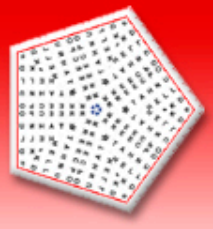


rmjt.ffzg.hr

Derivational Patterns in Slavic Languages

Marko Tadić

University of Zagreb
Faculty of Humanities and Social Sciences

marko.tadic@ffzg.hr

9th Mediterranean Morphology Meeting
Dubrovnik
2013-09-15

Talk overview

- building of Croatian WordNet (CroWN)
  - Croatian verbs as translations of English verbs
- grammaticality vs. lexicality continuum
  - position of derivation
- patterns of Croatian verbal prefixation
- patterns of other verbal derivations
- patterns in other Slavic languages

How did it all begin?

- in 2007 several long discussions with Karel Pala
- the level of isomorphism between
  - verbal derivational patterns (aspectual pairs) and
  - their meaning
- concrete reason
  - building of Croatian Wordnet started
  - using so called “expand model”
  - translation of Basic Concept Set (BCS)
  - from Princeton WN into Croatian
- detected morphological issues in Croatian following the semantic relations between words in CroWN
  - synonymy/antonymy
  - hyponymy/hyperonymy
  - meronymy/holonymy...

Croatian Wordnet (CroWN)

- Lexical Semantics in Building the Croatian WordNet
  - project run from 2007 by Ida Raffaelli
  - building of CroWN
- CroWN built on the basis of PWN 2.0 and synchronized later with PWN 3.0
  - 10,040 synsets
  - 31,300 literals
    - 53.54% nouns, 43.7% verbs, 2.73% adjectives, 0.02% adverbs
  - manually translated 98.8% of BCS1, 2, 3
  - more synsets added
  - freely downloadable from META-SHARE platform
    - within CESAR project
    - a part of META-NET Network of Excellence

Croatian vs. English verb in WN

- all Croatian verbs are always morphologically marked for aspect
  - perfective
  - imperfective
  - bi-aspectual (complementary distribution)
- English verbs have two or more translation equivalents in Croatian depending on the aspect
- en: *drink* or *imbibe*
  - PWN sense V_1: *take in liquids*
  - glossa: *The patient must drink several liters each day*.
  - *drink* can be translated with five Croatian verbs
    - *piti*, *popiti*, *ispiti*, *ispijati*, *poispijati*
    - all morphologically derived from the base form *piti*

