

Extracting dependency trees from Sanskrit texts

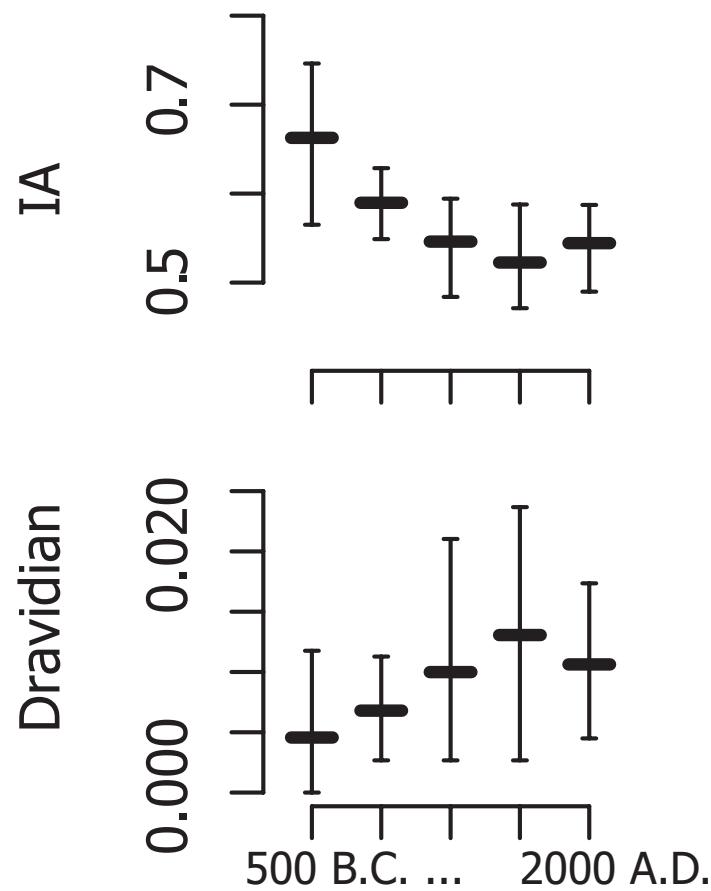
Oliver Hellwig

Institut für Sprachen und Kulturen Südasiens

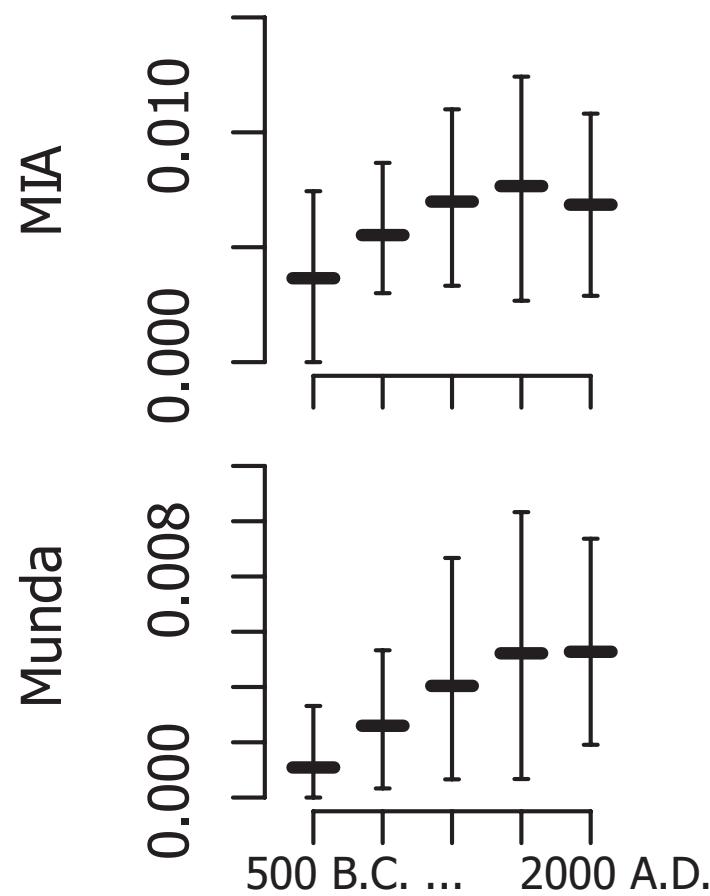
Freie Universität Berlin, Germany

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

SANSKRIT DEPENDENCY TREES



