možemo dobiti uvid u konkretniji slijed obradbe gradiva. Ovaj put pogledat ćemo samo program za prva dva razreda početnog programa[7] gimnazijskog učenja grčkog jer 7. i .8 razredi OŠ, koji su nam ključni za analizu nastavljačkog programa, trenutačno nisu usuglašeni s obzirom na kurikularnu reformu, odnosno, u 7. razredima se primjenjuju novi kurikulumi iz 2019., a u 8. razredima se radi prema HNOS-u[8] iz 2006 (MZO, 2020.). Budući da nam je najbitniji redoslijed, radi preglednosti sadržaje ćemo navesti kao niz natuknica.

Ovim su redoslijedom po tjednima raspoređeni sadržaji iz Jezične pismenosti za prva dva razreda početničkog programa učenja grčkog u klasičnim gimnazijama prema *Godišnjem izvedbenom kurikulumu za Grčki jezik za šk. g. 2020./2021.* (MZO, 2020.):

1. razred:

    1. Grčki alfabet
    2. Fonologija
    3. Pravila čitanja i pisanja
    4. Naglašavanje u grčkom jeziku
    5. O deklinacija m. / ind. prez. akt.
    6. O deklinacija n. / glagol biti
    7. A deklinacija, dugo alfa
    8. A deklinacija, kratko alfa
    9. A deklinacija, masculina
    10. Imp. i inf. prez. akt.
    11. Pridjevi O i A deklinacije
    12. Mediopasiv
    13. Augment, impf. akt.

[6] Inače je dužnost svakog nastavnika da ovakav dokument izradi sam za sebe za svaki razredni odjel. Zbog specifične epidemiološke situacije, Ministarstvo je odlučilo za šk. god. 2020./2021. objaviti ove dokumente te oni služe samo kao orijentir i predložak, a nastavnicima je ostavljena autonomija izrade vlastitih GIK-ova.
[7] Analiziramo samo prva dva razreda, budući da se u njima obrađuju gramatički sadržaji.
[8] HNOS je kratica za Hrvatski nacionalni obrazovni standard.

14. Impf. medpas.
15. III. deklinacija (kons. osnove)
16. III. deklinacija (kons. osnove)
17. Verba contracta
18. III. deklinacija (kons. osnove)
19. Pridjevi III. deklinacije
20. Particip prezenta
21. Osobitosti konsonantskih osnova
22. Pridjevi s osobitostima
23. III. deklinacija, osobna imena
24. III. deklinacija, vokalske osnove
25. Pridjevi vokalskih osnova
26. Glagoli na –μι

2. razred:

27. Zamjenice
28. Zamjenice
29. Zamjenice
30. Glagolski sustav
31. Futur
32. Futur
33. Aorist
34. Aorist
35. Aorist
36. Aorist
37. Aorist

Sada s oba dokumenta na umu možemo pristupiti analizi redoslijeda obradbe gradiva.

Čini se kako je dio naše pretpostavke točan – trenutni kurikulum prati tvorbenu logiku i dijelom slijedi konvencije školske gramatike (Musić i Majnarić, 1994.: III-V). Primjerice, najprije se uče tzv. vokalske deklinacije (O-deklinacija, pa A-deklinacija), a tek se kasnije uvodi III. deklinacija. Po načelu jednostavnosti tvorbe ovakav je pristup opravdan, budući da se kod vokalskih deklinacija korijen i nastavak analiziraju lakše nego kod imenica III. deklinacije, a to omogućuje i veću predvidljivost u tvorbi. Samim će time učenici lakše svladati dekliniranje imenica O-/A-deklinacije.

Kod glagola se najprije obrađuju tematski glagoli i to tako da se uče prezent, potom imperativ i infinitiv prezenta, a zatim se uči imperfekt. Ovakav slijed je opet, iz tvorbene perspektive, logičan: prezent, imperativ i infinitiv prezenta te imperfekt tvore se od prezentske osnove (kod imperfekta uz dodavanje augmenta). Ovo opet olakšava učenicima da nauče gradivo, jer im je za tvorbu tih glagolskih oblika dovoljno da na polaznu rječničku (prezentsku) osnovu dodaju odgovarajuće nastavke. Taj razlog opravdava i učenje mediopasiva navedenih glagolskih oblika, budući da se za tvorbu mediopasiva kod ovih glagola koriste ista osnova i isto načelo dodavanja nastavaka kao i kod aktivnih glagola pa su predvidljivi na osnovi već usvojenog znanja.

Glagoli na –μι pojavljuju se tek kasnije, vjerojatno jer pravila njihove konjugacije otežavaju već spominjanu predvidljivost. Uzmemo li, primjerice, glagole s prezentskom reduplikacijom (τίθημι, δίδωμι, ἵημι i ἵστημι) kod njih, osim rastavljanja na osnovu i nastavak, moramo svladati i koncept redupliciranja te moramo paziti na produljenje vokala osnove u singularu prezenta i imperfekta (dakle radi se o pravilu koje uključuje ne samo glasovnu pojavu, nego i glagolske kategorije lica i broja).

Za pretpostaviti je da iz sličnih razloga futur i aorist dolaze kasnije i to jedan za drugim. Njihova tvorba zahtijeva svladavanje više pravila (i iznimaka) te pamćenje slučajeva u kojima se primjenjuju. S druge strane, to što se uče jedan za drugim može se opravdati time da je nakon učenja futurske osnove vrlo jednostavno tvoriti slabi aorist, koji se od aoristā obrađuje prvi. Tako primjerice razlikujemo sigmatski i kontraktni futur, pa treba pamtiti kako se koji futur konjugira i kod kojih ćemo ga glagola koristiti. Sličan nas problem čeka i kod aorista gdje treba usvojiti koji glagoli tvore slabi, a koji jaki aorist i, naravno, kako izgledaju njihovi oblici. Ne treba ni spominjati kakve tek probleme sa sobom nose pasivni oblici aorista i futura.

Redoslijed obradbe gradiva provjerit ćemo i u recentnim udžbenicima za početno učenje grčkog u klasičnim gimnazijama. Posegnemo li, dakle, za udžbenicima *Prometej 1 MYTHOS* (Martinić-Jerčić, Matković i Gjurašin, 2019.) i *Prometej 2 EPOS* (Martinić-Jerčić, Matković i

Gjurašin, 2020.) i pogledamo kojim su redom poredane vježbe iz gramatike u drugom dijelu udžbenika, vidjet ćemo da vjerno prate *Godišnji izvedbeni kurikulum*.

Na kraju možemo zaključiti – gramatičko se gradivo zaista obrađuje po načelima tvorbene logike i jednostavnosti tvorbe.

5. Pogled u korpus

Nakon što smo utvrdili kakvo je stanje u aktualnom kurikulumu, trebamo provjeriti učestalost oblika i riječi u jeziku. U idealnom bismo slučaju prebrojali svaku riječ i svaki oblik u svim sačuvanim grčkim tekstovima. Međutim, u ovom ćemo se radu ipak zadovoljiti analizom malenog uzorka tih tekstova, odnosno korpusa. Razloga za takav pristup je više, a navest ćemo ih malo kasnije. Prethodno treba reći što je uopće korpus te objasniti kako izgleda korpus kojim se mi bavimo.

5.1. Nešto o korpusima

Budući da smo pojam korpus spomenuli već nekoliko puta i budući da ćemo ga još spominjati, valjalo bi iznijeti i definiciju. McEnery i Wilson (2001., 73. str.) u *Uvodu u korpusnu lingvistiku* navode kako bi se korpusom mogla smatrati skupina tekstova konačne veličine koja predstavlja reprezentativan uzorak nekog jezika ili jezične varijante i koja je strojno čitljiva. Uz to, korpusi su nerijetko anotirani, odnosno označeni raznim dodatnim informacijama (o vrstama riječi, sintaktičkim ulogama i sl.). Korpusima se bavi područje lingvistike koje se naziva korpusna lingvistika. Njezin je razvoj započeo u drugoj polovici 20. st. (McEnery i Wilson, 2001., 20. str.).

5.2. Nešto više o našem korpusu

Već smo spomenuli kako krucijalnu ulogu u ovom istraživanju ima skup grčkih tekstova koji su ušli u *Čitanku za Grčku morfologiju 1* (Bratičević i dr.., 2019.). Skup se može dohvatiti na mrežnoj stranici *https://github.com/nevenjovanovic/grcka-morfologija.git* (zadnji put preuzeto 28. rujna 2020.). Pogledajmo zadovoljava li taj skup gore navedenu definiciju korpusa.

Najprije, to je zaista skupina tekstova konačne veličine, odnosno radi se 58 tekstova[9] koje su napisala 23 različita autora i čija veličina iznosi 9005 pojavnica[10] (engl. *token*), ne računajući pravopisne znakove. Budući da se radi o XML datotekama, smijemo tvrditi i da su tekstovi strojno čitljivi. Nadalje, u istim tim datotekama nalaze se podatci dobiveni gramatičkim tagiranjem[11] (engl. *part-of-speech tagging*) i lematiziranjem[12], odnosno možemo reći da je naš

[9] Pogledamo li u verziju *Čitanke* navedene u literaturi naići ćemo na 60 tekstova, jer su ulomci iz Lukijanova *Kinika* i Gorgijina *Helenae encomium* dodani nakon nastanka korpusa.

[10] „pojavnica – sve što se nalazi između dva znaka koja služe kao graničnici (svako individualno pojavljivanje); svaka pojava jezične jedinice u korpusu, na razini riječi svaki oblik uključen u leksem“ (prema http://ihjj.hr/mreznik/page/pojmovnik/6/, pristupljeno 28. rujna 2020.).

[11] „gramatičko tagiranje - pridruživanje oznake za vrstu riječi pojavnicama u korpusu“ (prema http://ihjj.hr/mreznik/page/pojmovnik/6/, pristupljeno 28. rujna 2020.).

[12] “lematiziranje - uspostava kanonskoga oblika pojavnice” (prema http://ihjj.hr/mreznik/page/pojmovnik/6/, pristupljeno 28. rujna 2020.).

korpus anotiran. Dapače, jedna je od prednosti našeg korpusa što je on anotiran ručno, čime smo izbjegli greške koje se pojavljuju prilikom strojnog anotiranja korpusa.[13]

Ostaje nam još vidjeti predstavlja li naš korpus reprezentativan uzorak jezika. Smjesta se pojavljuje pitanje što uopće znači da je korpus reprezentativan. Ovdje ćemo prihvatiti odgovor koji donosi Leech (2007. : str. 135.), citirajući Manning i Schutze (1999. str. 119.): uzorak je reprezentativan ako ono što otkrijemo o uzorku vrijedi i za općenitu populaciju. Prema Leechu, dakle, korpus je reprezentativan ako proučavanje čitavog jezika možemo zamijeniti proučavanjem korpusa. Iz tog aspekta naš korpus ima i prednosti i mana. S jedne strane, moramo ga smatrati vrlo malenim. Mnogi korpusi broje riječi u milijunima, a možemo naći i tvrdnje da se svaki korpus koji sadrži manje od 5.000.000 riječi smatra malenim korpusom (O'Keeffe, McCarthy i Carter, 2007., str. 4.). Ipak, na više mjesta (npr. Leech, 2007. str. 136. ili McEnery i Hardie, 2012.: 8. str.) susrećemo upozorenje da veličina korpusa nije jedina relevantna za njegovu reprezentativnost.

Važno mjesto ima i balansiranost korpusa, odnosno raznolik sastav korpusa koji bi ravnomjerno zastupio populaciju. Valja imati na umu da naš korpus čine tekstovi čija je namjena „upoznavanje studenata prve godine studija grčkog jezika i književnosti s autentičnom starogrčkom prozom“ (Bratičević i dr., 2019., str. 4.).

Pogledajmo koji su autori zastupljeni u našem korpusu. To su, abecednim redom: Ahilej Tatije, Antifont, Aristotel, Demosten, Diodor Sicilski, Epiktet, Ezop, Filostrat, Flavije Arijan, Gorgija, Heliodor, Herodot, Izokrat, Ksenofont, Lukijan, Lizija, Marko Aurelije, Platon, Plotin, Plutarh, Polibije, Pseudo-Ksenofont i Tukidid.

