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

Anonymous ACL submission

Abstract

We introduce **USR Bank 1.0**, a multi-layered text-level semantic representation system specifically designed to capture speakers' intention as it is linguistically expressed, thus making the representation unique amongst all the existing ones. Universal Semantic Representation (USR) is rigorously modeled on Universal Semantic Grammar (USG), a foundational framework deeply inspired by Pāṇini and the rich Indian Grammatical Tradition (IGT). This work presents the development of the USR Bank, where initial USRs are automatically generated by a dedicated USR builder tool and then meticulously validated using a web-based validation interface. High inter-annotator agreement in dependency and discourse annotation, along with strong semantic similarity in USR-to-text generation, demonstrate the clarity, robustness, and downstream application of the framework for modern multilingual NLP applications.

1 Introduction

This paper introduces USR Bank 1.0, a new multilingual linguistic resource developed over three years, which formally represents the *speakers' desires for what to express, how much to express, and how to express* (vivakṣā) through a multi-layered semantic representation. Universal Semantic Grammar (USG), rooted in Pāṇinian grammar and the Indian Grammatical Tradition (IGT) (Sukhada and Paul, 2023; Garg et al., 2023), forms the theoretical basis for Universal Semantic Representation (USR).

USR is a text-level, multilayered representation that specifies disambiguated concepts along with their ontological semantic categories and morpho-semantic information, such as plurality, tense-aspect-modality, and causative. More interestingly, it captures the communicative meaning, meaning that the speaker desires to express, through its syntactico-semantic annotation schema of kāraka relations, inter-sentential discourse relations, and semantics of discourse particles rather than only featuring the predicate-argument structure meaning as existing Semantic Representations traditionally represent.

We have successfully demonstrated natural language generation from USR for both Hindi and English, establishing a strong foundation for multilingual generation. Ongoing efforts aim to extend generation capabilities to Tamil, Sanskrit, and other Indian languages. The strategic inclusion of Hindi and Sanskrit (Indo-Aryan), Tamil (Dravidian), and English (Germanic, part of the larger Indo-European family) serves to rigorously evaluate the completeness, universality, and language-agnostic nature of the information captured in the USR Bank 1.0.

In this paper, Section 2 introduces Universal Semantic Grammar (USG) and its theoretical foundation in the IGT. Section 3 provides a concise review of existing Semantic Representations and their theoretical orientations, contextualizing USR's distinct contribution. Section 4 elaborates on the multi-layered design principles of USR, including salient features that underscore its distinct contribution. Section 5 describes a comprehensive methodology employed for developing the USR Bank, detailing our semi-automatic annotation pipeline. Finally, Section 6 reports the inter-annotator agreement (IAA) for dependency and discourse relations and automatic evaluation of USR to text generation by human annotators, offering empirical validation of the representation and annotation scheme's reliability.

2 USG: The Theoretical Framework of USR

The IGT framework conceptualizes language as an inherently holistic phenomenon. Kiparsky (2002) pointed out that Pāṇini's grammar organization is a device that starts from meaning information and incrementally builds up a complete interpreted sentence. In more concrete terms, the derivation of a sentence is initiated by constructing the morphosyntactic analysis, i.e., the arguments of a predicate (or events) are assigned syntactico-semantic roles (kārakas) based on the ontology of the events and the speaker's wish to express certain features of it (vivakṣā). Bhartṛhari (Iyer, 1965) compares language communication to painting: the speaker starts with a unified idea and expresses it part by part, with words interconnected by the principles of semantic compatibility (sāmarthya) to form a coherent whole. While existing semantic representation focuses on predicate-argument structure (who did what, where, etc.) (Abend and Rappaport, 2017), it does not capture the speaker's intention (vivakṣā), which shapes how events are expressed from the perspective of the speaker. For example, in case of a simple event of "a boy's causing a glass to break", the conceptual structure grounded in the principle of semantic compatibility includes an agent ("the boy") and a patient ("the glass") of the event 'breaking'. But, how a speaker chooses to express this event depends on his/her communicative desire (vivakṣā):

- In *"The boy broke the glass"*, the speaker foregrounds the agent, *"the boy"*, who functions as the kartā, the most independent participant of the event *break-0*.
- In contrast, in *"The glass broke"*, the speaker emphasizes on the affected entity (*"the glass"*), thus making it kartā, the most independent participant of *break-1*.

