



Maitrey Mehta<sup>1</sup>, Mayank Jobanputra<sup>2</sup>, Çağrı Çöltekin<sup>2</sup>

<sup>1</sup> School of Computing, University of Utah

<sup>2</sup> Department of Linguistics, University of Tübingen

Relevant WGs: **WG1, WG4**

### Introduction

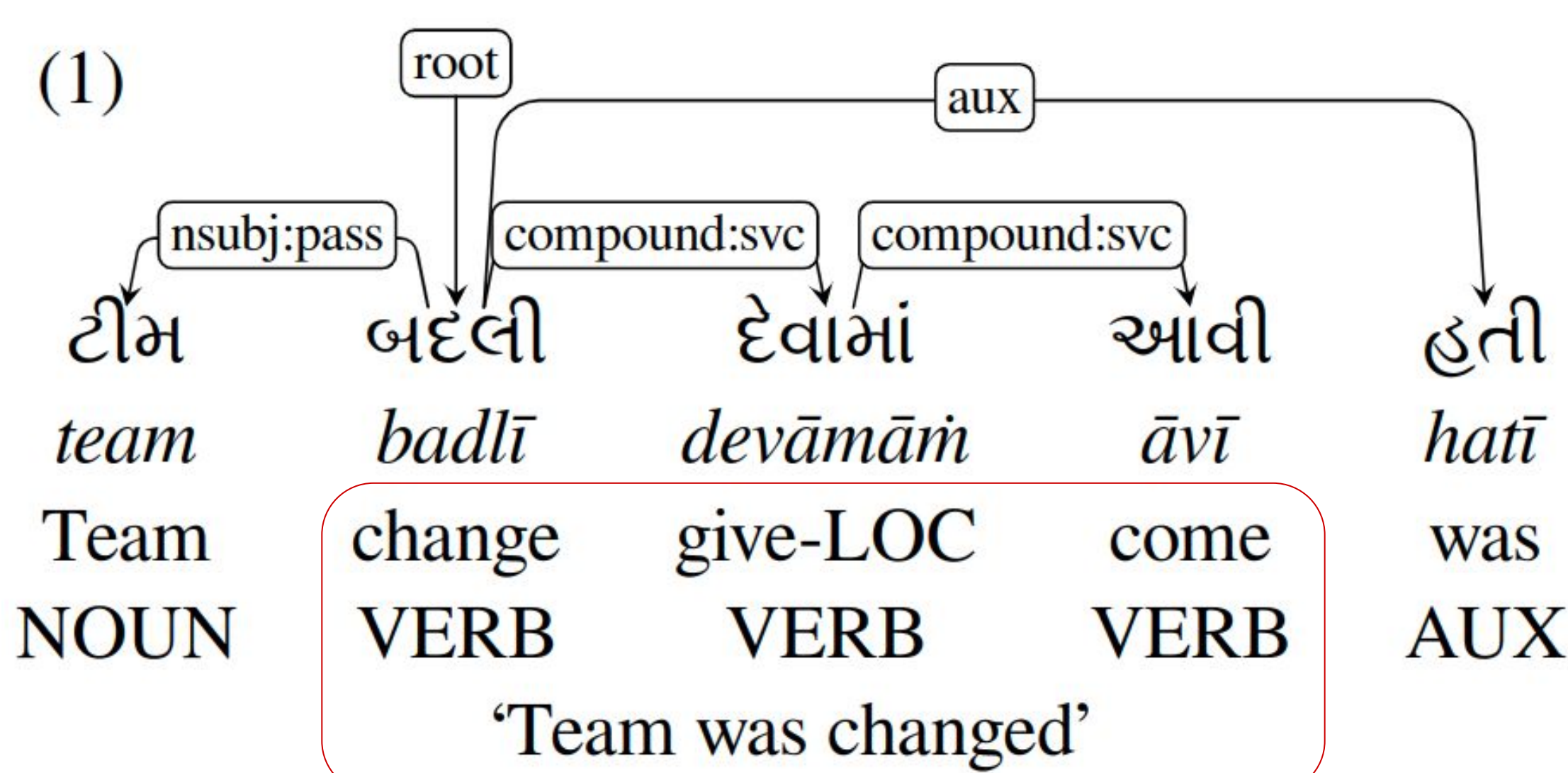
- We present an ongoing annotation effort for a new UD treebank in Gujarati, an Indo-Aryan language, widely spoken by over 56 million speakers.
- For our pilot study, we are doubly annotating a total of 300 sentences.
- We have taken sentences from Samanantar(news), UD Cairo(short), and Gujarati translations of the French novel – The Little Prince (fiction) for diversity purposes.

