



UPPSALA
UNIVERSITET

UNIVERSAL DEPENDENCIES: A PARSING COMPARISON FOR SWEDISH

**FILIP ANTONOV
UPPSALA UNIVERSITET
SVERIGE**



- *Universal Dependencies* is an open, community-driven project aiming to provide a consistent, language-independent grammatical annotation scheme for dependency treebanks



- *Universal Dependencies* is an open, community-driven project aiming to provide a consistent, language-independent grammatical annotation scheme for dependency treebanks
- Since *UD* are not related to any particular language/language group, what impact might *UD* have on parsing quality in comparison to traditional annotation schemes?



- *Universal Dependencies* is an open, community-driven project aiming to provide a consistent, language-independent grammatical annotation scheme for dependency treebanks
- Since *UD* are not related to any particular language/language group, what impact might *UD* have on parsing quality in comparison to traditional annotation schemes?
- The study presents parsing results for Swedish with 2 parsing systems: *MaltParser* & *Stanford NN Parser* – trained and evaluated on the novel *UD Treebank* and on the classical *Talbanken treebank*



- Some details on *Universal Dependencies*:
 - unified approaches – POS, morphosyntactic descriptions, dependency relations
 - currently, 17 POS tags divide into open & closed class words
 - DR between content words prioritized, since function words may correspond to e.g. morphological inflection
 - content words relate by DR, function words connect to the content word they specify, punctuation attaches to a phrase's head (Nivre et al, 2016)



- Results for Swedish (*UD Treebank – Talbanken Treebank*):

	<i>UD</i>	<i>Talbanken</i>
	<u>LAS / UAS \ TedEval</u>	<u>LAS / UAS \ TedEval</u>
<i>MaltParser</i>	83,2 / 86,5 \ 93,9	79,2 / 85,3 \ 90,3
<i>Stanford NN Parser</i>	80,1 / 82,8 \ 93,3	77,8 / 83,6 \ 89,7



- Results for Swedish (*UD Treebank – Talbanken Treebank*):

	<i>UD</i>	<i>Talbanken</i>
	<u>LAS / UAS \ TedEval</u>	<u>LAS / UAS \ TedEval</u>
<i>MaltParser</i>	83,2 / 86,5 \ 93,9	79,2 / 85,3 \ 90,3
<i>Stanford NN Parser</i>	80,1 / 82,8 \ 93,3	77,8 / 83,6 \ 89,7

- Possible concerns on using a language-independent annotation approach in a monolingual environment are not confirmed in this study
- *Universal Dependencies* in fact increase parsing scores for Swedish by a small margin, results which are consistent across both tested parsers



- Results for Swedish (*UD Treebank – Talbanken Treebank*):

	<i>UD</i>	<i>Talbanken</i>
	<u>LAS / UAS \ TedEval</u>	<u>LAS / UAS \ TedEval</u>
<i>MaltParser</i>	83,2 / 86,5 \ 93,9	79,2 / 85,3 \ 90,3
<i>Stanford NN Parser</i>	80,1 / 82,8 \ 93,3	77,8 / 83,6 \ 89,7

- Possible concerns on using a language-independent annotation approach in a monolingual environment are not confirmed in this study
- *Universal Dependencies* in fact increase parsing scores for Swedish by a small margin, results which are consistent across both tested parsers
- UD ver. 1.3 (May 2016) cover 40 languages, incl. 2 Russian treebanks:
 - 99.000 tokens (Википедия)
 - 1 million tokens (СинТагРус, НКРЯ)