

**Artur Šilić i Slobodan Hadžić**

Novena d.o.o.  
artursilicgmail.com

Sveučilište Sjever, Presscut  
slobodan.hadzic@presscut.hr

# **Novi analitički modeli za utvrđivanje standarda razumljivosti medijskog teksta u medijima masovnog komuniciranja**

## **Sažetak**

*Autori se u svom radu usredotočuju na razumljivost tekstova objavljenih u medijima masovnog komuniciranja u teorijskom kontekstu medijske pismenosti i medijske kompetencije sa nakanom ustanovljenja elemenata standarda razumljivosti medijskih tekstova.*

*Komparativnom analizom medijskih tekstova dnevnih novina, nacionalnog doseg, prije svega informativno političkog sadržaja kroz 11 kategorija /duljina teksta u rečenicama, duljina teksta u riječima, duljina teksta u znakovima, prosječnom duljinom rečenice, prosječnom duljinom riječi u slogovima, mjera leksičke gustoće, leksičke redundancije, mjera gustoće entiteta, kompleksnosti riječi po duljini slogova i po frekvenciji, mjere temeljene na procjeni složenosti riječi u novinskom korpusu/ autori potvrđuju hipotezu djelomične nerazumljivosti medijskih tekstova i značajne razlike u razumljivosti između medijskih nakladnika.*

*Fokus istraživanja temeljen je na grupama sastavljenim od različitih dobnih i obrazovnih i socijalnih skupina autori uočavaju bitne razlike u percepciji razumljivosti medijskih tekstova kako po dobnoj tako i po obrazovnoj strukturi što ih dovodi do zaključka da se standardi razumljivosti trebaju napraviti anticipirajući stečena istraživačka saznanja ali i nastaviti u pravcu utvrđivanja svih relevantnih faktora percepcije medijskog teksta i njihovog utjecaja na razumljivost istog.*

*Procjenom čitljivosti tekstova od strane "medijskih konzumenata" te uparivanjem s računalnim ocjenama složenosti tekstova, odabrane su najpogodnije mjere za automatsku ocjenu razumljivosti novinskih tekstova na hrvatskome jeziku. Pomoću odabranih automatskih mjera čitljivosti tekstova te ljudskom percepcijom čitljivosti tekstova, autori su pokušali doprinijeti izradi elemenata standarda tj. minimalnim zahtjevima na čitljivost teksta.*

**Ključne riječi:** *medijska pismenost, masovna komunikacija, medijske kompetencije, medijski jezik, analitički modeli.*

## UVOD I TEORIJSKA OSNOVA RADA

Čitljivost teksta (eng. readability) definira se kao lakoća razumijevanja i shvaćanja teksta što je posljedica stila pisanja (Klare, 1988, prema Al-Khalifa i Al-Ajlan, 2010). Cilj utvrđivanja razumljivosti teksta je odabrati tekstove primjerene određenom čitatelju, kako bi se utjecaj teksta na čitatelja optimalizirao. Čitljivost je ključan koncept u obrazovanju, industriji, medicini, medijima, te odnedavno stječe važnost i u internetskom prostoru.

Faktori čitljivosti su faktori koji utječu na razinu čitanja i razumijevanja teksta, a možemo ih podijeliti na faktore čitatelja i faktore teksta (Al-Khalifa i Al-Ajlan, 2010). Faktori čitatelja vezani su uz dob čitatelja i njegovu vještinu čitanja. Prethodnim istraživanjima je utvrđeno da prethodno znanje i iskustva čitatelja, njegovi interesi i motivacija utječu na čitljivost teksta (DuBay, 2004). Također, od iznimne važnosti su i sklonosti čitatelja, koje će ga ohrabriti ili obeshrabriti da se posveti tekstu iz određenog područja (Al-Dawsari, 2004, prema Al-Khalifa i Al-Ajlan, 2010).

Mnoge karakteristike samog teksta utječu na čitljivost a neki od njih su karakteristike riječi (dužina riječi, korištenje neuobičajenih riječi, frekvencija riječi), prosječna duljina rečenice, korištenje topologija, metafora i poredbi, korištenje pasiva i sl. (Al-Khalifa i Al-Ajlan, 2010).

Pokušaji kombiniranja ovih faktora rezultirali su težnjama za kvantificiranje čitljivosti teksta. U posljednjem stoljeću proučavanje čitljivosti teksta bilo je vrlo propulzivno područje. Započelo je istraživanjima Livelyja i Presseya (1923, prema Hiebert i Cruz, 2010), te njihovim prvim pokušajima osmišljavanja formula za izračunavanje čitljivosti teksta. Do 1980-ih je konstruirano preko 200 formula čitljivosti teksta (DuBay, 2004, prema Benjamin, 2011). Formule čitljivosti teksta su matematičke jednadžbe koje predviđaju težinu određenog teksta fokusirajući se primarno na lingvistička obilježja poput složenosti riječi ili rečenica. Uz spomenuta sintaktička obilježja, neke formule uzimaju u obzir i semantiku (DuBay, 2004). Neki od popularnijih indeksa čitljivosti su Fry indeks čitljivosti (Fry, 1977, prema Gallagher, Fazio i Gunning, 2012), Flesch-Kinkaid indeks (Flesch, 1948, prema Gallagher, Fazio i Gunning, 2012), Spache (1953, prema Gallagher, Fazio i Gunning, 2012), Dale-Chall (Dale i Chall, 1938, prema Gallagher, Fazio i Gunning, 2012) te Gunning Fog indeks (1952, prema Gallagher, Fazio i Gunning, 2012). Fry indeks čitljivosti te Flesch-Kinkaid indeks se baziraju na broju rečenica u odlomku, te broju slogova riječi, uz pretpostavku da su duge rečenice i višesložne riječi zaslužne za slabiju razumljivost teksta. Spache i Dale-Chall indeksi uzimaju u obzir frekvenciju korištenja određene riječi u jeziku te poznatost korištenog vokabulara. Gunning Fog indeks temelji se na kompleksnosti riječi u tekstu, uzimajući u obzir broj riječi duljih od tri sloga koje smatraju teškima.

Cloze test također spada u tradicionalne pristupe istraživanja čitljivosti teksta. Ovaj test procjenjuje čitljivost teksta mjereći individualno razumijevanje određenog teksta. U ovom testu, ciljnoj publici prezentira se tekst u kojem određene riječi nedostaju (primjerice, svaka peta riječ), a zadatak ispitanika je nadopuniti riječi koje nedostaju. Postotak točnih riječi predstavlja Cloze indeks (Gemoets, Rosemblat, Tse i Logan, 2004, prema Al-Khalifa i Al-Ajlan, 2010). U sklopu tradicionalnog

pristupa prisutna je i klasifikacija čitljivosti tekstova prema procjenama nastavnika, učenika ili eksperata, te primjena testa razumijevanja teksta u formi pitanja vezanih uz sadržaj (Al-Khalifa i Al-Ajlan, 2010).

S razvojem kognitivne psihologije 1970-ih godina utvrdili su da formule čitljivosti nisu toliko dobre koliko se mislilo. Naime, nizom eksperimenata u kojima se manipuliralo tekstom u skladu s postavkama formuli kako bi se dobio tekst određene težinske razine, utvrđeno je da je razumijevanje takvih tekstova otežano u odnosu na originalne tekstove koji su prema formulama bili teži (Beck, McKeown, Omanson i Pople, 1984, prema Hiebert i Cruz, 2010). Uslijedio je razvoj digitalnih formula čitljivosti, od kojih je najznačajnija Lexile skala (Smith i sur. 1989, prema Hiebert i Cruz, 2010). Kao i ranije formule bazira se na matematičkim algoritmima sintaktičkih i semantičkih elemenata. Sintaktičku mjeru predstavlja aritmetička sredina duljine rečenica u tekstu, a semantičku aritmetička sredina logaritma frekvencije riječi u jeziku. Formula daje rezultat od 0 (najlakši tekst) do 2000 (najteži tekst).

