

# Deriving the Graph

Using Affixal Senses for Building Semantic Graphs

Matea Filko ([matea.filko@ffzg.hr](mailto:matea.filko@ffzg.hr)), Krešimir Šojat ([ksojat@ffzg.hr](mailto:ksojat@ffzg.hr)), Vanja Štefanec ([vstefane@ffzg.hr](mailto:vstefane@ffzg.hr))

Faculty of Humanities and Social Sciences, University of Zagreb

# CroDeriv

- Phase 1 – morphological analysis (14.000 V, 5.500 N, 1.500 ADJ)
  - *učiteljica* ‘female teacher’ = *uč-i-telj-ic-a* (surface layer)  
*uk-i-telj-ic-a* (deep layer)
- Phase 2 – word-formation analysis
  - *učiteljica* ‘female teacher’ < *učitelj* ‘male teacher’ + *-ica*
    - word-formation process: suffixation; POS maintaining
- Phase 3 – semantic categories
  - generalized semantic categories marked on derivational affixes
  - focus of this presentation

# Semantic categories in existing derivational resources

- mainly regular semantic relations between derivationally connected words are marked
  - Démonette (Hathout and Namer, 2014)
    - indicates whether words denote an action, an agent or a property
  - Derivancze (Pala and Šmerk, 2015)
    - relations that have regular and transparent meanings, e.g. the relation between the action and the agent of the action
  - DeriNet (Ševčíková and Kyjánek, 2019)
    - semi-automatic procedure for assigning semantic relationships to units in DeriNet
    - diminutive and female (for nouns), possessive (for adjectives), and iterative and aspect (for verbs)

# Semantic categories in CroDeriv

- marked on derivational affixes
- **aim of the procedure:** to establish semantic paths, i.e. regular semantic shifts / patterns in the derivation
  - description of Croatian morphotactics
  - labelling of these categories in graphs within derivational families provides an insight into semantic processes taking part in derivation
- same semantic processes can be realized by various means of derivation and within different derivational families
  - e.g. **action > agent > female agent:**
    - *učiti* ‘to teach’ > *učitelj* ‘male teacher’ > *učiteljica* ‘female teacher’
    - *voziti* ‘to drive’ > *vozač* ‘male driver’ > *vozačica* ‘female driver’
    - *izdavati* ‘to publish’ > *izdavač* ‘male publisher’ > *izdavačica* ‘female publisher’

# Affixal meanings – theoretical background

- two basic approaches to affixal meanings:
  - homonimization of affixes – splits affixes into two or more separate units
    - multiplies the number of units
  - **polysemization of affixes** – reinterprets homonyms as a single unit
    - meaning networks of particular units – links between different affixal meanings
    - speakers recognize the links between meanings – economical and effective usage of vocabulary
    - polysemy is linguistically and cognitively more economical than homonymy
- meanings are determined synchronically
  - combination of semasiological and onomasiological approach (Bagasheva, 2017)

# Affixal meanings – theoretical background

- semasiological approach – analysis of polysemous structure of individual affixes
- e.g. nominal suffix *-ba*:
  - 1. action (*ploviti* ‘to sail’ > *plovid-ba* ‘sailing’)
  - 2. result (*skladati* ‘to compose’ > *sklad-ba* ‘composition’)
  - 3. event (*svat* ‘wedding guest’ > *svad-ba* ‘wedding’)
  - 4. location (*nastaniti* ‘to dwell’ > *nastam-ba* ‘dwelling’)
  - 5. non-transparent meaning (*opor* ‘harsh’ > *opor-ba* ‘political opposition’)
- difference between semantically transparent and non-transparent derivatives!

