



HR4EU - Using Language Resources in Computer Aided Language Learning

Daša Farkaš, Matea Filko, Marko Tadić

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

Outline



Investing
in future

1. HR4EU – web portal for e-learning of Croatian
2. HR4EU & LRs
 1. Croatian Dependency Treebank: the syntactic and the semantic layer
 2. SynSem Visualizer
 3. HR4EU sub-corpus of the HOBS
3. Conclusion



Investing
in future

HR4EU project

- project: *Mrežni portal za online učenje hrvatskoga jezika HR4EU / Web Portal for e-learning of Croatian (HR4EU)*
- institution: University of Zagreb, Faculty of Humanities and Social Sciences
- duration: 15 months (2015-07-01 –2016-09-30)
- funding: European Social Fund (*Research grants for the development of young researchers and postgraduates*)
- EU support: 177,000 EUR
- senior researcher: Marko Tadić
- PhD students: Daša Farkaš, Matea Filko, Diana Hriberski, Sara Librenjak, Jurica Polančec



Investing
in future

What is HR4EU?

- www.hr4eu.eu, www.hr4eu.hr
- first **free** online system for e-learning of Croatian language
- developed by **linguists – native speakers** of Croatian with the experience in language resources development and usage
- **visually attractive and interactive**
- based on empirical data – primarily on corpora and lexica for Croatian



Investing
in future

HR4EU portal

- **Courses:**
 - Beginner, Intermediate, Advanced
 - Croatian for business users / Croatian for students
- **About Croatia**
 - 9 interactive maps of Croatia illustrate social, cultural, historical and natural context
- **Living in Croatia**
 - institutions and forms relevant to foreigners
 - Croatian for business users / Croatian for students
- **Language Resources**
 - description of available resources and video-clips with instructions for use; pointing to LRs in lessons



Investing
in future

Courses @ HR4EU

- Course structure:
 - 10 lessons per course
 - 8 lessons with new content, 2 revisions
 - more than 1600 audio files, 200 digital illustrations
 - Dictionary
 - words and phrases used throughout the course
- Lesson structure:
 - lesson body
 - two quizzes
 - grammar book
 - *Practice your writing skills* tasks



Investing
in future

HR4EU technology

- courses (e-learning technology):
 - based on Moodle 2.9 with adapted theme
 - open source e-learning platform
 - features
 - interactive and intuitive interface
 - tracking and evaluation of users
 - changes needed in the core engine
- rest of the portal:
 - Wordpress with adapted theme

HR4EU & LRs



Investing
in future

- strongly connected with **language resources** in two directions:
 - ➔ directing to the use of LRs as the part of e-learning process
 - ➔ enlarging/enhancing/creating language resources
 - Croatian Morphological Lexicon (hml.ffzg.hr)
 - Croatian Dependency Treebank (hobs.ffzg.hr)
 - Croatian Wordnet (crown.ffzg.hr)
 - CroDeriV – Croatian Derivational Lexicon (croderiv.ffzg.hr)
 - Dictionary of Neologisms (neologizam.ffzg.unizg.hr)
 - Corpus of Croatian as L2
 - Semantic Role Labelling System



Investing
in future

HR4EU & LRs

- lessons and quizzes are designed to encourage the users to use LRs:
 - to find the appropriate word form or lemma in Croatian Morphological Lexicon
 - to learn semantically related words via lexical hierarchies or synsets in Croatian WordNet
 - to find derivationally related words in verbal derivational database CroDeriV
 - one part of the HR4EU portal is dedicated solely to Croatian language resources:

<http://www.hr4eu.hr/croatia/resources/>



HR4EU & LRs

- learners' activity will be used to enhance and enlarge existing LRs :
 - tracking their activity yielding empty results, and adding them to the respective resources
 - users' answers in *Practice your writing skills* tasks will be used to build the new LR, the corpus of Croatian as a second language
- some of the LRs were too complex their primary shape for language learners (HOBS)



HOBS before HR4EU

- approx. 4.500 sentences (Croatia Weekly 100kw, the newspaper sub-corpus of the Croatian National Corpus)
- 1st version – the sentences are manually tagged according to the PDT annotation guidelines
- 2nd version - modified PDT specification, adjusted to the syntactic structures of Croatian language

Syntactic sub-tagset for subordinate clause annotation in HOBS

tag	Description
Sub_Atr	attribute
Sub_Adv	adverbial
Sub_Obj	object
Sub_Pred	predicate
Sub_Sb	subject



Investing
in future

HOBS & HR4EU

- semantic layer of HOBS is developed - specification developed for the Croatian semantic role labelling
- SynSem Visualizer is developed – syntactic and semantic layer became available for search
- HR4EU sub-corpus was built



Investing
in future

HOBBS – semantic layer

- first steps towards automatic semantic role labelling in Croatian
- approx. 3500 sentences annotated with developed tagset for Croatian
- tags can be divided in two groups:
 - verbal arguments
 - adjuncts (adverbials and adverbial and attribute clauses) - can be very useful later in e.g. information extraction tasks

HOBs – semantic layer

Tag	Role	Tag	Role
AGT_anim	agent_animate	SRC	source
AGT_inanim	agent_inanimate	QUAN	quantity
PAT	patient	TMP	time
EXP	experiencer	LOC	location
BEN	beneficiary	LOC_ap	abstract location
RES	result	CAU/FIN	cause, intention
PART	participant	MNR	manner
TEM	theme	ATR	attribute clause
INS	instrument	SRC	source



Investing
in future

HOBS – semantic layer

- annotation principles:
 - semantic role is represented with one word → word is annotated with the SR tag
 - semantic role is represented with the whole phrase → SR tag is attached to the head node
 - dependent clause has semantic role → conjunction of the clause is marked with the SR tag



Investing
in future

SynSem Visualizer

- database-driven web application
- enables:
 - the queries on the syntactic and semantic level
 - the graphical representation of the hierarchical sentence structure
 - searching HOBS & HR4EU corpora independently and both at the same time
 - search by word form, lemma, morphosyntactic tag, syntactic tag or semantic role



Investing
in future

HR4EU sub-corpus

- 550 syntactically and semantically annotated sentences used in the Croatian language courses
- HOBBS model of annotation
- syntactic structure representations could improve the users' understanding of different grammatical relations
- semantic structure of the sentence can improve interpretation of semantic roles of the verb arguments

Conclusion



Investing
in future

- HR4EU - first free-of-charge e-learning portal for Croatian language
- more than 1.800 users from all around the world
- introduces users to Croatian language resources and their usefulness in learning a new language
 - LRs can be used when learning other languages as well – the same LRs are built and are freely available for dozens of languages



Investing
in future

Conclusion

- HOBS improvements:
 - semantic layer added
 - specification for SRL for Croatian
 - graphical representation of syntactic and semantic layer enabled via SynSem Visualizer
 - HR4EU sub-corpus added to HOBS
 - helps language learners to easily understand syntactic and semantic structure of the sentence



Thank you for your attention.



Investing
in future

www.hr4eu.hr | hr4eu@ffzg.hr