Mnogi autori zamjeraju spomenutim formulama artifičijelnost. Naime, kratke, često korištene riječi rezultiraju lakim tekstovima prema formulama iako empirijska istraživanja upućuju da to nije nužno tako. Druga zamjerka je da ove formule umjetno povećavaju težinu informativnih tekstova, a smanjuju težinu narativnih tekstova. Naime, informacijski tekstovi koriste precizan i rjeđe upotrebljavan rječnik, te se rijetke riječi često i ponavljaju. Takvo ponavljanje formule ne uzimaju u obzir, unatoč dokazima da nakon nekoliko ponavljanje riječi njena težina drastično opada (Finn, 1978, prema Hiebert i Cruz, 2010). Bailin i Grafstein (2011, prema Benjamin, 2011) ističu da ne postoji metoda da se jednim brojem odredi težina i čitljivost teksta, te da su faktori poput stila, vokabulara, prethodnog znanja, gramatike, tekstualne koherencije, te sposobnosti pojedinog čitatelja da „popravi“ nedostatke određenog teksta, najvjerojatnije u interakciji koja rezultira manje ili više teškim tekstom.

U posljednje vrijeme razumijevanje teksta je u fokusu kognitivnih psihologa, te se čini da je riječ o smjeru istraživanja koje najviše obećava (Benjamin, 2011). Autori iz ovog područja predlažu da težina teksta i čitljivost imaju više veze s koherencijom i međusobnim odnosima elemenata nego površinskim svojstvima samog teksta (McNamara i Kintsch, 1996). Temelj kognitivno psiholoških analiza tekstova su tzv. propozicije, koje Graesser i sur. (1997) definiraju kao kratke smislene jedinice koje zanemaruju aspekte kao što su glagolsko vrijeme ili lice. Sastoje se od predikata i najmanje jednog argumenta, a mogu sadržavati i druge propozicije. Argument ima funkcionalnu svrhu unutar propozicije, a definira odnos između smislenih riječi u rečenici (Benjamin, 2011). Kako bi tekst bio koherentan, mora postojati određeno preklapanje propozicija ili barem argumenata između rečenica koje slijede jedna iza druge. Također, mora postojati propozicijska dosljednost kroz veće dijelove teksta ili čitav tekst. Ukoliko se slijede ova pravila, rezultat je tekst koji čitatelju omogućava lako slijeđenje misli, te čitatelj ne mora popunjavati „rupe“ u tekstu svojim prethodnim znanjima. U protivnom, ako čitanje iziskuje zaključivanje i nadopunjavanje informacija u tekstu svojim znanjima,

to stavlja veći izazov pred slabijeg čitatelja. Stoga je i sam tekst teži za razumijevanje. Manipulacija propozicijama i inferencijama rezultira značajnim razlikama u razumijevanju pročitano (Britton i Gülgöz 1991; Britton i sur. 1993, prema Benjamin, 2011).

Latentna semantička analiza je automatizirano sredstvo koje prezentira sadržaj teksta kao vektor u semantičkom prostoru, a analizira semantičku povezanost između tekstova ili dijelova teksta (Benjamin, 2011). Sustav se može „podučiti“ kako bi prepoznao koje riječi se pojavljuju u kojim kontekstima. Temeljem tog znanja može procjenjivati povezanost među riječima, budući da se u tekstovima riječi poput *vozilo* i *cesta* često pojavljuju zajedno. Može analizirati i indirektno povezanosti. S matematičke strane, sustav barata matricama koje zahvaćaju riječi i njihove povezanosti s drugim riječima. Svaka riječ predstavljena je vektorom u multidimenzionalnom semantičkom prostoru. Pritom su redovi u vektoru kontekst u kojem se riječ pojavljuje, primjerice odlomak ili tekst. Kosinus između dva vektora ukazuje na stupanj povezanosti između dvije riječi. LSA sustav ima mogućnost procjene vjerojatnosti da će se određena riječ pojaviti u određenom tekstu, ako i odrediti važnost određene riječi za određeni kontekst. Pomoću ovog sustava može se procjenjivati težina teksta, ali i koherentnost, te posredno kvaliteta. Primjerice, sustav može odrediti stupanj sličnosti između određenog teksta i oglednog primjerka teksta iz područja. Što je stupanj sličnosti veći, pretpostavlja se da je tekst kvalitetniji. Sustav pomoću kojeg se provode latentne semantičke analize naziva se Coh-Matrix (Graesser i sur., 2004).

Empirijska potvrda kognitivno psihologijskih modela dolazi iz istraživanja Brittona i Gulgoza (1991). Uzeli su tekst korišten za obuku osoblja Američkih zračnih snaga, te su pomoću računalnog programa identificirali mjesta u tekstu gdje nedostaju poveznice odnosno zaključci između dijelova teksta ili rečenica. Modificirali su tekst povezujući prethodnu rečenicu sa sljedećom pomoću preklapanja prepozicija i argumenata, te eksplicitno navodeći važne implicitne zaključke. Sudionici su se bolje dosjećali takvog teksta, mjereno pomoću slobodnog dosjećanja i pitanja višestrukog izbora, iako su tradicionalni indeksi čitljivosti između originalnog i novog teksta ostali nepromijenjeni. Nalazi su obećavajući, s obzirom da su se modifikacije tekstova prema formulama čitljivosti pokazale kontraproduktivnima.

McNamara i sur. (1996) su utvrdili interaktivni odnos koherencije teksta i razine ekspertnosti čitatelja. Osobe niske ekspertnosti bolje rezultate postižu ukoliko im se na čitanje da tekst visoke koherentnosti u kojem su zaključci eksplicitno navedeni. S druge strane, eksperti imaju veću korist od tekstova niske koherentnosti. Objašnjenje ovog fenomena pruža Kintschev konstrukcijsko-integracijski model razumijevanja pročitano. Prema ovom modelu, ekspertni čitatelji trebaju određene „zapreke“ pred sobom kako bi se potaknulo dublje procesiranje pročitano. Tekstovi niske koherencije zahtjeva od ekspertnog čitatelja da stvara zaključke, što aktivira i jača kognitivne semantičke mreže. Budući da ne-ekspertni čitatelji ne posjeduju dovoljna prethodna znanja iz područja, za njih je potrebno ukloniti prepreke, jer znanja koja posjeduju nisu dovoljna da stvore cjelovitu sliku područja ili donesu točne zaključke. Važno je razlikovati pojmove koherencije i

kohezije. Koherencija je psihološki konstrukt, a odnosi se na način kako su propozicije povezane u čitateljevim mentalnim reprezentacijama, te kao takva nije mjerljiva. Kohezija predstavlja površinske indikatore povezanosti među rečenicama u tekstu, a što se postiže proučavanjem propozicijskih i argumentacijskih preklapanja. To se može postići pomoću programa Coh-Metrix, koji pruža uvid u više od 50 indikatora jezika, kohezije i težine teksta (Graesser i sur., 2004).

Lin i sur. (2009) su temeljem modela semantičkih mreža zaključili da za čitljivost teksta nije važna apsolutna već relativna dužina riječi. Krenuvši od bazične razine riječi, utvrdili su da su hipernimi u pravilu duži od polazne riječi, a hiponimi su gotovo uvijek duži. Stoga je relativna duljina riječi preciznija odrednica težine teksta. Nizom analiza utvrdili su da njihova metoda daje mnogo točnije procjene težine teksta u odnosu na tradicionalne formule čitljivosti. Formule čitljivosti koje su koristili bile su New Dale–Chall formula, Spache, Powers–Sumner–Kearl, Flesch–Kincaid, FOG, SMOG i FORECAST. Ovim formulama se često prigovarala pogreška do koje dolazi njihovim korištenjem, a to je nagli pad težine tekstova u udžbenicima za 7. i 8. razrede. Naime, formule su konstantno težima procjenjivale udžbenike za 4. do 6. razrede. Pristup Lina i suradnika (2009) nije bio opterećen ovakvim pogreškama.

Temeljem novijih saznanja o kognitivnoj podlozi čitljivosti osmišljen je DeLite/EnLite program (vor der Brück i sur., 2008). Ovaj program kombinira tzv. površinske i duboke indikatore težine teksta, što povećava psihologijsku valjanost procjena. Program barata znanjem o težini teksta kojem je „podučen“ temeljem 3000 procjena danih na skali Likertovog tipa sa sedam uporišnih točaka. Procjene se odnose na 500 tekstova, a dane su od strane 300 sudionika. Pomoću tih procjena, programeri su „naučili“ program da prepozna određene kombinacije semantičkih i sintaktičkih karakteristika prisutne na specifičnim razinama čitljivosti. Neki od indikatora koje ovaj program daje su dubina umetnute rečenice, kvaliteta semantičke mreže, broj propozicija po rečenici, te dužina kauzalnih lanaca (ima li previše ideja izraženih u premalo riječi). Iako je ovaj program pružio bolju predikciju težine teksta od tradicionalnih metoda, objašnjava svega 28% varijance težine teksta.

