

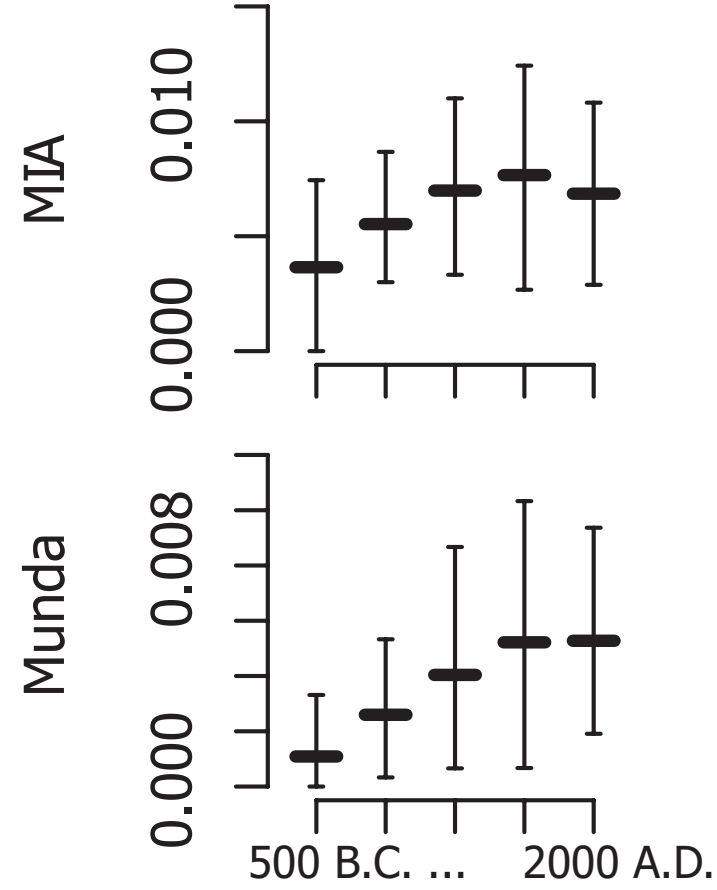
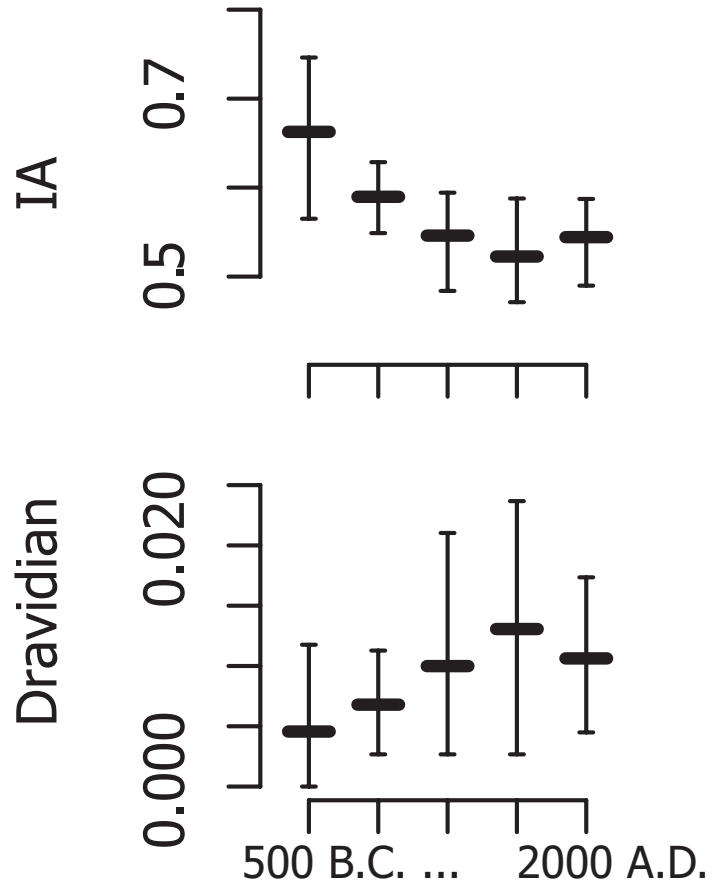
Extracting dependency trees from Sanskrit texts

