



# Comparing trees: a gentle introduction to treebanking comparison

Giuseppe G. A. Celano  
University of Leipzig  
at the Humboldt Chair in Digital Humanities  
23 May 2015



# overview

- Why evaluation
- How to evaluate
- Evaluation for treebanking
  - The case of the AGDT

# why evaluation

*good data*



*consistent data*



*reliable data*

# how to evaluate: premise

manual annotation is:

*time-consuming*

PCEDT: about 7,5 sentences per hour

AGDT 2.0: about 5 sentences per hour

LDT: about 6.5 sentences per hour

# how to evaluate: logic

inter-coder agreement

if annotators agree (to a certain extent),  
their annotations are **reliable**

if annotations are reliable, they are *probably*  
**valid**

# getting a high final inter-annotator agreement

- precise annotation guidelines
  - checking tests
- frequent calculations for inter-annotator agreements

# how much double annotation?

PDT 2.0 (tect. layer) : about 4% (of 50.000 tk)

English web treebank : about 2.6% (of 254.830 tk)

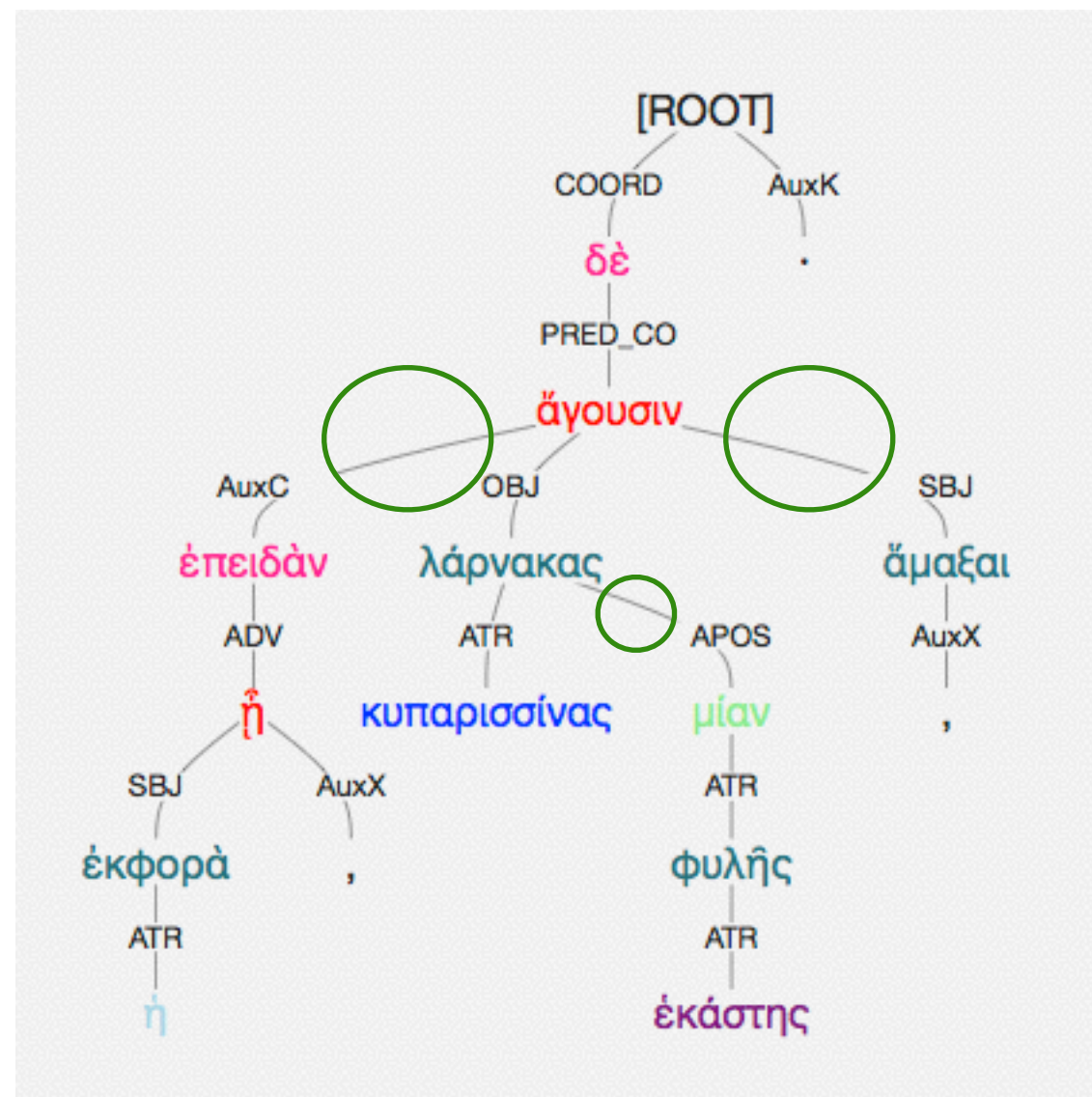


*then single annotation*



# evaluation metrics

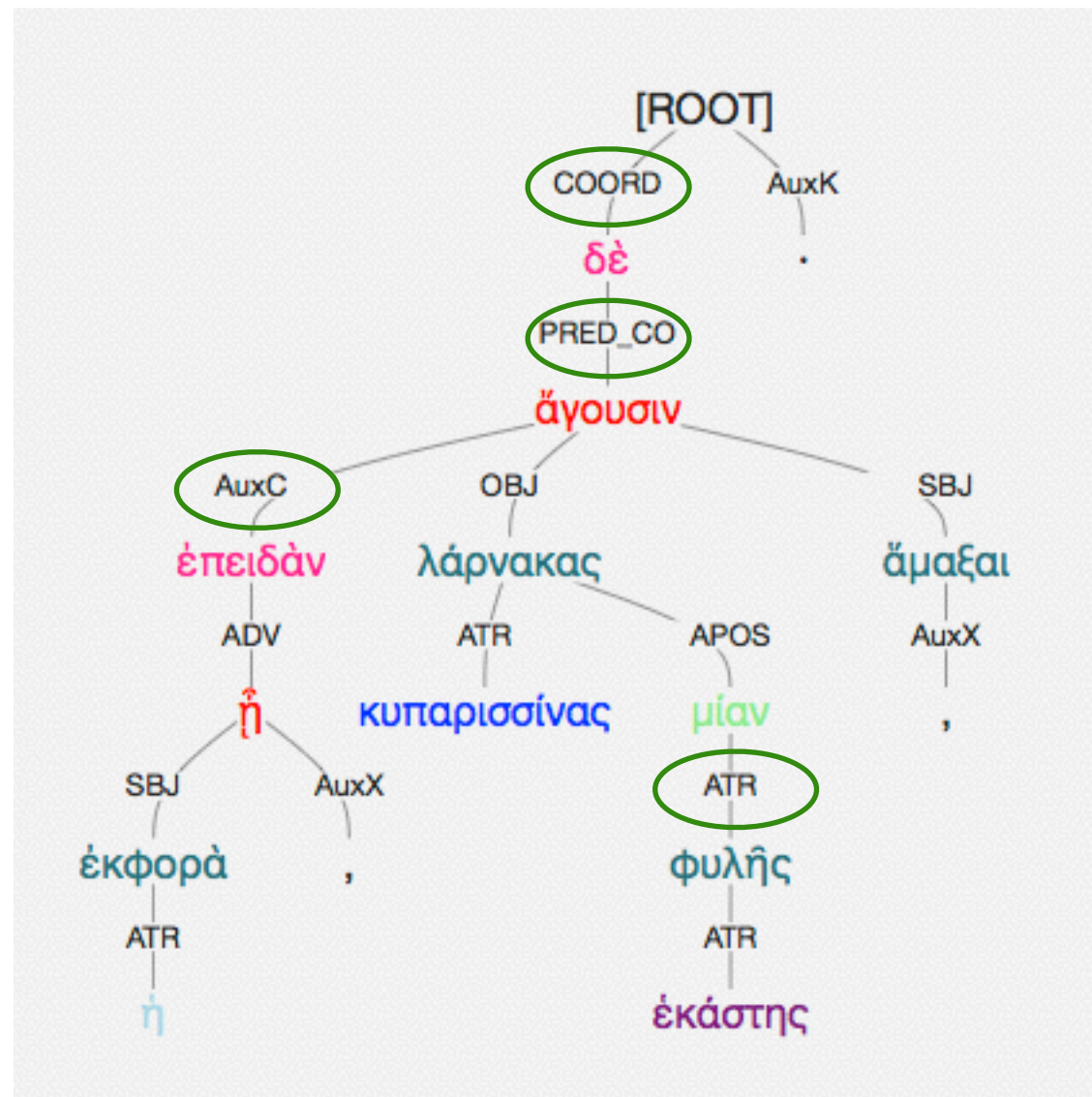
## *Unlabeled attachment agreement (UAA)*