Iz navedenih istraživanja vidljivo je da tradicionalne formule za određivanje čitljivosti nisu dovoljna niti pouzdana metoda utvrđivanja težine teksta. Saznanja iz kognitivne psihologije pružila su bolji uvid u procese koji leže u podlozi razumijevanja teksta, te omogućila nove analize. Kako je riječ o psihološkim konstruktima koje nije lako zahvatiti, potreban je daljnji razvoj kompjuterskih programa koji će precizno identificirati indikatore latentnih psiholoških konstrukata u podlozi čitljivosti teksta.

Problematika s kojom se autori bave je multidisciplinarna i pripada područjima komunikologije, računarske tehnologije, lingvistike, psihologije, masovnih komunikacija, publicistike a svrhu / korist/ ovakvih tipova istraživanja vidimo u području odgoja i obrazovanja te medijske pismenosti /kompetencija/.

## PROBLEM, CILJ I HIPOTEZE ISTRAŽIVANJA

Naš temeljni problem, cilj istraživanja bio je:

- Razmotriti odnos čitljivosti novinskih tekstova i razumljivosti istih
- te čitljivosti priopćenja za javnost i njihove prilagodljivosti stilu novinskog izvještavanja / čitljivosti novina/

Pri čemu smo pretpostavljali /hipoteze/

- Da postoji interaktivni odnos koherencije teksta i razine ekspertnosti čitatelja.
- Da postoji razlika u mjerama čitljivosti kod hrvatskih dnevnih novina što rezultira različitim stupnjevima razumljivosti
- Da priopćenja za javnost nisu prilagođena kriterijima stila novinskog pisanja /mjerama čitljivosti/

### Metodologija

U uzorak smo uzeli tri dnevne novine i to: Jutarnji list, Večernji list i Slobodnu Dalmaciju i to za period od 1.6. 2014. do 1.6. 2015. godine.

Rubrike u uzorku bile su: Unutarnja politika, vijesti dana, crna kronika i kultura.

Jedinica analize: novinski članak /objava/ bez natječaja i oglasa pri čemu je analizirano ukupno 69 274 medijske objave/članka.

**Za prvu pretpostavku postojanja interaktivnog odnosa koherencije teksta i razine ekspertnosti čitatelja** analizirati smo kolumniste iz Večernjeg lista, Jutarnjeg lista, Slobodne Dalmacije i Novog lista. /imena kolumnista navedena su u rezultatima istraživanja/. Ukupno je analizirano 4618 kolumni tijekom perioda istraživanja.

Korupus analize istraživali smo kroz 8 varijabli čitljivosti i to:

- Duljina teksta u rečenicama
- Duljina teksta u riječima
- Duljina teksta u znakovima
- Prosječna duljina rečenice
- Prosječna duljina riječi u slogovima
- Mjere leksičke gustoće
- Mjere leksičke redundancije
- Mjere gustoće entiteta

Kroz tri testa:

1. Gunningov indeks magle
2. Fleschov indeks i
3. Flesch – Kincidov indeks

**Razumljivost** smo provjeravali anketom I fokus grupama I to na niskoj I visokoj razini ekspertnosti. Aneka je sadržavala 12 čestica a fokusom smo provjeravali razumljivost.

### **Usporedba stilometrijskih značajki PR objava s prosječnim novinskim tekstom**

Cilj ovog eksperimenta je usporediti stilske značajke i testove čitljivosti PR objava kao posebne tekstne forme korištene u medijskoj produkciji s prosječnim tekstom dnevnih novina u koje se PR objave u konačnici kontekstno uklapaju. Korpus je sačinjen od 91 originalnih PR objava poslanih predstavnicima medija u razdoblju od početka 2011. do sredine 2015. g. Pri analizi izuzeti su naslovi, podnaslovi, nadnaslovi i kontaktne informacije s potpisom (a koje se često navode u PR objavama).

#### **REZULTATI ISTRAŽIVANJA**

##### **1. Usporedba tri izvora dnevnih novina**

Cilj ovog eksperimenta je usporediti tri dnevne novine i pokazati postoji li značajnih razlika u stilu i čitljivosti.

##### **1.1. Korpusi**

Pri ovom eksperimentu uzeta su tri korpusa sačinjena od ukupne produkcije praćenja tiskanih medija triju novinskih izvora – Jutarnjeg lista, Slobodne Dalmacije i Večernjeg lista u razdoblju od godine dana (od 6.1.2014. do 31.5.2015.).

<b>Korpus</b>	<b>Izvor</b>	<b>Broj članaka</b>
JL	Jutarnji list	12718
SD	Slobodna Dalmacija	14651
VL	Večernji list	14626

**Tablica T1.1.** Izvori i veličine korpusa jedne godine svih članaka dnevnih novina



## 1.2. Metodologija

Na svim tekstovima pomoću posebno napisanih programa za obradu tekstova izračunati su stilometrijski pokazatelji i tri testa čitljivosti opisani u dodatku A. Za račun stilometrije koristi se samo tijelo teksta bez naslova, podnaslova i ostale opreme.

## 1.3. Rezultati

U tablici T1.2. vide se prosječne vrijednosti pokazatelja na odabranim korpusima. Između korpusa ne postoji statistički značajnih razlika u bilo kojem od pokazatelja što znači da sva tri izvora imaju slične da prosječne tekstove prema računatim stilometrijskim značajkama i testovima čitljivosti.

	<b>JL</b>	<b>SD</b>	<b>VL</b>
<b>Broj riječi</b>	560,8 ( $\pm 465,5$ )	541,4 ( $\pm 382,8$ )	489,2 ( $\pm 413,5$ )
<b>Broj rečenica</b>	26,6 ( $\pm 27,3$ )	24,4 ( $\pm 22,0$ )	22,5 ( $\pm 23,8$ )
<b>Duljina riječi</b>	2,27 ( $\pm 0,14$ )	2,29 ( $\pm 0,16$ )	2,26 ( $\pm 0,13$ )
<b>Duljina rečenica</b>	23,7 ( $\pm 11,3$ )	24,6 ( $\pm 8,2$ )	24,4 ( $\pm 9,0$ )
<b>Leksička gustoća</b>	65,3 ( $\pm 4,6$ )	65,6 ( $\pm 4,4$ )	65,0 ( $\pm 4,7$ )
<b>Leksička redundancija</b>	35,3 ( $\pm 8,2$ )	34,9 ( $\pm 7,6$ )	35,1 ( $\pm 7,1$ )
<b>Gunningov indeks</b>	9,4 ( $\pm 4,5$ )	9,8 ( $\pm 3,3$ )	9,7 ( $\pm 3,6$ )
<b>Fleschov indeks</b>	-10,6 ( $\pm 15,9$ )	-14,1 ( $\pm 17$ )	-9,5 ( $\pm 15,2$ )
<b>Flesch-Kincaid indeks</b>	20,6 ( $\pm 4,5$ )	21,3 ( $\pm 3,9$ )	20,6 ( $\pm 3,9$ )

**Tablica T1.2.** Vrijednosti i standardne devijacije stilometrijskih pokazatelja i testova čitljivosti na korpusima JL, SD i VL

## 2. Usporedba kolumni četiriju dnevnih novina

Cilj ovog eksperimenta je usporediti tekstove kolumni četiriju dnevnih novina i pokazati postoji li značajnih razlika u stilu i čitljivosti. Korištena je ista metodologija kao i u prethodnom eksperimentu (vidi 1.2.).

### 2.1. Korpusi

Pri ovom eksperimentu uzeta su četiri korpusa dnevnih novina sačinjena isključivo od kolumnističkih tekstova odabranih autora u razdobljima i brojnosti kako je prikazano u tablici T2.1.

