

# Izrada tezaurusa digitalizirane hrvatske paremiološke građe

**Kristina Feldvari**

kfeldvari@ffos.com

**Irena Poje**

ipoje@ffos.hr

**Borna Petrović**

bpetrovic@ffos.hr

Filozofski fakultet Sveučilišta Josipa Jurja Strossmayera u Osijeku

## Sažetak

*Projekt “Digitalizacija hrvatske gramatikografske i paremiološke građe” projekt je Centra za istraživanje kulturnog identiteta i regionalne baštine Filozofskog fakulteta u Osijeku. Projekt je u prvoj fazi obuhvatio digitalizaciju zbirki poslovice Josipa Kekeza (1990.) i Tvrтка Čubelića (1975.). Druga faza, planirana za 2024. godinu, obuhvaća postavljanje sučelja tezaurusa kojim će digitalizirane poslovice (otprilike N=4000) postati dostupne široj javnosti i istraživačima na području paremiologije, usmenih žanrova, frazeologije i dijakronijskoga jezikoslovlja.*

*Nastavno na prvu fazu projekta, u kojoj je prikupljeno leksičko gradivo iz literature i drugih izvora, metodologija ove faze podrazumijeva razradu i unos klasifikacijskih kategorija i njihovih klasifikacijskih oznaka prema autorima Kekezu, Čubeliću i Kuusiju te unos svih poslovice iz digitalizirane literature. Za svaku pojedinu poslovicu unesen je izvor i broj izvora poslovice, napomene iz izvora te varijante poslovice (ukoliko postoje). Poslovica je potom povezana sa svim navedenim klasifikacijskim kategorijama kojima pripada, a dodane su joj i ključne riječi koje su povezane sa sinonimima i srodnim terminima. Ključne riječi oblikovane su kao imenice u nominativu jednine, pridjevi su pretvoreni u imenice, a glagoli u glagolske imenice.*

*Po završetku unosa svih poslovice (planirano za kraj 2024. godine) bit će izrađeno korisničko sučelje u kojem će poslovice biti pretražive po sljedećim kriterijima: ključne riječi i svi povezani sinonimi i srodni termini, klasifikacijske kategorije, značenje, te varijante poslovice (ukoliko postoje). Prednost ovakvog tezaurusa vidljiva je u poboljšanju pretraživanja kroz povećanje točnosti i odziva pretraživanja, kao i u mogućnosti pregledavanja abecednog prikaza ključnih riječi kao pomoći pri osmišljavanju strategija pretraživanja.*

**Ključne riječi:** *tezaurus, paremiologija, baština, digitalizacija, Hrvatska, poslovice.*

## Creation of a Thesaurus of Digitized Croatian Paremiological Material

**Kristina Feldvari**  
kfeldvari@ffos.com

**Irena Poje**  
ipoje@ffos.hr

**Borna Petrović**  
bpetrovic@ffos.hr

Faculty of Humanities and Social Sciences  
Josip Juraj Strossmayer University of Osijek

### Abstract

*The project “Digitization of Croatian Grammatical and Paremiological Material” is a project of the Center for Research of Cultural Identity and Regional Heritage at the Faculty of Humanities and Social Sciences in Osijek. In the first phase, the project included the digitization of proverb collections by Josip Kekez (1990) and Tvrtko Čubelić (1975). The second phase, planned for 2024, includes the setting up of the thesaurus interface through which digitized proverbs (approximately N=4000) will become available to the broader public and researchers in the field of paremiology, oral genres, phraseology and diachronic linguistics.*

*Following on the first phase of the project, in which lexical material from literature and other sources was collected, the methodology of this phase involves the development and entry of classification categories and their classification codes according to the authors Kekez, Čubelić and Kuusi, and the entry of all proverbs from the digitized literature.*

*For each individual proverb, the source and source number, notes from the source and proverb variants (if they exist) are entered. The proverb is then linked to all of the listed classification categories to which it belongs, and keywords that are linked to synonyms and related terms are also added to it. Keywords are formed as nouns in the nominative singular, adjectives are transformed into nouns, and verbs into verbal nouns.*

*Upon completion of the entry of all proverbs (planned for the end of 2024), a user interface will be created in which proverbs will be searchable according to the following criteria: keywords and all associated synonyms and related terms, classification categories, meaning, and variants of proverbs (if they exist). The advantage of such a thesaurus is visible in the improvement of searching by increasing the accuracy and recall of the search, as well as in the ability of browsing an alphabetical display of keywords as an aid in designing search strategies.*