Croatian vs. English verb in WN

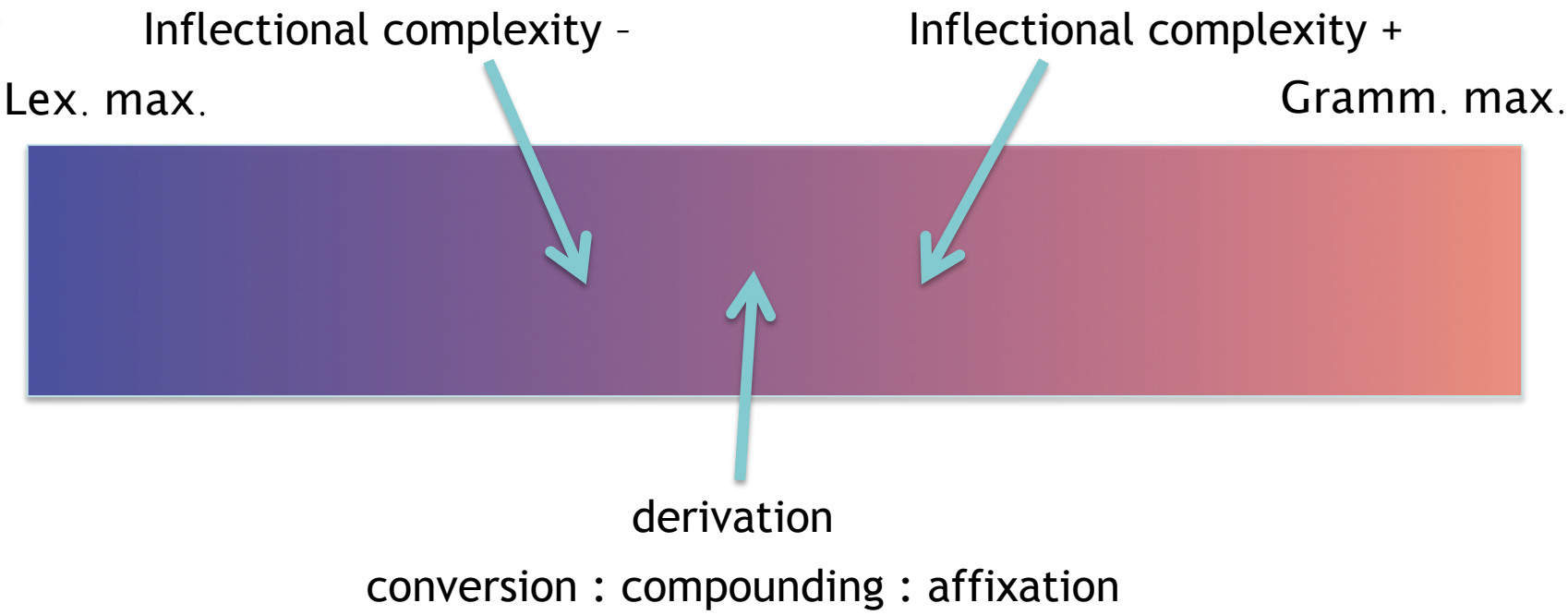
- five Croatian translational equivalents of *drink*
  - *piti (drink)*
    - neutral base form, general action
  - *popiti* (*drink until the end*)
    - act until the action is finished
  - *ispiti* (*drink to the bottom*)
    - the action is finished and there is no liquid left
  - *ispijati* (*drink several times to the bottom*)
    - includes both of previous components and the additional component of iterativity of action
  - *poispijati (drink several times to the bottom by more of us)*
    - all previous components are present, but the action is performed either by several subjects or on several objects

Croatian verb in CroWN

- Fellbaum et al. (2007) discuss cross-POS relations among words belonging to different synsets sharing a stem with the same meaning
- 14 “morphosemantic links”
  - introduced into PWN 2.0.
  - relations based on suffixation patterns between verbs and nouns
  - each relation semantically labeled
    - Agentive, Instrument, Vehicle, Location etc.
- none of the morphosemantic links include verb-verb pairs
- Raffelli et al. (2008) already spotted the problem of aspectual pairs in CroWN and possible derivation relation between verbs

Grammaticality vs. Lexicality

- grammaticality vs. lexicality seen as continuum
- typologically we can map different languages in this continuum
  - according to their morphological complexity



Theoretical background

- distant theoretical background
  - Construction Morphology
- we are using patterns as methodological construct to explain where and how the two layers are mapped to each other in a predictive manner
- our approach was the result of intention to apply bottom-up direction in order to complement the initial “expand model” (simple translation of PWN)

How we proceeded?