Korpus	Izvor	Razdoblje	Broj članaka
JL-kol	Jutarnji list	1.1.2014.-31.5.2015.	905
NL-kol	Novi list	1.1.2014.-30.6.2014.	625
SD-kol	Slobodna Dalmacija	1.6.2014.-31.5.2015.	127
VL-kol	Večernji list	1.1.2014.-31.5.2015.	830

**Tablica T2.1.** Izvori i veličine korpusa članaka kolumni

## 2.2. Rezultati

	JL-kol	NL-kol	SD-kol	VL-kol
<b>Broj riječi</b>	815.2 (±547.0)	654.8 (±414.2)	826.5 (±217.9)	684.2 (±559.5)
<b>Broj rečenica</b>	38.5 (±30.8)	25.8 (±22.9)	37.8 (±15.2)	34.2 (±37.0)
<b>Duljina riječi</b>	2.31 (±0.12)	2.26 (±0.13)	<b>2.45</b> (±0.27)	2.26 (±0.10)
<b>Duljina rečenica</b>	23.8 (±8.7)	<b>28.1</b> (±9.9)	23.5 (±6.8)	22.9 (±6.5)
<b>Leksička gustoća</b>	64.8 (±3.6)	63.0 (±3.8)	64.8 (±3.3)	63.3 (±3.4)
<b>Leksička redundancija</b>	36.3 (±8.0)	38.8 (±6.4)	36.6 (±5.4)	38.0 (±7.0)
<b>Gunningov indeks</b>	9.5 (±3.5)	<b>11.2</b> (±4.0)	9.4 (±2.7)	9.2 (±2.6)
<b>Fleschov indeks</b>	-15.2 (±13.4)	-14.0 (±16.0)	<b>-26.1</b> (±23.8)	-9.3 (±12.0)
<b>Flesch-Kincaid indeks</b>	21.3 (±3.7)	22.2 (±4.4)	22.7 (±4.2)	<b>20.2</b> (±3.0)

**Tablica T2.2.** Vrijednosti i standardne devijacije stilometrijskih pokazatelja i testova čitljivosti na korpusima JL-kol, NL-kol, SD-kol i VL-kol; (podebljane su komentirane vrijednosti)

Tekstovi kolumnista Novog lista imaju prosječno dulje rečenice. Tekstovi kolumnista Slobodne Dalmacije u prosjeku koriste nešto dulje riječi. Usporedbom tablica T1.2. i T2.2. čitamo kako tekstovi kolumnista imaju blago veću leksičku redundanciju od općenitih tekstova. Razlog tome je što kolumnisti pišu nešto dulje tekstove i detaljnije obrađuju pojedine teme.

Tekstovi kolumnista Večernjeg lista imaju u prosjeku nešto manji Flesch-Kincaidov indeks čitljivosti. Tekstovi kolumnista na Novom listu u prosjeku imaju veći Gunning Fog indeks od ostalih. Tekstovi kolumnista Slobodne Dalmacije imaju nešto niži Fleschov indeks čitljivosti.

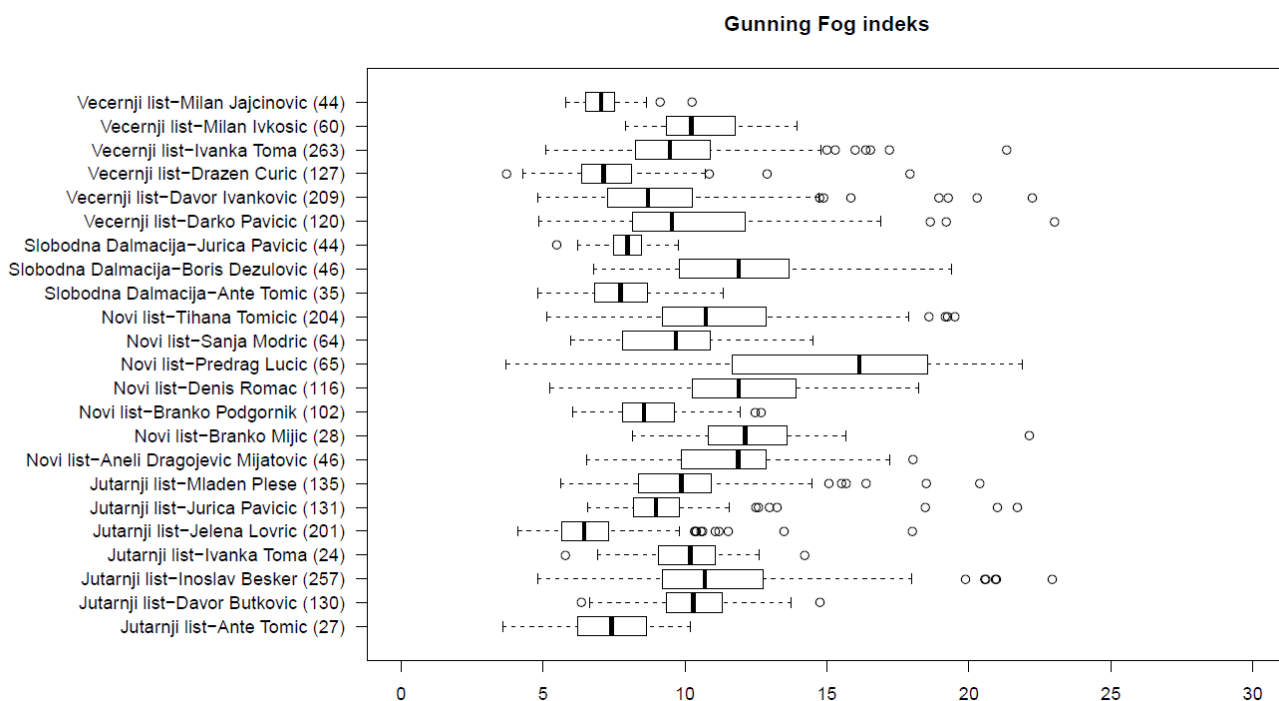
Kao zaključak se nameće da kolumnistički tekstovi u pojedinim tiskovinama imaju manje stilometrijske razlike kao i razlike u čitljivosti, ovisno razmatranom testu čitljivosti.

### 3. Usporedba čitljivosti tekstova pojedinih kolumnista

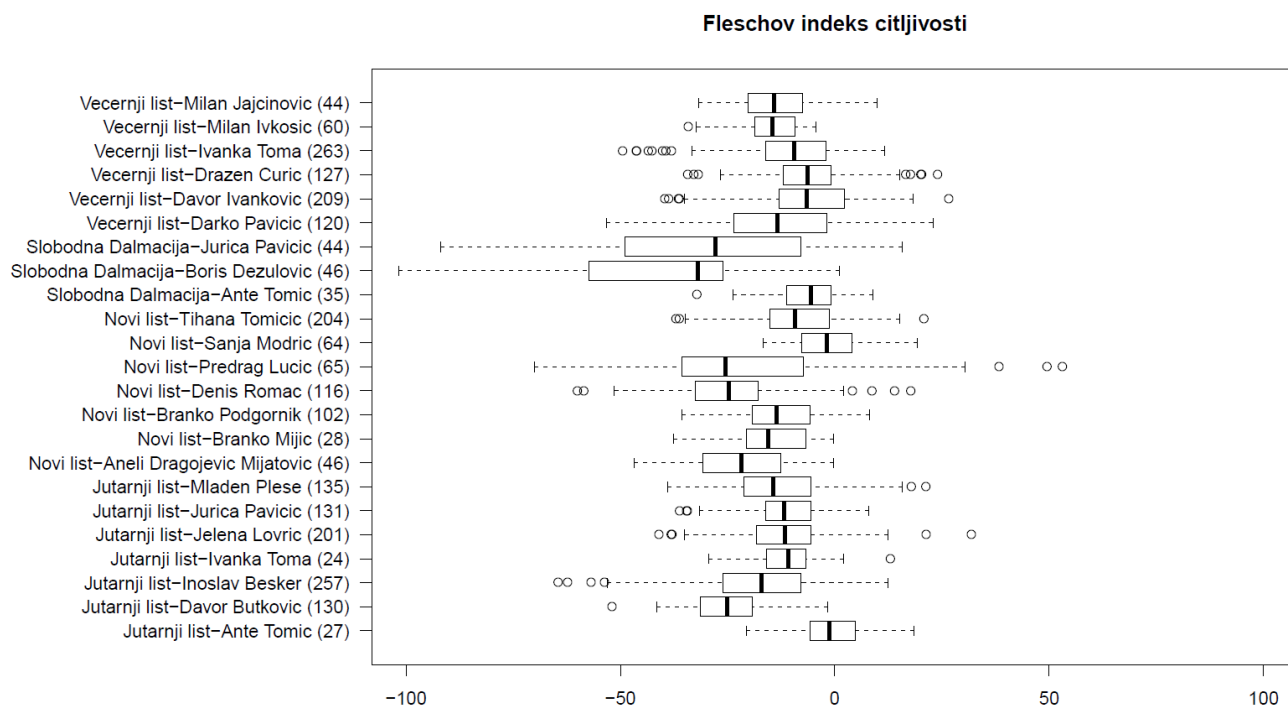
Cilj ovog eksperimenta je usporediti tekstove pojedinih kolumnista četiriju dnevnih novina. Korišten je isti korpus kao i u prethodnom poglavlju (vidi 2.1.) te ista metodologija (vidi 1.2.).

