



Towards building a preposition database for Croatian – a constructional perspective

Ivana Matas Ivanković, Institute for Croatian Language

imatas@ihjj.hr

Daniela Katunar, University of Zagreb, Faculty of Humanities and Social Sciences

dkatunar@ffzg.unizg.hr

Project *Croatian Prepositions in Use – Semantic and Syntactic Analysis (HRPA)*

IP-2022-10-6867, Croatian Science Foundation

21. rakenduslingvistika kevadkonverents 18.-19. aprill 2024

Tallinn, Estonia

Project *Croatian Prepositions in Use – Semantic and Syntactic Analysis* (HRPA)

- IP-2022-10-6867, Croatian Science Foundation
- Institute for Croatian Language
- Start date: December 31, 2023
- End date: December 30, 2027
- <https://hrpa.jezik.hr/>
- [dr. sc. Ivana Matas Ivanković](#), **project PI**
- [dr. sc. Goranka Blagus Bartolec](#)
- Institute for Croatian Language, Zagreb
- [dr. sc. Maja Glušac](#),
- Faculty of Humanities and Social Sciences
Josip Juraj Strossmayer University
of Osijek
- [dr. sc. Vedran Juričić](#),
- [dr. sc. Daniela Katunar](#),
- [dr. sc. Darko Matovac](#),
- Faculty of Humanities and Social Sciences
University of Zagreb
- [dr. sc. Ljiljana Šarić](#),
- University of Oslo, Norway

HRPA – project goals

- 1) research semantic and syntactic features of prepositions on the basis of a unified comprehensive methodology
- 2) develop a database of prepositions representing prepositions in a uniform way
- 3) research complex preposition constructions and build a repository of such constructions
- 4) study the relationship between basic and figurative meanings of prepositions in use
- 5) relate contemporary theoretical preposition research with applied linguistic research → Croatian as L2

Why prepositions?

- Synsemantic, function words
 - integral point of study for the lexicon-grammar interaction
 - SYNTAX: Actionality and Aktionzart (e.g., *for/in* test, Vendler 1957, Van Valin and La Polla 1997), verb valency, (Tesnière 1959, Samardžija 1988) complement/adjunct distinction (Dowty 2000), argument structure constructions (Goldberg 1995, 2006, Rohde 2001), modification (Croft 2001, Katunar and Raffaelli 2018)
 - SEMANTICS: polysemy studies (Brugman 1981, Lakoff 1987, Šarić 2008), grammaticalization (Lehmann 1995), semantic role labelling (Fillmore 1968, 1977 onwards)

Why (Croatian) prepositions?

- Heterogenous word class
- primary: *do*.GEN, *na*.ACC/LOC ‘on, to, at’, *za*.ACC/INST ‘for’...
- secondary (complex):

Semantic (and syntactic) representation?

- *čelo*.GEN ‘front’ < *čelo* ‘forehead’
- *iznad*.GEN < *iz* + *nad* ‘from out of + over/above’
- *nadohvat*.GEN < *na*+*dohvat* ‘within reach, close’
- *na osnovi*.GEN ‘on the basis of’, *s obzirom na*.ACC ‘regarding’, *u vrijeme*.GEN ‘during, lit. in the time of’
- preposition expressions e.g. *posredstvom koje X* ‘by means of which, through which’

Lexicalization and grammaticalization?

Lexicographic representation/inclusion?

Why Croatian prepositions?

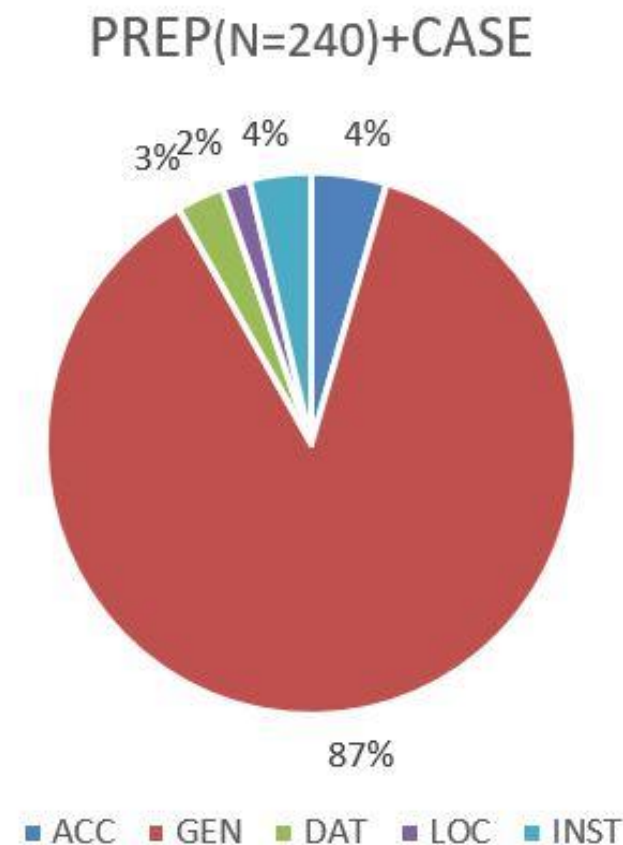
- Slavic language
- rich in fleective and derivational morphology
- preposition – **case** constructions
- grammars: prepositionless / prepositional cases
- „merging of the preposition-case semantics” (Pranjković 2002)
- „discontinuous morphemes” (Kuryłowicz 1948)