# Affixal meanings – theoretical background

- onomasiological approach – affixes used in the formation of the same semantic categories
- e.g. semantic category ‘property, quality’:
  - 1. -ina (*bijel* ‘white’ > *bjelina* ‘whiteness’)
  - 2. -oća (*slijep* ‘blind’ > *sljepoća* ‘blindness’)
  - 3. -ost (*slab* ‘weak’ > *slabost* ‘weakness’)

# Affixal meanings – theoretical background

- lexical meaning vs. derivational meaning of words
  - e.g. *stolar* ‘carpenter’ < *stol* ‘table’ + *-ar*
    - **derivational meaning:** ‘the one who makes/produces tables’
    - **lexical meaning:** ‘the one who produces any kind of wooden furniture, window frames or doors’
  - the generalized meaning ‘male agent’ is provided by the suffix *-ar*
  - clear from the onomasiological perspective: *kuh-ar* ‘cook’, *kip-ar* ‘sculptor’, *mes-ar* ‘butcher’...
- affixal meanings are determined wrt generalized semantic categories
  - the most extensive list in Bagasheva (2017) – adapted and additionally specified and divided into subcategories when needed
    - e.g. location > top-down, bottom-up, over, around...

# Affixal meanings

NOUNS	NOUNS2	ADJECTIVES	VERBS	VERBS2
agent - male	product / food	quality	secondary imperfective	TIME
agent - female	result	possessive	reflexivity	time_inchoativity
inhabitant	state	relation	action - resultative	time_finitiveness
patient	event	definiteness		time_distributivity
action	game	privative	LOCATION	time_preceding
location / space	letter	ability	loc_bott_up	time - continuity
instrument	diminutive		loc_top_down	
time	pejorative		loc_prox	QUANTITY
abstraction	lexicalized meaning		loc_through	quan_sufficiency
quality / property			loc_apart	quan_excessiveness
holder of a property			loc_to_toward	quan_intensity
movement / group / institution / body			loc_over	quan_exceeding
activity			loc_into	quan_deprivation
member of the movement/group...			loc_around	quan_addition
object / substance			loc_under	
animal			loc_reloc	MANNER
plant			loc_behind	mann_inter-connection
sum / totality			loc_across	mann_change of property
part of the body / organ			loc_from	

# Affixal meanings – theoretical background

- only one meaning of the polysemous word is taken into account
  - e.g. *upravljač*
    - an agent ‘controller’
    - an object ‘control device’
  - however > same morphological and derivational properties > we didn’t want to multiply the number of entries
- first meaning listed by the extensive online dictionary of the Croatian language ([hjp.znanje.hr](http://hjp.znanje.hr)) was marked in CroDeriv

# Affixal meanings – theoretical background

- three types of suffixes:
  - 1) suffixes that change POS of the stem
    - pjevati<sub>V</sub> ‘to sing’ – pjevanje<sub>N</sub> ‘singing’
  - 2) suffixes that modify the meaning of stems without changing their POS
    - lav<sub>N</sub> ‘lion’ – lavica<sub>N</sub> ‘lioness’
  - 3) suffixes that modify the meaning of stems and change their POS
    - voziti<sub>V</sub> ‘to drive’ – vozač<sub>N</sub> ‘driver’

# Affixal meanings in CroDeriv

- three (four) layers of description for each lemma in CroDeriv:
  - morphological
    - morphemes (roots, prefixes, derivational and inflectional suffixes, interfixes) + allomorphs
  - word-formation
    - clusters = multimorphemic units – stems and prefixing, interfixing or suffixing formants (1 or more derivational affixes + inflectional affix)
    - link between morphological and word-formation layer
    - meanings are assigned to clusters – each cluster can have multiple meanings due to the polysemous nature of affixes, but only one of these meanings can be assigned to the particular word