#### 3.1. Rezultati

Na slici S3.1. vidljivo je kako se tekstovi pojedinih kolumnista značajno razlikuju po Gunning Fog indeksu. Na primjer, Predrag Lucić ima najveću devijaciju od svih i u prosjeku najsloženije tekstove po Gunning Fog indeksu. Slijede ga Boris Dežulović, Denis Romac, Branko Mijić i Aneli Dragojević Mijatović. Po Gunning Fog indeksu najjednostavnije tekstove pišu Milan Jajčinović, Dražen Ćurić, Ante Tomić i Jelena Lovrić.

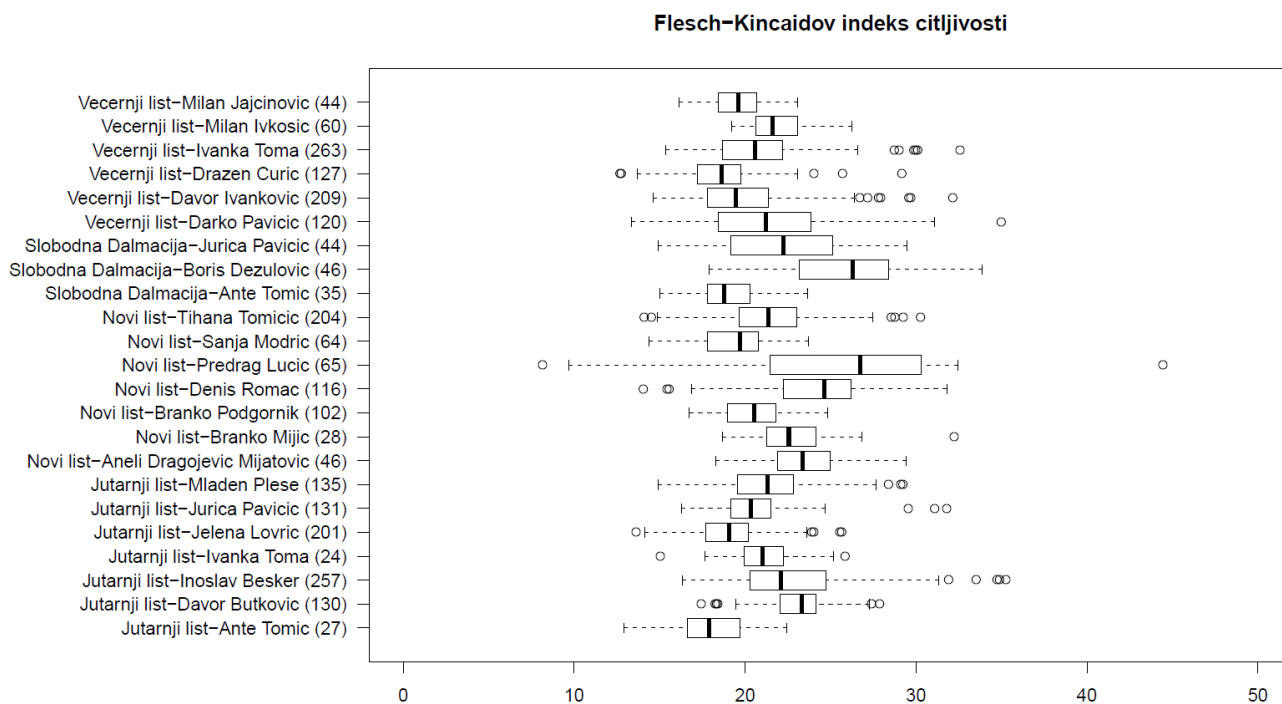


**Slika S3.1.** Distribucija vrijednosti Gunning Fog indeksa čitljivosti po kolumnistima; u zagradi je naveden broj obrađenih tekstova pojedinog autora; nacrtani su **box-and-whiskers** dijagrami na kojima se vide kvartili i medijan svake od distribucija



**Slika S3.2.** Distribucija vrijednosti Fleschovog indeksa čitljivosti po kolumnistima

Fleschov indeks čitljivosti najsloženijim ocjenjuje tekstove J. Pavičića, Dežulovića i Lucića, dok kao najjednostavnije identificira tekstove A. Tomića i S. Modrić.



**Slika S3.3.** Distribucija vrijednosti Flesch–Kincaid indeksa čitljivosti po kolumnistima

Flesch–Kincaidov indeks čitljivosti izdvaja tekstove Dežulovića i Lucića kao nasloženije, a Ante Tomića kao najjednostavnije.

### 3.2. Zaključak

Iz ove analize je jasno kako postoje razlike među autorima za svaki od indeksa čitljivosti. Kako se ove mjere razlikuju, očekivano daju različite rezultate, tj. za određen skup autora različit poredak, no postoje i određene korelacije. [interpretacija mjera]

[komentar: očekivano je da Lucić i Dežulović pobjeđuju po složenosti jer su etablirani kao dobri kolumnisti, ovakve komentare i interpretacije treba podebljati]

## 4. Usporedba stilometrijskih značajki PR objava s prosječnim novinskim tekstom

Cilj ovog eksperimenta je usporediti stilske značajke i testove čitljivosti PR objava kao posebne tekstne forme korištene u medijskoj produkciji s prosječnim tekstom dnevnih novina u koje se PR objave u konačnici kontekstno uklapaju.

### 4.1. Korpus i metodologija

Korpus je sačinjen od 91 originalnih PR objava poslanih predstavnicima medija u razdoblju od početka 2011. do sredine 2015. g. Pri analizi izuzeti su naslovi, podnaslovi, nadnaslovi i kontaktne informacije s potpisom (a koje se često navode u PR objavama).

### 4.2. Rezultat

Uspoređujući tablice T4.1. i T1.2. vidimo kako PR objave imaju značajno veću leksičku gustoću od prosječnog novinskog članka ili kolumne. Činjenični razlog za to nalazimo u tome što PR objave imaju manje funkcijskih riječi, pomoćnih glagola, zamjenica i veznika. Leksička redundancija ima slične vrijednosti onima na prosječnim tekstovima.

<b>Broj riječi</b>	467.5 ( $\pm 316.7$ )
<b>Broj rečenica</b>	20.2 ( $\pm 17.8$ )
<b>Duljina riječi</b>	<b>2.45</b> ( $\pm 0.13$ )
<b>Duljina rečenica</b>	27.2 ( $\pm 13.6$ )
<b>Leksička gustoća</b>	<b>71.4</b> ( $\pm 3.9$ )
<b>Leksička redundancija</b>	35.7 ( $\pm 8.6$ )
<b>Gunningov indeks</b>	10.9 ( $\pm 5.4$ )
<b>Fleschov indeks</b>	-25.9 ( $\pm 16.2$ )
<b>Flesch-Kincaid indeks</b>	23.6 ( $\pm 5.2$ )

*Tablica T4.1. Vrijednosti i standardne devijacije stilometrijskih pokazatelja i testova čitljivosti na korpusu PR objava; podebljane su vrijednosti značajno veće nego u tablici T1.2.*

Nadalje, PR objave koriste složenije (tj. duže) riječi, što se očituje i u mjerama čitljivosti. Naime, Gunningov indeks i Flesch-Kincaid indeks blago su veći, a Fleschov indeks blago manji, što znači da svaki od njih indicira nešto veću složenost PR objava u usporedbi s prosječnim novinskim tekstom.

## 5. Usporedba veličine vokabulara dnevnih novina 2015. g. i 1980. g.

Cilj eksperimenta je uspostaviti vezu između istraživanja Dubravka Škiljana [ ] i vidjeti ima li razlika u osnovnim kvantitativnim značajkama tekstova dnevnih novina danas i prije 35 godina.