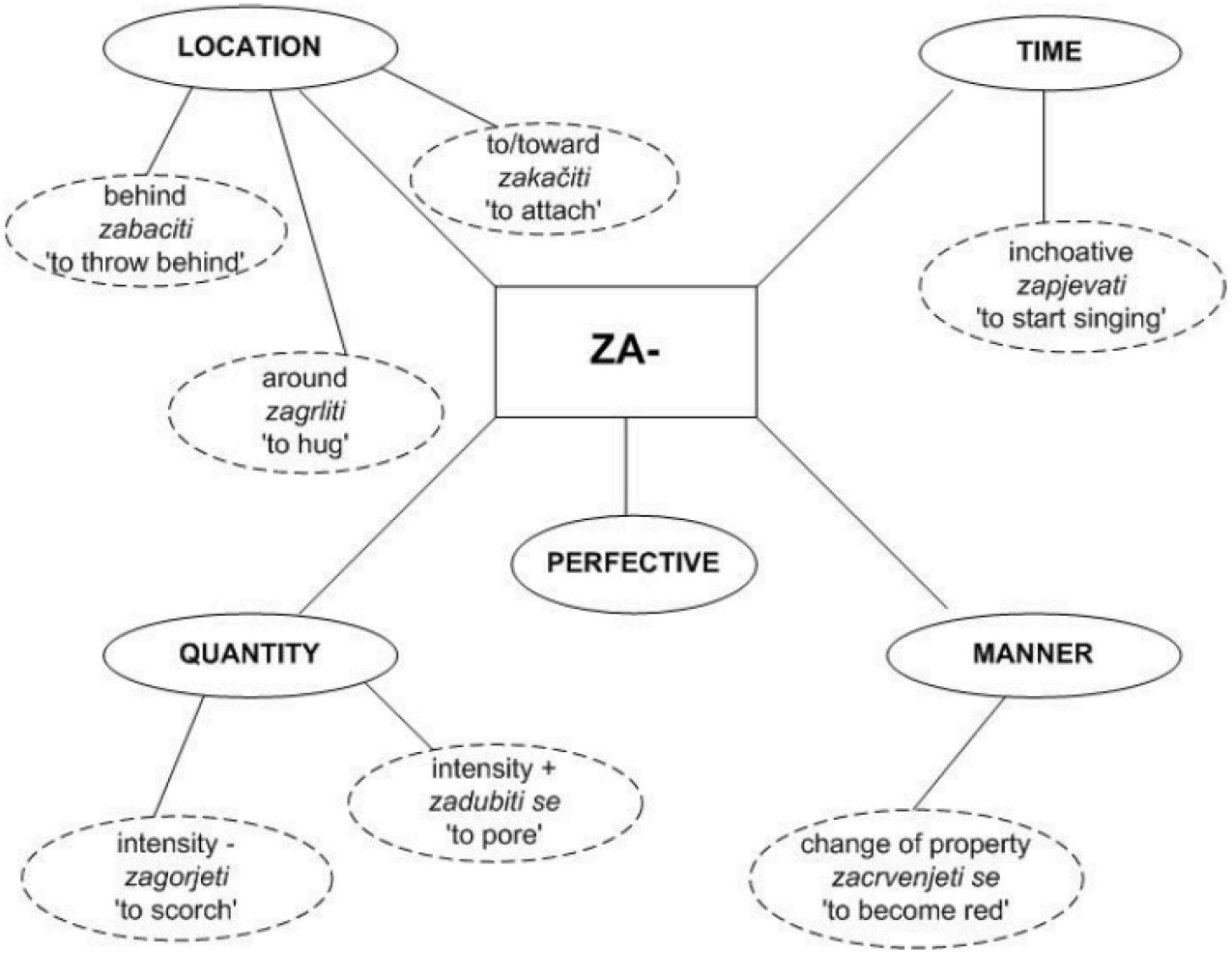
- investigation of aspectual pairs in Croatian verbs
- CroDeriV database of ca 14,000 verbs was built
  - details in Šojat et al. (2012), Šojat et al. (2013) and more papers to come
- function of prefixes
  - primary: mark the aspect
  - secondary: encode additional conditions of the performance of action (Aktionsart)
- regularities in semantic modifications of base forms by one prefix
  - regularly reoccuring meaning alternations
  - prefixation patterns

Croatian verbal prefixation

- 19 prefixes in Croatian verbs
  - non-Latin origin
  - generally morphologically predictable
  - although there are some verbs that turn into completely idiosyncratic constructions
- objective
  - find the set of prefixal meanings that reoccur in different semantic fields
  - find which prefixes carry the same meanings
- general schema
  - beside the pure aspectual meaning
  - four general additional conditions of action (Aktionsart) were detected in the meaning of prefixes

Croatian verbal prefixation

- meaning of prefix “za-”