Neki od ovih autora tradicionalna su srednjoškolska lektira za učenike grčkog – Platon, Ksenofont, Herodot., Demosten ili Aristotel. U aktualnom kurikulumu još možemo naći Plutarha i Polibija. Ovo nam pokazuje da je korpus dijelom sačinjen od lektirnih autora, odnosno onih autora koji se obrađuju u školi. To nam, naravno odgovara, jer čini naš korpus još primjerenijim za naše istraživanje. Nadalje, mnogi autori u našem korpusu poznati su kao tipični predstavnici atičkog grčkog (uz neke koje smo naveli kao lektirne, to su i Antifont, Gorgija, Izokrat, Lukijan, Lizija, Tukidid), a i ostali autori koje nailazimo uglavnom ne odskaču

[13] Primjerice, problem s kojim se susrela Mahoney prilikom pretraživanja strojno anotiranog korpusa bio je što automatski gramatički tageri koje je koristila nisu mogli odabrati koji je od više mogućih gramatičkih opisa za određenu riječ ispravan, pa je morala pretpostaviti da su se svi ti oblici pojavljivali jednakom učestalošću (Mahoney, 2004., str. 101.), što je zasigurno utjecalo na rezultate.

značajno od atičkog. Ovo nam je važno jer je upravo taj dijalekt onaj koji se danas koristi kao standard za učenje grčkog.

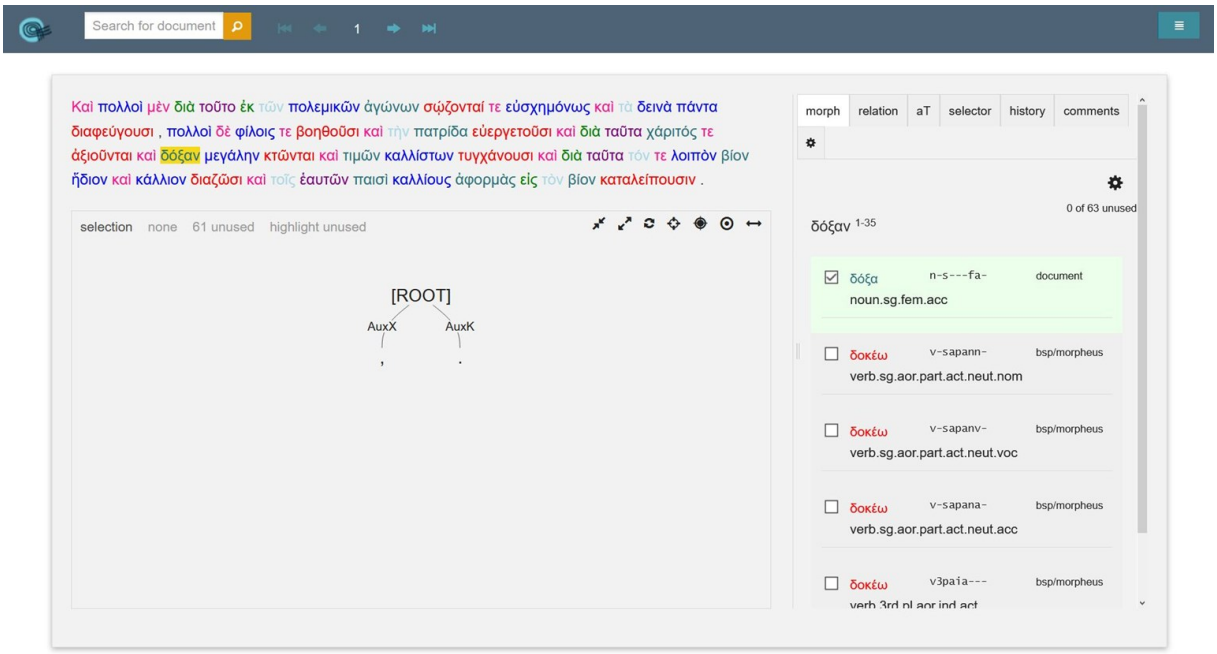
Vidimo, dakle, da je naš korpus odabran tako da većinski predstavlja onaj grčki koji nas najviše zanima – standardni atički grčki koji se uči u školi.[14] A upravo to – reprezentativnost s obzirom na onaj varijetet jezika koji nas zanima – navode McEnery i Wilson (2001.) kao jednu od važnih odlika reprezentativnog korpusa.

Dodatnu sigurnost u točnost naših rezultata može nam pružiti usporedba s rezultatima sličnih istraživanja, kakva su proveli npr. Mahoney (2004.) ili Major (2008.). Oba su se istraživanja koristila pretraživanjem daleko većeg korpusa *Perseus*. Osim broja pojavnica, naše se istraživanje u odnosu na njihovo razlikuje i u načinu analize riječi. U našem su korpusu, kao što smo već spominjali, riječi ručno anotirane, dok je taj posao u dva spomenuta istraživanja odrađen računalno, koristeći program *Morpheus* (Mahoney), odnosno alate korpusa *Perseus* (Major). Više o tim istraživanjima bit će riječi nešto kasnije.

5.3. Analiza korpusa i izvlačenje podataka

Sljedeći je korak pogledati što nam zapravo govori naš korpus. Ono što nas najviše zanima je statistička obradba anotacija, odnosno jezičnih informacija dobivenih označavanjem korpusa. Pogledajmo najprije kako je naš korpus anotiran. Skupina studenata studija Grčkog jezika i književnosti na FFZG-u u raznim je periodima od 2018. do 2019. ručno anotirala tekstove koje su pripremili nastavnici istog studija. Anotacija je obavljena pomoću alata *Arethusa* na platformi *Perseids*. Svaki je student dobio neku količinu rečenica koje je trebao lematizirati i na njima obaviti gramatičko tagiranje. Prilikom uvoza teksta u aplikaciju, moguće je birati tzv. *tagset*, odnosno standard prema kojem će se tekstovi gramatički tagirati. U svim tekstovima u našem korpusu koristi se tzv. *ALDT* (što je kratica za Ancient Language Dependency Treebank). Odabir *tagseta* bitniji je za sintaktičko anotiranje nego za morfološko, pa nam ovom prilikom nije od ključnog značaja. *Arethusino* sučelje omogućava da korisnik samostalno odredi lemu i gramatički tag ili da to za njega obavi *Arethusa*, a da korisnik rezultate samo provjeri.

[14] Iznimku možemo pronaći u Herodotu, koji piše jezikom baziranim na jonskom dijalektu koji se često naziva „šarena jonština“. Herodot je prilično zastupljen i u aktualnom kurikulumu (MZO, 2020.), zbog njegove općepovijesne važnosti, ali i važnosti za grčku književnost. Budući da je jonski dijalekt bitan i za neka druga važna književna djela (npr. Homerove epove), a s obzirom na to da je vrlo blizak atičkom, Herodotovo mjesto u kurikulumu i u našem korpusu je sasvim razumljivo.



Slika 1. Primjer okruženja "Arethusa"

5.3.1. Zapis korpusa

Rezultati anotiranja spremaju se u obliku XML[15] dokumenta. Dobiveni dokument u stvari tokenizira tekst i to tako da svaki token dobije svoj XML element. Nadalje, anotacije se za svaku riječ bilježe kao vrijednost atributa uz element riječi, pa se tako leme bilježe kao vrijednosti atributa `@lemma`, a gramatički tagovi kao vrijednosti atributa `@postag`.

Ovako izgleda primjer zapisivanja jedne riječi i njenih anotacija:

```
<word id="10" form="οἰκουμένη" lemma="οἰκέω"
postag="v-sppefn-" relation="" head=""/>
```

Premda se leme u *Arethusi* mogu unositi samostalno, ipak se gotovo uvijek odabire jedna od onih koje aplikacija sama ponudi. Tada leme odgovaraju rječničkim natuknicama koje možemo naći u rječniku *Liddle-Scott-Jones* (*LSJ*). Ovo će biti posebno važno u slučajevima u kojima je moguće odabrati između više lema koje su naizgled iste, ali je brojkom označeno da se odnose na drugačije rječničke natuknice. Primjerice, *Arethusa* nam može ponuditi da nekom obliku pridružimo leme *λέγω1*, *λέγω2* ili *λέγω3*, a naš će odabir ovisiti o tome koju od rječničkih natuknica iz *LSJ-a* smatramo najprikladnijom.

Što se tiče gramatičkog tagiranja, on se zapisuje u obliku „šifre“ koja se sastoji od 9 mjesta. Svako mjesto označuje određenu morfološku kategoriju (npr. vrsta riječi, lice, broj, rod, padež

[15] Za više o XML-u vidi https://www.w3schools.com/xml/default.asp (pristupljeno 5. listopada 2020.).

itd.), a informacija o vrijednosti morfološke kategorije (npr. glagol, 3. l., jednina itd.) zapisuje se određenim simbolom. Precizniji opis ovog sustava nalazi se u Prilogu 1.

Kako bismo lakše obradili podatke koji su zapisani na ovaj način, treba ih najprije prikazati u nekom preglednijem formatu. Pomoću upitnog jezika *XQuery* možemo napraviti upit koji će izlistati sve leme i gramatičke tagove iz našeg korpusa. Skripte za obavljanje tog zadatka nalaze se u Prilogu 2.

5.3.2. Statistički temelji analize

Prije nego prikažemo obradbu gramatičkih tagova i lema i njene rezultate, ukratko ćemo prikazati koje su statističke mjere korištene prilikom analize korpusa. Kao glavna referenca za statistiku korišten je priručnik *Statistics in Corpus Linguistics* Vaclava Brezine (2018.). Temeljni je pojam frekvencija i odnosi se na učestalost neke pojave u korpusu. Razlikujemo apsolutnu frekvenciju (AF) koja prikazuje koliko se puta neka riječ ili pojava pojavljuje u korpusu i relativnu frekvenciju (RF) koja predstavlja omjer apsolutne frekvencije i broja tokena u korpusu (odnosno udio neke riječi ili pojave u svim tokenima); taj omjer može biti i pomnožen nekom bazom (primjerice, bazom 10) radi normalizacije, odnosno kako bi brojke bile preglednije. Relativna frekvencija izračunava se tako da podijelimo apsolutnu frekvenciju s brojem tokena. U našem radu prikazivat ćemo relativnu frekvenciju kao postotak, odnosno množit ćemo s bazom 100.

Osim što nas zanima koliko se često neka riječ pojavljuje u korpusu, zanima nas i je li ta riječ podjednako zastupljena u svim dijelovima korpusa ili je koncentrirana u pojedinim dijelovima. Drugim riječima, zanima nas disperzija neke riječi ili pojave u korpusu. Disperzija označuje kako su riječ ili obilježje raspoređeni kroz korpus te je, uz frekvenciju, jedan od važnih faktora u analizi korpusa za izradu kurikuluma (Leech, 2001., str. 3). U našem slučaju, dijelove korpusa kroz koje mjerimo disperziju predstavlja 58 individualnih tekstova od kojih je korpus sačinjen. Naime, nije nam svejedno pojavljuje li se neka riječ velik broj puta u svega nekoliko tekstova ili se pojavljuje u gotovo svim tekstovima podjednako. Disperziju možemo izraziti različitim statističkim mjerama. U ovom su radu korištene mjere standardna devijacija, koeficijent varijacije i koeficijent varijacije prikazan kao postotak. Standardna devijacija (SD) pokazuje koliko pojedini podatci u našem skupu odstupaju od njihove aritmetičke sredine. Koeficijent varijacije omjer je standardne devijacije i srednje vrijednosti. Koeficijent varijacije prikazan kao postotak omjer je izmjerenog koeficijenta varijacije i maksimalnog koeficijenta varijacije. Ovo je mjera koju ćemo prikazivati u našim rezultatima. Ona prikazuje koliku je disperziju neka riječ ili pojava postigla u odnosu na maksimalnu moguću disperziju. Maksimalna je

vrijednost u ovoj mjeri 100 % i događa se kada se sve pojavnice koje mjerimo nalaze u samo jednom dijelu korpusa, odnosno (u našem slučaju) samo jednom tekstu.

U izračunavanju statistika koristile su se matematičke formule koje prikazuje Brezina (2018.). Sve su one navedene u Prilogu 3. Za izračun podataka korišten je program *Office 365 Excel.*

Unatoč tome što smo već spominjali koju ulogu u reprezentativnosti rezultata igra veličina korpusa, ovdje još jednom treba naglasiti da je naš korpus zbog samo 9005 pojavnica iz perspektive statistike limitirano reprezentativan.

5.3.3. Obradba gramatičkih tagova i rezultati

Budući da nas zanima učestalost gramatičkih oblika u tekstu, trebamo analizirati dobiveni popis tagova. To možemo napraviti tako da napišemo skriptu koja će proći kroz već spomenute šifre i čitajući vrijednosti na određenim mjestima prebrojati koliko se koji oblik pojavljuje. Tako možemo ispitati koliko u tekstu ima npr. veznika, akuzativa, imperfekta, participa itd.