Oliver Hellwig

Institut für Sprachen und Kulturen Südasiens

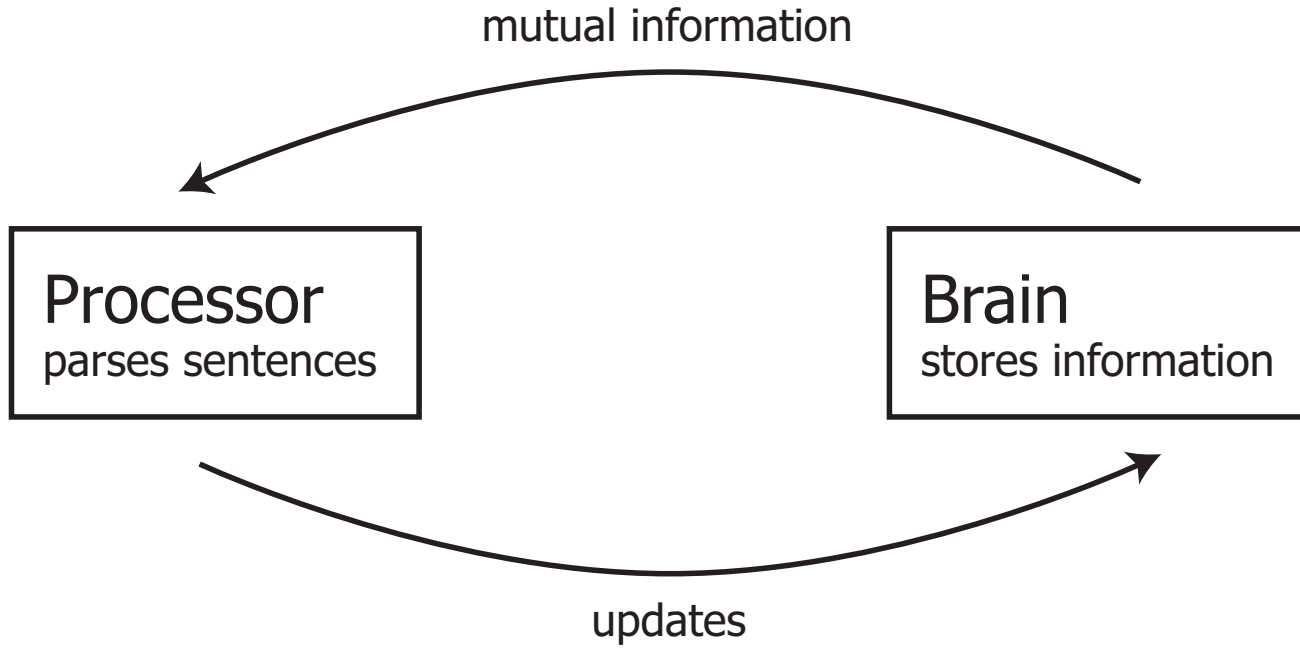
Freie Universität Berlin, Germany

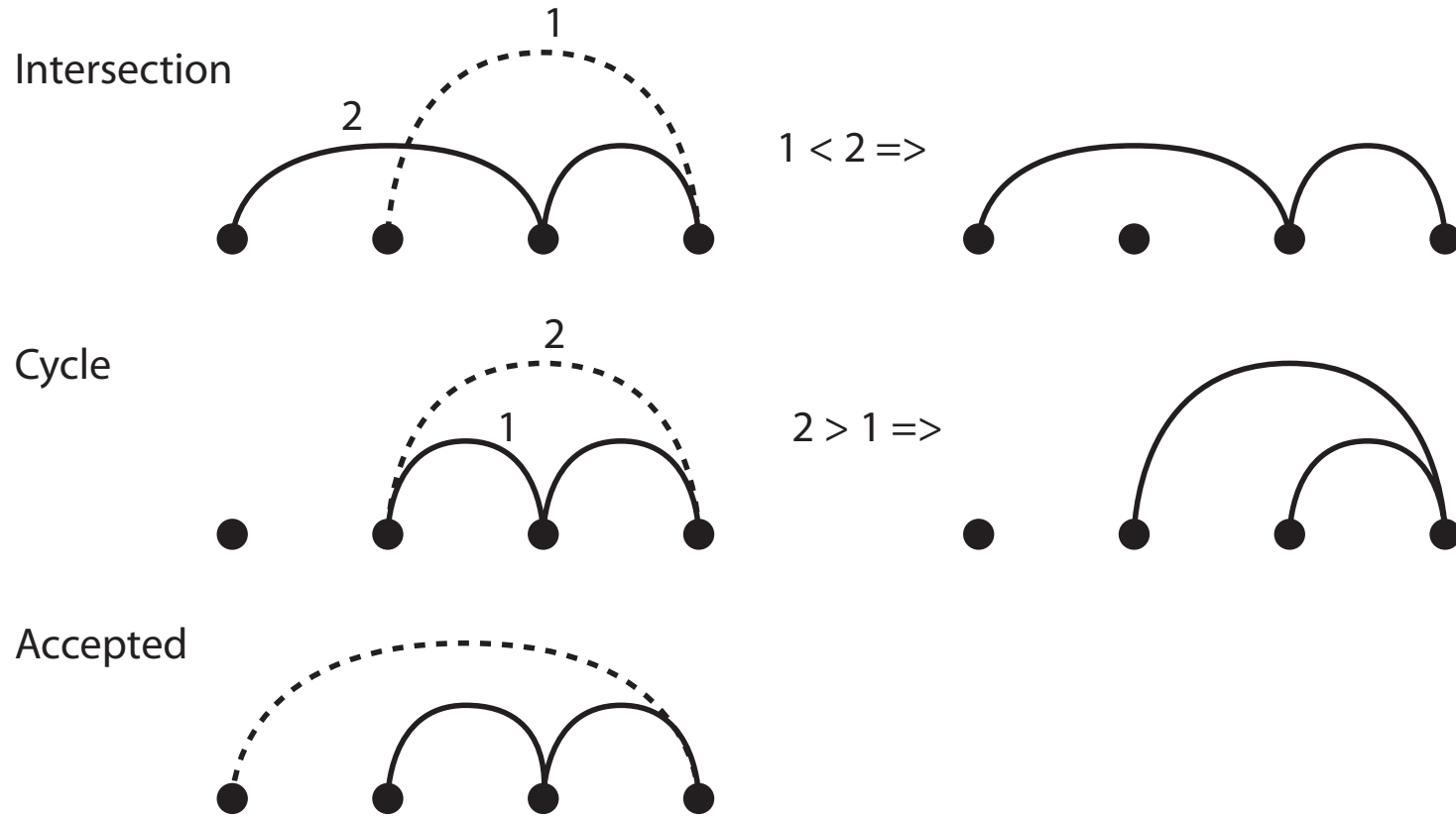
1. Collecting data:
 - (a) **Dependency trees**
 - (b) Dep. trees $\rightarrow$ syntactic trees
2. Comparing structures by graph matching
3. Deriving syntactic time markers?



Basic principles of Yuret's parser:

- Dependency structure defined by mutual lexical information → undirected graph
- Graph should be acyclic (one parent per child).
- Links should not cross (projectivity).