The sub-eventive explanation of Parsons (1990) accounts for this analysis. Both break-0 and break-1 are subevents of the larger event 'break'. Hindi uses two different lexical items for the two events: *toḍa* (break-0) and *ṭūṭa* (break-1).

3 Related Work

Most Semantic Representations (SRs) generally abstract away from the surface-level grammatical and syntactic idiosyncrasies of natural languages, focusing instead on representing their underlying meanings. A detailed account of various SR parameters and a comparative analysis can be found in Boguslavsky (2019). Certain SRs adhere to specific linguistic frameworks, which determine their representational choices and theoretical commitments. For instance, Minimal Recursion Semantics (MRS) (Copestake et al., 2005) is rooted in Head-driven Phrase Structure Grammar (HPSG); the Prague Dependency Treebank (PDT) (Hajic et al., 2006) align with Functional Generative Description (FGD); Framenet (Baker et al., 1998) is based on Frame Semantics and; the Parallel Meaning Bank (PMB) (Abzianidze et al., 2017) is based on Discourse Representation Theory (DRT); and Abstract Meaning Representation (AMR) (Banarescu et al., 2013) adopts a neo-Davidsonian event-semantic representation. Similarly, USR is built upon the USG framework (Sukhada and Paul, 2023), as discussed in Section 2. A comparative overview of these frameworks is presented in table 1, which shows that USRs, unlike other SRs, capture elaborate discourse level information, which we will explain in the following sections.

Name of SR	**Multi-layered**	**Discourse Level Connectives**	**Extra-propositional Semantics**	**Abstract Concepts such as Construction**	**Participant Roles**
Prague Dependency Treebank (PDT)	Yes	No	No	No	Thematic role
Framenet	No	No	No	Yes	Semantic Roles
Uniform Meaning Representation (UMR)	Yes	No	No	Yes	Complements underspecified (ARG1, ARG2, ..., ARGn)
USR	Yes	Yes	Yes	Yes	Pāṇinian kāraka-based Syntactico-semantic relations

Table 1: A comparative study of various Semantic Representations

4 Design of USR

USR is conceptualized as a multi-layered system designed for comprehensive meaning encoding. This system operates at three primary levels: (a) lexico-conceptual- focusing on disambiguated concepts along with their semantic category; (b) intra-sentential - detailing semantic relationships between head and dependents within a single sentence; and (c) discourse - capturing inter-sentential coherence and anaphora (Garg et al., 2023). Additionally, USR incorporates an emerging pragmatic layer to capture linguistically expressed speaker's attitude or specific focus.

4.1 Lexico-Conceptual Layer

Every USR consists of a list of concepts - either Simple or Complex Concepts (CC). Only entities, events, and modifiers, including quantifiers, are concepts. Complex Concepts represent higher-order cognitive schema that structure meaning independently of surface linguistic forms (Langacker, 1987; Evans and Green, 2018). For example, 10 inches (or 10”) is [height_meas] CC in table 2:

(1) The boy who came from Pune is 10 inches taller than my brother.

Concept	**Index**	**Sem_ cat**	**Morpho_ sem**
boy_1	1	anim	
be_1-pres	2		
10	3	numex	
inch	4		
[height_meas_1]	5		
tall_1	6		comparmore
$speaker	7		
brother_1	8		
$yad	9		
come_1-past	10		
Pune	11	place	
[ne_1]	12		

Table 2: Concepts with Semantic Category and Morpho-semantic information

$yad represents the relative pronoun. Every simple concept is assigned a unique identifier (ID) that unambiguously represents that concept. The digit with CC indicates the serial number of that CC in the USR.