Primjerice, ako uzmemo gore naveden primjer riječi iz našeg korpusa čija pojavnica glasi οἰκουμένη, vidjet ćemo da šifra njenog gramatičkog taga glasi `"v-sppefn-"`. Naša će skripta napraviti sljedeće: najprije će provjeriti vrijednost prvog mjesta u šifri koje sadrži informaciju o vrsti riječi, prepoznati da se radi o slovu „v“ i zaključiti da se radi o glagolu. Zatim će pročitati drugo mjesto. Budući da se radi o participu, drugo mjesto u šifri, koje označuje lice, nema vrijednost. Naš program čita dalje i na trećem mjestu nalazi vrijednost „s“ po čemu zna da se radi o jednini, odnosno singularu. Na četvrtom mjestu nalazi znak „p“ pa oblik ubraja među prezente. Dalje, na petom mjestu opet nalazi vrijednost „p“, a budući da se radi o mjestu koje označuje glagolski način, zna da se radi o participu. Šesto mjesto sadrži informaciju o glagolskom stanju i nosi vrijednost „e“, što označuje da je riječ o mediopasivu. Sedmo mjesto nam govori o gramatičkom rodu, ima vrijednost „f“ i znamo da se radi o ženskom rodu. I na osmom mjestu čitamo vrijednost „n“, znamo da se radi o mjestu koje čuva vrijednost padeža pa zaključujemo da se radi o nominativu. Zadnje mjesto rezervirano je za izražavanje stupnja pridjeva, ali u našim anotacijama ta informacija nije navedena. Za zadatak iteriranja kroz velik broj ovakvih šifri i provjeru vrijednosti određenog mjesta najprikladniji je programski jezik *Python*. Prebrojavanja gramatičkih tagova možemo obaviti uz pomoć skripte koja se nalazi u Prilogu 4.

U tablicama su prikazani rezultati našeg pretraživanja. U prebrojavanju tagova izostavljeni su interpunkcijski znakovi jer nisu od značaja za naše istraživanje. Statističke mjere navedene u tablici objašnjene su u zasebnom dijelu rada.

Vrsta riječi	Apsolutna frekvencija	Relativna frekvencija	Koeficijent varijacije (%)
Imenice	1498	16,63 %	3,57 %
Zamjenice	763	8,47 %	5,15 %
Pridjevi	836	9,28 %	5,20 %
Brojevi	18	0,20 %	34,16 %
Glagoli	1836	20,38 %	2,68 %
Članovi [16]	1149	12,76 %	4,00 %
Prilozi	1222	13,57 %	5,07 %
Prijedlozi	521	5,78 %	5,24 %
Veznici	1138	12,63 %	4,10 %
Usklici	20	0,22 %	35,52 %
Nepravilni oblici	6	0,07 %	40,56 %
Ukupno [17]	9007	100 %	/

Tablica 3. Frekvencija vrsta riječi

[16] Pogledamo li u tablicu s lemama vidjet ćemo da su tamo prebrojana dva člana više, zbog toga što su dva člana u tekstu označena kao zamjenice, vjerojatno pokazujući izvorno deiktičko značenje (Musić i Majnarić, 1994.: str. 174.).

[17] Razlika između 9007 tagova i 9005 lema objašnjiva je tagiranjem dvaju pojavnica interpunkcijskih znakova (koje smo izostavili iz brojanja tagova i lema) kao usklika. Tako su oni ubrojeni u tagove, ali ne i u leme.

Padež	Apsolutna frekvencija	Udio među imenskim riječima	Koeficijent varijacije (%)
Nominativ	1339	27,49 %	5,57 %
Genitiv	1128	23,16 %	6,16 %
Dativ	652	13,39 %	7,28 %
Akuzativ	1730	35,52 %	4,23 %
Vokativ	22	0,45 %	26,75 %

Tablica 4. Frekvencija padeža u svim riječima koje imaju tu kategoriju

Vrijeme	Apsolutna frekvencija	Udio među vremenima	Koeficijent varijacije (%)
Prezent	959	52,23 %	5,79 %
Imperfekt	177	9,64 %	13,43 %
Aorist	512	27,89 %	9,35 %
Futur	78	4,25 %	20,22 %
Perfekt	99	5,39 %	11,34 %
Pluskvamperfekt	11	0,60 %	51,21 %
Futur egzaktni	0	0,00 %	/

Tablica 5. Frekvencija vremena u svim glagolskim oblicima

Način	Apsolutna frekvencija	Udio među načinima	Koeficijent varijacije (%)
Indikativ	746	40,63 %	4,49 %
Konjunktiv	75	4,08 %	19,66 %
Optativ	59	3,21 %	16,95 %
Imperativ	25	1,36 %	45,95 %
Infinitiv	323	17,59 %	8,08 %
Particip	608	33,12 %	5,92 %

Tablica 6. Frekvencija glagolskih načina

Stanje	Apsolutna frekvencija	Udio među stanjima	Koeficijent varijacije (%)
Aktiv	1246	67,86 %	3,34 %
Medij	157	8,55 %	10,67 %
Mediopasiv	371	20,21 %	6,44 %
Pasiv	61	3,32 %	17,58 %
Deponentni	1	0,05 %	100,00 %

Tablica 7. Frekvencija glagolskih stanja svih glagolskih oblika

Particip	Apsolutna frekvencija	Udio među participima	Koeficijent varijacije (%)
Particip prezenta	364	59,87 %	7,66 %
Particip aorista	181	29,77 %	13,22 %
Particip futura	10	1,64 %	38,14 %
Particip perfekta	53	8,72 %	14,52 %

Tablica 8. Frekvencija participa po vremenima

Infinitiv	Apsolutna frekvencija	Udio među infinitivima	Koeficijent varijacije (%)
Infinitiv prezenta	209	64,71 %	11,58 %
Infinitiv aorista	94	29,10 %	14,74 %
Infinitiv futura	10	3,10 %	31,41 %
Infinitiv perfekta	10	3,10 %	34,06 %

Tablica 9. Frekvencija infinitiva po vremenima

Valja iznijeti nekoliko napomena.

U Tablici 3. u kategoriju Nepravilni oblici (engl. *Irregular*) ubraja se šest pojava upitne zamjenice τίς.[18] Sama *Arethusa* tu zamjenicu automatski prepoznaje i svrstava u navedenu kategoriju. Razlog tomu autoru ovog rada nije sasvim jasan, ali moguće je da je to zbog njena nepravilna naglaska, odnosno zbog akuta koji nikad ne prelazi u gravis (Musić i Majnarić, 1994.: str. 59.).

U Tablici 4. u obzir su uzete sve riječi koje imaju kategoriju padeža, dakle imenske riječi (imenice, pridjevi, zamjenice i brojevi), uključujući participe, i članovi.

Treba nešto razjasniti i u Tablici 6., u kojoj su među glagolskim načinima navedeni particip i infinitiv. Njih inače ne bismo smatrali glagolskim načinima, već nefinitnim glagolskim oblicima (Emde Boas, Rijskbaron, Huitnik i Bakker, 2019.). Ipak, budući da se u sustavu *Arethusa* participi i infinitivi zapisuju na isto mjesto u šifri kao i načini (v. str. 14. i Prilog 1), ovdje ih sve navodimo u istoj tablici.

U Tablici 7., koja prikazuje glagolska stanja, samo je jedan glagol označen kao deponentan. To, međutim, ne oslikava stvarnu sliku korpusa, već se radi o svojevrsnoj anomaliji u označavanju riječi. Naime, svi drugi primjeri deponentnih glagola označeni su kao mediopasivni (jer se u označavanju gleda oblik) pa sve deponentne treba pribrojiti mediopasivnima. To također znači da nisu svi mediopasivi pravi mediopasivi, već da stanovit dio njih predstavljaju deponentni glagoli.

Tablice 8. i 9. primjer su detaljnije analize korpusa. Mogli smo napraviti analizu učestalosti nekog oblika s obzirom na vrijednost svih kategorija koje on ima (npr. mogli smo istražiti koliko imamo glagola u konjunktivu trećeg lica jednine aorista pasivnog), ali ti nam podatci za naše istraživanje ne bi bili pretjerano korisni. Participi i infinitivi izabrani su kao ilustracija jer često imaju istaknutu ulogu u učenju grčkog zbog manjka pandana u hrvatskom jeziku.

5.3.4. Obradba lema i rezultati

Obradba lema nešto je kompleksnija od obradbe gramatičkih tagova. Nakon što smo dobili popis svih lema u našem korpusu, najprije je trebalo izdvojiti jedinstvene vrijednosti, odnosno

[18] To je ujedno i uzrok uočljivog koeficijenta varijacije od 100%.

pogledati koje se jedinstvene leme pojavljuju u našem korpusu. Ovo je moguće obaviti na više načina, a u našem slučaju i ovaj ćemo zadatak obaviti pomoću *Pythona*.

Skripta najprije pročita popis svih lema koje se pojavljuju u korpusu, onoliko puta koliko se pojavljuju. Koristimo prednosti dva podatkovna tipa koja nalazimo u *Pythonu*, a to su lista i skup. Liste su uređene kolekcije i možemo brzo i lako dohvatiti njene elemente (Kalafatić, 2012. str. 91), a skup je neuređena kolekcija čija nam je trenutno najzanimljivija karakteristika jedinstvenost elemenata, tj. dodavanjem elemenata koji su već u skupu ne mijenja se sadržaj skupa (Kalafatić, 2012. str. 103.). Naš program uzima popis lema i od njega stvara listu, a zatim od liste i skup. Budući da se u dobivenom skupu nalaze samo jedinstvene vrijednosti, možemo ga koristiti da usporedimo skup i listu, odnosno da svaki element skupa iteriramo kroz listu. Tako možemo prebrojati koliko se puta svaka lema pojavljuje u korpusu. Skripta koja je korištena u ovom radu kako bismo obradili leme nalazi se u Prilogu 5.

U našem se tekstu pojavljuje ukupno 2005 lema u 9005[19] pojavnica.

Leme smo izdvajali kako bismo mogli vidjeti koje se od njih najčešće pojavljuju u korpusu. Pokušaj izrade popisa najčešćih lema nalazi se u Tablici 10. Prikazano je 108 najčešćih riječi – namjera je bila navesti njih 100, ali je odlučeno da će se uvrstiti sve riječi koje se u tekstu pojavljuju najmanje 11 puta pa se broj popeo na 108. Takav popis uključuje otprilike 5,5 % jedinstvenih lema u korpusu.

[19] O razlici između 9005 pojavnica kod lema i 9007 kod tagova, vidi fusnotu 19.

	Lema	Apsolutna frekvencija	Relativna frekvencija	Koeficijent varijacije (%)
1.	ὁ	1152	12,79 %	3,94 %
2.	καί	565	6,27 %	4,82 %
3.	δέ	339	3,76 %	6,24 %
4.	εἰμί	176	1,95 %	9,85 %
5.	αὐτός	168	1,87 %	10,30 %
6.	μέν	147	1,63 %	9,46 %
7.	οὐ	142	1,58 %	14,26 %
8.	οὗτος	113	1,25 %	10,93 %
9.	τε	106	1,18 %	16,70 %
10.	γάρ	92	1,02 %	11,16 %
11.	ὅς	77	0,86 %	14,05 %
12.	ἐν	70	0,78 %	16,23 %
13.	πρός	69	0,77 %	14,38 %
14.	ἐγώ	67	0,74 %	22,09 %
15.	τις	65	0,72 %	13,89 %
16.	εἰς	62	0,69 %	12,74 %
17.	ἐπί	61	0,68 %	18,28 %
18.	ὡς	56	0,62 %	14,96 %
19.	ἀλλά	54	0,60 %	14,13 %
20.	γίγνομαι	53	0,59 %	15,32 %
21.	ἤ1[20]	52	0,58 %	22,55 %
22.	ἄλλος	50	0,56 %	17,26 %
23.	ἄν1	50	0,56 %	20,91 %
24.	πᾶς	47	0,52 %	18,45 %
25.	ἔχω	46	0,51 %	17,12 %
26.	πολύς	45	0,50 %	20,48 %
27.	εἰ	44	0,49 %	20,42 %
28.	ὅτι2[21]	44	0,49 %	17,46 %
29.	διά	44	0,49 %	20,78 %

[20] ἤ1 je disjunktivna čestica, značenja „ili“.
[21] ὅτι2 je veznik koji uvodi subjektne i objektne ili uzročne rečenice, prevodimo ga s „da“ ili „jer“.