u šumi.LOC ‘in the forest’

u šumu.ACC ‘to the forest’

(oči/dlaka/njuh...) u mačke.GEN

‘(eyes/fur/smell...) of a cat’



cf. Matas Ivanković (2006)

- verbal morphosyntactic properties – prepositions (cf. Talmy 1985, 2000, Goldberg 1995, Katunar et al. 2012, Katunar 2021)

*boca je **u**-plutala **u** pećinu.ACC*

‘the bottle floated into the cave’

dubiti (bundevu.ACC) ‘to hollow (a pumpkin)’

> **za**-*dubiti se u.ACC* ‘to delve into SMTH’

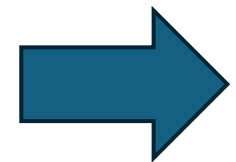
zagrijati se X ‘to warm up’

> **zagrijati se za.ACC** ‘to have the hots for; to be into’

- sentence and discourse-level structures – prepositions
 - passive (*od*.GEN ‘by, lit. from’), adjuncts (adverbials), NP modification, discourse markers (e.g., Hagège 2010)

Why a database of Croatian prepositions?

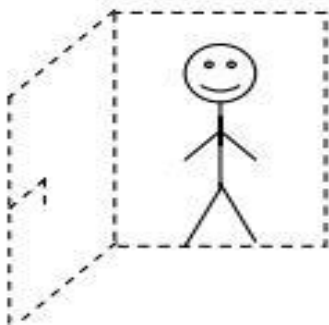
- Preposition database – unification of extant and novel data
- currently dispersed across:
 - traditional lexicographic sources and reference grammars
 - case studies of selected prepositions
 - valency dictionaries and the Croatian dependency treebank
 - NO FrameNet (cf. *Preposition project*, Litkowski & Hargraves 2021)
 - PHASE 1
- Croatian Web Corpus, Croatian National Corpus, Croatian Language Corpus
 - sampling preposition uses
- Identifying relevant metalinguistic data for annotation
 - Evaluating extant online and offline resources → preposition semantics
 - expanding data on frequent and productive usage patterns of prepositions → a constructional perspective (Fillmore, Kay & O'Connor 1988, Goldberg 1995, 2006, Boas 2008, Perek 2015)



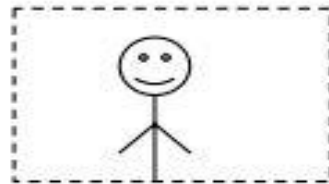
Preposition semantics

- How many senses?
- Contextual (in)dependence of senses?
 - core sense approach (Bennet 1975, Hottenroth 1993)
 - polysemy approach (Brugman 1981, Lakoff 1987, Tyler & Evans 2003)
- *na*.LOC 'on'