Prefixation in CroWN

- we have marked all the prefixed verbs in CroWN for these aspectual relations
  - 30% of all verbs in CroWN
- this inventory of relations
  - defined as prefixal patterns
  - applicable in CroWN
  - but, also in wordnets for other Slavic languages
- related work: Pala & Hlavačkova (2007), Koeva (2008), Marziarz et al. (2012) etc.
  - all talk about aspectual pairs, but treat them differently
  - didn't include such a detailed analysis

Additional patterns

- aspectual markedness in Croatian is realised by
  - prefixes (*za-baci-ti*) or suffixes (*bac-a-ti*)
  - combination (*za-bac-iva-ti*)
- suffixation and circumfixation play additional role
  - Šojat et al. (2013) describe the overall derivational structure of Croatian verbs
  - again using CroDeriV database
  - e.g. overall Croatian verbal structure is defined as
    - (prefix4) + (prefix3) + (prefix2) + (prefix1) + (derivational stem2 + interfix) + derivational stem1 + (sufix1) + (sufix2) + (sufix3) + infinitive sufix
- can we find other patterns useful in CroWN or other Slavic wordnets?
  - not dealing just with verbal aspect, but wider

Derivational relations in PWN 3.0

Verb

- S: (v) **teach**, learn, instruct (impart skills or knowledge to) *"I taught them French"; "He instructed me in building a boat"*
  - *direct troponym* / *full troponym*
  - *direct hypernym* / *inherited hypernym* / *sister term*
  - *cause*
  - ***derivationally related form***
    - W: (adj) teachable [Related to: teach] (ready and willing to be taught) *"docile pupils eager for instruction"; "teachable youngsters"*
    - W: (n) teacher [Related to: teach] (a person whose occupation is teaching)
    - W: (n) teacher [Related to: teach] (a personified abstraction that teaches) *"books were his teachers"; "experience is a demanding teacher"*
    - W: (n) teaching [Related to: teach] (the activities of educating or instructing; activities that impart knowledge or skill) *"he received no formal education"; "our instruction was carefully programmed"; "good classroom teaching is seldom rewarded"*
    - W: (adj) instructive [Related to: instruct] (serving to instruct or enlighten or inform)
    - W: (n) instruction [Related to: instruct] (the activities of educating or instructing; activities that impart knowledge or skill) *"he received no formal education"; "our instruction was carefully programmed"; "good classroom teaching is seldom rewarded"*