30.	ἐκ	43	0,48 %	17,56 %
31.	περί	41	0,46 %	17,65 %
32.	ἐκεῖνος	40	0,44 %	18,19 %
33.	λέγω3[22]	39	0,43 %	32,42 %
34.	ποιέω	37	0,41 %	20,08 %
35.	ἑαυτοῦ	34	0,38 %	28,21 %
36.	μή	32	0,36 %	27,03 %
37.	μέγας	32	0,36 %	19,98 %
38.	κατά	30	0,33 %	22,86 %
39.	λόγος	30	0,33 %	22,34 %
40.	οὐδείς	30	0,33 %	23,47 %
41.	σύ	29	0,32 %	27,10 %
42.	φημί	29	0,32 %	25,09 %
43.	ἄνθρωπος	28	0,31 %	29,91 %
44.	ὁράω	27	0,30 %	24,33 %
45.	φίλος	25	0,28 %	48,98 %
46.	ὅσος	25	0,28 %	34,22 %
47.	ἀγαθός	24	0,27 %	24,53 %
48.	οὖν	23	0,26 %	20,34 %
49.	ὑπό	23	0,26 %	25,19 %
50.	ἐάν	22	0,24 %	26,29 %
51.	θεός	22	0,24 %	25,69 %
52.	λαμβάνω	21	0,23 %	30,68 %
53.	δή	20	0,22 %	23,09 %
54.	παρά	20	0,22 %	26,30 %
55.	νῦν	19	0,21 %	27,47 %
56.	πόλις	18	0,20 %	23,72 %
57.	ἀνήρ	18	0,20 %	30,76 %
58.	μετά	18	0,20 %	28,14 %
59.	σῶμα	17	0,19 %	47,53 %
60.	οὐδέ	17	0,19 %	27,64 %
61.	οὕτως	16	0,18 %	31,80 %
62.	δύναμαι	16	0,18 %	28,09 %

[22] λέγω3 je glagol u značenje „govoriti“.

63.	εἷς	15	0,17 %	38,31 %
64.	οἷος	15	0,17 %	33,46 %
65.	τοιοῦτος	15	0,17 %	34,75 %
66.	ἔτι	15	0,17 %	28,61 %
67.	ὦ	15	0,17 %	28,18 %
68.	καλός	15	0,17 %	35,62 %
69.	βίος	14	0,16 %	32,23 %
70.	γῆ	14	0,16 %	36,11 %
71.	δοκέω	14	0,16 %	23,76 %
72.	μόνος	14	0,16 %	28,10 %
73.	μᾶλλον	14	0,16 %	27,39 %
74.	πράσσω	14	0,16 %	28,75 %
75.	φαίνω	14	0,16 %	27,69 %
76.	χρόνος	14	0,16 %	26,38 %
77.	ἁλώπηξ	14	0,16 %	52,71 %
78.	ἀρχή	14	0,16 %	27,82 %
79.	ἅπας	14	0,16 %	31,65 %
80.	γε	13	0,14 %	25,58 %
81.	μηδείς	13	0,14 %	36,15 %
82.	οἶδα	13	0,14 %	31,33 %
83.	πάρειμι1[23]	13	0,14 %	36,32 %
84.	πρᾶγμα	13	0,14 %	31,44 %
85.	φύσις	13	0,14 %	49,92 %
86.	ἐθέλω	13	0,14 %	34,47 %
87.	ἕκαστος	13	0,14 %	31,25 %
88.	ἕτερος	13	0,14 %	39,52 %
89.	ὥστε	13	0,14 %	31,92 %
90.	Ἀθηναῖος	13	0,14 %	44,07 %
91.	βούλομαι	12	0,13 %	33,53 %
92.	δεῖ	12	0,13 %	35,36 %
93.	κακός	12	0,13 %	57,51 %
94.	πρῶτος	12	0,13 %	28,88 %
95.	ἅμα	12	0,13 %	33,76 %

[23] πάρειμι1 je glagol u značenju „biti nazočan“.

96.	ὥσπερ	12	0,13 %	29,71 %
97.	ἀπό	12	0,13 %	26,49 %
98.	θάνατος	11	0,12 %	34,89 %
99.	μηδέ	11	0,12 %	55,83 %
100.	παῖς	11	0,12 %	47,50 %
101.	πόλεμος	11	0,12 %	36,61 %
102.	τότε	11	0,12 %	34,24 %
103.	ἀλλήλων	11	0,12 %	39,18 %
104.	ἔργον	11	0,12 %	34,23 %
105.	ἔρως	11	0,12 %	52,19 %
106.	ἡγέομαι	11	0,12 %	29,61 %
107.	ἤδη	11	0,12 %	33,22 %
108.	εἶπον	11	0,12 %	42,94 %
	UKUPNO	5513	61,22 %	/

Tablica 10. Popis 109 najčešćih lema u korpusu poredanih po frekvenciji

I ovdje trebamo dati par napomena. Jedan od problema prilikom obradbe lema bilo je pitanje kodiranja znakova. Naime, prilikom stvaranja korpusa na nekim su mjestima identični grafemi kodirani drugačijim *Unicode* šiframa. Ti grafemi naoko izgledaju isto, ali ih računalo prepoznaje kao dva različita znaka. Primjerice, riječi ἀλλά i ἀλλά čine nam se identičnima, ali slovo ά u prvoj riječi ima kôd U+03AC, dok slovo ά u drugoj riječi ima kôd U+1F71. Prvo alfa u kodnoj tablici *Unicode* nosi ime *Greek Small Letter Alpha with Tonos*, a drugo *Greek Small Letter Alpha with Oxia*. Ovo stvara probleme kada želimo strojno prebrojati leme, jer će računalo naše dvije gore spomenute riječi računati kao dva različita znakovna niza i neće prepoznati da se radi o istoj riječi. Kako bi se taj problem riješio, pretraživanjem i zamjenom provedena je normalizacija simbola.[24] Postupak je proveden za sve leme u kojima je primijećena ova pojava. Time smo omogućili izračunavanje stvarnog broja lema.

Još jedan problem bile su leme s gravisom. Naime, nekim su pojavnicama pridružene leme koje umjesto akuta imaju gravis. Primjerice, među lemama možemo naići na oblik ἀλλὰ, premda je standardan način zapisivanja oksitona ἀλλά. Budući da se radi o grešci prilikom anotiranja i ove smo riječi sveli na isti oblik kako bismo ih mogli pravilno izbrojati.

[24]Za više o ovom problemu i mogućim automatiziranim rješenjima vidi https://docs.cltk.org/en/latest/greek.html#normalization (pristupljeno 13. listopada 2020.).

Uz navedene, postojao je i problem krivo shvaćenih lema. Naime, već smo spomenuli da neke leme imaju više značenja koja se označuju brojkama Primjerice, glagolu λέγω mogu se pridružiti leme λέγω1 (u značenju „skupljati, ubrati“), λέγω2 (u značenju „brojiti“) ili λέγω3 (u značenju „govoriti“). Budući da ovo nije jasno vidljivo prilikom anotacije, lako je moguće da dođe do zabune oko ispravne leme. Tako ćemo pretraživanjem našeg korpusa saznati da postoji 21 lema λέγω1, 17 lema λέγω3 i 1 lema λέγω2. Takvo bi nas stanje moglo začuditi, budući da valja očekivati da će glagol „govoriti“ biti češći od glagola „skupljati“. Ručnim pregledom svih pojavnica kojima su pridodane ove tri leme ustanovljeno je da je zaista u svim slučajevima njihova pojavljivanja najprikladnije pridružiti lemu λέγω3. Stoga smo prilikom prebrojavanja sve tri leme zbrojili pod lemu λέγω3. Isti smo postupak proveli i za leme ἄν1 i ἄν2, svevši ih obje na ἄν2. Postupak je proveden samo za leme koje su ušle u popis 108 najčešćih riječi.

Treba spomenuti i pitanje leme εἶπον. Naime, u gramatikama se taj oblik navodi kao aorist glagola λέγω, ali u rječnicima nerijetko ima zasebnu natuknicu (npr. Liddle-Scott-Jones ili Senc). Iz tog razloga pripisana mu je vlastita lema. S obzirom na to da ovo nije greška, već planirana osobina sustava, lemu εἶπον nismo pripisali lemi λέγω3. Ono što ipak jest učinjeno, pripisivanje je leme λέγω (bez broja!), lemi εἶπον, budući da je lema λέγω iskorištena 4 puta, svaki puta za neki od aoristnih oblika i ručno je dodana, što znači da je sustav nije predvidio (već je to bila odluka označivača).

Još treba reći da su postojale i neke krivo unesene leme: npr. lema Ἀθήναιος umjesto leme Ἀθηναῖος ili lema ἀν umjesto leme ἄν2. Sve takve leme svedene su na ispravan oblik.

Primjećujemo da je prilikom obradbe lema obavljeno više intervencija u same podatke, što nije bio slučaj kod obradbe gramatičkih tagova. Naime, premda bismo i normalizacijom gramatičkih tagova postigli bolju sliku stanja u korpusu (primjerice, pribrajanjem jednog pojavljivanja deponentnog glagola mediopasivnim glagolima), statistika nam tog područja koristi kako bismo donosili zaključke o kreiranju kurikuluma. U slučaju popisa lema, tj. najčešćih riječi, ne samo da donosimo zaključke na temelju rezultata, već nam sami rezultati (tj. sam popis) služe kao nastavno sredstvo. Iz tog razloga, želja je autora da prikaz nalaza bude čim „čišći“.

Premda će o tome još biti riječi, i ovdje treba pripomenuti nešto o koeficijentu varijacije. On nam, naime, govori koliko je riječ raspršena po korpusu, odnosno, što je taj koeficijent manji, to je riječ raspršenija. Promatrajući njegovu vrijednost možemo uočiti riječi koje imaju relativno slabu raspršenost (pogotovo u odnosu na riječi s istom apsolutnom frekvencijom) Primjerice, riječ ἀλώπηξ koja označuje lisicu, ima koeficijent varijacije od 52,71 %, dok ostale

riječi koje se u tekstu pojavljuju 14 puta imaju prosječni koeficijent varijacije od 30 %. Pogledamo li bolje u kojim se tekstovima u našem korpusu spominje riječ ἀλώπηξ, vidjet ćemo da se pojavljuje 13 puta u tri Ezopova teksta te jedanput kod Aristotela (u tekstu u kojem prepričava što je Ezop govorio na sudu). Dakle, ova je riječ prilično koncentrirana u tekstovima koji pripadaju žanru basne. Sličan je slučaj i s pridjevom κακός čijih se 7 od 12 pojavnica nalazi u jednom tekstu Marka Aurelija i time dobiva koeficijent varijacije od 57,51 %. Da smo odlučili u naš korpus uključiti i leme s AF vrijednosti 10, vidjeli bismo da lema θυμός ima koeficijent varijacije od 100 % jer se svih 10 pojavnica nalazi u tekstu Ahileja Tatija. Koeficijent varijacije može nas, dakle, upozoriti da dodatno provjerimo i razmislimo pripada li lema u naš popis osnovnih riječi.

5.4. Prikaz sličnih istraživanja

Kako bismo mogli bolje procijeniti kvalitetu naših rezultata, pogledat ćemo još dva istraživanja koja se bave nekim pitanjima dotaknutim u ovom radu. Mahoney (2004.) je provela istraživanje o frekvenciji oblika u grčkom i latinskom na korpusu *Perseus*, a Major (2008.) je analizom istog tog korpusa pokušao konstruirati popis osnovnog vokabulara za grčki jezik. Uz njegov, prikazat ćemo i još jedan pokušaja stvaranja takvog popisa.

Mahoney (2004.) je, vođena željom da sazna koji su oblici koje njeni studenti nužno moraju znati, istražila učestalost morfoloških oblika koristeći se programom za morfološku analizu *Morpheus*. U analizi su obuhvaćene 4 175 533 pojavnice. Problem s kojim se susrela i koji je utjecao na njene rezultate je što *Morpheus* u većini slučajeva nije mogao razlučiti između dvaju ili više morfoloških obilježja čiji su oblici identični. Primjerice, nije mogao raspoznati nominativ od akuzativa srednjeg roda, budući da uvijek imaju isti oblik. U tim je slučajevima autorica rezultate pribrajala kao razlomke, npr. ako su bila dva moguća oblika, oba su prebrojena kao ½ oblika. U Tablici 11. možemo vidjeti usporedbu autoričinih i naših rezultata.