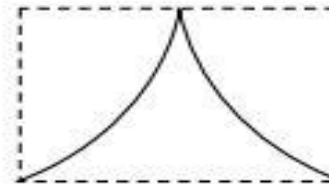
čovjek na vratima



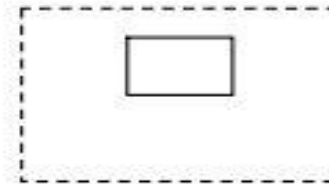
čovjek na prozoru



zavjese na prozoru



portret na zidu



- Kerovec 2012: 113

Lexicographic sources and reference grammars

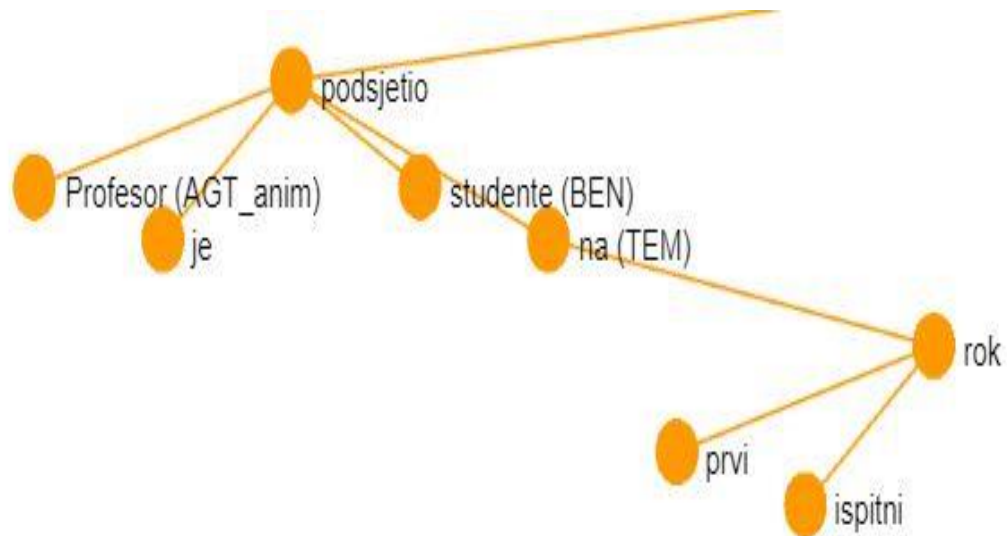
- differ in the number of senses; descriptive glosses
- *na*.LOC/ACC
- *Dictionary of Croatian Language* (Šonje 2000) - 2 senses
- *Croatian Language Portal* – 3 senses
- *Great Dictionary of Standard Croatian Language* (Jojić et al. 2015) - > 10 senses
- *Gramatika hrvatskog jezika za gimnazije i visoka učilišta* (Silić and Pranjković 2007)
 - ACC + *na* - 10 senses, LOC + *na* – 8 senses
 - supralocality – *na*.LOC/ACC, *po* ‘by, in, over’, *iznad* ‘above’, *povrh* ‘over, on top of’

Valency dictionaries

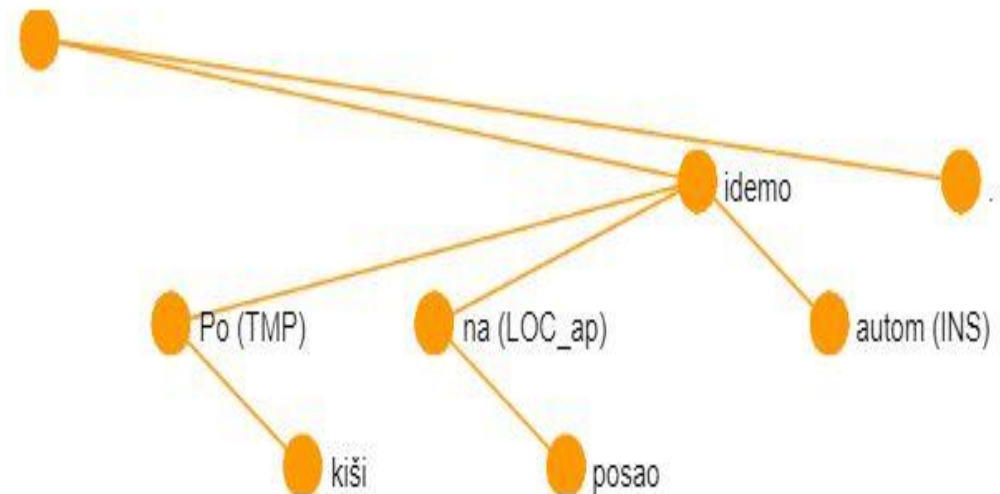
- *e-Glava* – valency of psychological verbs
- *koncentrirati se* ‘concentrate’
- A) NomComp + PrepComp = *na*.ACC (object of attention; w/o restrictions)
 - *Jahači se trebaju koncentrirati na svoj položaj tijela* ‘riders should concentrate on their body position’
- B) NomComp + AdvComp = *na*.LOC/*u*.LOC/*oko*.GEN (where something gathers; place)
 - *Gotovo ukupna masa atoma koncentrira se u jezgri*. ‘almost the entire mass of the atom is concentrated in the core’
- limited verb inventory, no clear PREP sense inventory, omission of sentence periphery, e.g., adverbials and NP modification