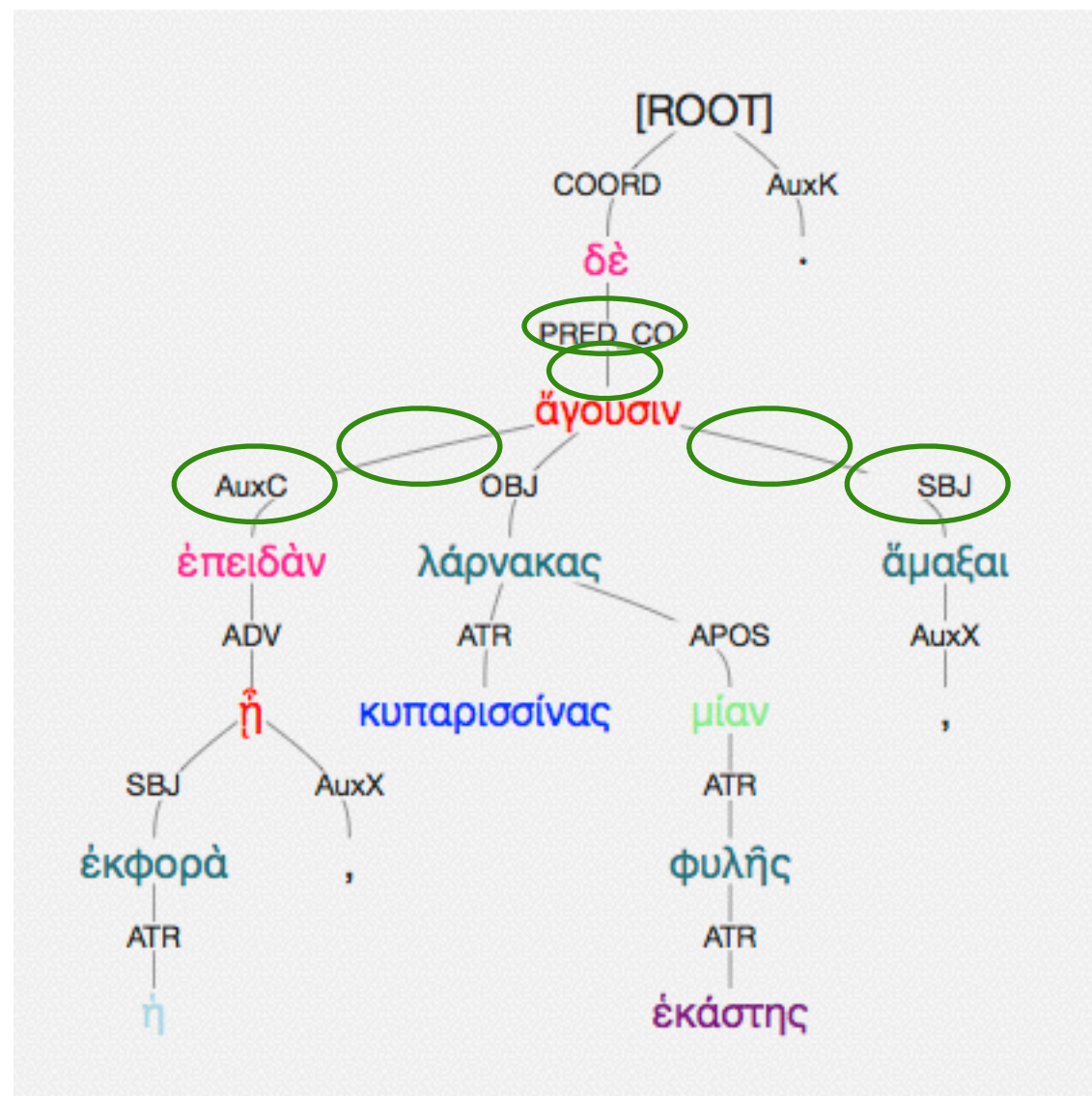
# evaluation metrics

## *label only agreement (LOA)*



# evaluation metrics

## *labeled attachment agreement (LAA)*



# evaluation metrics

*word-based (micro-average)*

*sentence-based (macro-average)*

# evaluation metrics

*percentage of agreement*

*chance-corrected coefficients*

*two coders:  $S$ ,  $\pi$ ,  $\kappa$*

# open questions

*do all nodes count the same when comparing?*

*which values for agreement coefficients are acceptable?*

# the case of the AGDT

*(morphological and syntactic annotation)*

standard method/double annotation:

Homer (232.569 tk)

Hesiod (18.881 tk)

single annotation:

e.g., The Deipnosophists,  
Antigone, Euthyphro



# the case of the AGDT 2.0

*(morphological, syntactic, and semantic  
annotation)*

*standard method/double annotation:*

*Aesop (4924 tk)*

*Ps.-Apollodorus (1265 tk)*



# inter-coder agreement for the AGDT 2.0

*annotators within ERASMUS program*

*ICA documented online (multiple  
calculations):*

*original annotated files*

*inter-coder agreement files*

# inter-coder agreement for the AGDT 2.0

## 1. Morphology

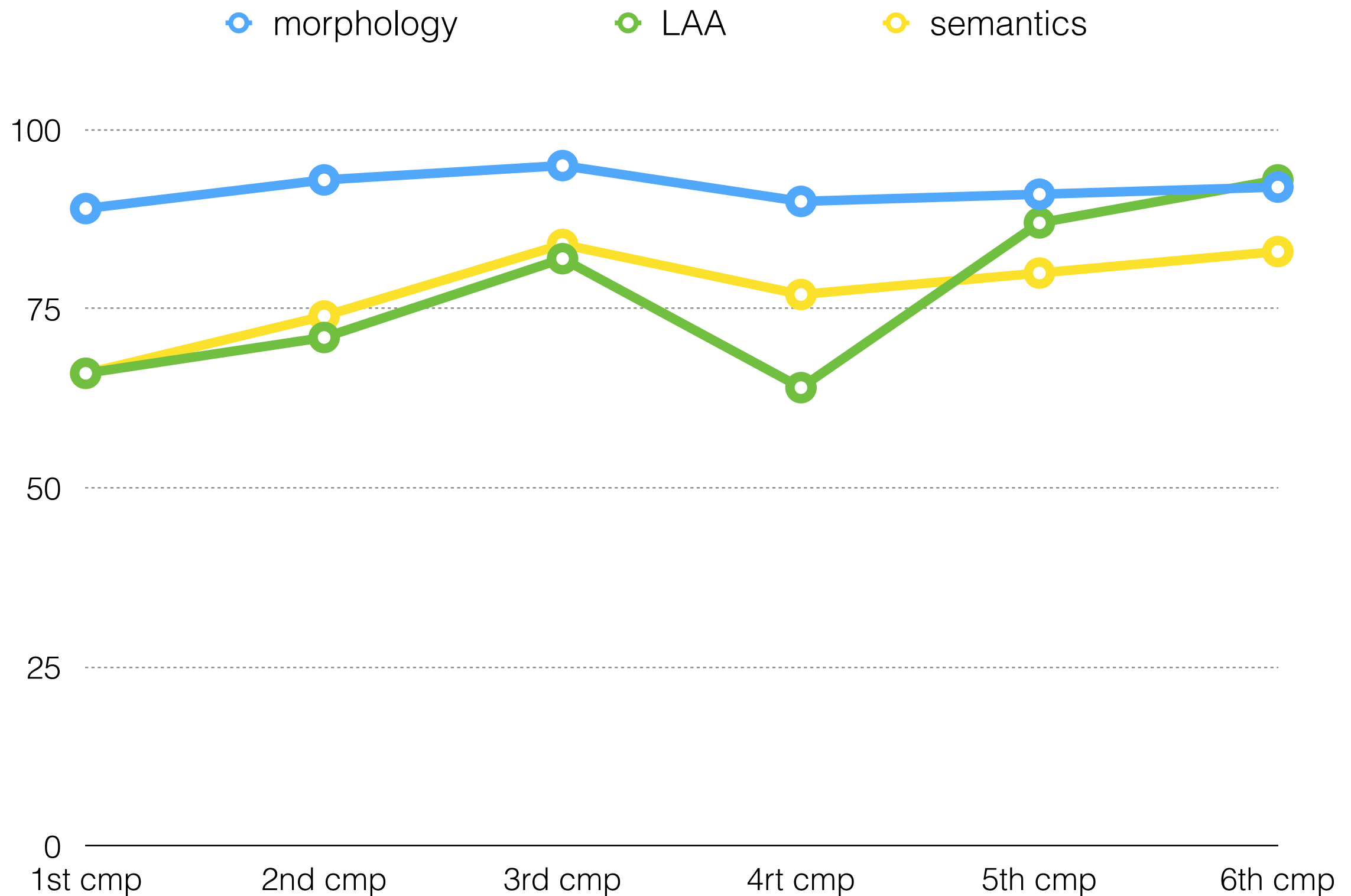
### Metrics

Word number	Words with equal value	Words with unequal value	Percentage of agreement	Cohen's kappa
1870	1700	170	0.9090909090909091	0.9062025517390737