This layer also captures ontological categories of the concepts and semantically derived information attested on a concept. For example, the comparative degree of adjectives.

4.2 Intra-Sentential Layer

This layer encodes two kinds of information: (a) dependency relations among heads and dependents; (b) semantic tags for the components of Complex Concepts. Table 3 illustrates the intra-sentential relations for 1.

According to IGT, there are two kinds of dependency relations: (a) kāraka relations, (b) kāraketara (other than kāraka) relations (Kulkarni and Sharma, 2019; Begum et al., 2008). *kāraka* roles include *kartā* (the most independent participant, often agentive), *karma* (the most desired object/patient), instrument, beneficiary, source and temporal-spatial. There are 73 dependency relations in the current USR Guidelines V 4.2.1.

The main clause ‘the boy is 10 inches taller than my brother’ in (1) is a copulative sentence. Unlike most other SRs that treat such predicative adjectives as a functor and the subject as its argument, as in tall (boy) for ‘the boy is tall’, Pāṇini’s grammar treats the copula as the main predicate. That is why be_1-pres is assigned *0:main*. The noun that agrees with the copula is considered *expressed* (abhihita) and occupies the subject position (Bharati and Kulkarni, 2009), which is annotated as *kartā* in USR. The predicative adjective is annotated as *kartāsamānādhikaraṇa*, which implies that the ‘boyhood’ and ‘tallhood’ exist in the same entity.

This layer also specifies the internal structure of Complex Concepts (CCs). For example, the CC [height_meas] has two components: *count* and *unit* as specified in the ‘CxN component’ column. The next layer is the Discourse Layer.

4.3 Discourse Layer

We add sentence 2b with 1 repeated here as 2a.

(2) a. The boy who came from Pune is 10 inches taller than my brother.

b. Besides that, he is very strong.

In the discourse layer, we capture the semantics of discourse connectives. In 2b, the author could have used the connective "and", which would have retained the discourse coherence of 2a and 2b. However, the author has chosen the phrase "besides that" by which the author desires to express conjunction and *something more*. In PDTB 3.0 Annotation Manual (Prasad et al., 2019), the relation attested for ‘besides’ is Expansion.Conjunction, same as for "and", "additionally". Such an annotation schema does not capture the speaker’s communicative desire. We propose capturing the speaker’s intention to convey *something additional* beyond the basic ‘Conjunction’ meaning within the Speaker’s View layer, which will be discussed in detail in the following section.

4.4 Speaker’s View

This layer, currently in its preliminary stage of development, aims to capture extra-propositional information that is overtly expressed in languages,

Concept	Index	Sem_cat	Morpho-Semantic	Dependency	CxN Components
boy_1	1	anim		2:kartā	
be_1-pres	2			0:main	
10	3	numex			5:count
inch	4				5:unit
[height_meas_1]	5			6:rmeas[1]	
tall_1	6		comparmore	2:kartā samānādhikaraṇa	
$speaker	7			8:genitive	
brother_1	8			1:rv[2]	
$yad	9			10:kartā	
come_1-past	10			1:rcdelim[3]	
Pune	11	place			12:begin
[ne_1]	12			10:source	

Table 3: Representation of Inter-sentential Layer along with Lexico-conceptual layer.
[1]rmeas – **r**elation **meas**urement; measurement of event or entity.
[2]rv – **r**elation *vibhājana*; inequalities between two compared entities.
[3]rcdelim – **r**elative **c**lause **delim**itation; when the relative clause delimits the head noun.

rather than implicitly inferred, mainly capturing the meaning of speakers' vivaksā or choice of expression. In the above table, for example, the semantics of 'inclusive' entails that the speaker has already expressed some qualities of the boy in previous sentences. Similarly, the semantics of 'additional' is specified in the speaker's view row of the verb. Thus, information encoded in discourse and speaker's view layer together constitute the meaning of 'besides that'. Some more tags of the speaker's view layer are 'definiteness' (e.g., 'the' vs. 'a'), expressions of respect or formality, informal address, 'exclusive' for only, 'inclusive' for also and so on. The research here focuses on understanding how these nuanced pragmatic meanings are lexicalized and grammaticized across different languages. Initial comparative studies on Hindi and English have revealed systematic and consistent behavioral patterns for many of these pragmatic categories, suggesting a promising potential for universal modelling and application.