Genre	sentences	tokens
news	240	2666
short	20	173
fiction	50	658

Pilot data statistics in the Gujarati UD corpus

- We discuss some interesting cases we have encountered in our pilot study.

### Compound Verb with Locative Markers

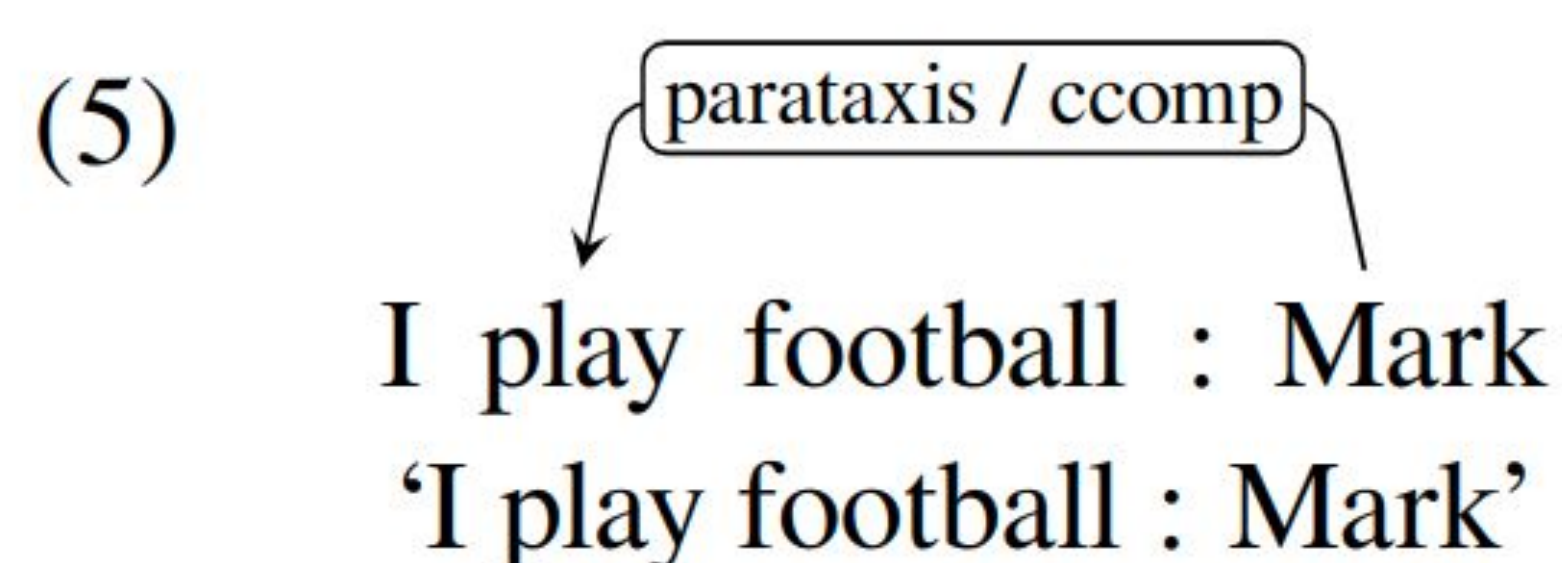


- The functionality of the locative marker and the morphological features it merits remains unclear.

### Genuine disagreements

- This disagreement arises for the selection of the ROOT due to focus ambiguity in the sentence. Instead of mitigating this disagreement, we keep both annotations in adjudicated treebank.