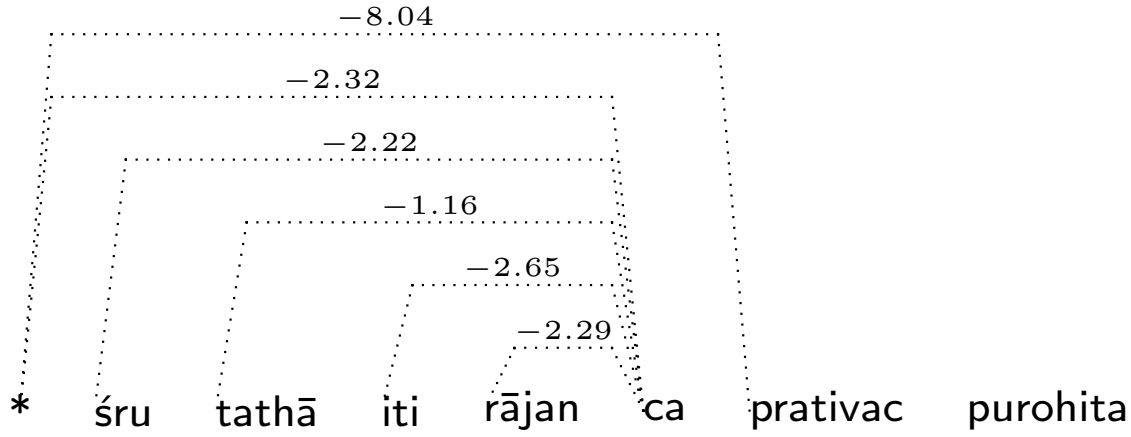
Improving Yuret's parser:

1. Grammatical information
2. Verb-valence dictionary
3. Smoothing cooccurrence frequencies

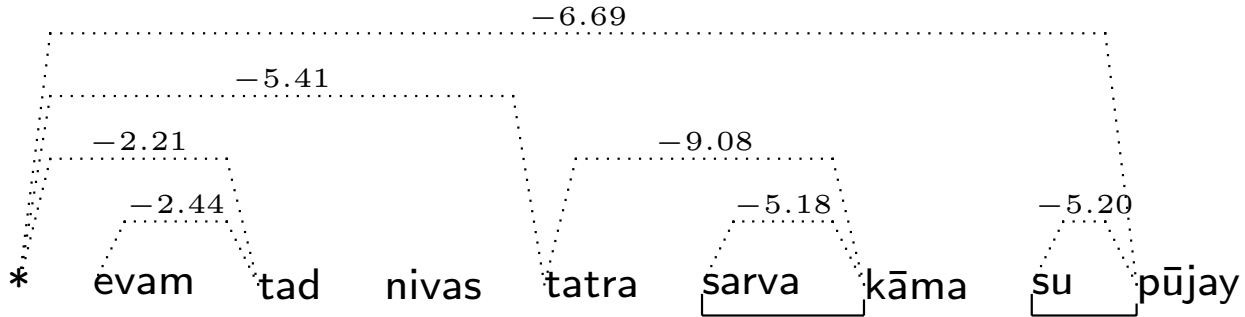
$$(w_1, w_2) \Rightarrow \sum \begin{matrix} (w'_{11}, w_2) \\ (w'_{12}, w_2) \\ \dots \end{matrix}$$

4. Syntactic rules (fixed, enforcements)

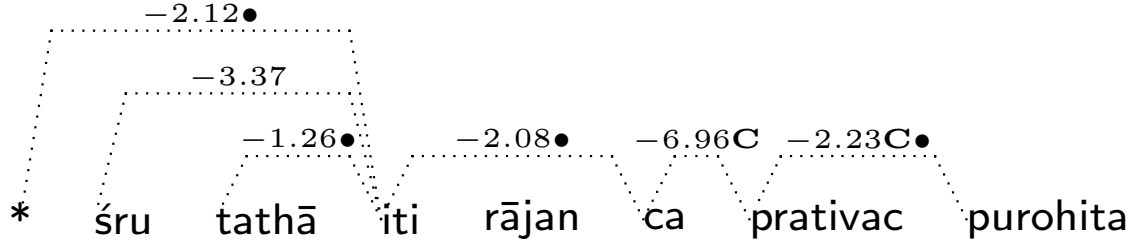
śrutvā tatheti rājā ca pratyuvāca purohitam