<sent_id=2>

Concept	Index	Dependency	Discourse	Speaker's View
$tyad	1	2:kartā		
be_1-pres	2	0:main	1.2:conjunction	additional
very_1	3	4:intf		
strong_1	4	2:kartā samānādhikaraṇa		inclusive

</sent_id>

Table 4: USR representation of 2b including discourse information

5 Developing the USR Bank 1.0

This section describes the stages of the creation of USR Bank 1.0 and presents statistics of USR created so far.

5.1 Tool and Annotation

The development of the USR Bank 1.0 follows a structured, three-phase pipeline to ensure accuracy and efficiency:

5.1.1 Segmentation of Complex Sentences

USR employs a principled segmentation strategy to manage the complexity of natural language, particularly in syntactically dense or information-heavy sentences. Instead of treating the sentence as a whole, USR breaks it down into semantically coherent segments, typically each containing one finite clause. Each segment is assigned a unique ID. Segment IDs accommodate *titles, headings*, and *fragments*, ensuring structural clarity throughout the annotation of a text. Segmentation adheres to consistent rules, such as - splitting at discourse connectives, postulating elided elements, not segmenting relative clauses if the head noun is modified by one relative clause and so on. Evaluated against 500 gold-standard sentences, our segmentor tool achieved a 96.3% success rates. An example of segmented output is available in Table 5 for a sentence taken from NCERT Geography textbook:

<sent_id=Geo_11stnd_13ch_0055>
Tide: The periodic rise and fall of the sea

level, occurring once or twice a day, is called a tide and this movement of water is caused by certain meteorological effects, such as changes occurred in winds and atmospheric pressure; by the gravitational attraction of the sun and the moon.

Sentence ID	Text
Geo_11stnd_13 ch_0055H	Tide
Geo_11stnd_13 ch_0055a	The periodic rise and fall of the sea level, occurring once or twice a day, is called a tide.
Geo_11stnd_13 ch_0055b	This movement of water is caused by certain meteorological effects.
Geo_11stnd_13 ch_0055cF	Changes occurred in winds and atmospheric pressure.
Geo_11stnd_13 ch_0055dF	By the gravitational attraction of the sun and the moon.

Table 5: Segmented Output with appended specific segment ID

5.1.2 Automatic USR Generation (USR-builder)

A USR-builder tool for Hindi has been developed to automatically generate USRs. The following tools have been integrated into the USR-builder to support this automatic construction:

- Simple Concept Identifier
- Complex Concept (CC) Identifier
- Morphological Analyzer
- Named Entity Recognizer
- Dependency Parser + Mapper
- Discourse Relation Marker

The Simple Concept Identifier Tool, CC Identifier Tool and the Discourse Relation Marker tool, which specify the semantics of the explicit discourse connectives, have been developed in-house. The Complex Concept Identifier currently achieves an accuracy of 84.26%, while the Discourse Relation Marker demonstrates an accuracy of 94%.

The USR-builder applies a set of heuristics to process the output of these tools and create layered output. A schematic flowchart illustrating the overall architecture and data flow of the USR-builder is presented in Figure 1 in Appendix A.

5.1.3 Manual Validation via TAT Interface

Once the USRs are automatically created, they are uploaded in the PostgreSQL database (Stonebraker et al., 1990). PostgreSQL is a powerful open-source relational database known for its robust support for complex queries, data integrity, and scalability. This makes it ideal for managing interconnected linguistic data and the semantic layers of USRs.

