

Croatian Dependency Treebank 2.0: New Annotation Guidelines for Improved Parsing



Željko Agić*, Daša Berović++, Danijela Merkle++, Marko Tadić+

*Linguistics Department, University of Potsdam, Germany, +Department of Linguistics, Faculty of Humanities and Social Sciences, ++University of Zagreb, Croatia
zagic@uni-potsdam.de, {dberovic, dmerkler, mtadic}@ffzg.hr

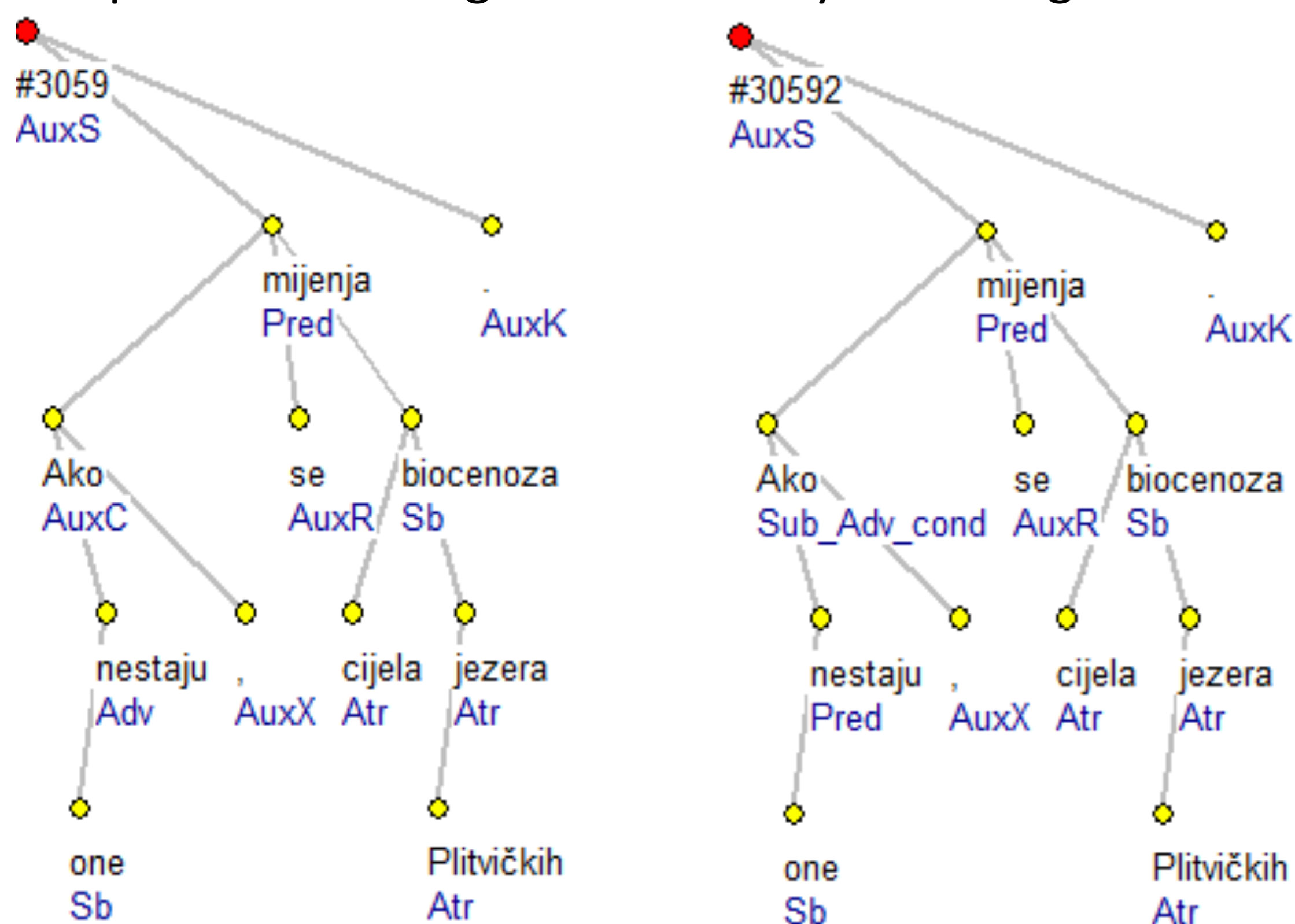
LREC2014, Reykjavik
2014-05-29

What?

- new version of the Croatian Dependency Treebank 2.0 (HOBS 2.0)
 - derived from the HOBS 1.0 (Tadić, 2007)
 - closely following PDT analytical layer annotation guidelines (Hajič et al., 2001)
 - number of issues with the manual annotation of subordinate clauses encountered
- departure from the original PDT annotation guidelines, but adapted better for Croatian

Where?