Croatian Dependency Treebank (Tadić 2007)

- PREP sense inventory $\approx$ semantic role labelling



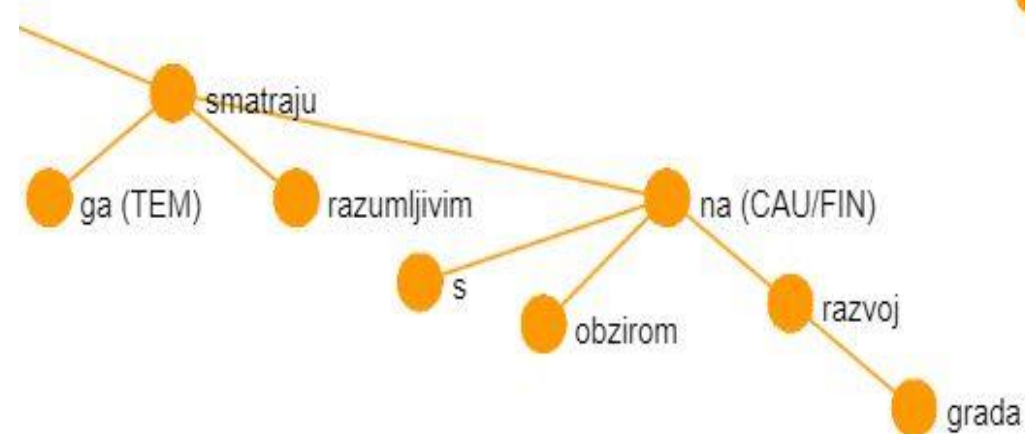
Professor reminded the students about the 1st exam term



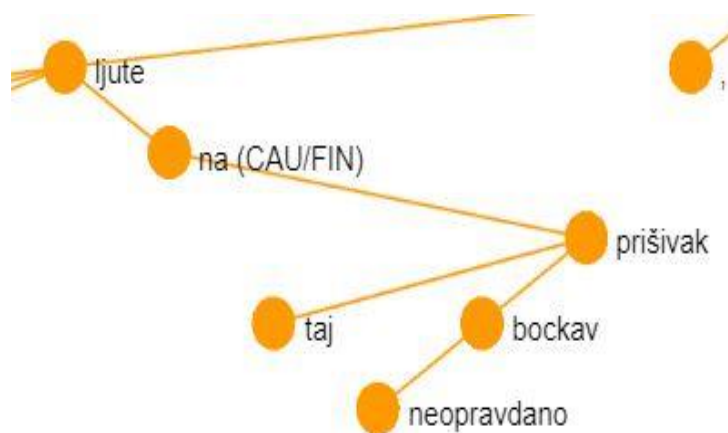
In the rain, we're going to work by car



with regards to the list of plenaries



...with regards to city development



Be angry at that...prickly patch...