Historical developments in the Sanskrit vocabulary

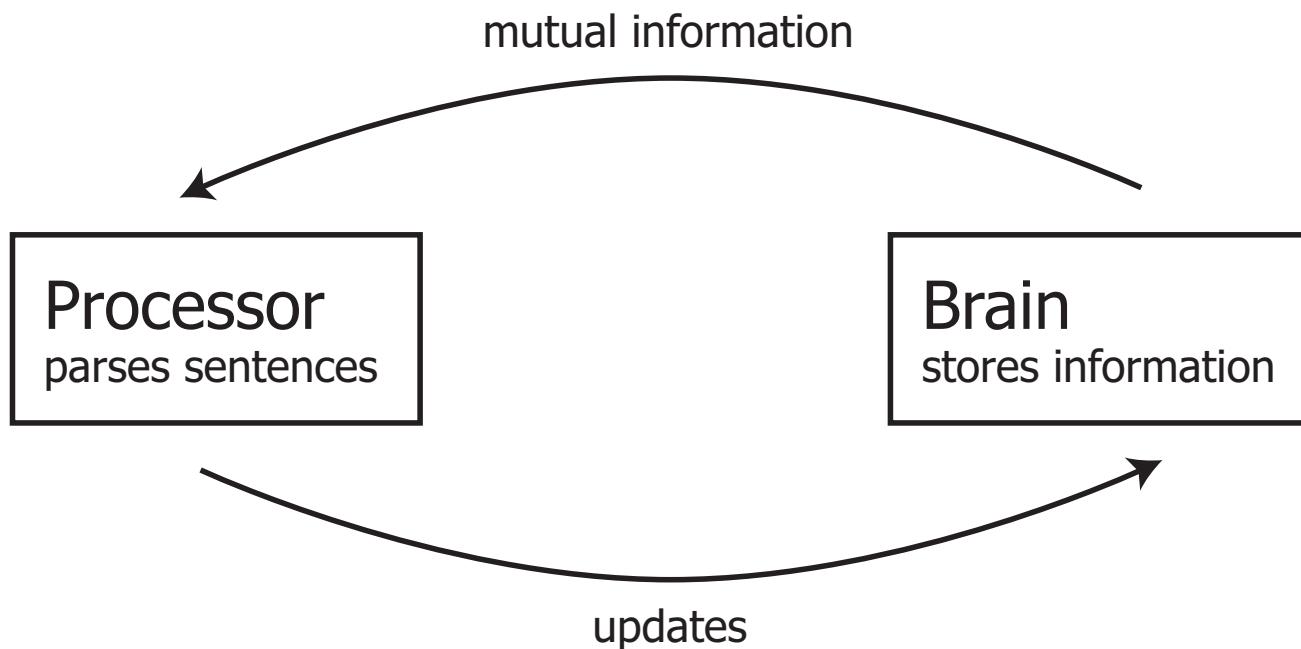


Basic principles of Yuret's parser:

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

SANSKRIT DEPENDENCY TREES

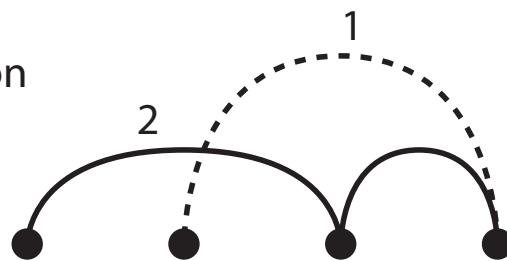
Yuret's parser (II)



SANSKRIT DEPENDENCY TREES

Yuret's parser (III)