The database schema is hierarchical, linking Chapter to Sentences, Sentences to Segments, and each Segment forming the base for Lexico-Conceptual, Construction, Relational, and Discourse tables. Manual validation of these USRs is performed by trained annotators using the Text Annotation Tool (TAT), a custom-built, web-based interactive interface. TAT significantly streamlines the validation process by adopting a multi-layered approach for organizing information into separate, intuitive tabs. This allows annotators to efficiently correct tags (e.g., Semantic_category, Morpho-Semantic, Speaker's View) via dropdown menus, validate dependency relations by selecting head indices and relation names, and confirm Complex Concept components (which are color-coded across tables for clarity). Furthermore, TAT features integrated visualizers for dependency trees and discourse graphs, providing immediate visual feedback that greatly aids in accurate validation. Figure 1 in Appendix A presents the output generated by the TAT Visualizer, which displays the layered semantic representation constructed by the USR-builder.

5.2 Data

To develop and test the USR framework and USR-builder tool, two different dataset were used. The first dataset is used to train the tool by showing it how to represent detailed linguistic features in a controlled setting. The second dataset is used to test how well the framework and tool perform when working with real-world texts from specific domains.

5.2.1 First Data: Manually Curated Simple Sentences

The primary corpus for USR Bank 1.0 comprises 659 simple and small sentences. This data was created manually, with the focus on encoding information at various linguistic levels. The primary goal of this dataset is to provide a controlled environment for detailed linguistic annotation. Sentences were carefully crafted to include distinctions such as animacy and gender, named entities at the semantic-category level, and features like plurality and causativization at the morpho-semantic level. Complex concepts (such as noun compounds or different measurements) and different dependency relations (such as- *kartā, karma, karaṇa*) and usages of different T(ense)-A(spect)-M(odality) in natural language have been taken care of.

5.2.2 Second data- Domain-specific text (Health and education)

The second data set has been taken from two different sources from two different domains, namely, health and education. The health domain data is derived from consent forms used for patients and their relatives undergoing specific medical procedures by Christian Medical College, Vellore. The data set for the education domain is sourced from NCERT (National Council of Educational Research and Training) and NIOS (National Institute of Open Learning) geography textbooks in Hindi, ranging from Classes 6 through 12. This dataset offers domain-specific, thematically coherent material, ideal for evaluating the adaptability and depth of the USR framework across real-world contexts. All texts are carefully annotated and manually validated by experts trained in the USR schema.

5.2.3 Annotated Data Statistics

The current statistics for the annotated health data in USR Bank 1.0 are given in Table 6 and the statistics of the top 5 most frequently annotated dependency relations are given in Table 7.

Statistics	First Data	Health Domain	Quantity
Number of sentences	659	168	5727
Number of segments	659	261	7029
Number of Simple concepts	2809	2131	56734
Number of Complex concepts	356	437	6888

Table 6: Statistics of the annotated data in the USR bank

Dependency Relation	Frequency
Modifier (mod)	7579
Genitive relation (r6)	6888
kartā (k1)	6655
karma (k2)	3031
Location (k7p)	2563

Table 7: Statistics of the top 5 most frequent Dependency relations annotated

6 Evaluation

The USR Bank 1.0 is evaluated in this paper using two parameters: (i) ease of annotation and consistency in the annotation schema and (ii) effectiveness of USR for a downstream application, namely natural language text generation. For the former, we have calculated the Inter-Annotator Agreement (IAA) score and reported it in 6.1. For the latter, we conducted an automatic similarity measure between the texts generated by human annotators from USRs and the original text. The closeness of the text generated from the USR with that of the original text will prove the correctness of the meaning representation in USR.

6.1 Evaluation Parameter 1: Inter-Annotator Agreement (IAA)

The Inter-Annotator Agreement (IAA) experiment was conducted based on the annotation guidelines for dependency and USR discourse-layer tagging. Two highly experienced expert annotators participated in this study, both with a background in Indian Grammatical Tradition (IGT) and over one year of dedicated experience in dependency and discourse-level annotation within the USR framework, ensuring a high level of annotation expertise.

6.1.1 Dataset

The experiment was carried out on a carefully selected dataset comprising 70 unique segments. These segments, with an average length of 11 words, were extracted from the NIOS geography textbook corpus and preprocessed using our Segmentor Tool. The USR Builder generated initial USRs for these segments, which were then uploaded to the database, ready for independent validation by human annotators.