Derivation in Czech Wordnet

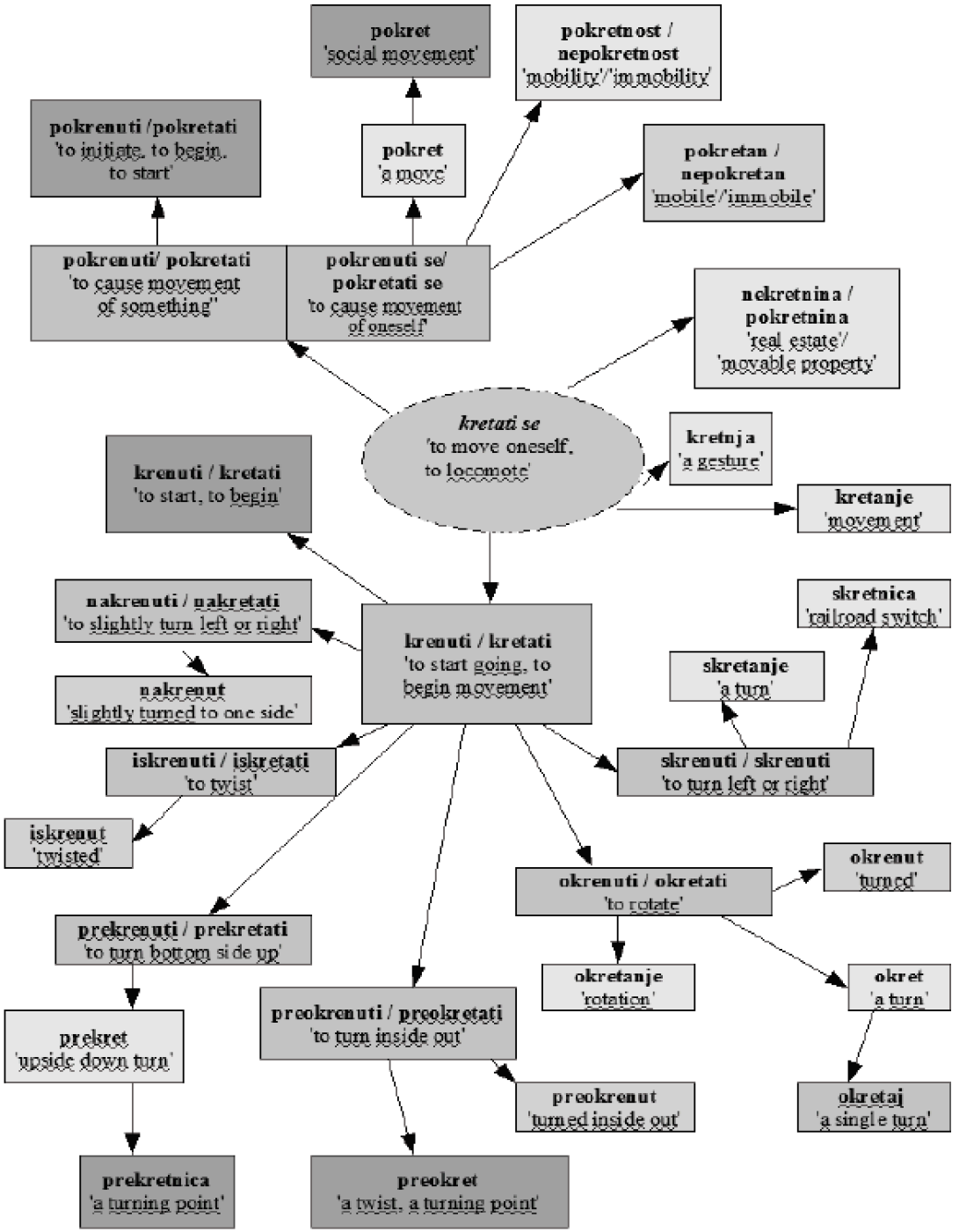
- Pala & Hlavačkova 2007, after Pala & Sedlaček 2005
- transfer 14+2 derivational processes from Czech grammars
- conflated to 12 derivational relations inserted in Czech Wordnet
  1. deriv-na: noun -> adjective (property)
  2. deriv-ger: verb -> adjective (property)
  3. deriv-dvrb: verb -> noun (activity as a noun)
  4. deriv-pos: noun -> adjective (possessive relation)
  5. deriv-pas: verb -> adjective (passive relation)
  6. deriv-aad: adjective -> adverb (property of property)
  7. deriv-an: adjective -> noun (property)
  8. deriv-g: noun -> noun (gender relation)
  9. deriv-ag: verb -> noun (agentive relation)
  10. deriv-pat: verb -> noun (patient relation)
  11. deriv-dem: noun -> noun (diminutive relation)
  12. deriv-loc: noun -> noun (locative relation)

Derivation in Czech Wordnet

- derivational nests
  - all possible derivatives using different patterns
  - around the same root or lexical morpheme
- e.g. root: *-prác-/-prac-e-*
  - deriv-act: *prac-ova-t* (to work)
  - deriv-loc1: *prac-ov-iště* (workplace)
  - deriv-loc2: *prac-ov-na* (study)
  - deriv-ag1: *prac-ov-ník* (worker),
  - deriv-g: *prac-ovn-ice* (she-worker),
  - deriv-ag2: *prac-ant* (plodder)
  - deriv-ger: *prac-uj-ící* (working-person)
  - deriv-pro: *prac-ov-ní* (professional, working)
  - deriv-pro: *prac-ov-i-t-ý* (diligent, hardworking)
  - deriv-pro: *prac-ov-i-t-ost* (diligence)...