### 5.1. Korpusi

Korpusi VL-2015, JL-2015 i SD-2015 nastali su tako se uzela u obzir cijela produkcija praćenja<sup>184</sup> tiskanih izdanja odgovarajućih novina u razdoblju od 21.3. do 1.4.2015. Razdoblje ovih triju korpusa namjerno je odabrano da bude identično razdoblju starih korpusa kako bi se uklonio efekt sezonalnosti koja postoji u medijima (npr. godišnji odmori i blagdani imaju specifičnosti koje se oslikavaju u medijskom prostoru) tj. kako bi usporedba bila što vjerodostojnija. Iz navedene produkcije uzeti su oni članci koji imaju barem 500 znakova (oko 4 rečenice). Donja granica od 500 znakova postavljena je kako bi se izuzele vrlo kratke tekstne forme (npr. izvještaji, izvješća burzovnih kretanja ili pak slikovne vijesti kojima je kratki tekst samo dodatna oprema slikovnom sadržaju). Broj članaka dobiven ovom metodom za svaki korpus prikazan je u tablici T5.1.

Korpus	Izvor	Broj članaka
VL-1980	Večernji list	120
V-1980	Vjesnik	120
VL-2015	Večernji list	631
JL-2015	Jutarnji list	565
SD-2015	Slobodna Dalmacija	637

**Tablica T5.1.** Izvori i veličine korpusa

Korpusi VL-1980 i V-1980 nisu bili direktno dostupni tijekom ovog istraživanja; naime, mjerenja veličine fonda riječi na tim korpusima proveli su kolege Škiljan i ostali [A]. U ovom se radu njihova mjerenja uspoređuju s našima. Korpusi VL-1980 i V-1980 nastali su na temelju 10 izdanja od 21.3. do 1.4.1980. Iz svake od šest novinskih rubrika ručno su odabrana po dva članka („prosječne duljine“, kako navodi autor), čime se dolazi do 120 ručno odabranih tekstova za oba korpusa.

<sup>184</sup> Praćenje medija koje uključuje skeniranje, OCR i stvaranje digitalne arhive obavila je tvrtka Presscut d.o.o. iz Zagreba

## 5.2. Metoda usporedbe veličine vokabulara dvaju korpusa

Veličina vokabulara tj. ukupan broj lema ovisi o veličini korpusa tj. o broju pojavnica. Jasno, što je korpus veći, očekuje se veći broj jedinstvenih lema. Ovom temom za hrvatski jezik bavio se Tuđman [B]. Veličina vokabulara nelinearno ovisi o veličini korpusa specifično po domeni i vrsti tekstova.

Kako se korpusi VL-1980, V-1980, VL-2015, JL-2015, SD-2015 koje uspoređujemo znatno razlikuju u veličini, nije moguće direktno uspoređivati njihovu veličinu vokabulara. Kako bismo izbjegli bilo kakve vrste heurističkih procjena prilikom usporedbe (a kako je to radio Škiljan [A]), odlučili smo se za usporedbu slučajnim odabirom. Naime, prilikom usporedbe vokabulara dvaju korpusa različitih veličina, za manji korpus broj se računa izravno. Iz većeg se korpusa slučajnim odabirom izdvaja broj tekstova koji odgovara veličini manjeg korpusa te se potom računa veličina vokabulara. Kako se bi se ispravila pristranost uzorkovanja, postupak se ponavlja više puta i rezultat se iskazuje kao srednja vrijednost sa standardnom devijacijom.

## 5.3. Metoda mjerenja veličine vokabulara

Prilikom brojanja jedinstvenih lema za pojedini korpus, potrebno je svaku pojavnicu svesti na normalni oblik što praktično znači uklanjanje padeža, broja, roda, vremenskog oblika, stupnja, itd. To je učinjeno pomoću dostupnog morfološkog rječnika za hrvatski jezik [C]. Za riječi koje se nisu našle u rječniku, rabljen je algoritam korjenovanja [D]. Brojevi i riječi s točkom u sebi se izostavljaju, a riječi s crticom se normiraju na dio prije crtice (npr. „HNS-u“ -> „HNS“). Kod homografnih pojavnica slučajnim odabirom se uzima jedna lema od svih mogućih. Utjecaj OCR grešaka nastalih sljepljivanjem riječi ublažava se preskakanjem riječi duljih od 24 znaka. Kao i kod Škiljana, u račun su ušle samo riječi tijela teksta, bez naslova i podnaslova.

Kako tekstovi naših korpusa (VL-2015, JL-2015 i SD-2015) sadrže određeni broj OCR grešaka koje neželjeno podižu broj jedinstvenih lema izračunatih strojnim normiranjem ( $H_1$ ), pribjegava se dodatnim brojanjem samo onih lema čiji se pojavni oblici nalaze u morfološkom rječniku ( $H_2$ ). U tom se rječniku pored običnih riječi u svim formama nalaze i osobna imena, prezimena te određeni skup toponima. Kako nijedan rječnik nije potpuno sveobuhvatan u kontekstu konstantno mijenjajućeg jezika, broj  $H_2$  nešto je manji od stvarne veličine vokabulara korpusa jer su nužno izostavljene pojedine rijetke riječi te pojedina imena. Iz opisanih razloga možemo reći da se stvarni broj lema nalazi između  $H_1$  i  $H_2$ .

#### 5.4. Rezultat usporedbe veličine vokabulara

U prva dva retka tablice T5.2. nalaze se originalni brojevi iz Škiljanova istraživanja, a u sljedeća tri retka nalaze se brojevi izračunati nad suvremenim korpusima. Iz tablice T5.2. čita se da Večernji list, Jutarnji list i Slobodna Dalmacija iz 2015. g. na slučajnih 120 članaka imaju veličinu vokabulara u prosjeku sličnu Večernjem listu iz 1980. g. (za VL-2015, JL-2015 i SD-2015 vrijedi  $H_2 < H_{VL-1980} < H_1$ ). Za iste korpusse vrijedi da imaju vrlo slične veličine u broju riječi što govori da su tekstovi tih korpusa sličnih duljina. U odnosu na sve ostale korpusse, Vjesnik iz 1980. ima veći broj pojava te shodno i veći broj lema zbog prosječno duljih tekstova.