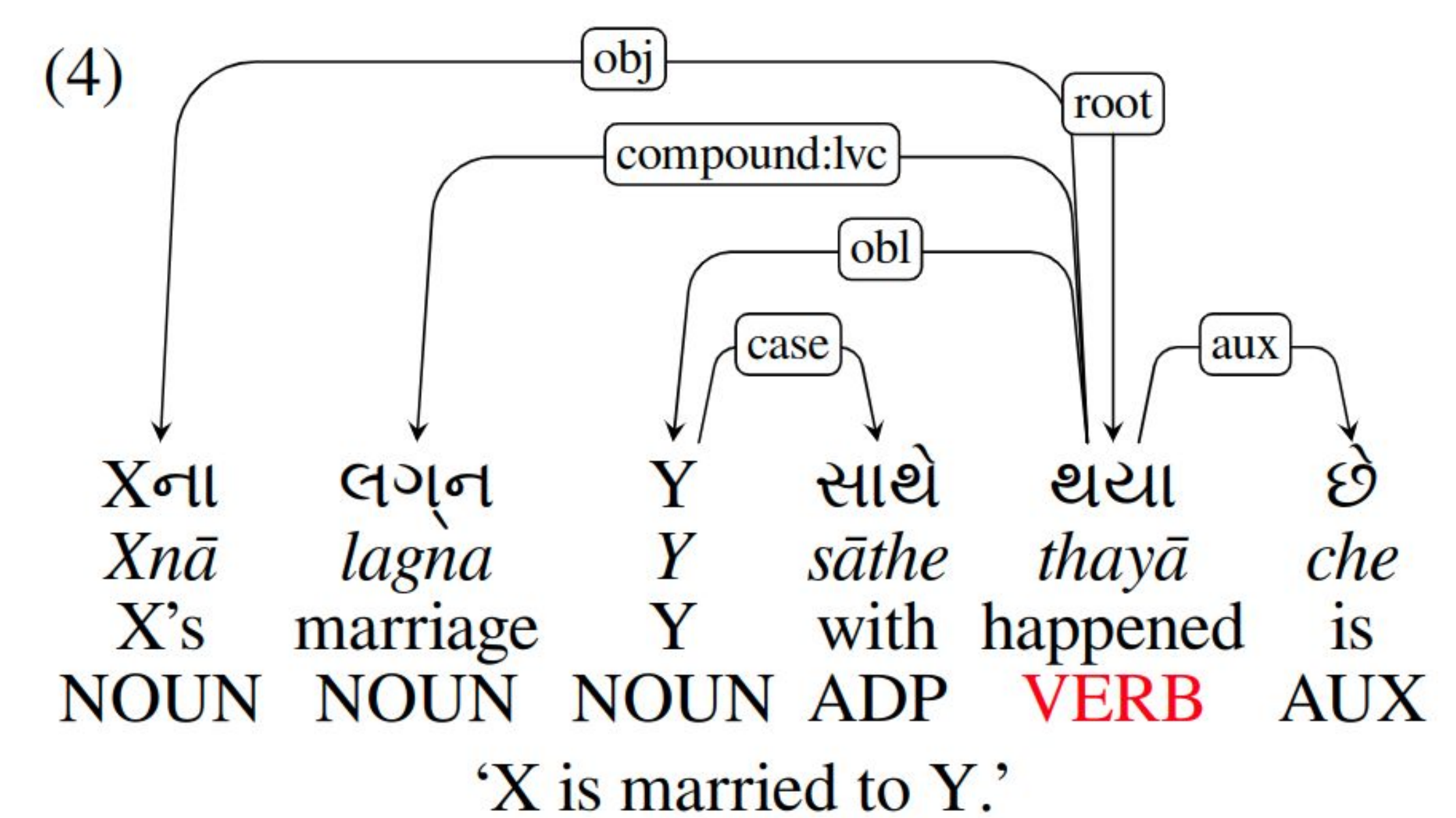
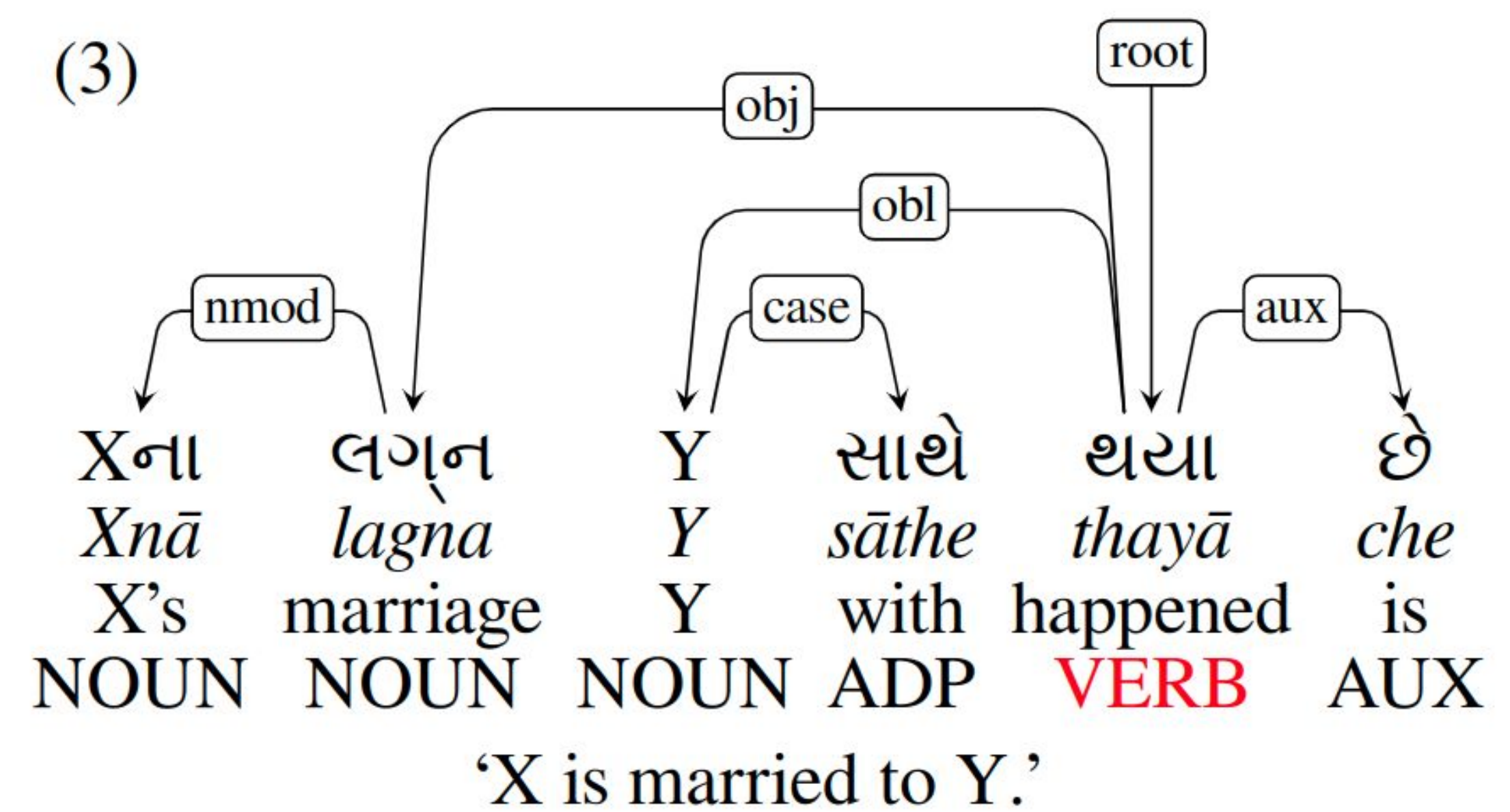
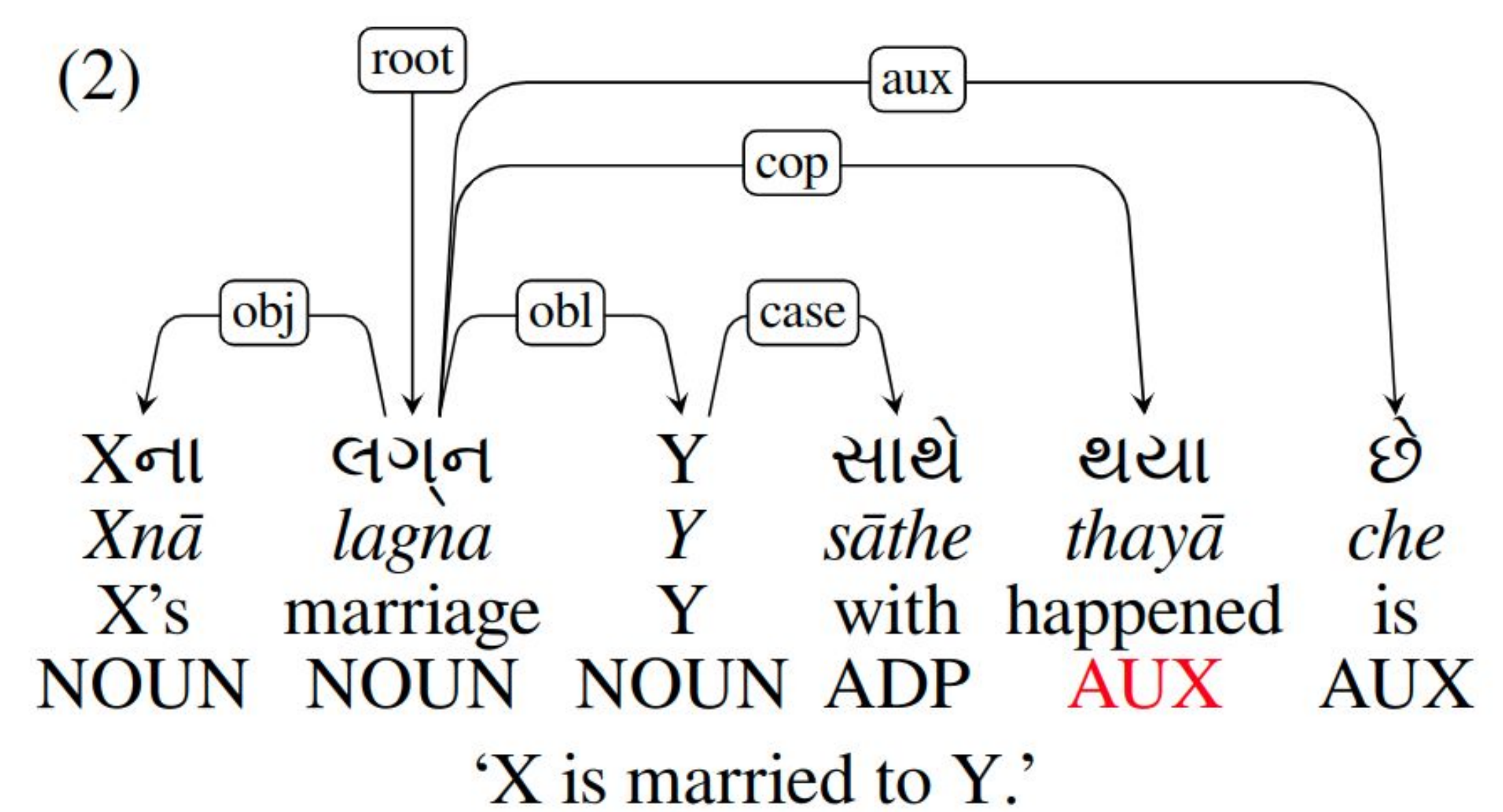
### Quoter and Quotation



- Early UD literature suggests solutions for cases with speech verbs. Recent guidelines recommend ccomp over parataxis for reported speech.

### Aspectual Verb vs. Independent Verb

- We face challenges in deciding the role of a multifaceted verb *thayā* ('happened').
- In Gujarati, *thayā* can occur in the text as an independent verb as well as a light verb.
- We want to investigate two major questions:
  - Do we consider *thayā* as a VERB or an AUX (2)?
  - If we consider *thayā* to be a VERB, should we annotate it as an independent verb (3) or as an aspectual light verb (4)?



### Splitting Genitive Markers

- According to Gujarati grammars, demonstrative pronouns like એ, તે, પેલું behave differently when attached to a nominal, versus when used independently.
- When occurring independently, we treat them as pronouns.
- Gujarati grammars do not discuss determiners as such. However, we see this usage closer to the UD definition of determiners and hence use the same.