- subordinate clause annotation
- predicate in the subordinate clause is annotated with the syntactic tag of a syntactic element whose place it takes in the main clause
 - e.g., if subordinate clause = subject clause, its predicate is assigned with the syntactic tag **Sb**

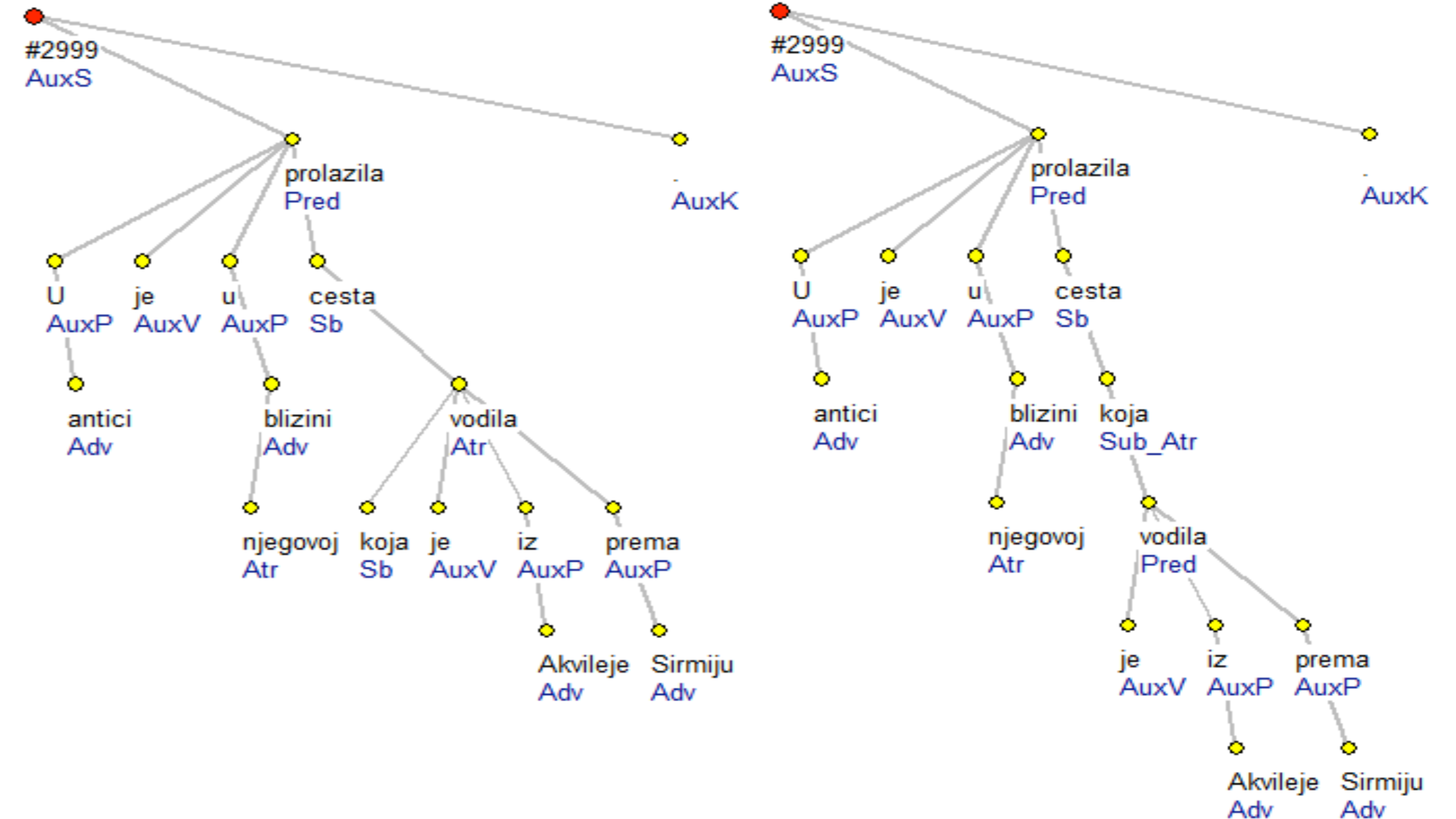


How?

- new annotation guideline proposal
 - conjunctions in the tree should depend on the predicate in the main clause
 - predicate in the subordinate clause should depend on the conjunction that introduces it
- new syntactic sub-tagset for subordinate clause annotation in HOBS 2.0

Clause type	Syntactic tag
attribute	Sub_Atr
adverbial	Sub_Adv
object	Sub_Obj
predicate	Sub_Pred
subject	Sub_Sb

example of an attribute clause



for adverbial clauses additional subclassification introduced

Adverbial clause	Sub_Adv
local	Sub_Adv_loc
temporal	Sub_Adv_temp
modal	Sub_Adv_mod
causative	Sub_Adv_caus
consequential	Sub_Adv_cons
final	Sub_Adv_fin
conditional	Sub_Adv_cond
concessive	Sub_Adv_cons

possible search in the HOBS 2.0 for

- specific label: **Sub_Adv_mod**
- more abstract levels: **Sub_Adv** or **Sub** only

treebanks statistics

Trebank	Tagset	Tokens	Sentences	MSDs	Tags
HOBS 1.0	basic	117369	4626	914	27
	full	117369	4626	914	70
HOBS 2.0	basic	117369	4626	911	28
	full	117369	4626	911	81

Experiment with parsing

- used graph-based MSTParser generator system (McDonald et al., 2005)
 - backward compatibility with previous parsing experiments on HOBS 1.0 (Agić 2012, Agić and Merkle 2013, Agić et al. 2013)
 - tenfold cross-validation parsing experiment
- | Trebank | Tagset | LAS | UAS | LA |
|----------|--------|-------|-------|-------|
| HOBS 1.0 | basic | 71.93 | 79.98 | 84.65 |
| | full | 71.71 | 80.34 | 81.75 |
| HOBS 2.0 | basic | 74.50 | 81.41 | 86.87 |
| | full | 73.04 | 81.10 | 82.85 |

Learning curves (LAS), two versions of HOBS