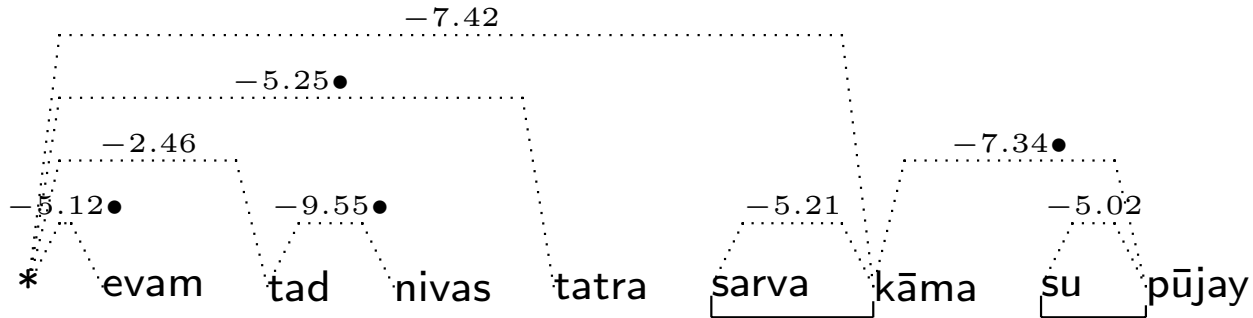
evaṃ sa nyavasat tatra sarvakāmaiḥ supūjitaḥ



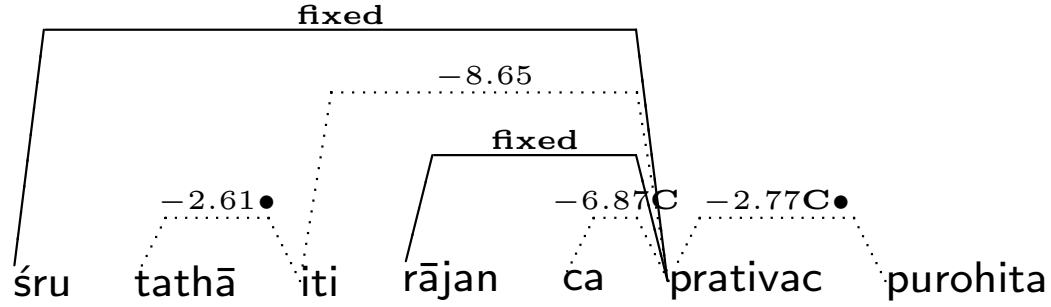
śrutvā tatheti rājā ca pratyuvāca purohitam



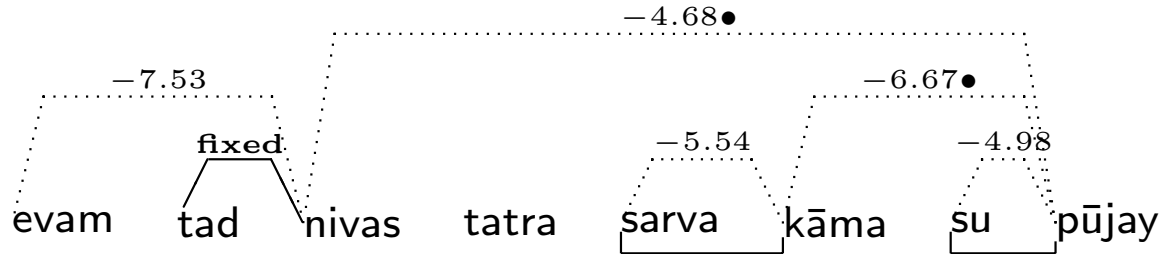
evam sa nyavasat tatra sarvakāmaiḥ supūjitaḥ



śrutvā tatheti rājā ca pratyuvāca purohitam



evam sa nyavasat tatra sarvakāmaiḥ supūjitaḥ



[illegible]