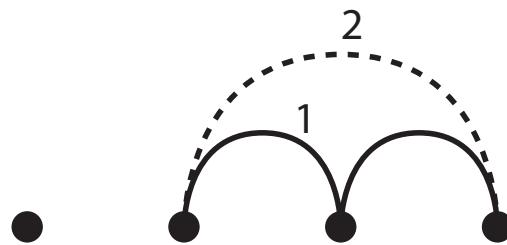
Intersection



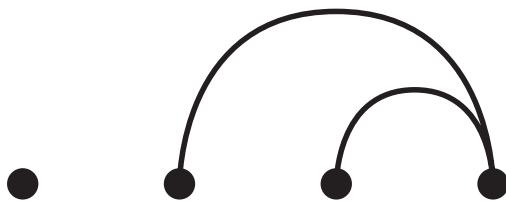
$1 < 2 \Rightarrow$



Cycle



$2 > 1 \Rightarrow$



Accepted



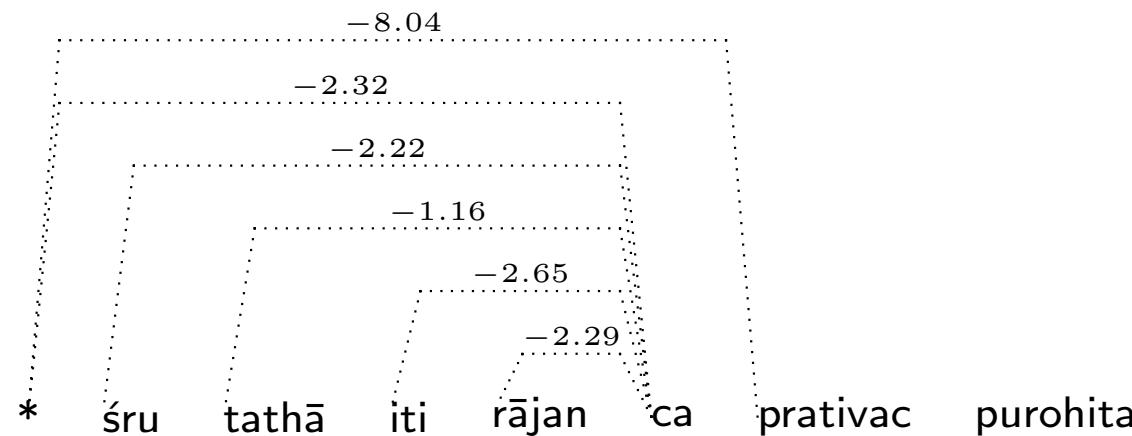
Improving Yuret's parser:

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

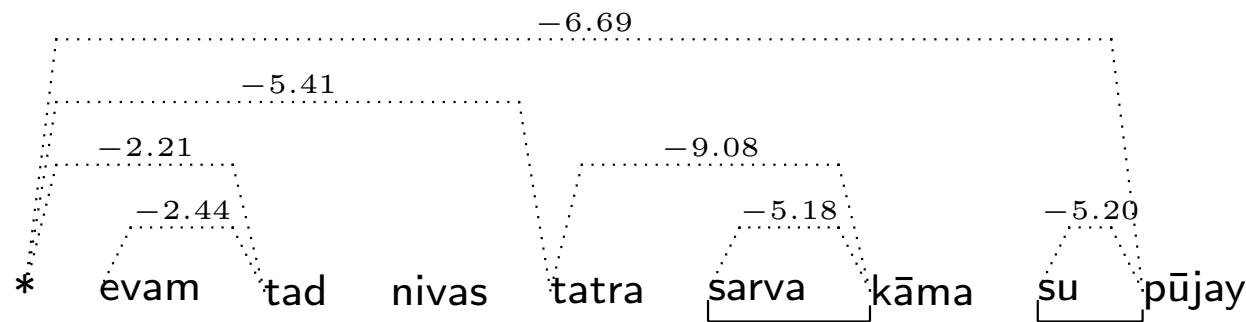
$$(w'_1, w_2) \\ (w_1, w_2) \Rightarrow \sum (w'_{12}, w_2) \\ \dots$$

4. Syntactic rules (fixed, enforcements)

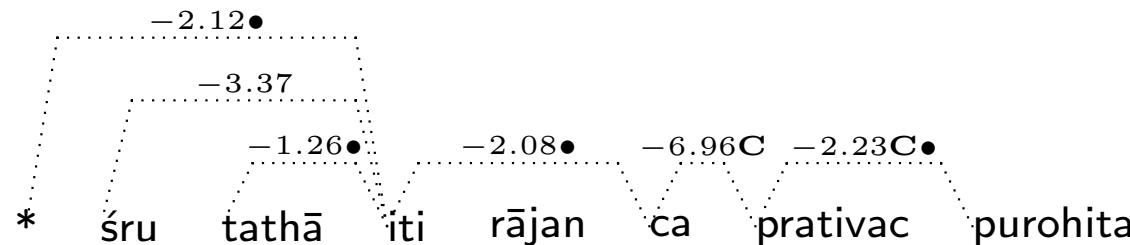
śrutvā tatheti rājā ca pratyuvāca purohitam