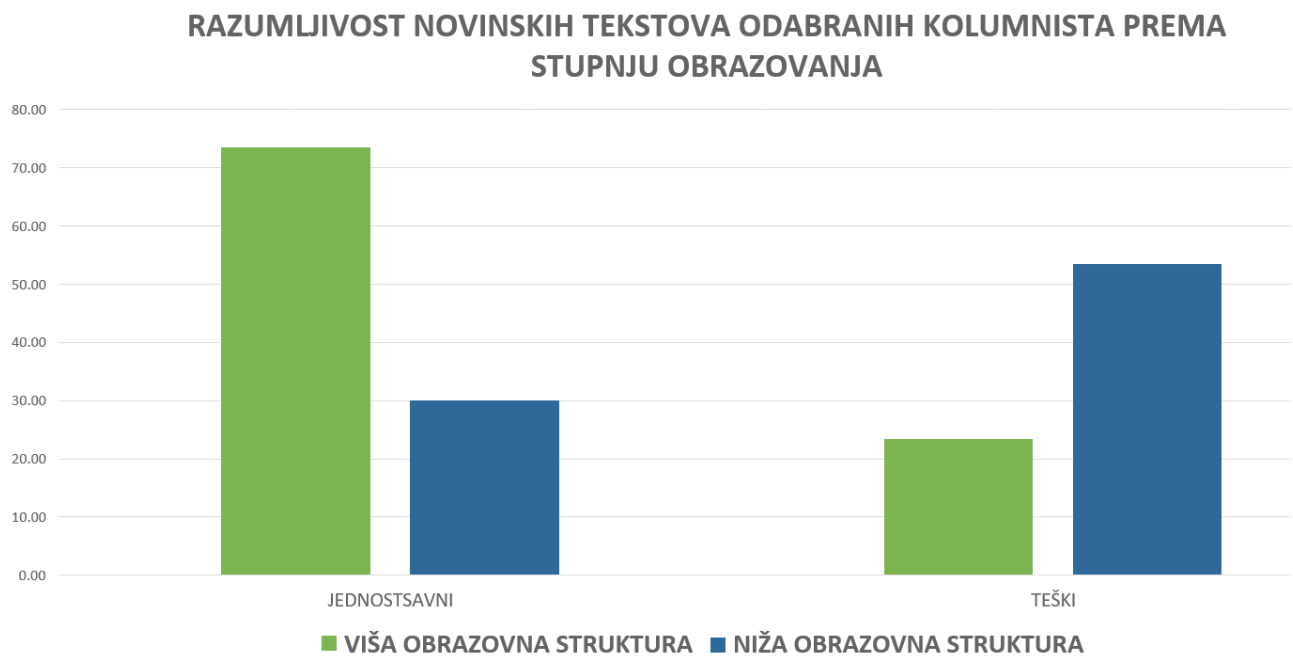
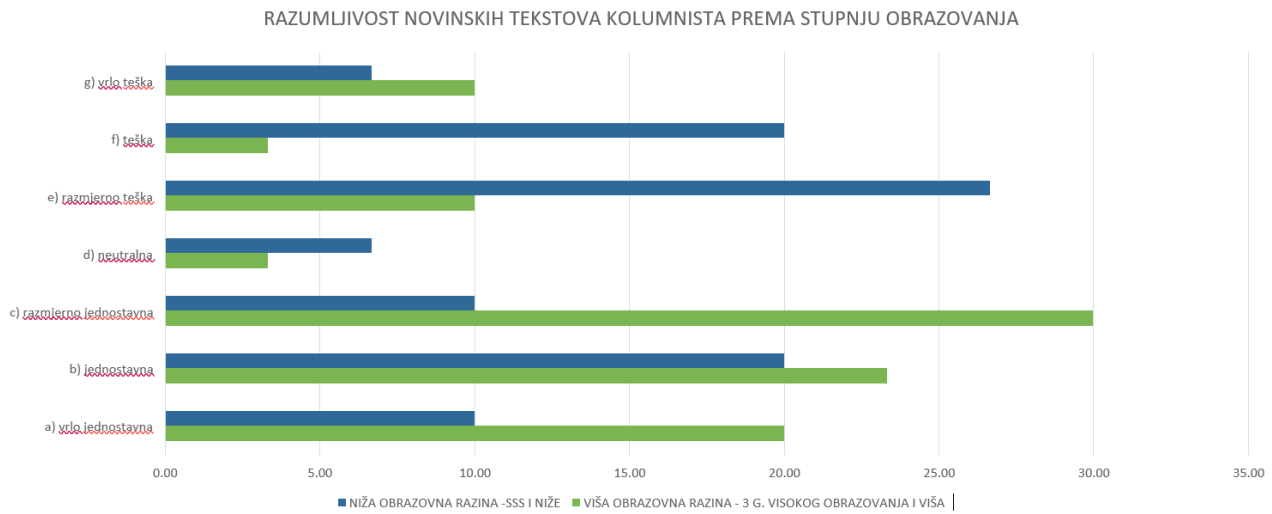
Korpus	N	H	H <sub>1</sub>	H <sub>2</sub>
VL-1980	54525	8653		
V-1980	75754	10124		
VL-2015	51503 (±4069)		9005 (±414)	6925 (±280)
JL-2015	54263 (±3727)		9621 (±484)	7286 (±315)
SD-2015	55993 (±3758)		10552 (±458)	7719 (±288)

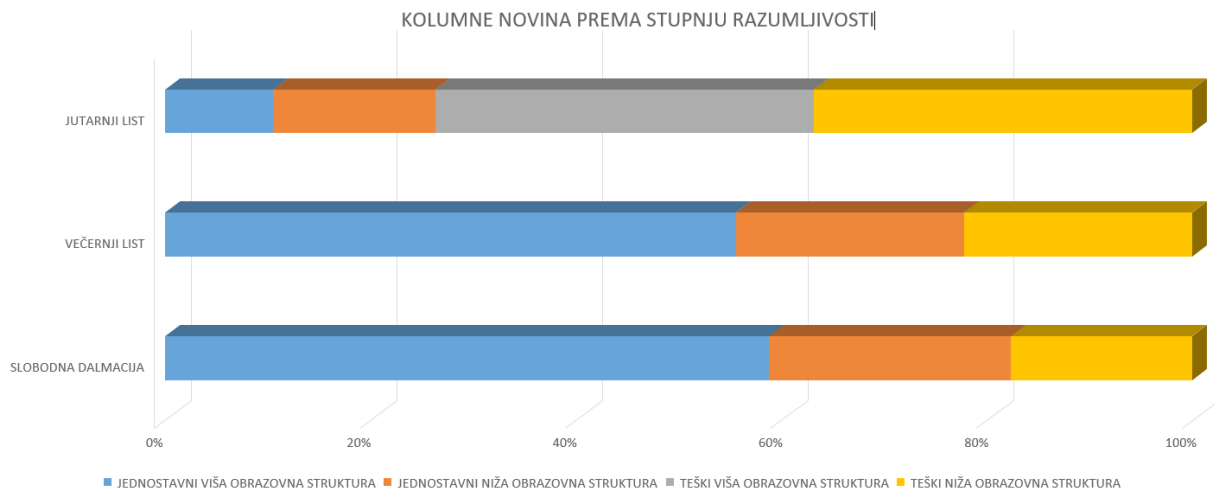
**Tablica T5.2.** Veličina vokabulara po korpusima za 120 tekstova, gdje su **N**—broj pojava korpusa, **H**—broj ručno normiranih jedinstvenih lema, **H<sub>1</sub>**—broj lema dobiven strojnim normiranjem, **H<sub>2</sub>**—broj lema dobiven strojnim normiranjem gdje se pribrajaju samo leme s poznatim pojavnicama

U suštini, za popularne dnevne novine Večernjeg lista ovi brojevi govore da su unatoč društvenim promjenama, digitalnoj revoluciji i novim medijima, osnovne značajke tekstne forme novinskog članka – duljina u broju riječi i raznolikost rječnika ostale slične. Također, Jutarnji list i Slobodna Dalmacija iz 2015. prate osnovne značajke prosječnog članka Večernjeg lista iz 1980. g. Za preciznija mjerenja bit će potrebno digitalno pripremiti originalne korpusse iz 1980., ali i u suvremenima ručno ispraviti OCR greške.



## 6. Razumljivost





Zbog ograničenja prostora ne možemo prikazati sve istraživane čestice razumljivosti.

## 7. Zaključci vezani uz hipoteze:

H1 – Postoji interaktivni odnos koherencije teksta i razine ekspertnosti čitatelja

- Hipoteza potvrđena
- Podaci su indikativne naravi jer istraživanje razumljivosti nije provedeno na nacionalno reprezentativnom uzorku zbog nedostatka sredstva /terensko istraživanje/

H2 – Postoji razlika u mjerama čitljivosti kod hrvatskih dnevnih novina

- Hipoteza nije potvrđena
- Ne postoji statistički značajna razlika u mjerama čitljivost kod hrvatskih dnevnih novina testiranih kroz 8 varijabli i kroz 3 testa čitljivosti
- Postoji značajna razlika u mjerama čitljivosti između tekstova hrvatskih dnevnih novina i tekstova kolumnista u njima što utiče na razumljivost potonjih sukladno različitim obrazovnim nivoima ispitanika

H3 – Priopćenja za javnost nisu prilagođena kriterijima stila novinskog pisanja /mjerama čitljivosti/

- Hipoteza potvrđena
- Rezultat dobiven ovom hipotezom može imati velike praktične koristi za odnose s javnostima u smislu poboljšanja efekta komunikacije sa medijima

## 8. Rasprava

Brojni su mogući pravci budućih istraživanja, što proizlazi iz teorijskog uvoda, I to I u segmentu I čitljivosti I razumljivosti.

### Dodatak A – stilometrijski pokazatelji i testovi čitljivosti

#### A.1. Broj riječi

Prebrojava se ukupan broj riječi bez obzira na interpunkcije i razmake. Pritom se vodi računa o specifičnostima hrvatskog jezika poput točaka u kraticama ili crtica u riječima.

#### A.2. Broj rečenica

Prebrojava se ukupan broj rečenica vodeći računa o raznolikosti interpunkcija koje završavaju rečenice te korištenju točaka u kraticama i rednim brojevima. Novi redak prethodne interpunkcije uzima se kao početak nove rečenice.