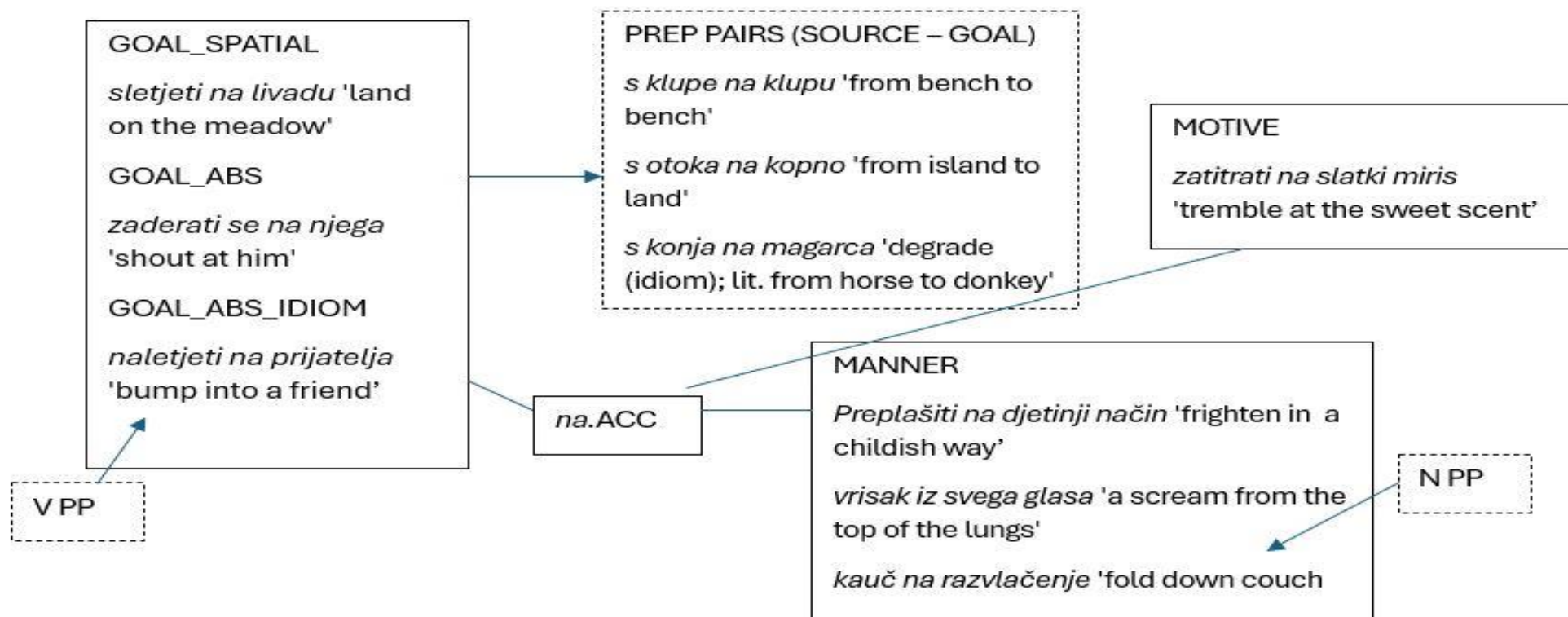
- Secondary preposition annotation?
- Semantic granularity problem
- Exclusion of complex prepositions and various non-compositional language structures
- e.g., MWEs (cf. Šojat et al. 2016)

CxG and PPs

- Constructional meaning $\approx$ preposition sense
- Grammatical underpinnings of lexicalization
 - Lexicalization patterns model (Raffaelli & Kerovec 2017, Raffaelli & Katunar 2018, Raffaelli, Katunar & Kerovec 2019)
 - N PP constructions
 - N *na*.ACC[MANNER] N $\rightarrow$ *ekran na dodir* ‘touch screen’
 - N *od*.GEN[MATERIAL] N $\rightarrow$ *čaj od borovnice, mente, kamilice...* ‘blueberry, mint, chamomile...tea’
 - V-P constructions
 - *držati meso do povrća* ‘hold meat next to veggies’ $\rightarrow$ *držati do X* ‘value’
- Stable and productive preposition pair schemas (Kovačević & Matas Ivanković 2007, Katunar & Berović 2015)
 - *od*.GEN – *do*.GEN ‘from – to’, *s*.GEN – *na*.ACC ‘off of – onto’, *iz*.GEN – *u*.ACC ‘out of – into’
 - *s otoka na kopno* ‘from island to mainland’, *s klupe na klupu* ‘from bench to bench’, *od rođenja do smrti* ‘from birth till death’

PREP Sense Inventory (so far)

- 36 sense candidates - GOAL, DISTRIBUTION, ADDITION, COMPANY, MATERIAL, EMPHASIS, AVOIDANCE, PARTITIVE, EVASION, SOURCE, CONDITIONAL, CRITERIUM, MEASURE, MOTIVE, MANNER, RELATION, CIRCUMSTANCE, RATIO, DESCRIPTION, STARTING POINT, SUBORDINATION, EQUATION, ORIGIN, REASON, TOPIC, POSSESSION, SPACE, RANGE, MEANS, SUPERIORITY, CONTRAST, REPRESENTATION, COMPARISON, CAUSE, TIME, AGENT



In conclusion - CxG perspective

- Prepositions as anchors of complex constructions
 - formal idioms (Fillmore, Kay & O'Connor 1988)
 - substantive idioms – phraseological units
 - free/bound preposition choice → reflected in constructional meaning and preposition sense
 - Repository of constructions – focused on prepositions
- Prepositions as constructions – secondary prepositions
- Potential for a two-level approach to semantic analysis
 - Preposition senses
 - Network of preposition construction senses – antonymy, synonymy, inheritance links



Thank you for your attention!

*Hvala Vam **na** pozornosti!*

HRPA = Cro. ‘thrown together and not neatly arranged pile of different things’