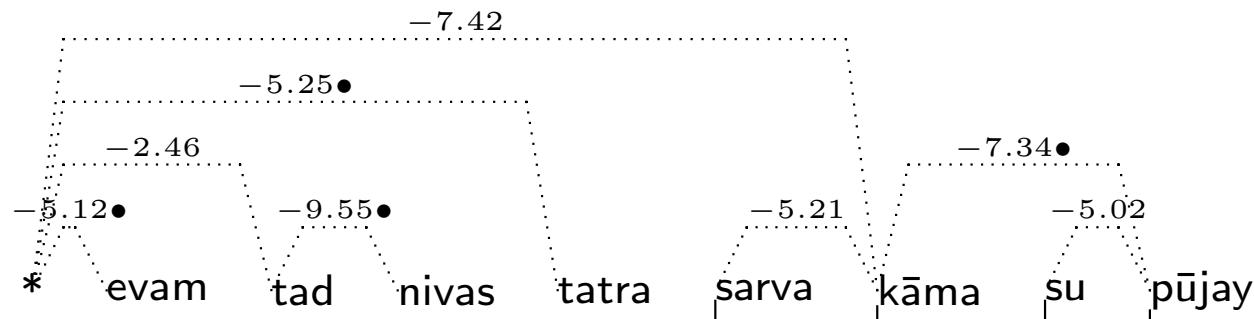
evam sa nyavasat tatra sarvakāmaih supūjitaḥ



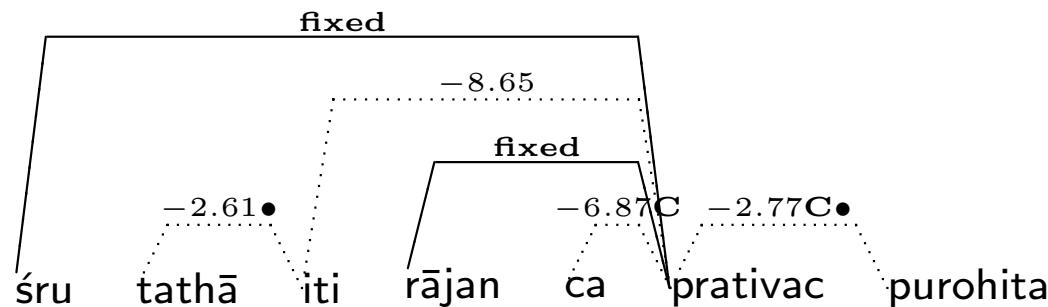
śrutvā tatheti rājā ca pratyuvāca purohitam



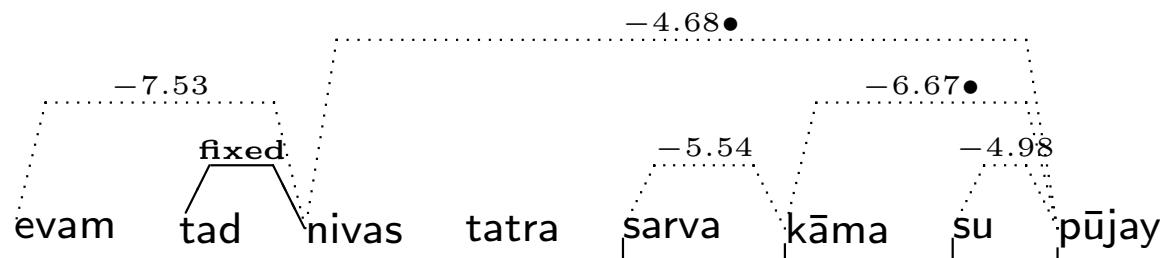
evam sa nyavasat tatra sarvakāmaiḥ supūjitaḥ



śrutvā tatheti rājā ca pratyuvāca purohitam



evam sa nyavasat tatra sarvakāmaiḥ supūjitaḥ



pramāṇabhūta ācāryo darbhapavitrāpanīḥ prāñmukhaḥ śucau avakāśe
upaviṣya mahatā yatnena sūtram praṇayati sma