#### A.3. Prosječna duljina riječi u slogovima

U tekstu se pronalaze sve riječi i za njih se računa broj slogova, a potom i prosjek za cijeli tekst. Slogovi su osnovni fonološki blokovi riječi i njihov broj određen algoritmom napisanom prema Hrvatskoj gramatici [E]. Pri računu broja slogova pazi se na spojene samoglasnike, dvojaku ulogu slova „r“ te refleks jata „ije“<sup>185</sup>. Broj slogova u riječima preciznije određuje složenost od broja znakova. Očekuje se da je tekst složeniji tim više čim je veća prosječna duljina riječi mjerena u broju slogova.

#### A.4. Duljina rečenica

Tekst se razlama na rečenice i za svaku rečenicu se računa duljina u riječima, a potom se računa prosjek za cijeli tekst. Interpunkcije i praznine se ne uzimaju u obzir. Dulje rečenice čine tekst složenijim.

#### A.5. Leksička gustoća

Nakon pronalaska riječi u tekstu, one se svrstavaju u tzv. leksičke i funkcijske riječi. Funkcijske riječi uključuju čestice, veznike, prijedloge, nepromjenjive priloge (koji nemaju pridjevski par), zamjenice, uzvike i sve oblike glagola „biti“ i „htjeti“. Leksičke riječi uključuju sve preostale kategorije, tj. pridjeve, imenice, glagole i promjenjive priloge (koji imaju pridjevski par).

Konačna mjera računa se kao  $100 * \text{broj leksičkih riječi} / \text{broj svih riječi}$ . Tekstovi s nižom leksičkom gustoćom jednostavniji su za razumijevanje, no ugrubo rečeno, nose manje informacija. Primjerice, govorni jezik ima mnogo manju leksičku gustoću od pisanog jezika.

<sup>185</sup> Tako su primjerice riječi „mat“ i „vrh“ jednosložne, „auto“ i „mlijeko“ dvosložne, a riječ „krtičnjak“ trosložna.

## A.6. Leksička redundancija

Leksička redundancija kazuje koji je udio riječi koje se u tekstu ponavljaju tj. pojavljuju barem dva puta. Konačna mjera računa se kao  $100 * \text{broj riječi s ponavljanjem} / \text{broj svih riječi}$ . Tekstovi koji imaju višu leksičku redundanciju detaljnije i koncentriranije pišu o nekoj temi.

## A.7. Gunningov indeks magle

Gunningov indeks magle (engl. Gunning Fog index) mjeri čitljivost engleskih tekstova. Često se koristi za procjenu godina formalnog obrazovanja potrebnih za čitanje teksta u prvom prolazu. Iako razvijen za engleski jezik, možemo ga koristiti i za hrvatski jezik ako nam je bitna samo relativna složenost teksta u nekim izvorima, a ne točna procjena godina obrazovanja potrebnih za razumijevanje.

Formula kojom se računa Gunningov indeks magle je:  $0.4 * [(\text{broj riječi}/\text{broj rečenica}) + 100 * (\text{broj kompleksnih riječi} / \text{broj riječi})]$ .

Određivanje je li riječi kompleksna računa se po sljedećem kriteriju: ako riječ bez zadnjeg samoglasnika ima tri ili više slogova i ako nije vlastita imenica, onda je kompleksna. Popis vlastitih imenica sastavljen je iz velikog popisa hrvatskih i stranih imena te popisa hrvatskih toponima.

Što je Gunningov indeks magle veći, to je tekst teži za brzo razumijevanje.

## A.8. Fleschov indeks

Slično kao i Gunningov indeks magle, Fleschov indeks indicira težinu čitanja teksta, a računa se po formuli:  $206,835 - 1,015 * (\text{broj riječi}/\text{broj rečenica}) - 84,6 * (\text{broj slogova}/\text{broj riječi})$ .

Niža vrijednost indicira tekst kojeg je teže čitati.

## A.9. Flesch-Kincaid indeks

Slično kao i Gunningov indeks magle, Flesch-Kincaidov indeks indicira težinu čitanja teksta. Često se koristi u području obrazovanja. Računa se po formuli:  $0,39 * (\text{broj riječi}/\text{broj rečenica}) + 11,8 * (\text{broj slogova}/\text{broj riječi}) - 15,59$ .

Viša vrijednost indicira složeniji tekst.

## **Bibliografija:**

[A] Lingvističko istraživanje Večernjeg lista, Zagreb, 1980.

[B] Miroslav Tuđman, ...

[C] Croatian Morphological Lexicon v5.0, HML5, dostupno na <http://meta-share.ffzg.hr/repository/browse/croatian-morphological-lexicon-v50/2d429672703d11e28a985ef2e4e6c59e27b37c59b92d42a5be839f7daff7ecfb/>

[D] Ljubešić, Nikola, Damir Boras, and Ozren Kubelka. "Retrieving information in Croatian: Building a simple and efficient rule-based stemmer." *Digital information and heritage/Seljan, Sanja (2007): 313-320.*

[E] „Barić; Lončarić; Malić; Pavešić; Peti; Zečević; Znika: Hrvatska gramatika“

## **Literatura:**

Al-khalifa, H. S., Al-ajlan, A. A. (2010). Automatic Readability Measurements of the Arabic Text: An Exploratory study. *The Arabian Journal for Science and Engineering*, 35(20). 103-125.

Benjamin, R.G. (2011). Reconstructing Readability: Recent Developments and Recommendations in the Analysis of Text Difficulty. *Educational Psychology Review*, 24(1), 63-88.

Britton, B., & Gülgöz, S. (1991). Using Kintsch's computational model to improve instructional text: Effects of repairing inference calls on recall and cognitive structures. *Journal of Educational Psychology*, 83(3), 329–345. doi:10.1037/0022-0663.83.3.329.

DuBay, W.H. 2004. The principles of readability. Skinuto 11.7.2014. s <http://www.impact-information.com/impactinfo/readability02.pdf>.

Gallagher, T.L., Fazio, X., Gunning, T.G. (2012). Varying Readability of Science-based Text in Elementary Readers: Challenges for Teachers . *Reading Improvement*, 49(3), 93-112.

Graesser, A. C., Gernsbacher, M. A., & Goldman, S. R. (1997). *Cognition*. U: T. A. van Dijk i T. A. van Dijk (Ur.). *Discourse as structure and process: Discourse studies: A multidisciplinary introduction*, vol. 1 (pp. 292–319). Thousand Oaks: Sage

Graesser, A., McNamara, D., Louwerse, M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments & Computers*, 36(2), 193–202.

Hiebert, E.H., Cruz, S. (2010). Beyond Single Readability Measures : Using Multiple Sources of Information in Establishing Text Complexity. *Journal od Education*, 191(2), 33-42.

Lin, S., Su, C., Lai, Y., Yang, L., & Hsieh, S. (2009). Assessing text readability using hierarchical lexical relations retrieved from WordNet. *Computational Linguistics and Chinese Language Processing*, 14(1), 45–84.

- McNamara, D., Kintsch, E., Songer, N., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, 14(1), 1–43.
- vor der Brück, T., Hartrumpf, S., & Helbig, H. (2008). A readability checker with supervised learning using deep indicators. *Informatica*, 32(4), 429–435.

## New Analytical Models for determining Text Comprehensibility Standards in Mass-Media Communication

### **Abstract**

*In this paper we are focused on the comprehensibility of mass-media texts within the theoretical context of media literacy and media competence, in order to determine the elements of standards for their comprehensibility.*

*Our comparative analysis of texts (primarily political news reports) from Croatian national daily newspapers confirmed the hypothesis about their partial incomprehensibility and considerable variation in comprehensibility between different media outlets. We analysed 11 different categories: the amount of text in a sentence, word length, symbol length, average sentence length, average word length in syllables, lexical density, lexical redundancy, the density of entities, the complexity of words by their syllable length and frequency, and measures based on the estimated complexity of words in the newspaper corpus.*

*By observing groups of individuals with different age, level of education, and social background, we found significant differences in their perception and understanding of mass-media texts, depending on their age and education. This leads us to conclude that standards for comprehensibility need to be determined by taking into account the previously obtained research findings, while continuing work to establish all relevant factors of the perception of media text, as well as ways in which they affect its comprehensibility.*

*We selected those measures that are best suited for automated assessment of comprehensibility of newspaper reports in the Croatian language by pairing comprehensibility assessments from the consumers of media with the results of text complexity computer analysis. By using selected automated measures of text readability and the human perception of text readability, we attempted to make a contribution to determining the elements of text readability standards, as the minimum requirement for text readability.*

**Key words:** *media literacy, mass communication, media competencies, media language, analytical models.*



This journal is open access and this work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.