Dependency Relations for Sanskrit Parsing and Treebank

Amba Kulkarni\textsuperscript{1} Pavankumar Satuluri\textsuperscript{2} Sanjeev Panchal\textsuperscript{1} Malay Maity\textsuperscript{1} Amruta Malvade\textsuperscript{1}

\textsuperscript{1}Department of Sanskrit Studies
University of Hyderabad

\textsuperscript{2}School of Linguistics & Literary Studies
Chinmaya Vishwavidyapeeth
Ernakulam

19\textsuperscript{th} International Workshop on Treebanks and Linguistic Theories 2020
Introduction

1 Pānini’s Grammar
Introduction

1. Pānini’s Grammar

2. Saṁsādhanī Tagset
Introduction

1 Pānini’s Grammar

2 Saṁsādhanī Tagset

3 Enhanced Tagset
Introduction

1. Pānini’s Grammar
2. Samāsādhanī Tagset
3. Enhanced Tagset
4. Samāsādhanī Parser
Introduction

1. Pāṇini’s Grammar
2. Saṁsādhanī Tagset
3. Enhanced Tagset
4. Saṁsādhanī Parser
5. Sanskrit Treebank
1 Pānini’s Grammar
2 Saṁsādhanī Tagset
3 Enhanced Tagset
4 Saṁsādhanī Parser
5 Sanskrit Treebank
6 Evaluation
Introduction

1. Pānini’s Grammar
2. Saṁsādhanī Tagset
3. Enhanced Tagset
4. Saṁsādhanī Parser
5. Sanskrit Treebank
6. Evaluation
7. Conclusion
Introduction

1. Pānini’s Grammar
2. Saṁsādhanī Tagset
3. Enhanced Tagset
4. Saṁsādhanī Parser
5. Sanskrit Treebank
6. Evaluation
7. Conclusion
Pāṇini’s Grammar

- Around 400 BCE
- Full-fledged grammar of the then prevalent Sanskrit
- Covering both Vedic and Non-Vedic texts
- Only full-fledged, manually written, computational grammar for any natural language
The first tagset for Sanskrit (Ramakrishnamacharyulu 2009)
- Approximately 100 relations
- Inter-sentential relations
- Intra-sentential relations
Examined from computational viewpoint

- Found to be fine-grained
- Fine-grain relations - 100
Kartā

- anubhayi (experiencer)
- amūrtaḥ (abstract)
- prayojakaḥ (causative agent)
- prayojyah (causee)
- abhiprerakaḥ (cause of temptation)
- karmakartṛ
- karaṇakartṛ
- ṣaṣṭīkarta
Skt: Rāmaḥ(Kartā) pacati
Eng: Rama cooks

Skt: Ghaṭaḥ(anubhavī kartā) naśyati
Eng: Pot is experiencing the destruction

Coarse-grain relations - 31
Lacunae in the Samśādhanī Tagset

Relations without any associated semantics
  • Relations due to morphological requirement
  • Complementiser

Inconsistent with Pāṇinian Grammar
  • Coordinating conjuncts
Relations due to morphological requirement

grāmaṃ village  paritah surrounding  vṛkṣāḥ trees  santi are.

**Figure:** Old annotation

grāmaṃ village  paritah surrounding  vṛkṣāḥ trees  santi are.

**Figure:** New annotation
ahaṃ gṛhaṃ gacchāmi iti rāmaḥ avadat.
I home goes that Rama said

Figure: Complementiser: old version

ahaṃ gṛhaṃ gacchāmi iti rāmaḥ avadat.
I home goes that Rama said

Figure: Complementiser: new version
Coordinating Conjuncts

Figure: Conjuncts: Old version

Figure: Conjuncts: New version
The current version has 54 relations and classified into following categories.

- Predicate-argument relations
- Non-Predicate argument relations
  - Verb-verb relations
  - Verb-noun relations
  - Noun-noun relations
- Relations due to special words
- Conjunctions and Disjunctions
- Miscellaneous
1 First full-fledged Parser for Sanskrit.

2 Follows Pāṇinian Grammar and the Theories of Verbal cognition.
   - ākāṅkṣā (expectancy)
   - yogyatā (meaning congruity)
   - sannidhi (proximity)

3 Implemented as an edge-centric binary join to build a dependency tree, in bottom-up approach, with local and global constraints on the edges and the edge labels.
Sanskrit Treebank

1. SHMT (Sanskrit-Hindi Machine Translation System) Consortium - 3000 Sentences.
2. Pañcatantra - 230 sentences. (Dwivedi and Guha, 2017)
3. Treebank of Vedic Sanskrit (Hellwig et al., 2020)
Samṣādhanī Platform

1. Saṅkṣepa Rāmāyaṇam
2. Śrīmad-Bhagavad-Gītā
3. Śiśupālavadham
1. Grammar Books
2. NCERT (National Council for Education, Research and Training) 9th grade - 284 sentences
3. Independent sentences from various Sanskrit books.
4. 130 Short stories
Ambiguities during annotation

Skt: mārgaḥ avaruddhāḥ bhavanti.
Gloss: Road{pl,nom} blocked{pl,nom} be{pres,pl,3p}.
Eng: The roads are blocked.

Figure: Inflectional Information
Ambiguities during annotation

Skt: mitrāṇi kathayanti.
Gloss: friend{pl,nom} / friend{pl,acc} tell{pres,pl,3p} / tell{pres,pl,3p [causative]}
Eng: Friends tell / (They) tell friends / Friends make (somebody) tell / (They) make (somebody) tell friends.
Skt: rāmaḥ pustakaṃ kṛītvā paṭhati.
Gloss: Rama{sg,nom} book{sg,acc} purchase{abs} read{pres,sg,3p}.
Eng: Rama reads a book after purchasing it.
## Table: Performance of Parser

| Source         | Sentences | Exact Match | Failed | Partial Match | LAS  | UAS  |
|----------------|-----------|-------------|--------|---------------|------|------|
| Grammar 9th grade | 468       | 343         | 2(.4%) | 123           | 89%  | 97%  |
| Skt Learner    | 284       | 183         | 15(.6%)| 87            | 82%  | 89%  |
| BhG sample     | 1070      | 817         | 66(6%) | 181           | 88%  | 92%  |
| Skt Learner    | 36        | 7           | 3(8%)  | 26            | 70%  | 76%  |
| Average        | 1858      | 1350        | 86 (4.6%) | 417            | 85.5% | 91.5% |
## Evaluation

| machine→ manual↓ | kartā (agent) | karma (goal) | adjective | pred adj | .. | Total |
|------------------|--------------|--------------|-----------|----------|----|-------|
| kartā (agent)    | **1322**     | 14           | 10        | 26       | .. | 1523  |
| karma (goal)     | 31           | **883**      | 7         | ..       |    | 1069  |
| adjective        | 29           | 12           | **260**   | ..       |    | 406   |
| pred adj.        | 23           | ..           | ..        | **114**  | .. | 162   |
| ..               | ..           | ..           | ..        | ..       | .. | ..    |
| Total            | **1460**     | **952**      | **306**   | **140**  | .. | **6226** |

**Table:** Confusion Matrix
Conclusion

- Improvised version of dependency relations
- Significant improvements
Thank You