**Keywords:** *thesaurus, paremiology, heritage, digitization, Croatia, proverbs.*

# IZRADA TEZAUUSA DIGITALIZIRANE HRVATSKE PAREMIOLOŠKE GRAĐE

izv. prof. dr. sc. Kristina Feldvari  
kfeldvari@ffos.hr

Irena Poje  
ipoje@ffos.hr

Borna Petrović  
bpetrovic@ffos.hr

Filozofski fakultet  
Sveučilište J. J. Strossmayera u Osijeku

## CILJEVI

Područje digitalizacije hrvatske paremiološke građe u ovoj godini podrazumijeva **postavljanje sučelja** kojim bi **digitalizirane poslovice** skupljene od 12. do 19. stoljeća postale dostupne široj javnosti i istraživačima na području paremiologije, usmenih žanrova, frazeologije i dijakronijskoga jezikoslovlja. Za razliku od nekih drugih jezika u kojima već postoje paremiološke zbirke ili rječnici poslovice u tiskanom ili elektroničkom obliku, većina je hrvatskih poslovice dokumentirana u zbirama iz 19. i 20. stoljeća (točnije od 1846. do 1987. godine) i **do sada ne postoji nikakva zbirka hrvatskih poslovice koja bi odgovarala očekivanjima suvremene paremiologije i paremiografije.**

Cilj je digitalizacije paremiološke građe ove godine digitalizirati zbirke poslovice autora Josipa Kekeza (1990.) *SVAKI JE KAMEN DA SE KUĆA GRADI* i Tvrтко Čubelića (1975.) *USMENE NARODNE PITALICE, POSLOVICE I ZAGONETKE*, koje su dostupne u PDF formatu u digitalnoj knjižnici Filozofskoga fakulteta u Osijeku **učiniti dostupnima široj javnosti**, učenicima, studentima te svim istraživačima narodnoga blaga **putem mrežnog tezaurusa hrvatskih poslovice.**

Krajnji je cilj izrade tezaurusa osim očuvanja narodnoga blaga i dijakronijskog istraživanja hrvatske jezične baštine njegova didaktička funkcija i primjena na području podučavanja hrvatskoga jezika.



## AKTIVNOSTI

- 1 filološka analiza prikupljene građe:** jezikoslovna analiza poslovice (otprilike N=4000) podrazumijeva u ovome projektu opis značenja poslovice te dodavanje ključnih riječi povezanih sa značenjem svake poslovice
- 2 sadržajna analiza (intelektualna analiza i predmetno označavanje) prikupljene građe:** izbor i izrada termina-deskriptora, pripadajućih kategorija-faseta i podkategorija-podfaseta
- 3 izrada mrežnog sučelja:** kreiranje mrežnog mjesta i sustava za pretraživanje poslovice od 12. do 19. st. pomoću prirodnog jezika i kontroliranog tezaurusa te postavljanje na mrežne stranice Centra
- 4 diseminacija i promocija rezultata projektne aktivnosti:** izrada publikacija, sudjelovanje na znanstvenim i stručnim skupovima, promocija projektne aktivnosti

## METODOLOGIJA

Nastavno na prvu fazu projekta, u kojoj je prikupljeno leksičko gradivo iz literature i drugih izvora, metodologija ove faze podrazumijeva:

- **razradu i unos klasifikacijskih kategorija i njihovih klasifikacijskih oznaka** prema autorima Kekezu, Čubeliću i Kuusiju
- **te unos svih poslovice iz digitalizirane literature**

Za svaku pojedinu poslovice unesen je:

- **izvor i broj izvora** poslovice
- **napomene** iz izvora
- **varijante poslovice** (ukoliko postoje)

Poslovice je potom:

- **povezana** sa svim navedenim klasifikacijskim kategorijama kojima pripada
- **označena ključnim riječima** koje su **povezane sa sinonimima i srodnim terminima**

Ključne riječi oblikovane su kao imenice u nominativu jednine, pridjevi su pretvoreni u imenice, a glagoli u glagolske imenice.

Po završetku unosa svih poslovice (planirano za kraj 2024. godine) bit će izrađeno **korisničko sučelje** u kojem će **poslovice biti pretražive** po sljedećim kriterijima: ključne riječi i svi povezani sinonimi i srodni termini, klasifikacijske kategorije, značenje, te varijante poslovice (ukoliko postoje). Prednost ovakvog tezaurusa vidljiva je u **boljšanju pretraživanja** kroz povećanje točnosti i odziva pretraživanja, kao i u mogućnosti pregledavanja abecednog prikaza ključnih riječi kao pomoći pri osmišljavanju strategija pretraživanja.

## OČEKIVANI REZULTATI

Mrežnim tezaurusom s korpusom poslovice od 12. do 19. stoljeća u Hrvatskoj će se stvoriti **jedinstvena paremiološka baza koja će se moći proširiti** i suvremenim poslovicama, antiposlovicama te srodnim oblicima.

Time će se ostvariti temelji za ne samo dijakronijsko istraživanje hrvatske jezične baštine, već će se ispuniti preduvjeti za didaktiziranje hrvatskoga paremiološkoga blaga, čime se **uvelike doprinosi očuvanju i popularizaciji hrvatskoga paremiološkog naslijeđa** među populacijom školske djece.

Rezultatima istraživanja pridonijet će se i **popularizaciji znanosti i povećavanju vidljivosti** znanstvenih aktivnosti Filozofskoga fakulteta u Osijeku.