6.1.2 Experiment

The two annotators independently annotated the entire 70-segment dataset, without any prior consultation. Upon completion, their annotations were systematically compared to quantify inter-annotator consistency.

6.1.3 Result

Inter-Annotator Consistency was quantitatively measured using both raw agreement percentage and Cohen's Kappa (k). Cohen's Kappa provides a more robust measure of agreement by adjusting for the proportion of agreement that would be expected by chance. For composite annotations (like dependency relations, which involve both a head-dependent pair and a specific label), Cohen's Kappa is calculated by considering each possible combination of head, dependent, and relation label as an annotation unit, allowing for a standard application of the formula. The results, summarized below, demonstrate a remarkably high level of consistency between the annotators for both dependency-level and discourse-level annotations. This strong agreement empirically affirms the clarity, unambiguous nature, and semantic groundedness of the USR guidelines and its tagset.

Features	**Cohen's Kappa**	**Agreement %**
Dependency	0.8465	0.8912
discourse	0.8817	0.9978

Table 8: IAA results using Cohen's Kappa (k) and Agreement Percentage

The high Kappa scores, particularly for discourse relations, indicate that the annotation scheme is well defined and consistently applied across annotators, which is crucial for building a reliable and high-quality semantic resource for machine learning and linguistic analysis.

6.2 Evaluation Parameter 2: USR-to-Text Generation

In addition to the Inter-Annotator Agreement (IAA), an automatic evaluation of the correctness of USR annotation was conducted by comparing the human generated text from the USR with the original text from which these USRs were obtained.

6.2.1 Dataset

For this experiment, we used a manually validated, gold-standard set of USR representations comprising fifty-nine sentences drawn from the health domain. Each USR representation was carefully annotated and cross-validated by expert annotators to ensure consistency and correctness across all annotated layers.

6.2.2 Experiment

Three annotators participated in the USR-to-text generation task, each independently producing texts from the same set of USR. These three annotators are new annotators who have been trained in USR annotation for only one month at the time of the experiment.

6.2.3 Result

A multilingual sentence transformer model (paraphrase-multilingual-MiniLM-L12-v2) has been used to evaluate the quality and consistency of the texts generated by the three annotators. Cosine similarity was employed to measure semantic closeness between the original reference sentences and the annotator-generated alternatives, as well as between the paired outputs of different annotators. The results, summarized below, demonstrate strong agreement between the texts generated by the annotators and the original text, with all three annotators achieving high mean similarity scores above 80%.

The overall mean similarity across all annotators indicates high semantic consistency in the annotated USR. Inter-annotator agreement was also robust, with pairwise similarities consistently above 80%, demonstrating that all three annotators maintained comparable levels of semantic fidelity to the source material while providing linguistically diverse alternatives. These results suggest that the annotation protocol successfully captured the meaning of the original text.

Metric	**A1**	**A2**	**A3**	**Overall Mean**
Cosine-Similarity	0.8866	0.8277	0.8065	0.8403

Table 9: Semantic Similarity by annotators: A1, A2, A3

7 Conclusion

The USR Bank 1.0 advances the field of semantic representation by systematically integrating principles from the Indian Grammatical Tradition. Anchored in the Universal Semantic Grammar (USG) framework, it captures core concepts from IGT—namely, samarthya (semantic compatibility) and vivakṣā (speaker intention)—to offer a multi-layered, coherent, and cognitively grounded model of textual meaning representation. Evaluations through inter-annotator agreement and USR-to-text generation demonstrated the framework's reliability and semantic consistency. Its successful application in Hindi and ongoing efforts to extend it to Tamil, Sanskrit, and English demonstrate its potential for cross-linguistic and multilingual generation. This work bridges classical linguistic theory with modern language technology, offering a scalable, language-agnostic semantic model. Future efforts will focus on expanding the treebank across more languages and refining automatic USR construction tools to enhance multilingual NLP capabilities.