Unatoč 500 puta većem broju riječi, rezultati do kojih je došla Mahoney ne razlikuju se drastično od naših, osobito po pitanju ranga pojedinog oblika (a taj je podatak za izradu kurikuluma izrazito bitan). Jedina razlika po tom pitanju nalazi se kod optativa i imperativa. Optativ je kod Mahoney rangiran kao zadnji, a imperativ kao predzadnji, dok je kod nas obratno. Treba upozoriti i da Mahoney ne razlikuje medij i mediopasiv te ne prepoznaje kategoriju deponentnih glagola.

Po pitanju padeža, najosjetnije je odstupanje kod vokativa. Razlog ovome mogao bi biti malen uzorak na kojem je rađeno naše istraživanje, ali je moguće i da je zbog ranije navedenog

problema sa strojnim označavanjem kod Mahoney mnogo vokativa netočno prebrojano. Sumnja je potaknuta time što je vokativ padež koji se oblikom često podudara s nominativom i akuzativom te bi stvarno stanje trebalo dodatno istražiti.

	Mahoney (2004.)		Naše istraživanje	
Rang	***Padež***	***RF***	***Padež***	***RF***
1.	Akuzativ	28,8 %	Akuzativ	35,52 %
2.	Nominativ	23,9 %	Nominativ	27,49 %
3.	Genitiv	20,8 %	Genitiv	23,16 %
4.	Dativ	13,4 %	Dativ	13,39 %
5.	Vokativ	13,2 %	Vokativ	0,45 %
	Vrijeme	***RF***	***Vrijeme***	***RF***
1.	Prezent	46,7 %	Prezent	52,23 %
2.	Aorist	28,0 %	Aorist	27,89 %
3.	Imperfekt	13,2 %	Imperfekt	9,64 %
4.	Perfekt	6,4 %	Perfekt	5,39 %
5.	Futur	4,8 %	Futur	4,25 %
6.	Pluskvamperfekt	0,8 %	Pluskvamperfekt	0,60 %
7.	Futur egzaktni	0,1 %	Futur egzaktni	0,00 %
	Način[25]	***RF***	***Način***	***RF***
1.	Indikativ	41,6 %	Indikativ	40,63 %
2.	Particip	30,6 %	Particip	33,12 %
3.	Infinitiv	13,4 %	Infinitiv	17,59 %
4.	Konjunktiv	5,7 %	Konjunktiv	4,08 %
5.	Imperativ	3,9 %	Optativ	3,21 %
6.	Optativ	2,8 %	Imperativ	1,36 %
	Stanje	***RF***	***Stanje***	***RF***
1.	Aktiv	85,5 %	Aktiv	67,86 %
2.	/	/	Mediopasiv	20,21 %
2./3.	Medij	10,2 %	Medij	8,55 %
3./4.	Pasiv	4,3 %	Pasiv	3,32 %
5.	/	/	Deponentni	0,05 %

Tablica 11. Usporedba frekvencija u Mahoney (2004.) i u našem istraživanju

[25] O tome što participi i infinitivi rade među načinima vidi str. 20. Uz to, Mahoney ih u svojem radu navodi u tablici s načinima pa je radi očuvanja njene tablice njihov smještaj ovakav.

Major (2008.) u svojem istraživanju ne razmatra učestalost morfoloških oblika, već učestalost riječi. Koristeći se spomenutim korpusom *Perseus* i njegovim alatima, na temelju analize više od 4 100 000 riječi stvara dva popisa najčešćih riječi: tzv. „pedesetpostotni" popis i „osamdesetpostotni" popis. Prvi popis čini 61 riječ koja sačinjava 50 % pregledanog korpusa, dok se drugi popis sastoji od 1100 riječi koje čine 80 % korpusa. O nekim aspektima Majorova rada bit će govora kasnije, a za sad ćemo usporediti njegov i naš popis riječi. Radi jednostavnosti usporedbe i sličnijeg broja riječi, uzet ćemo pedesetpostotni. Riječi koje se nalaze na Majorovu popisu, ali ne i na našem su (abecednim redom): ἀνά, βασιλεύς, οὔτε, πρότερος i τίς, τί. Radi se, dakle, o 5 riječi, odnosno 8,20 % riječi na pedesetpostotnom popisu ne nalazi se na našem popisu. Uzmemo li u obzir činjenicu da Majorov popis sadrži gotovo dvostruko manje riječi od našeg te da su razlike našeg i Mahoneyjina istraživanja osjetno manje, mogli bismo tvrditi da ta razlika nije sasvim zanemariva. To nas i ne bi trebalo pretjerano čuditi, budući da se ovdje radi o leksiku, a on osjetnije varira s obzirom na veličinu i sastav uzorka. Sjetimo se samo da smo u našem popisu naišli na riječ ἀλώπηξ zbog relativno velike prisutnosti Ezopa u našem korpusu. Ipak se čini da smo „ulovili" većinu najvažnijih riječi.

Major nije jedini koji se okušao u takvom pothvatu. *Dickinson College Commentary* (*DCC*) mrežna je platforma Dickinson Collegea iz Pennsylvanije namijenjena objavljivanju komentara i sličnih resursa za čitanje latinskih i grčkih tekstova, čiji su autori stručnjaci iz cijelog svijeta (Francese, 2019.). Jedan od njihovih projekata je upravo *Greek Core Vocabulary*, odnosno popis osnovnih riječi, koji je dostupan na njihovim mrežnim stranicama. Popis sadrži oko 500 lema koje bi trebale odgovarati oko 65 % riječi u tipičnom grčkom tekstu. Izvori za popis bili su odabrani tekstovi iz *Thesaurus Linguae Graecae* te korpus tekstova grčkih autora koji se nalaze na platformi *Perseus under PhiloLogic*. Oba su izvora automatski lematizirana dvama različitim lematizatorima te je usporedbom rezultata dobiven trenutni popis.

Uzmemo li 108 najčešćih riječi s popisa *DCC*-a, vidjet ćemo da se od našeg popisa razlikuje u 27 riječi, odnosno 25 % popisa je drugačije. Ovo nije zanemariv udio, a vjerojatno se može objasniti dvama razlozima: s jedne strane popis *DCC*-a je temeljen na daleko većem korpusu, a s druge strane njihov je popis ručno dorađivan.. O tome ćemo detaljnije govoriti kasnije.

6. Nacrt kurikuluma

Temeljem dobivenih rezultata pokušat ćemo opisati kurikulum nastave grčkog jezika koji bi pratio učestalost pojedinih morfoloških oblika i riječi. U fokusu će biti redoslijed i intenzitet obrade gradiva, odnosno bavit ćemo se pitanjima kojim redom poučavati oblike/riječi te koliko vremena uložiti u njihovo uvježbavanje. Treba imati dvije stvari na umu: prvo, da se radi o nacrtu kurikuluma iz čega slijedi da je on okviran i ne ulazi u detalje te drugo, da ćemo prilikom izrade koliko je god moguće poštivati frekvenciju do koje smo došli istraživanjem. Zaključke o realnoj izvedivosti i primjerenosti takvog kurikuluma, kao i moguće izmjene ili odstupanja od principa frekventnosti, iznijet ćemo kasnije.

Najprije ćemo prikazati redoslijed gramatičkih sadržaja, a nakon toga se pozabaviti pitanjem osnovnog vokabulara. Pokušat ćemo uz svaku promjenu u odnosu na trenutni kurikulum navesti eventualne metodičke probleme i, bude li moguće, njihova rješenja.

6.1. Glagoli

6.1.1. Vremena

Na samom početku nailazimo na poklapanje – prezent je najčešće vrijeme i ujedno se prvi obrađuje. Jedna od najuočljivijih razlika između trenutnog redoslijeda obrade nastavnog sadržaja i situacije kakvu nalazimo u korpusu jest položaj aorista. Aorist se obrađuje na drugoj ili trećoj godini učenja (v. str. 5. – 8.), odnosno nakon imperfekta i futura, premda je osjetno češće vrijeme. S obzirom na učestalost, aorist bi se trebao učiti odmah nakon prezenta te prije imperfekta. Svima koji su se okušali u svladavanju grčkog aorista jasno je zašto se imperfekt obrađuje ranije – imperfekt se vrlo jednostavno tvori od prezentske osnove, dok za tvorbu aorista moramo svladati razlike između jakog, slabog i korjenitog aorista te posebnu pažnju pridati „nepravilnim“ glagolima. Drugim riječima, obrađujući imperfekt ranije poštuje se didaktički princip sistematičnosti i postupnosti (Poljak, 1991.).

Metodički problem kod rane obrade aorista proizlazi iz kompleksne tvorbe koju treba pomno objasniti nekome tko se s grčkim prvi put susreće. Taj bi se problem mogao ublažiti time da se prije prelaska na aorist dobro obrade imenske riječi tako da se učenici priviknu na glasovne promjene grčkog jezika. Pomicanjem aorista bliže početku učenja omogućili bismo dulje uvježbavanje i upoznavanje s tim vremenom. Osim toga, rano učenje aorista daje nam priliku da ranije učenicima objasnimo značenja različitih vremenskih osnova u grčkom jeziku te važnost aspekta kod grčkih glagola. Budući da je imperfekt vrijeme koje po učestalosti slijedi odmah nakon aorista, mogli bismo ih obraditi istovremeno (ili otprilike istovremeno) te tako od

početka naglašavati i s učenicima uvježbavati odnose svršenosti i nesvršenosti u grčkom jeziku, kao i razlike između prezentske i aoristnih osnova.

Još jedna promjena koja bi se mogla uvesti jest učenje perfekta prije futura. Ovdje bismo opet mogli govoriti o kršenju principa jednostavnosti tvorbe, budući da je tvorba perfektne osnove (poglavito zbog reduplikacije) učenicima vjerojatno neobičnija i kompliciranija od tvorbe futurske osnove. Ipak, nadovezujući se na priču o glagolskim osnovama i njihovim značenjima koju smo spomenuli kod aorista, opet ima smisla obraditi perfekt ranije upravo da se uvježba značenje gotove radnje. Jednostavna tvorba futura i značenje koje imamo i u hrvatskom (kao i niska frekvencija futura u tekstovima) mogu sugerirati kako buduće vrijeme nije šteta obraditi na samom kraju jer je lako shvatljivo, a nije pretjerano često.

Pluskvamperfekt zadržava svoje mjesto u kurikulumu, dok se futur egzaktni ionako i ne obrađuje u osnovnoškolskoj i srednjoškolskoj nastavi grčkog jezika.

6.1.2. Načini

Opet na početku nailazimo na podudaranje jer se indikativ, koji je najučestaliji glagolski način, obrađuje prvi. Najveću razliku uočavamo kod imperativa koji se obrađuje vrlo rano, kao drugi po redu način, iako je frekvencijski posljednji (ili pretposljednji, ako sudimo prema Mahoney). Trenutni kurikulum, vidimo, i ovdje prati jednostavnost tvorbe.

Premda se participi i infinitivi ne smatraju glagolskim načinima (već neodređenim glagolskim oblicima), bit će najpraktičnije ovdje i jedne i druge razmotriti zajedno, budući da ni participi ni infinitivi nemaju kategoriju načina, a imaju kategoriju vremena i stanja. Participi se obrađuju nakon indikativa, imperativa i infinitiva, premda su po frekvenciji češći pa bi ih tako trebalo smjestiti i u našem kurikulumu. Infinitiv se trenutno uči prilično rano, nakon indikativa i imperativa, a takav položaj ima i kad gledamo frekvenciju. Konjunktiv i optativ, koji imaju nisku frekvenciju, obrađuju se na kraju.

Najveća je promjena, dakle, što bismo imperative morali obraditi prilično kasnije, dok bismo s participima trebali početi ranije i dulje ih obrađivati. To znači da bismo indikative prezenta i aorista obradili rano, a njihove imperative kasno. Pomicanje obradbe imperativa (i rastavljanje imperativa od indikativa) djelomično bi olakšalo situaciju u kojoj rano obrađujemo aorist, jer smo smanjili broj nastavaka koje učenici moraju naučiti. Osim toga, imperative bismo mogli obrađivati kada i konjunktive, usredotočujući se na mogućnost konjunktiva da izraze slična značenja kao i imperativi (tj. adhortativno ili prohibitivno značenje konjunktiva). Kod participa bi eventualni problem bio što moramo pričekati obradbu pridjeva i –ντ osnova kako bismo ih

mogli istumačiti, no, budući da se i u trenutnom kurikulumu te jedinice obrade rano, to i nije neka prepreka. Više vremena uloženo u obradbu participa dalo bi nam priliku da objasnimo veliko bogatstvo njihova značenja, što je tema kojoj se, usput budi rečeno, ne pridaje mnogo pažnje u aktualnom kurikulumu.

