



# Verbal Multiword Expressions in Croatian

**Krešimir Šojat, Matea Filko, Daša Farkaš**

University of Zagreb, Faculty of Humanities and Social Sciences  
{ksojat, msrebaci, dberovic}@ffzg.hr



Investing  
in future

# Why?

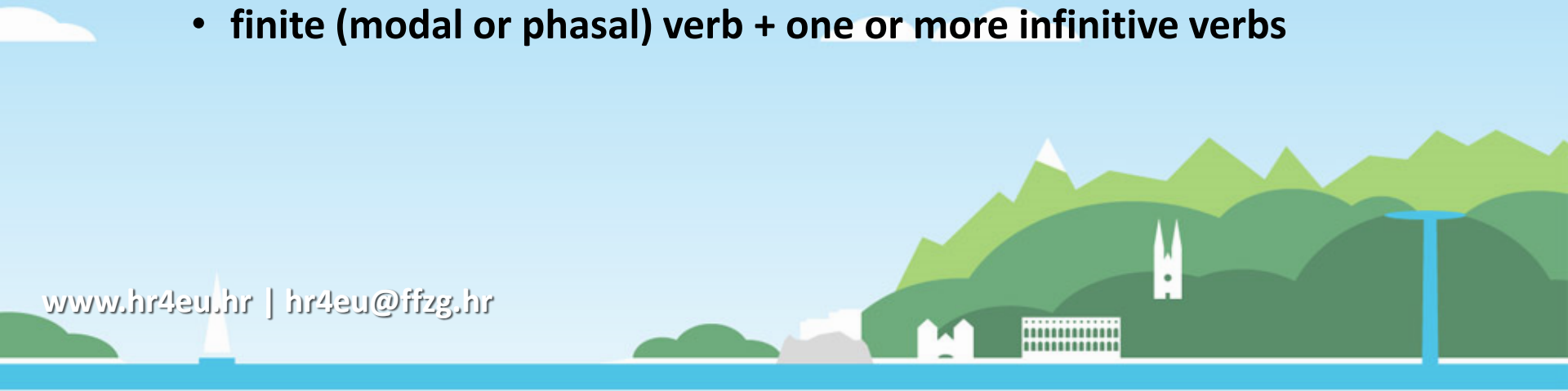
1. there is no previous research done on the identification and annotation of verbal MWEs in Croatian language resources / based on corpus data
2. there is no resource which would enable an extensive research of (verbal) MWEs in Croatian
3. there is no clear linguistic criteria for the classification of verbal MWEs



Investing  
in future

# What?

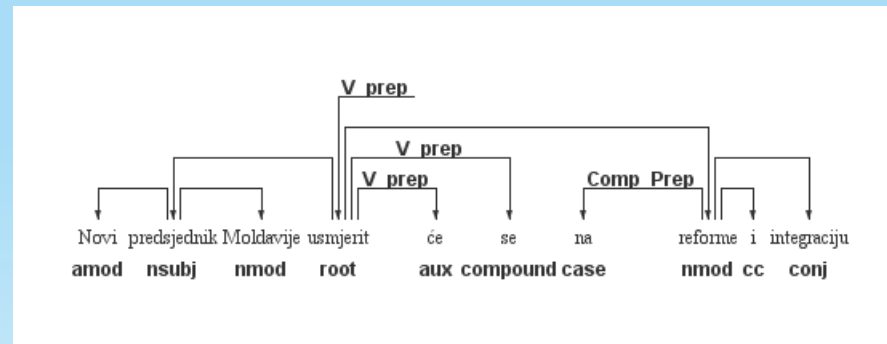
- four types of MWEs were annotated:
  - **prepositional verb constructions**
    - verb + PP
  - **light verb constructions**
    - light verb + NP / PP
  - **verbal idioms**
    - phrasemes
  - **complex predicate constructions**
    - finite (modal or phasal) verb + one or more infinitive verbs



Investing  
in future

# Where?

- Universal Dependency Treebank (UDT) for Croatian
  - 3557 sentences from SETimes parallel corpus
  - preliminary research > more diverse texts than in the HOBS – more MWEs



# Annotation process



Investing  
in future

1. initial list of 80 verbal MWEs from the existing literature – only 14 of them were found in UDT

- MWEs in the literature are not extracted from the real language data / corpora

2. manual annotation of verbal MWEs in UDT

- to obtain the list of verbal MWEs based on the real language data

# Results



Investing  
in future

- database containing all extracted instances of four types of verbal MWEs
  - two biggest groups:
    - phrasal / prepositional verbs – 371
    - light verb constructions – 252
- for the discussion visit our poster area!



Thank you for your attention.



Investing  
in future

[www.hr4eu.hr](http://www.hr4eu.hr) | [hr4eu@ffzg.hr](mailto:hr4eu@ffzg.hr)