Limitations

The annotators require a good amount of training in Universal Semantic Grammar before starting the annotation. Retaining good annotators is an expensive affair.

References

Omri Abend and Ari Rappaport. 2017. The universal anaphora annotation framework. *Transactions of the Association for Computational Linguistics*, 5:561–577.

Lasha Abzianidze, Johannes Bjerva, Kilian Evang, Hessel Haagsma, Rik Van Noord, Pierre Ludmann, Duc-Duy Nguyen, and Johan Bos. 2017. The parallel meaning bank: Towards a multilingual corpus of translations annotated with compositional meaning representations. *arXiv preprint arXiv:1702.03964*.

Collin F Baker, Charles J Fillmore, and John B Lowe. 1998. The berkeley framenet project. In *COLING 1998 Volume 1: The 17th International Conference on Computational Linguistics*.

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract meaning representation for sembanking. In *Proceedings of the 7th linguistic annotation workshop and interoperability with discourse*, pages 178–186.

Rafiya Begum, Samar Husain, Arun Dhwaj, Dipti Misra Sharma, Lakshmi Bai, and Rajeev Sangal. 2008. Dependency annotation scheme for indian languages. In *Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-II*.

Akshar Bharati and Amba Kulkarni. 2009. ‘subject’ in english is abhihita. Presented at the 14th World Sanskrit Conference, September 1–5, 2009.

Igor Boguslavsky. 2019. Interlingual lexical ontology: Design, construction, applications. *Cham: Springer*.

Ann Copestake, Dan Flickinger, Carl Pollard, and Ivan A Sag. 2005. Minimal recursion semantics: An introduction. *Research on language and computation*, 3:281–332.

Vyvyan Evans and Melanie Green. 2018. *Cognitive linguistics: An introduction*. Routledge.

Kirti Garg, Soma Paul, Sukhada Sukhada, Fatema Bawahir, and Riya Kumari. 2023. Evaluation of Universal Semantic Representation (USR). In *Proceedings of the Fourth International Workshop on Designing Meaning Representations*, pages 13–22, Nancy, France. Association for Computational Linguistics.

Jan Hajic, Jarmila Panevová, Eva Hajicová, Petr Sgall, Petr Pajas, Jan Štepánek, Jiří Havelka, Marie Mikulová, Zdeněk Žabokrtskỳ, Magda Ševcíková-Razímová, and 1 others. 2006. Prague dependency treebank 2.0. *CD-ROM, linguistic data consortium, LDC Catalog No.: LDC2006T01, Philadelphia*, 98.

K. A. Subramania Iyer. 1965. *The Vākyapadīya of Bhartṛhari (Kāṇḍa II)*. Motilal Banarsidass.

Paul Kiparsky. 2002. On the architecture of pāṇini’s grammar. Three lectures delivered at the Hyderabad Conference on the Architecture of Grammar, Jan. 2002, and at UCLA, March 2002.

Amba Kulkarni and Dipti Misra Sharma. 2019. Pāṇinian syntactico-semantic relation labels. In *Proceedings of the Fifth International Conference on Dependency Linguistics (Depling, SyntaxFest 2019)*, pages 198–208.

Ronald W Langacker. 1987. *Foundations of cognitive grammar: Volume I: Theoretical prerequisites*, volume 1. Stanford university press.

Terence Parsons. 1990. *Events in the Semantics of English: A Study in Subatomic Semantics*. MIT Press.

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2019. The penn discourse treebank 3.0. In *Proceedings of the 13th Linguistic Annotation Workshop*, pages 1–10, Florence, Italy. Association for Computational Linguistics.

Michael Stonebraker, Lawrence A. Rowe, Michael Ling, and Jeffery W. Du Bois. 1990. The design of postgres. *ACM Transactions on Database Systems (TODS)*, 15(2):174–188.

Sukhada Sukhada and Soma Paul. 2023. Theory of sāmarthya in Indian Grammatical Tradition: The foundation of Universal Semantic Representation. In *Int J Sanskrit Res*.