6.1.3. Stanja

Možda je najuočljivije kod glagolskih stanja to što je aktiv osjetno češći od medija ili pasiva. Prema Mahoney 85,5 % svih oblika označno je kao aktivno, dok prema našem istraživanju taj broj iznosi 67,86 %. Premda se aktiv trenutno uči prvi, mediopasiv slijedi vrlo brzo za njim (prema *Godišnjem izvedbenom kurikulumu* 7 sati nakon upoznavanja s aktivnim glagolima obrađuje se mediopasiv). Ovdje se opet, vjerojatno, radi o jednostavnosti tvorbe – za mediopasiv možemo koristiti prezentsku osnovu. Međutim, uzmemo li u obzir promjene nastale u obradbi vremena (među kojima je najuočljivija rani aorist), pomicanje mediopasiva kasnije u kurikulum smanjilo bi broj oblika koje bi učenici morali naučiti, pogotovo s obzirom na činjenicu da bismo kod aorista morali medijalne i pasivne oblike obrađivati zasebno. Još je jedan razlog za ovakav položaj mediopasiva to što bi se mogao svrstati u značenjski kompleksnije oblike. Naime, Leech (2001., str. 4.) prema Clark i Clark (1977.) objašnjava kako su psiholingvistička istraživanja pokazala da su pasivne konstrukcije teže za procesiranje od aktivnih. Osim toga, medij je stanje koje, kao takvo, u hrvatskom nemamo te je samim time učenicima teže dokučivo.

Dodatan su faktor koji treba uzeti u obzir prilikom smještanja stanja u kurikulum deponentni glagoli. Analizom našeg popisa najčešćih lema možemo vidjeti da se među njima nalaze četiri deponentna glagola, s time da je γίγνομαι drugi među glagolima po učestalosti. On ulazi i u Majorov pedesetpostotni popis riječi, dok se na osamdesetpostotnom nalazi 51 deponentni glagol (odnosno, deponentni glagoli čine oko 4,6 % svih riječi na tom popisu). To može poslužiti kao argument za rano učenje mediopasiva, što spominje i Major (2008.), smatrajući da bi se upravo zbog učestalosti glagola γίγνομαι učenici mogli obeshrabriti susretnu li nepripremljeni neke od njegovih oblika u tekstu.

Rješenje bismo mogli pronaći u tome da γίγνομαι izdvojimo i obradimo zasebno, poput glagola εἶναι. Osim toga možemo razmisliti i o svojevrsnom kompromisu, da mediopasive poučavamo kasnije u odnosu na trenutni kurikulum (primjerice nakon obradbe aorista), ali pazeći da dođu dovoljno rano (primjerice prije obradbe drugih vremena poput futura ili perfekta).

6.1.4. Konjugacije

Treba se još osvrnuti i na pitanje konjugacija. Naime, na nešto proširenom popisu od 206 najčešćih lema u našem korpusu nalaze se četiri atematska glagola, dok se na Majorovu osamdesetpostotnome popisu nalazi njih 23, što ukazuje na to da se, unatoč malenom ukupnom broju, susreću relativno često. Atematski se glagoli trenutno obrađuju nakon prezenta, imperfekta, imperativa prezenta i infinitiva prezenta (aktivnog i mediopasivnog), ali s određenim vremenskim odmakom. Iznimka je, naravno, glagol *εἰμί*, koji se obrađuje kao izniman glagol. Niti u našem kurikulumu ne bismo atematske glagole smjeli zapostaviti. Metodički je problem što smo učenike već opteretili nezanemarivom količinom gradiva na početku, a uvođenje atematskih glagola može stvari dodatno otežati pogotovo jer u konjugaciji pokazuju zasebne završetke i često dulje vokale u prezentu.

Kompromis bismo mogli naći u njihovoj djelomičnoj obradbi. Mogli bismo primjerice, najprije obrađivati samo najčešća glagolska lica, odnosno 3. lice singulara i plurala. Tada bismo žrtvovali strukturalnu preglednost, ali bismo uvelike smanjili broj oblika koje treba naučiti. Još jedan primjer djelomične obradbe donosi Major (2008.) koji predlaže semantičko grupiranje nekih izrazito frekventnih glagola, poput glagola λέγω, φημί i εἶπον, koji se svi mogu svesti na glagol „govoriti“. Major tvrdi kako bismo za prezent tog glagola mogli koristiti prezentske oblike glagola λέγω, oblike imperfekta glagola φημί za imperfekt, a εἶπον kao aorist. Takvim prikazom bismo čest atematski glagol φημί mogli obraditi u sklopu glagola „govoriti“. Osim toga, aoriste nekih atematskih glagola mogli bismo obraditi prilikom obrade aorista, budući da se u njemu ponašaju u skladu s paradigmom (npr. glagol δείκνυμι), a njihov prezent ostaviti za nekad kasnije.

Ovdje valja spomenuti i pitanje *verba vocalia*. Njih se na popisu iz Tablice 10. nalazi četiri, dok se na proširenom popisu od 206 najfrekventnijih lema nalazi njih 12. Najviše je glagola s osnovnom na -έω, njih osam, glagola na -άω su tri, dok glagola na -όω nema nijednog. Major u svojem osamedestpostotnom popisu nalazi isti poredak osnova po učestalosti. Možemo zaključiti da su glagoli s vokalskim osnovama prilično frekventni i da bi ih trebalo rano obraditi. Njih bi se moglo u kurikulum smjestiti nakon indikativa prezenta glagola s osnovom koja završava na konsonant, a prije obradbe prošlih vremena. Tako ne samo da bismo ranije mogli uvesti neke vrlo frekventne glagole, već bismo učenike mogli priviknuti na glasovne promjene (poput stezanja vokala) koje igraju važnu ulogu u grčkom.

6.2. Imenske riječi

6.2.1. Padeži

U trenutnom se kurikulumu jednako obrađuju svi padeži u pojedinoj deklinacijskoj paradigmi, premda se njihova frekvencija značajno razlikuje. Najuočljivije je što se akuzativ nalazi na prvom mjestu, dok je u našem korpusu vokativa jako malo. Nominativ je drugi po frekvenciji i nije puno češći od genitiva. Dativ, pak, ima prilično nižu frekvenciju. Ovome možemo pristupiti na dva načina. Mogli bismo najprije obraditi najčešće padeže u nekom skupu deklinacijskih paradigmi (primjerice obraditi najprije akuzativ i nominativ A- i O- deklinacije te zamjenica αὐτός i οὗτος). Treba napomenuti da je promjena poretka padeža pomalo prijeporno pitanje. Dukat (1990.) je u *Gramatici grčkoga jezika* padežima promijenio mjesta i obradio ih ovim redoslijedom: nominativ-vokativ-akuzativ-genitiv-dativ. Čini se da se taj slijed ipak nije udomaćio u poučavanju grčkog (kao i neke druge inovacije poput termina infektum, konfektum i perfektum za glagolski vid), a Šešelj (1984.) čak izrijekom kritizira ovaj pristup, predbacujući mu da uvodi pomutnju i predstavlja teškoće đacima kojim je poredak nominativ-genitiv-dativ-akuzativ-vokativ prirodniji jer ga poznaju iz učenja drugih predmeta.[26] Sličan pristup u poučavanju latinskog koristi Hans H. Ørberg u svom udžbeniku *Lingua Latina per se illustrata* (2019.), u kojem se deklinacijske klase ne obrađuju jedna po jedna, već se obrade pojedini ključni oblici. Primjerice, u prvoj se lekciji obrađuje nominativ jednine i množine A- i O-deklinacije.[27] Alternativno, deklinacije bismo mogli obraditi kao u trenutnom kurikulumu, s time da više pažnje posvetimo obradbi, uvježbavanju i vrednovanju frekventnijih padeža.

6.2.2. Deklinacijski razredi

U našem korpusu među 206 najčešćih lema, koje čine više od 2/3 svih pojavnica u korpusu (s tim da je 206. lema po redu zadnja koja se u korpusu pojavljuje 6 puta), njih 155 su imenice ili pridjevi, od čega imamo 56 imenica/pridjeva A-deklinacije, 62 imenice/pridjeva O-deklinacije i 41 imenicu/pridjev 3. deklinacije. Sudeći prema tome, mogli bismo zaključiti da su A- i O-deklinacija sličnih frekvencija i da su obje frekventnije od 3. deklinacije. Ovdje su, međutim,

[26] S druge strane, trebali bismo pripomenuti i da neki udžbenici Hrvatskog jezika za 5. razred osnovne škole padeže redaju klasičnim nizom (nominativ-genitiv-dativ...) kad ih nabrajaju tablično, ali pojedine padeže obrađuju drugačijim redoslijedom, primjerice: nominativ-akuzativ, genitiv, dativ-lokativ, instrumental, vokativ (Levak, Močibob, Sandalić, Pettö i Budija, 2020.). O poretku padeža mogli bismo dodati još neke misli i navesti neke primjere, ali tomu ovdje nije mjesto - bitno je samo da ostane napomena kako je poredak padeža delikatno pitanje.
[27] Sličan pristup učenju grčkog jezika možemo naći u projektu Seumasa Macdonalda *Ἡ Ἑλληνικὴ γλῶσσα καθ' αὑτὴν φωτιζομένη* (*Lingua Graeca Per Se Illustrata*) (2019.) koji je dostupan na https://seumasjeltzz.github.io/LinguaeGraecaePerSeIllustrata/ (pristupljeno 23. listopada 2020.). Treba imati na umu da tekst na poveznici nije recenziran i nerijetko se u njemu mogu naći greške.

brojane leme, a ne pojavnice, tako da oblici poput participa ili glagolskih pridjeva, za čije nam razumijevanje trebaju deklinacije, nisu ušli u računicu. Za otkrivanje prave frekvencije trebalo bi provesti detaljnije istraživanje. Majorov popis pokazuje još veću ujednačenost između deklinacija, točnije, u njegovu osamdesetpostotnom popisu nalazimo 114 imenica A-deklinacije, 103 imenice O-deklinacije i 108 imenica 3. deklinacije. S pridjevima je drugačije, budući da je pridjeva A- i O- deklinacije više (ukupno 119) od pridjeva 3. deklinacije (ukupno 23).

Ono što bi se moglo zaključiti na temelju naših nalaza jest da nijedna deklinacija nije značajno podzastupljena. Sjetimo li se uz to da su nam sve tri deklinacije potrebne za objašnjavanje raznih gramatičkih oblika, možemo zaključiti da sve tri treba obraditi dovoljno rano.

6.2.3. Zamjenice

Pogledamo li u popis najčešćih riječi, vidjet ćemo da se među njima nalazi nemali broj zamjenica, među kojima su najčešće αὐτός, οὗτος, ὅς, ἐγώ, τις i ἐκεῖνος. Međutim, u kurikulumu se one obrađuju prilično kasno (tek u drugoj godini učenja). Deklinacijske paradigme zamjenica αὐτός, ὅς i ἐκεῖνος ne razlikuju se značajno od deklinacijskih paradigmi imenica A- i O- deklinacije. Deklinacija zamjenice οὗτος premda je neobična, u stvari je lako shvatljiva kada se svlada član. Zamjenica τις prati 3. deklinaciju, dok je zamjenica ἐγώ prilično specifična. Zamjenice su, dakle, deklinacijski uglavnom vrlo bliske imenicama. Ako uz to uzmemo u obzir mogućnost da se detaljnije ili jednostavno ranije obrađuju njihovi najfrekventniji padeži, njihovo pomicanje ranije ne bi trebalo stvarati posebne teškoće, a dalo bi nam priliku da ih uključimo u vježbe već na samom početku učenja.

6.3. Metodička pitanja

Najuočljiviji metodički problem prilikom poštivanja frekvencije u redoslijedu nastavnog sadržaja je problem obradbe složenijeg gradiva prije jednostavnijeg gradiva. Time se krši već spominjani didaktički princip sistematičnosti i postupnosti te bi se moglo prigovoriti da se onemogućuje sagledavanje grčkog jezika kao sustava. S druge strane, ono što dobivamo jest realnija slika stanja kakvo ćemo zateći u izvornim tekstovima i intenzivnije vježbanje onih oblika na koje ćemo češće nailaziti. S obzirom na to da je naglasak u trenutnom kurikulumu više na razumijevanju teksta, a manje na proizvodnji određenih gramatičkih oblika, manjak strukture ne predstavlja toliki problem, dok mogućnost da razumijemo veći dio teksta ne samo da nam olakšava čitanje, već nas može dodatno motivirati. U krajnjoj liniji, sistematičnost

možemo postići naknadnim ustustavljivanjem prethodno obrađenog gradiva u paradigme kakve poznaje klasična gramatika (u satima predviđenim za ponavljanje).