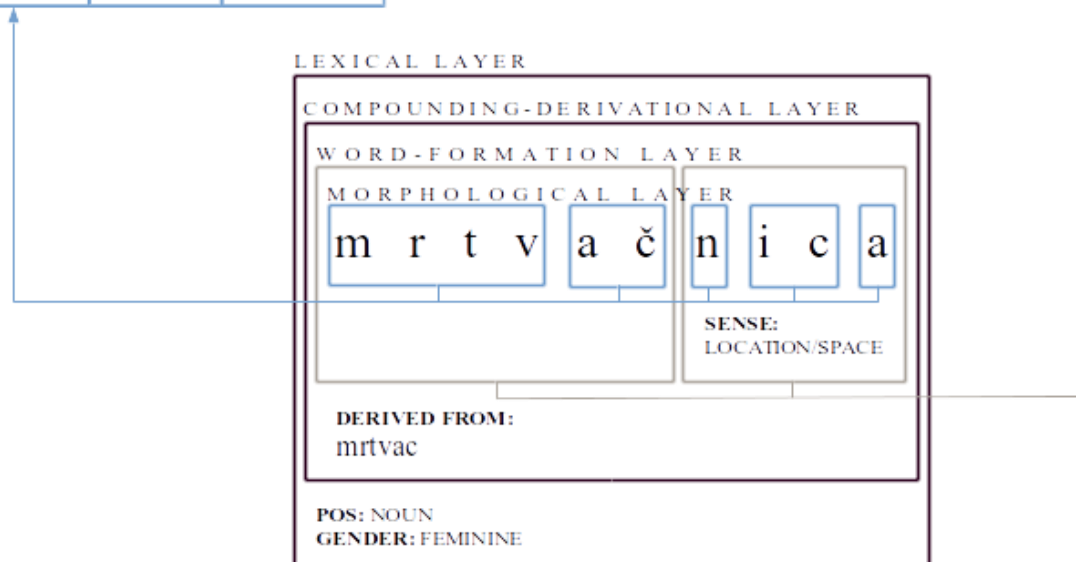
# Affixal meanings in CroDeriv

- three (four) layers of description:
  - compounding-derivational
    - compounding segments – serve as child members in derivational relations; consist of clusters
    - *naredbodavac* ‘commander’ [lit. command-giver] – *naredbo-* and *-davac*
    - contains information about **the base word** in the derivational process
    - link between word-formation and compounding-derivational layer
    - compound words can be a part of multiple derivational families
  - (fourth layer – grammatical information)
    - e.g. gender and number for nouns, aspect for verbs etc.

# Affixal meanings in CroDeriv

MORPHEME INVENTORY		
SURFACE LAYER	DEEP LAYER	MORPHEME TYPE
...	...	...
mrtv	mrtv	ROOT
ač	ac	SUFFIX
n	n	SUFFIX
ic	ic	SUFFIX
a	a	ENDING
...	...	...

WORD-FORMATION CLUSTER INVENTORY			
SURFACE LAYER	DEEP LAYER	CLUSTER TYPE	CLUSTER SENSES
...	...	...	...
mrtvač	mrtv-ac	STEM	
nica	n-ic-a	SUFFIXING FORMANT	INSTRUMENT OBJECT/SUBSTANCE LOCATION/SPACE
...	...	...	...



# Semantic graphs

- the interconnection of different layers of description in CroDeriv facilitates exploring other, not explicitly annotated, phenomena that occur at the intersections of these layers
- e.g. derivational semantics can be studied with the help of semantic graphs
  - two different graph representations:
    - lexeme-semantic representation
    - structure-semantic representation

# Lexeme-semantic representation

- directed acyclic graph with labeled nodes representing lexemes, and labeled edges representing derivational meaning of the lexeme they connect to
- informative representation of derivational families and the semantic motivation for the expansion of the derivational tree
- combining information from the word-formation and compounding-derivational layer enables labeling of the derivational link between the base word and the derivative as

*mrtvac*  $\xrightarrow{\text{LOCATION/SPACE}}$  *mrtvačnica*



# Structure-semantic representation

- derivational generalizations of semantic build-up
- directed acyclic graph with semantic labels attached to edges, but here the nodes are labeled with derivational affixes involved in the last derivational step of a particular lexeme
- combining information from the word-formation layer of both base word and the derivative, and from the compounding-derivational layer of the derivative, the following derivational mechanism is obtained:

$$-ac \xrightarrow{\text{LOCATION/SPACE}} -nica$$



# Structure-semantic representation

- structure-semantic graphs allow for:
  - 1) quantifying the derivational similarity between derivational trees
  - 2) exploring the distribution of particular affixes involved in certain semantic change and vice-versa
  - 3) calculating causal distribution of derivational mechanisms
  
- it is possible to match two not completely isomorphic graphs – comparison of derivational families and finding families with large derivational similarity