Derivation

- derivation nest of Croatian root *-kret-*
- Katunar & Šojat (2011)



Derivation nests in Slavic Ls

- investigate whether the similar derivative patterns could be detected for the same set of derivative relations established in Czech Wordnet
- taken the root -*uč*- (teach, learn)

ACT	cz	učit
	sk	učiť
	sl	učiti
	hr	učiti
	pl	uczyć, nauczyć
	ru	učit’
	en	teach/learn

Derivation patterns in Slavic Ls

AG	cz	učitel
	sk	učiteľ
	sl	učitelj
	hr	učitelj
	pl	nauczyciel
	ru	učitel'
	en	teacher
AG	cz	učenec
	sk	učenec
	sl	učenjak
	hr	učenjak
	pl	uczony
	ru	učonyj
	en	scholar

Derivation patterns in Slavic Ls

PAT	cz	učedník
	sk	učeník
	sl	učenec
	hr	učenik
	pl	uczyć, uczeń
	ru	učenik
	en	disciple
PAT	cz	učeň
	sk	učeň
	sl	(vajenec)
	hr	naučnik
	pl	uczyć, uczeń
	ru	učenik
	en	apprentice

Derivation patterns in Slavic Ls

INST	cz	učebnice
	sk	učebnica
	sl	učbenik
	hr	udžbenik
	pl	(podręcznik)
	ru	učebnik
	en	textbook
LOC	cz	učebna
	sk	učebňa
	sl	učilna, učilnica
	hr	učionica
	pl	(sala, klasa)
	ru	(klass)
	en	classroom

Derivation patterns in Slavic Ls

LOC	cz	učiliště
	sk	učilište
	sl	učilišče, vseučilišče
	hr	učilište, sveučilište
	pl	uczelnia
	ru	učilišče
	en	training institution
GEN	cz	učitelka
	sk	učiteľka
	sl	učiteljica
	hr	učiteljica
	pl	nauczycielka
	ru	učitel'nica
	en	she-teacher

Derivation patterns in Slavic Ls

GEN	cz	učednice
	sk	učeníčka
	sl	učenka
	hr	učenica
	pl	uczennica
	ru	učenica
	en	she-disciple
GEN	cz	učnice
	sk	učeníčka
	sl	(vajenka)
	hr	naučnica
	pl	uczennica
	ru	učenica
	en	she-apprentice

Derivation patterns in Slavic Ls

PRO	cz	učitelský
	sk	učiteľský
	sl	učiteljski
	hr	učiteljski
	pl	nauczycielski
	ru	učitel'skij
	en	preceptorial
DVRB	cz	učení
	sk	učenie
	sl	učenje
	hr	učenje
	pl	uczenie
	ru	učenije
	en	teaching

Derivation nests in Slavic Ls

- some patterns do not exhibit the full symetry

PRO	cz	učebnicový
	sk	učebnicový
	sl	učbeniški
	hr	udžbenički
	pl	?
	ru	?
	en	textbook adj

