

An Indian-Grammatical-Tradition Based Discourse Relation Marking Tool for Hindi

Soma Paul IIIT-Hyderabad, India	Sukhada IIT(BHU), Varanasi, India	Kumari Riya IIT(BHU), Varanasi, India
Adepu Varshith Kumar KITSW(Warangal), India	Sanyam Jha IIT(BHU), Varanasi, India	Isma Anwar IIIT-Hyderabad, India

July 10, 2024

Abstract

This demo paper presents a Hindi discourse relation marking tool along with its interface aligned with the Universal Semantic Representation (USR) system, rooted in Indian Grammatical Tradition. The user-friendly interface allows for easy validation and editing of marked relations for accuracy.

1 Introduction

The demonstrated tool is designed to mark relations between clauses and sentences in Universal Semantic Representation (USR) which is currently being developed as part of one of the Bhashini projects. USR is a semantic representation system that is based on Indian Grammatical Tradition (IGT) and Paninian Grammar[4]. USR aims to represent the meaning of the whole text. Thus, It captures information at three levels: Lexico-Conceptual, Syntactico-Semantic, and Discourse[1].

Discourse information currently being annotated are pronominal co-reference, speaker's view and discourse relations among the clauses and contiguous sentences that establish the coherence of the text. Kulkarni et al. (2012) have postulated discourse relations which have further been extended in our work especially for handling cases where the relation is not marked by a discourse connective but by pronoun+Post position constructions as discussed below.

2 Strategy Adopted in the Discourse Relation Marking Tool

For discourse relation marking, first sentences are simplified because complex sentences can have more than one sentence (or clauses) connected through discourse connectives. Let us take the following small discourse as an example

1. rāṃabābū bahuta bīmāra haiṃ. phira bhī unake beṭe unheṃ ekabāra bhī dekhane ke liye nahīṃ āye. isa kāraṇa ve bahuta dukhī haiṃ.
'Rambabu is very sick. Even then his sons did not come even once to meet him. That is why he is very sad.'
2. rāṃabābū bahuta bīmāra haiṃ. isake bāvajūda bhī unake beṭe unheṃ ekabāra bhī dekhane ke liye nahīṃ āye. isa kāraṇa ve bahuta dukhī haiṃ.
'Rambabu is very sick. Despite that his sons did not come even once to meet him. That is why he is very sad.'

3. rāmabābū bahuta dukhī haiṃ kyom̐ki unake bahuta bīmāra hone ke bāvajūda bhī unake beṭe unheṃ ekabāra bhī dekhane ke liye nahīṃ āye.
 'Rambabu is very sad because despite his illness his sons did not come even once to meet him.'

The text in (1), (2) and (3) convey the same meaning although the syntactic representation is different. (3) has one sentence in which two finite clauses are connected through a discourse connective *kyom̐ki*; while (1) and (2) have three independent sentences. In them, the meaning of *kyom̐ki* is expressed through *isa kāraṇa* which is identical to *AltLex* in PDTB. Similarly, *phira bhī* is a discourse connective construction whose equivalent *isake bāvajūda* is a pronoun+post-position construction and plays the role of discourse connective in (2).

Since USR is a semantic representation, it can be argued that the above three discourses will have one canonical USR because they express the same meaning. But we intend to maintain in the USR the discourse planning information. The uniqueness of USR is that it represents the interplay of *samarthaḥ padavidhiḥ* and *vivakṣā* - the former constrains or licenses “who did what to whom, where, when, why, i.e., events, their participants and relations between them” [3] and the latter belongs to the domain of the speaker’s thought/idea – speaker’s intention[1][5][4].

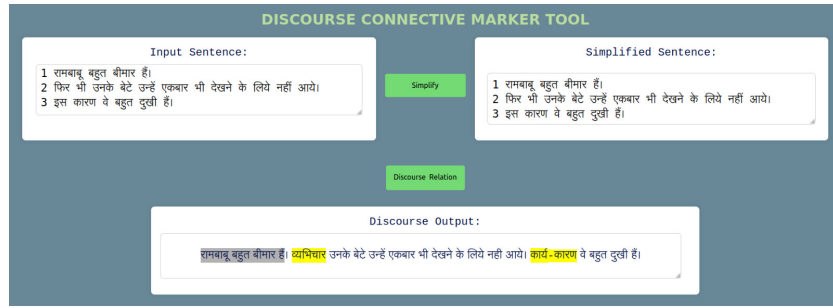


Figure 1: Output of Discourse Relation Marker Tool

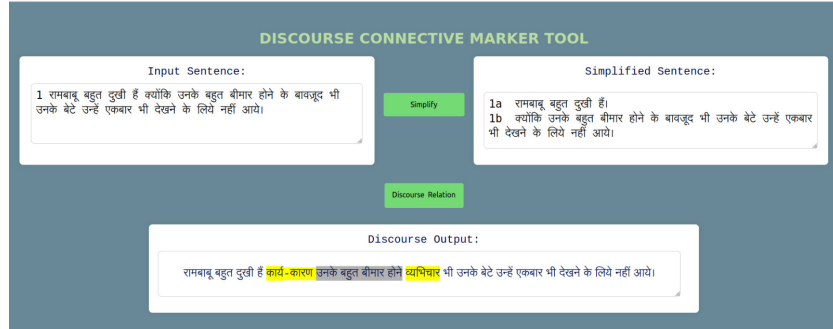


Figure 2: Output of Discourse Relation Marker Tool

3 Interface of Discourse Relation Marking Tool

Figure 1 and 2 are screenshots of the discourse relation marking tool. Conventions followed are:

- Relations are marked yellow.
- The sentences that are not mandatory for continuing the flow of the discourse is marked gray.

The simplified output box as well as the discourse relation box are editable so that human validators can correct the output at this stage.

References

- [1] Kirti Garg, Soma Paul, Sukhada Sukhada, Fatema Bawahir, and Riya Kumari. Evaluation of universal semantic representation (usr). In Proceedings of the Fourth International Workshop on Designing Meaning Representations, pages 13–22, 2023.
- [2] Amba Kulkarni and Monali Das. Discourse analysis of sanskrit texts. In Proceedings of the workshop on advances in discourse analysis and its computational aspects, pages 1–16. Nancy, France. Association for Computational Linguistics, 2012.
- [3] Rashmi Prasad, Bonnie Webber, Alan Lee, and Aravind Joshi. Penn discourse treebank version 3.0. LDC2019T05, 2019.
- [4] Sukhada and Soma Paul. Theory of sāmānthya in Indian Grammatical Tradition: the Foundation of Universal Semantic Representation. International Journal of Sanskrit Research, 2023.
- [5] Sukhada Sukhada, Sirisipalli Veera Hymavathi, and Soma Paul. Generation of MRS Abstract Predicates from Paninian USR. In Stefan Müller and Elodie Winckel, editors, Proceedings of the 30th International Conference on Head-Driven Phrase Structure Grammar, University of Massachusetts Amherst, pages 122–142, Frankfurt/Main, 2023. University Library.