BHASHA WORKSHOP

December 22, 2025

Indian Grammatical Tradition-Inspired Universal Semantic Representation Bank (USR Bank 1.0)

Soma Paul, Sukhada Sukhada, Bidisha Bhattacharjee, Kumari Riya,Sashank Tatavolu,Kamesh R, Isma Anwar, Pratibha Rani
IIIT Hyderabad | IIT (BHU), Varanasi | SIST, Chennai

ABSTRACT

Universal Semantic Representation (USR) is a multi-layered semantic framework that captures not only predicate–argument structure but also speaker's intention and discourse meaning.

Grounded in Universal Semantic Grammar (USG) and inspired by Pāṇinian grammar and the **Indian Grammatical Tradition (IGT)**, USR systematically distributes semantic and pragmatic information across layers, preserving surface linguistic structure while enabling deep meaning representation.

Automatic USRs are generated using a USR-builder tool and validated through a web-based interface **(SAVI)**, ensuring high annotation consistency. Evaluation on Hindi texts demonstrates strong inter-annotator agreement and high semantic fidelity in USR-to-text generation, establishing USR as a robust, language-agnostic framework for multilingual NLP.

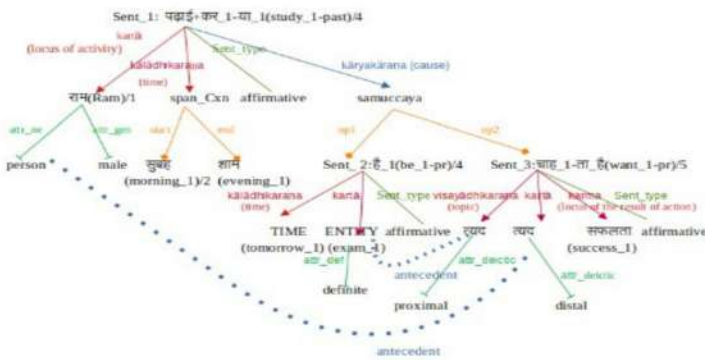## Universal Semantic Representation

USR is:
- A multi-layered, language-agnostic, document-level semantic representation
- Built on Universal Semantic Grammar (USG)
- Grounded in Pāṇinian grammar and the Indian Grammatical Tradition (IGT)
- Designed to capture predicate–argument structure and speaker intention

LAYERS OF USR

USR captures meaning at four levels:
- Lexico-Conceptual Layer
- Intra-Sentential (Syntactico-Semantic) Layer
- Discourse Layer
- Speaker's View (Pragmatic) Layer
✦This layered design enables fine-grained modeling of meaning, discourse relations, and speaker perspective.

Publications

Sukhada, S., Hymavathi, S. V., & Paul, S. (2023).
Generation of MRS Abstract Predicates from Paninian USR.
Proceedings of the 30th International Conference on Head-Driven Phrase Structure Grammar, 122–142.
https://doi.org/10.21248/hpsg.2023.7

Garg, K., Paul, S., Sukhada, S., Bawahir, F., & Kumari, R. (2023).
Evaluation of Universal Semantic Representation (USR).
Proceedings of the Fourth International Workshop on Designing Meaning Representations, ACL.

Sukhada, S. & Paul, S. (2023).
Theory of samarthya in Indian Grammatical Tradition: The foundation of USR.
International Journal of Sanskrit Research.

Data Curation

Domains Covered
- Curated linguistic data
- Health (medical consent forms)
- Education (NCERT & NIOS textbooks)

Statistics
- 7,000+ segments
- 56,000+ simple concepts
- 6,800+ complex concepts

Methods and Approaches

USR Bank Creation Pipeline
1. Sentence Segmentation
   - Complex sentences split into meaningful segments
   - Accuracy: 96.3%
2. Automatic USR Generation
   - Dependency Parser
   - Morphological Analyzer
   - Named Entity Recognition
   - Discourse Marker Tool
   - Rule-based integration (USR-builder)
3. Manual Validation
   - PostgreSQL-backed annotation
   - SAVI web interface with visualizers

Evaluation

Inter-Annotator Agreement
- Dependency: $\kappa \approx 0.85$
- Discourse: $\kappa \approx 0.88$
- ✦Confirms annotation consistency

USR → Text Generation
- Tested on Hindi health & education data
- Human + **Gemini 2.5 Pro** generation
- Semantic similarity: 80–93%
- ✦Demonstrates semantic completeness and reliability of USR

Applications

- Multilingual Natural Language Generation (NLG)
- Machine Translation
- Question-answering System

Conclusion

USR Bank 1.0 integrates classical Indian linguistic theory with modern NLP, offering a multi-layered, cognitively grounded, and scalable semantic representation. Strong evaluation results and multilingual applicability position USR as a powerful interlingua for next-generation NLP systems.