

JAKA ČIBEJ

STATISTIČNA ANALIZA IZGOVORA ČRKE *l* V SLOVENSKEM OBLIKOSLOVNEM LEKSIKONU *SLOLEKS*

COBISS: 1.01

[HTTPS://DOI.ORG/10.3986/JZ.31.1.03](https://doi.org/10.3986/JZ.31.1.03)

Dvounost izgovora črke *l* v položaju pred soglasniškim grafemom (*polža*, *alge*, *volilca*) predstavlja problem v grafemsko-fonemski pretvorbi za slovenščino in kljub večkratni obravnavi v slovenskih jezikovnih virih še vedno ni razrešena. Zaradi pomanjkanja empiričnih strojno berljivih podatkov o izgovoru rojenih govorcev in govork smo ročno označili množico približno 6.000 leksemov (in njihovih pregibnih oblik) iz *Slovenskega oblikoslovnega leksikona Sloleks* glede na izgovor (/l/, /ɫ/ ali variantno) ter opravili statistično analizo, ki razkrije najbolj problematične točke. Izsledki bodo v pomoč pri razvoju modela, ki napoveduje izgovor črke *l* pred soglasniškim grafemom.


Ključne besede: izgovor črke *l*, grafemsko-fonemska pretvorba, oblikoslovni leksikon, slovenščina, statistična analiza

A Statistical Analysis of the Pronunciation of the Letter *l* in the *Sloleks Morphological Lexicon of Slovene*

The ambiguous pronunciation of the letter *l* before consonant graphemes (e.g., *polža*, *alge*, *volilca*) poses a problem for grapheme-to-phoneme conversion for Slovenian. Despite having been addressed in multiple Slovenian language resources, the problem is still unresolved. Because of a lack of empirical and machine-readable data on the pronunciation of native speakers, a dataset of approximately six thousand lexemes (and their inflected forms) was manually annotated from the *Sloleks Morphological Lexicon of Slovene* according to their pronunciation (/l/, /ɫ/, or both) and a statistical analysis was performed to reveal the most problematic points. The findings will be useful in developing a model to predict the pronunciation of the letter *l* before a consonant grapheme.

Keywords: pronunciation of the letter *l*, grapheme-to-phoneme conversion, morphological lexicon, Slovenian, statistical analysis

Jaka Čibej ■ Univerza v Ljubljani, Filozofska fakulteta ■ jaka.cibej@ff.uni-lj.si ■

 <https://orcid.org/0000-0002-3037-6848>

Prispevek je nastal v okviru raziskovalnega projekta *Temeljne raziskave za razvoj govornih virov in tehnologij za slovenski jezik* (MEZZANINE, J7-4642) in raziskovalnega programa *Jezikovni viri in tehnologije za slovenski jezik* (P6-0411), ki ju financira Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije (ARIS), ter projekta *Razvoj slovenščine v digitalnem okolju* (RSDO), ki sta ga med letoma 2020 in 2023 sofinancirali Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj (operacija se je izvajala v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014–2020). Iskrena hvala tudi vsem označevalcem in označevalkam.