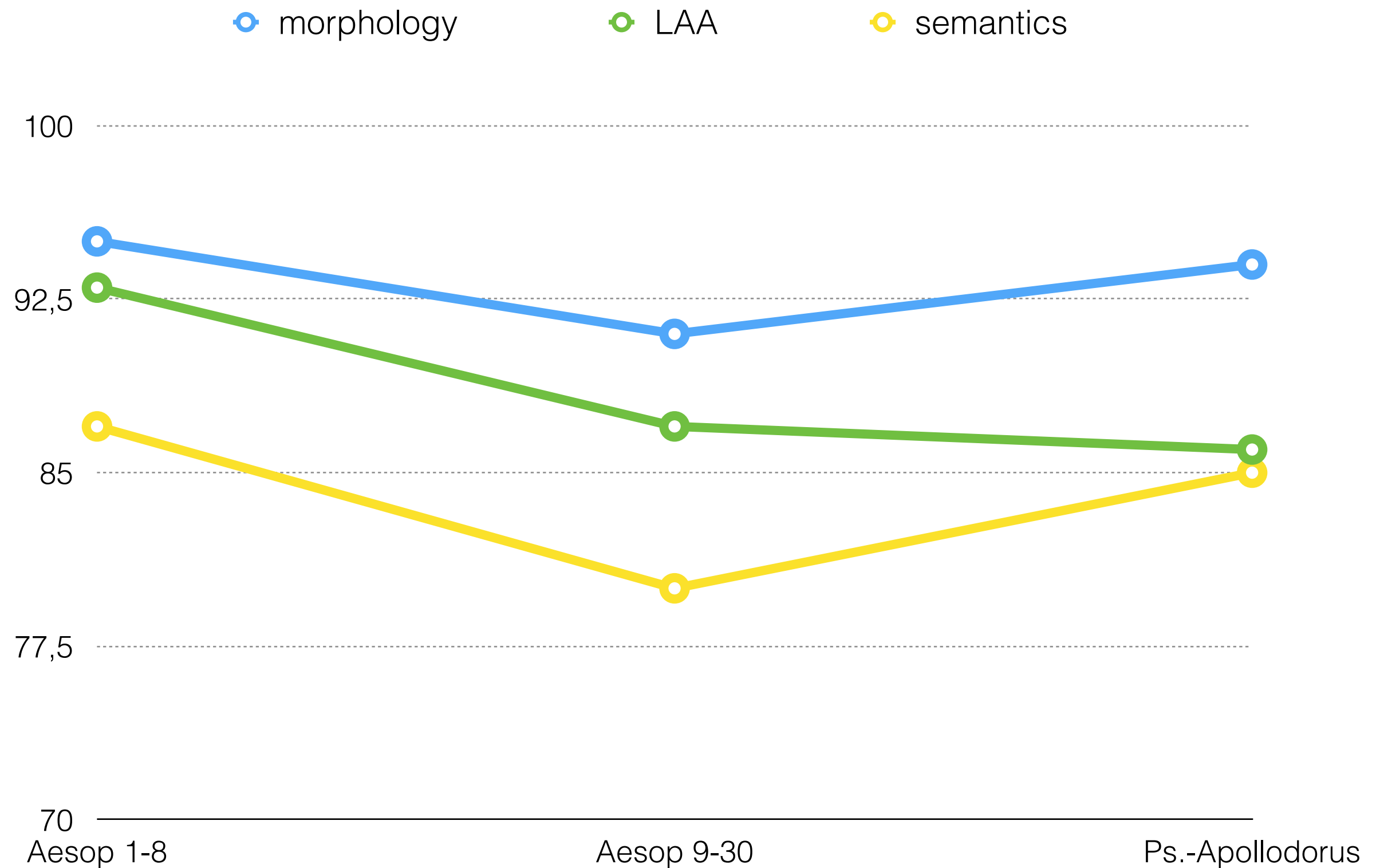
### Differences

	Sentences	Value in 1st document	Value in 2nd document
1	<b>1st</b> Ἀηδόνι συνεβούλεψε χελιδὼν τοῖς ἀνθρώποις εἶναι <b>ὁμόροφον</b> καὶ σύνοικον ὥς αὐτὴ <b>2nd</b> Ἀηδόνι συνεβούλεψε χελιδὼν τοῖς ἀνθρώποις εἶναι <b>ὁμόροφον</b> καὶ σύνοικον ὥς αὐτὴ	a-s---ma-	a-s---fa-
2	<b>1st</b> Ἀηδόνι συνεβούλεψε χελιδὼν τοῖς ἀνθρώποις εἶναι ὁμόροφον καὶ <b>σύνοικον</b> ὥς αὐτὴ <b>2nd</b> Ἀηδόνι συνεβούλεψε χελιδὼν τοῖς ἀνθρώποις εἶναι ὁμόροφον καὶ <b>σύνοικον</b> ὥς αὐτὴ	a-s---ma-	a-s---fa-
3	<b>1st</b> Ἀηδόνι συνεβούλεψε χελιδὼν τοῖς ἀνθρώποις εἶναι ὁμόροφον καὶ σύνοικον <b>ὥς</b> αὐτὴ <b>2nd</b> Ἀηδόνι συνεβούλεψε χελιδὼν τοῖς ἀνθρώποις εἶναι ὁμόροφον καὶ σύνοικον <b>ὥς</b> αὐτὴ	d-----	c-----

# work-in-progress inter-coder agreement for the AGDT 2.0



# (almost-)final inter-coder agreement for the AGDT 2.0



Thank you for your attention!