Još bi se jedan argument mogao povući kada govorimo o ranijoj obradbi kompleksnijih oblika, a to je pitanje kognitivne zrelosti. Naime, mogli bismo tvrditi da su učenici na prvoj godini učenja još uvijek nedovoljno kognitivno zreli da uspješno svladaju koncepte i oblike koje im predstavimo. Radi se, dakle, o didaktičkom principu primjerenosti i napora (Poljak, 1991.). Ipak ovakav prigovor traži dodatno razmatranje i istraživanje, a odgovor na njega mogao bi biti pažljiviji metodički pristup gradivu (npr. sporija obrada gradiva u početku ili veća količina uvježbavanja) i ne tako značajna razlika u zrelosti između prve i druge godine učenja.

Svi potencijalni problemi i prednosti koje smo naveli temelje se na promišljanju strukture i karakteristika nastavnog sadržaja te iskustvu iz nastave (kako učeničkom, tako i nastavničkom). Premda takav pristup ima određeno teorijsko utemeljenje, jedini način da ispitamo koliko su problemi i prednosti zaista prisutni jest da to praktično provjerimo. Dakle, sljedeći bi korak u ispitivanju funkcionalnosti frekvencijski utemeljenog kurikuluma bilo provođenje istraživanja među učenicima grčkog koje bi usporedilo dva moguća pristupa učenju.

6.4. Osnovni vokabular

Frekvencijska analiza korpusa omogućuje ne samo drugačije raspoređivanje gramatičkog sadržaja, već i određivanje najučestalijeg vokabulara. Analizirajući najčešće leme možemo stvoriti popis osnovnih riječi, poznat i pod engleskim nazivom *core-vocabulary*. Ideja je da učenici od početka uče one riječi koje će im prilikom čitanja izvornih tekstova najviše trebati i poznata je u svijetu poučavanja stranih jezika. O takvom popisu najčešćih riječi govore Leech (2001.) i Major (2008.), obojica naglašavajući njegovu važnost u poučavanju jezika. I jedan i drugi autor oslanjaju se na statistički fenomen ključan u proučavanju najčešćih riječi, Zipfov zakon. Zipfov zakon tvrdi da će na popisu riječi nekog korpusa, poredanih po njihovoj frekvenciji, frekvencija rapidno padati s rangom, odnosno da će prva najčešća riječ biti otprilike dvostruko frekventnija od druge najčešće riječi, tri puta frekventnija od treće, itd. (Brezina, 2018.). To znači da vrlo malim brojem riječi možemo pokriti velik dio nekog jezičnog izraza (pisanog ili govornog). Takav je slučaj i u našem popisu najčešćih lema (Tablica 10.), gdje nalazimo 108 lema, odnosno 5,5 % jedinstvenih lema u tekstu, a možemo ih povezati s 61,22 % posto svih pojavnica u korpusu. Drugim riječima, sa 108 lema pokrili smo 5513 riječi.

Dok je Leech dobra referenca za općenito razmišljanje o frekvenciji riječi, Major nam je zanimljiv jer se bavi upravo popisom najčešćih riječi u grčkom jeziku. On tvrdi da je popis

najčešćih riječi upravo jedna od snaga grčkog jezika. Smatra da grčki ima popis najčešćih riječi gotovo upola manji od mnogih drugih jezika i tvrdi da takvi popisi unaprjeđuju učenje jezika (Major, 2008., str. 1.). Već smo spominjali pedesetpostotni i osamdesetpostotni popis koje Major primjenjuje u nastavi, kao i popis načinjen u sklopu platforme *Dickinson College Commentary* (str. 28.).

Treba ipak primijetiti da su oba navedena popisa nakon frekvencijske obradbe dodatno uređivana. Major s popisa uklanja osobna imena i krivo lematizirane riječi (primjerice riječ *ἔχις* koja označava zmiju, a nalazi se na njegovu osamdesetpostotnom popisu jer joj je prilikom automatske lematizacije oblik ἔχει neispravno pripisan) te iz raznih razloga neke riječi dodaje (npr. riječ ἀγγέλλω jer se na neobrađenom popisu pojavljuje samo u složenicama ili riječ δαίμων zbog kulturnog značaja; Major, 2008. str. 6.). S popisa *DCC*-a izostavljene su riječi koje se pojavljuju učestalo samo kod nekih autora, ali ne i u ostatku korpusa (primjerice, riječ κίνησις izrazito je česta kod Aristotela i u određenim filozofskim djelima), dok su i ovdje dodavane riječi koje su od kulturnog značaja, poput riječi ἐλεύθερος, te glagoli koji su česti u složenicama (npr. βαίνω*;* Francese, 2019.).

Jasno je da se ni naš popis najčešćih lema ne može uzeti zdravo za gotovo kao popis osnovnih riječi. Sjetimo se da smo morali normalizirati mnogo toga kako bismo dobili čim „čišći“ popis. I mi bismo mogli ukloniti riječi poput Ἀθηναῖος budući da se radi o etniku ili riječ ἀλώπηξ koja je na naš popis zalutala zahvaljujući Ezopu. Upravo u takvim slučajevima koristan je podatak o disperziji riječi u korpusu. Trebalo bi razmisliti i o veličini osnovnog vokabulara. Naš bi pedesetpostotni vokabular sadržavao 43 leme, dok bi ih osamdesetpostotni imao 481. Kao što lijepo navodi C. Francese (2019.) pišući o *DCC*-ovom popisu, „cilj je postići ravnotežu između (neostvarive) statističke savršenosti s jedne i pedagoške korisnosti s druge strane“.

U našem bi kurikulumu popis osnovnog vokabulara bio najkorisniji ako bi se učenici njime koristili od početka učenja. Bilo bi idealno da do trenutka u kojem se sreću s izvornim tekstom sasvim ovladaju pedesetpostotnim popisom i da su barem upoznati s većim dijelom osamdesetpostotnog popisa. Osim davanja popisa učenicima da ga nauče, valjalo bi se njime koristiti čim više i u zadatcima i tekstovima za vježbu. Ne bi bilo loše i najfrekventnije riječi koristiti kao primjer za paradigme pa odabrati, recimo, riječi ἀρχή ili ἡμέρα kao primjere za A-deklinaciju.

7. Zaključak

Istraživanjem smo pokazali da trenutni kurikulum nastave grčkog jezika ne prati frekvenciju oblika, već se oslanja na strukturu jezika kakvu nalazimo u klasičnim gramatikama, poštujući pri tome svojevrsnu „tvorbenu logiku“, odnosno didaktički princip sistematičnosti i postupnosti. Do tih smo rezultata došli usporedbom *Kurikuluma nastavnog predmeta Grčki jezik za osnovne škole i gimnazije*, detaljnije razrađenog u obliku *Godišnjeg izvedbenog kurikuluma za šk. g. 2020./2021*, s podatcima koje smo dobili obradbom anotiranog korpusa tekstova iz *Čitanke za Grčku morfologiju 1*. Obradba je uključivala ne samo dohvaćanje podataka i statističku obradbu, već i razne procese normalizacije. Kurikulum temeljen na frekvencijskoj analizi korpusa tražio bi preraspodjelu sadržaja u odnosu na trenutni kurikulum. Posebno bi važno bilo znatno ranije obraditi aorist i particip kod savladavanja glagolskih oblika te dati prednost akuzativu i nominativu prilikom obrade imenica. Razmotrili smo i mogućnost stvaranja popisa osnovnog vokabulara za grčki jezik i njegove implementacije u nastavi. Osnovni vokabular mogao bi se koristiti kao izvor za gramatičke paradigme i za izbor riječi koje učenici trebaju naučiti i koje će se vrednovati.

Trebamo dati još nekoliko važnih napomena. Na nekoliko smo mjesta spominjali kako je rezultate ovog istraživanja odredio relativno malen korpus pa bi valjalo istraživanje ponoviti na većem, ali jednako kvalitetnom uzorku.[28] Osim toga, upozorili smo i da su naše usporedbe prednosti i mana dvaju razmotrenih kurikuluma temeljene samo na promišljanju te bismo za prave zaključke morali provesti empirijsko istraživanje u učionicama. I jedna i druga opaska neka služe kao poticaji na daljnje proučavanje problematike postavljene u ovom radu.

Međutim, autor ovog rada nada se da vrijednost rada ne leži samo u rezultatima, već i u postavljenim pitanjima i predloženim metodama. Nema sumnje da samim propitivanjem ustaljenih struktura u nastavi grčkog, kao i razmišljanjem o odnosu cilja i sadržaja, otvaramo vrata poboljšanju nastave. S druge strane, konkretne metode prikazane u ovom radu mogle bi se iskoristiti čak i ako na umu nemamo radikalnu promjenu kurikuluma. Mogli bismo, primjerice, stvoriti računalni korpus srednjoškolske lektire iz grčkog jezika i analizom tog korpusa izraditi popise riječi i vježbe koje bi učenicima olakšale susret s izvornim tekstom. U svakom slučaju, premda se radi o jednom od najdugovječnijih nastavnih predmeta u nas, o nastavi grčkog jezika još mnogo toga treba otkriti, ispitati i propitati.

[28] Za jedan takav uzorak mogao bi nam poslužiti veliki, ručno anotirani korpus Vanesse Gorman, kojemu se može pristupiti preko sljedeće poveznice: https://vgorman1.github.io/.

8. Literatura

Bratičević, I., Čengić, N., Jovanović, N., Rezar, V., Šoštarić, P., & Zubović, N. (rujan 2019.). *Čitanka za Grčku morfologiju 1.* Zagreb: Filozofski fakultet Sveučilišta u Zagrebu.

Brezina, V. (2018.). *Statistics in Corpus Lingustics: A Practical Guide.* Cambridge: Cambridge University Press. doi:10.1017/9781316410899

Cindrić, M., Miljković, D., & Strugar, V. (2010.). *Didaktika i kurikulum.* Zagreb: IEP-D2.

Dukat, Z. (1990.). *Gramatika grčkog jezika.* Zagreb: Školska knjiga.

Emde Boas, E., Rijksbaron, A., Hutinik, L., & Bakker, M. (2019.). *The Cambridge Grammar of Classical Greek.* Cambridge: Cambridge University Press.

Francese, C. (30.. srpnja 2019.). *Core Vocabulary*. Dohvaćeno iz Dickinson College Commentary: http://dcc.dickinson.edu/vocab/core-vocabulary

Godišnji izvedbeni kurikulum za šk. g. 2020./2021. (rujan 2020.). Ministarstvo znanosti i obrazovanja. Preuzeto 5. listopada 2020. iz https://mzo.gov.hr/vijesti/okvirni-godisnji-izvedbeni-kurikulumi-za-nastavnu-godinu-2020-2021/3929

Jovanović, N. (7. listopada 2020.). *nevenjovanovic/grcka-morfologija*. Preuzeto 28. rujna 2020. iz GitHub: https://github.com/nevenjovanovic/grcka-morfologija

Kalafatić, Z. (2012.). *Skritpni jezici - materijali za predavanja u elektroničkom obliku.* Zagreb: FER-2.

Kurikulum nastavnog predmeta Grčki jezik za osnovne škole i gimnazije. (2019.). Ministarstvo znanosti i obrazovanja. Preuzeto 5.. listopada 2020. iz https://mzo.gov.hr/istaknute-teme/odgoj-i-obrazovanje/nacionalni-kurikulum/predmetni-kurikulumi/539

Leech, G. (2001.). The role of frequency in ELT: New corpus evidence brings a re-appraisal. In H. Wenzhong (Ed.), *ELT in China 2001: Papers presented at the 3rd International Symposium on ELT in China* (pp. 1.-23.). Foreign Language Teaching and Research Press.

Leech, G. (2007.). New resources, or just better old ones? The Holy Grail of representativeness. *Language and Computers*, 133.-149. doi:10.1163/9789401203791_009

Levak, J., Močibob, I., Sandalić, J., Pettö, I., & Budija, K. (2020.). *Hrvatski bez granica 5 - integrirani udžbenik hrvatskog jezika i književnosti za peti razred osnovne škole.* Zagreb: Školska knjiga.