A Architecture of the USR Builder

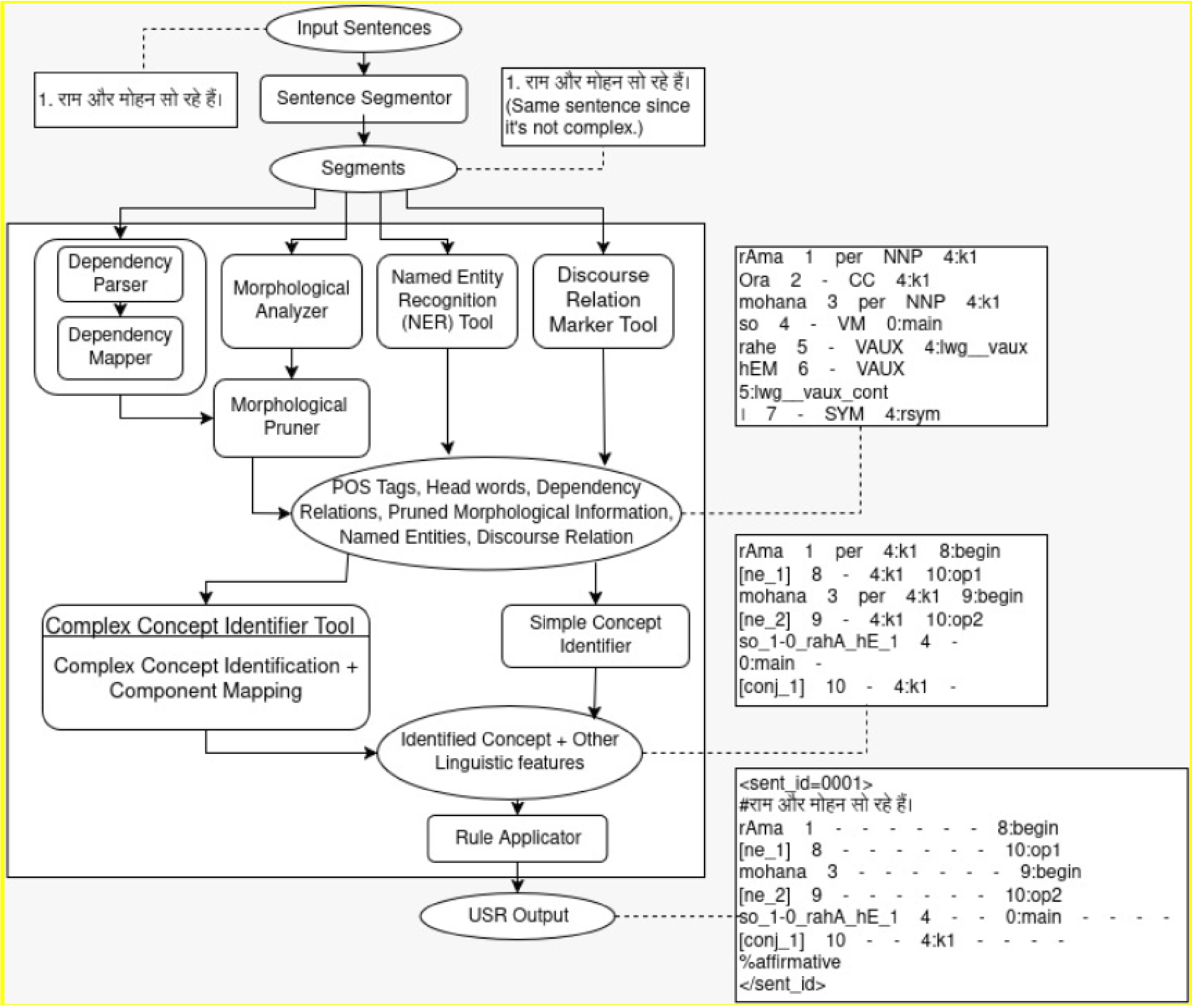

Figure 1: USR Builder