# Structure-semantic representation

- facilitates distributional analysis of both derivational affixes and semantic changes in a direct derivational sequence
- highly polysemous affixes – it is possible to determine their productivity in each of the semantic categories across derivational families
- semantic categories – it is possible to examine their distribution within affixes that can carry their meaning

# First experiment

- affixes in 10 large derivational families annotated for meanings (1.254 lemmas)
  - most of them are verbocentric
- most prominent semantic paths, i.e. regular semantic shifts / patterns in the derivation are detected
- longest path: 7 lemmas from the base word to the final derivative

(base) N	(base) action V, IPF	V, PF, modified meaning	V, secondary imperfective	N, agent male	N, agent female	ADJ, possessive
stav	staviti	sastaviti	sastavljati	sastavljač	sastavljačica	sastavljaččin
'attitude, opinion; posture'	'to put IPF'	'to put together, develop... PF'	'to put together, develop... IPF'	'developer, male'	'developer, female'	'(female) developer's'

<b>(base) N</b>	<b>(base) action V, IPF</b>	<b>V, PF, modified meaning</b>	<b>V, secondary imperfective</b>	<b>N, agent male</b>	<b>N, agent female</b>	<b>ADJ, possessive</b>
stav	staviti	sastaviti	sastavljati	sastavljač	sastavljačica	sastavljačičin
'attitude, opinion; posture'	'to put IPF'	'to put together, develop... PF'	'to put together, develop... IPF'	'developer, male'	'developer, female'	'(female) developer's'
	pisati	upisati	upisivati	upisivač	upisivačica	upisivačičin
	'to write IPF'	'to type PF'	'to type IPF'	'typist, male'	'typist, female'	'(female) typist's'
<b>(base) N</b>	<b>N, agent male</b>	<b>N, activity</b>	<b>ADJ, relation</b>	<b>N, agent male</b>	<b>ADJ, relation</b>	
red	redar	redarstvo	redarstveni	redarstvenik	redarstvenički	
'order'	'keeper of order; bouncer'	'police'	'pertaining to police'	'policeman'	'policeman's'	
<b>(base) action V, IPF</b>	<b>V, PF, modified meaning</b>	<b>secondary imperfective</b>	<b>N, agent male</b>	<b>ADJ, relation / possessive</b>		
pisati	upisati	upisivati	upisivač	upisivačev / upisivački		
'to write IPF'	'to type PF'	'to type IPF'	'typist, male'	'typist's relational / poss.'		
<b>(base) action V, IPF</b>	<b>V, PF, action - resultative</b>	<b>ADJ, ability</b>	<b>N, quality / property</b>			
gledati	sagledati	sagledljiv	sagledljivost			
'comprehend, lit. to watch IPF'	'comprehend PF'	'comprehensible'	comprehensibility			
<b>(base) action V, IPF</b>	<b>V, PF, manner – change of property</b>	<b>ADJ, quality</b>	<b>ADV, manner</b>			
pisati	propisati	propisan	propisno			
'to write IPF'	'to regulate PF'	'regulated'	'as regulated'			
<b>(base) N</b>	<b>(base) action V, IPF</b>	<b>V, PF, action - resultative</b>	<b>N, abstraction</b>			
stav	staviti	ostaviti	ostavka			
'attitude, opinion; posture'	'to put IPF'	'to put PF'	'(to put) a resignation'			
<b>(base) action V, IPF</b>	<b>N, agent male</b>	<b>N, location / space</b>				
pisati	pisar	pisarnica				
'to write'	'copyist, male'	'writing-office'				

## Conclusion

- novel way of representing CroDeriv data for the purpose of investigating word-formation mechanisms in the Croatian language
- new line of research on Croatian derivational morphology with the help of generalized morphological data in the form of semantic graphs
- the data annotated for word-formation so far was chosen on the criterion of belonging to large and productive derivational families
- goal: to annotate the entire CroDeriv dataset in this way

# Thank you for your attention!

Matea Filko ([matea.filko@ffzg.hr](mailto:matea.filko@ffzg.hr))  
Krešimir Šojat ([ksojat@ffzg.hr](mailto:ksojat@ffzg.hr))  
Vanja Štefanec ([vstefane@ffzg.hr](mailto:vstefane@ffzg.hr))