Mahoney, A. (2004.). The Forms You "Really" Need to Know. *The Classical Outlook, 81.*, str. 101.-105.

Major, W. E. (2008.). It's Not the Size, It's the Frequency: The Value of Using a Core Vocabulary in Beginning and Intermediate Greek. *CPL Online*, 1.-24.

Martinić-Jerčić, Z., Matković, D., & Gjurašin, M. (2019.). *Prometej 1 MYTHOS.* Zagreb: Školska knjiga.

Martinić-Jerčić, Z., Matković, D., & Gjurašin, M. (2020.). *Prometej 2 EPOS.* Zagreb: Školska knjiga.

McEnery, T., & Hardie, A. (2012.). *Corpus Linguistics: Method, Theory and Practice.* Cambridge: Cambridge University Press.

McEnery, T., & Wilson, A. (2001.). *Corpus Linguistics: An Introduction.* Edinburgh: Edinburgh University Press.

Musić, A., & Majnarić, N. (1994.). *Gramatika grčkog jezika.* Zagreb: Školska knjiga.

O'Keeffe, A., McCarthy, M., & Carter, R. (2007.). *From Corpus to Classroom: language use and language teaching.* Cambridge: Cambridge University Press.

Poljak, V. (1991.). Didaktika. Zagreb: Školska knjiga.

Šešelj, Z. (1984.). Zdeslav Dukat: Gramatika grčkoga jezika. *Latina et Graeca, 1*(23), 154.-158.

9. Popis slika

10. Popis tablica

11. Prilozi

Prilog 1. : Ključ za dešifriranje zapisa gramatičkih tagova u *Arethusi*

Vrijednost mjesta	Znak	Vrijednost znaka
Vrsta riječi	"a--------"	pridjev
	"c--------"	veznik
	"d--------"	prilog
	"i--------"	uzvik
	"l--------"	član
	"m--------"	broj
	"n--------"	imenica
	"p--------"	zamjenica
	"r--------"	prijedlog
	"u--------"	interpunkcija
	"v--------"	glagol
	"x--------"	nepravilno
Lice	"v1-------"	1. lice
	"v2-------"	2. lice
	"v3-------"	3. lice
Broj	"v-s------"	singular
	"v-p------"	plural
Vrijeme	"v--p-----"	prezent
	"v--i-----"	imperfekt
	"v--a-----"	aorist
	"v--f-----"	futur
	"v--r-----"	perfekt
	"v--l-----"	pluskvamperfekt
	"v--t-----"	futur egzaktni

Vrijednost mjesta	Znak	Vrijednost znaka
Način	"v---i----"	indikativ
	"v---s----"	konjunktiv
	"v---o----"	optativ
	"v---m----"	imperativ
	"v---n----"	infinitiv
	"v---p----"	particip
Stanje	"v----a---"	aktiv
	"v----m---"	medij
	"v----e---"	mediopasiv
	"v----p---"	pasiv
	"v----d---"	deponentni
Padež	"n------n-"	nominativ
	"n------g-"	genitiv
	"n------d-"	dativ
	"n------a-"	akuzativ
	"n------v-"	vokativ
	"n------l-"	lokativ
Stupanj komparacije	"a-------p"	pozitiv
	"a-------c"	komparativ
	"a-------s"	superlativ

Prilog 2. : XQuery kod za upit lema i gramatičkih tagova

```
(: Za leme :)
for $lemmata in //*:word/@lemma/string()
order by $lemmata
return $lemmata

(: Za tagove :)
for $forms in //*:word/@postag/string()
order by $forms
return $forms
```

Prilog 3. : Statističke formule

$$\text{relativna frekvencija} = \frac{\text{apsolutna frekvencija}}{\text{ukupni broj pojavnica}}$$

$$\text{standardna devijacija} = \sqrt{\frac{\text{zbroj kvadrata razlika RF od srednje vrijednosti}}{\text{ukupni broj dijelova korpusa}}}$$

$$\text{koeficijent varijacije} = \frac{\text{standardna devijacija}}{\text{srednja vrijednost}}$$

$$\text{koeficijent varijacije \%} = \frac{\text{koeficijent varijacije}}{\sqrt{\text{broj dijelova korpusa - 1}}} \times 100$$

Prilog 4. : Python kod za prebrojavanje gramatičkih tagova

```
# !/usr/bin/env python3
import sys
#Define variables to be used in loops
#Tense variables
pres = 0
impf = 0
aors = 0
futI = 0
perf = 0
plpf = 0
ftpf = 0
#Mood variables
indc = 0
subj = 0
optt = 0
impr = 0
infn = 0
ptcp = 0
#Ptcp variables
ptpr = 0
ptar = 0
ptft = 0
ptpf = 0
#Infn variables
inpr = 0
inar = 0
inft = 0
inpf = 0
#Voice variables:
actv = 0
medd = 0
```

```
mdps = 0
pass = 0
depn = 0
#Case variables:
nomi = 0
geni = 0
dati = 0
accu = 0
voca = 0
#POS variables:
adjc = 0
conj = 0
advb = 0
intr = 0
artc = 0
numr = 0
noun = 0
pron = 0
prep = 0
verb = 0
irrg = 0

#Open the file containing POS tags
with open(sys.argv[1], 'r', encoding='utf8') as file:

#Make a list where each line is an element
     omnes = file.readlines()

#Check each line and count how many times POS occurs
     for line in omnes:

          if line[0] == 'a':
```

```
        adjc = adjc + 1
    elif line[0] == 'c':
        conj = conj + 1
    elif line[0] == 'd':
        advb = advb + 1
    elif line[0] == 'i':
        intr = intr + 1
    elif line[0] == 'l':
        artc = artc + 1
    elif line[0] == 'm':
        numr = numr + 1
    elif line[0] == 'n':
        noun = noun + 1
    elif line[0] == 'p':
        pron = pron + 1
    elif line[0] == 'r':
        prep = prep + 1
    elif line[0] == 'v':
        verb = verb + 1
    elif line[0] == 'x':
        irrg = irrg + 1


    if line[3] == 'p':
        pres = pres + 1
    elif line[3] == 'i':
        impf = impf + 1
    elif line[3] == 'a':
        aors = aors + 1
    elif line[3] == 'f':
        futI = futI + 1
    elif line[3] == 'r':
```

```
     perf = perf + 1
elif line[3] == 'l':
     plpf = plpf + 1
elif line[3] == 't':
     ftpf = ftpf + 1

if line[4] == 'i':
     indc = indc + 1
elif line[4] == 's':
     subj = subj + 1
elif line[4] == 'o':
     optt = optt + 1
elif line[4] == 'm':
     impr = impr + 1
elif line[4] == 'n':
     infn = infn + 1
     if line[3] == 'p':
          inpr = inpr + 1
     elif line[3] == 'a':
          inar = inar + 1
     elif line[3] == 'f':
          inft = inft + 1
     elif line[3] == 'r':
          inpf = inpf + 1
elif line[4] == 'p':
     ptcp = ptcp + 1
     if line[3] == 'p':
          ptpr = ptpr + 1
     elif line[3] == 'a':
          ptar = ptar + 1
     elif line[3] == 'f':
          ptft = ptft + 1
```

```
                elif line[3] == 'r':
                    ptpf = ptpf + 1

            if line[5] == 'a':
                act = act + 1
            elif line[5] == 'm':
                mdd = mdd + 1
            elif line[5] == 'e':
                mdp = mdp + 1
            elif line[5] == 'p':
                pas = pas + 1
            elif line[5] == 'd':
                dep = dep + 1

            if line[7] == 'n':
                nom = nom + 1
            elif line[7] == 'g':
                gen = gen + 1
            elif line[7] == 'd':
                dat = dat + 1
            elif line[7] == 'a':
                acc = acc + 1
            elif line[7] == 'v':
                voc = voc + 1

print (   pres, "\r\n",
          impf, "\r\n",
          aors, "\r\n",
          futI, "\r\n",
          perf, "\r\n",
          plpf, "\r\n",
          ftpf, "\r\n",
```

```
indc, "\r\n",
subj, "\r\n",
optt, "\r\n",
impr, "\r\n",
infn, "\r\n",
ptcp, "\r\n",
ptpr, "\r\n",
ptar, "\r\n",
ptft, "\r\n",
ptpf, "\r\n",
inpr, "\r\n",
inar, "\r\n",
inft, "\r\n",
inpf, "\r\n",
act, "\r\n",
mdd, "\r\n",
mdp, "\r\n",
pas, "\r\n",
dep, "\r\n",
nom, "\r\n",
gen, "\r\n",
dat, "\r\n",
acc, "\r\n",
voc, "\r\n",
noun, "\r\n",
pron, "\r\n",
adjc, "\r\n",
numr, "\r\n",
verb, "\r\n",
artc, "\r\n",
advb, "\r\n",
prep, "\r\n",
```

```
conj, "\r\n",
intr, "\r\n",
irrg, "\r\n",
)
```

Prilog5. : Python kod za prebrojavanje lema

```
# !/usr/bin/env python3
import sys

with open(sys.argv[1], 'r', encoding='utf8') as dat1: #Open
the file containing lemmas

#Create a list object of all lemmas
          lemlist = dat1.readlines()
          lemlist_nnl = []



#Remove new lines
          for i in lemlist:
               lemlist_nnl.append(i.rstrip())

#Create a set object from the list object-unique values remain
          lemset = set(lemlist_nnl)
          lemlistunq = list(lemset)
          lemlistunq.sort()



#For each unique lemma count how many times it occurs in total
#lemma list
          for ulemma in lemlistunq:
               occurence = lemlist_nnl.count(ulemma)
               print (ulemma,",",occurence, sep="")
```

Sažetak

Nacrt kurikuluma grčkog jezika temeljen na frekvencijskoj analizi korpusa

Kurikularna reforma hrvatskog školstva nije zaobišla ni klasične jezike. Glavna je karakteristika novog kurikuluma za grčki jezik to što kao središnji ishod postavlja razumijevanje teksta, što znači da učenike treba čim bolje pripremiti za susret s izvornim tekstom. Stoga nas zanima kakav je odnos stvarnog jezičnog stanja i nastave. U ovom se radu ispituje mogućnost stvaranja kurikuluma nastave grčkog jezika temeljenog na frekvencijskoj analizi korpusa. Najprije se ispituje prati li aktualni kurikulum za grčki jezik, donesen 2019. godine, učestalost oblika u grčkom jeziku, odnosno daje li se prednost onim oblicima koji se najčešće pojavljuju. Utvrđeno je da situacija u kurikulumu nije takva, već da se kurikulum zasniva na tradicionalnim gramatikama te da gramatičke sadržaje obrađuje tako da grupira one oblike koji se slično tvore, najprije obrađujući najjednostavnije oblike. Zatim je u radu prikazana analiza korpusa tekstova iz *Čitanke za Grčku morfologiju 1*. Opisuju se struktura, izvlačenje i statistička obradba podataka, a rezultati se uspoređuju s rezultatima sličnih istraživanja (Mahoney, 2004. i Major 2008.). Na kraju se na temelju našeg istraživanja opisuju glavne karakteristike kurikuluma koji prati učestalost oblika.

Ključne riječi: kurikulum grčkog jezika, korpus, frekvencijska analiza, osnovni vokabular, metodika nastave grčkog jezika

Summary

A draft of curriculum for Ancient Greek based on a corpus frequency analysis

The classical languages have been affected by the curricular reform in Croatia. The new Ancient Greek language curriculum sets text comprehension as its main goal, which means that students should be better prepared for reading original texts. In that context, we want to study the relation between the real state of the Ancient Greek, as attested by its texts, and its teaching in schools. In this paper we examine the possibility of creating a curriculum for Ancient Greek based on corpus frequency analysis. We first examine whether the current curriculum, introduced in 2019, is based on the frequency of forms in Ancient Greek or not. In other words, we are trying to see if it prepares the students to read texts by favouring those linguistic forms that occur more frequently. We have determined that the current curriculum is not based on frequency of forms, but rather on traditional grammars and that it groups together forms with similar word-formation or inflection rules, teaching first what is regarded as simplest. Then we analyse a corpus of texts from The Reader for Greek Morphology 1, prepared at the University of Zagreb, Faculty of Humanities and Social Sciences. We describe the structure, extraction and the statistical analysis of the data and compare the results with those of similar studies (Mahoney, 2004 and Major, 2008). In the end, we describe the main features of a frequency-based curriculum according to the results of our analysis.

Key words: Ancient Greek curriculum, corpus, frequency analysis, core vocabulary, Ancient Greek teaching methods