Rel	Czech	Slovak	Slovenian	Croatian	Polish	Russian	English
act	učit	učiť	učiti	učiti	uczyć, nauczyć	učit'	teach
ag	učitel	učiteľ	učitelj	učitelj	nauczyciel	učitel'	teacher
ag	učenec	učenec	učenjak	učenjak	uczony (1)	učonyj	scholar
pat	učedník	učeník	učenec	učenik	uczyć, uczeń	učenik	disciple
pat	učeň	učeň	(vajenec)	naučnik	uczyć, uczeń	učenik	apprentice
instr	učebnice	učebnica	učbenik	udžbenik	(podręcznik)?	učebnik	textbook
loc	učebna	učebňa	učilna, učilnica	učionica	(sala, klasa)?	(klass)	classroom
loc	učiliště	učilište	učilišče, vseučilišče	učilište, sveučilište	uczelnia	učiliščе	training inst
gen	učitelka	učiteľka	učiteljica	učiteljica	nauczycielka	učitel'nica	she-teacher
gen	Učitelová	?	?	učiteljevica	nauczycielowa	?	teacher's wife
gen	učednice	učeníčka	učenka	učenica	uczennica	učenica	she-disciple
gen	učnice	učeníčka	(vajenka)	naučnica	uczennica	učenica	she-apprentice
dim	učitýlek	učitelík	učiteljček	učiteljčić	?	?	small teacher
dim	učebnička	učebnička	?	udžbeničić	?	?	small textbook
dim	učedníček	učeníček	učenček	učeničić	uczniak	?	small disciple
aug	učitelisko	?	?	učiteljčina	?	?	big teacher
ag	úča, pejor.	učka, pejor.	?	učo, pejor. masc.	nauczycielina	učilka	pej. she-teacher
pro	učební	učebný	učen, učilen	poučan	uczony (2)	?	didactic
pro	učitelský	učiteľský	učiteljski	učiteljski	nauczycielski	učitel'skij	preceptorial
pro	učňovský	učňovský	vajeniški?	učenički	uczniowski	učeničeskij	apprentice adj
pro	učenecký	učenecký	učenjaški	učenjački	uczony	?	scholarly
pro	učednický	učenícky	učenski	učenički	?	?	discipular
pro	učený	učený	učen	učen	?	?	learned, scholarly
pro	učenlivý	učenlivý	učljiv	naučljiv	nauczalny	(obučaemyj)	docile
pro	učebnicový	učebnicový	učbeniški	udžbenički	?	?	textbook adj
pro	učenlivost	učenlivosť	učljivost	učljivost	nauczalność	obučaemosť	docility
poss	učitelův	učiteľov	učiteljev	učiteljev	nauczyciela (gen.sg)	gen. učitelja	teacher's
poss	učitelčin	učiteľkin	učiteljičin	učiteljičin	nauczycielki (gen.sg)	gen. učitel'nicy	she-teacher's
poss	učencův	učencov	učenjakov	učenjakov	?	gen. učonogo	scholar's
poss	učedníkův	učeníkov	učenčev	učenikov	ucznia (gen.sg)	gen. učenika	disciple's
poss	učňův	učňov	vajenčev?	učenikov	ucznia (gen.sg)	gen učenika	apprentice's
ger	učící	učiaci	učeč	učeći	uczący	učaščij	teaching
pass	učen	?	učen	učen	?	učon	taught
dvrb	učení	učenie	učenje	učenje	uczenie	učenije	teaching
dvrb	učivo	učivo	učivo	gradivo	(program nauczania)	(programma obučenija)	curriculum
dvrb	učednictví	učeníctvo	?	učeništvo	uczniostwo	učeničestvo (?)	discipleship
dvrb	učenost	učenosť	učenost	učenost	uczoność	učonost'	erudition
dvrb	učitelství	učiteľstvo	učiteljstvo	učiteljstvo	nauczycielstwo	učitelja	teaching profession
dvrb	učitelstvo	učiteľstvo	učiteljstvo	?	nauczycielstwo	učitelja	teaching staff

Future directions

- add frequency of patterns (or even sub-patterns) to CroDeriV from
  - Croatian National Corpus (231 Mw) / hrWaC (1.2 Bw)
- insert the Aktionsart information from Croatian verbal prefixes in CroWN and enable its search
- link the CroDeriV with Croatian Morphological Lexicon (HML)
  - a comprehensive Croatian inflectional lexicon
  - ca 125,000 lemmas yielding +6M wordforms
  - generation of word-forms, lemmatisation and POS/MSD-tagging
  - freely accessible at http://hml.ffzg.hr
  - also downloadable from META-SHARE
  - soon to be accessible as web service as well

Future directions

- the CroDeriV database could be used as a role model for modelling overall structure of derivational verbal systems of other Slavic languages
- add examples from more Slavic languages
  - Bulgarian, Serbian, Ukrainian, Sorabian, etc.
- add frequency information from respective lexica and corpora

Thank you for your attention.