

# Evaluation of Discourse Relation Annotation in the Hindi Discourse Relation Bank

Sudheer Kolachina\*, Rashmi Prasad†, Dipti Misra Sharma\*, Aravind Joshi‡

\* LTRC, IIT-Hyderabad  
sudheer.kpg08@research.iiit.ac.in, dipti@iiit.ac.in

† HIA, University of Wisconsin-Milwaukee  
prasadr@uwm.edu

‡ IRCS, University of Pennsylvania  
joshi@seas.upenn.edu

## Abstract

We describe our experiments on evaluating recently proposed modifications to the discourse relation annotation scheme of the Penn Discourse Treebank (PDTB), in the context of annotating discourse relations in Hindi Discourse Relation Bank (HDRB). While the proposed modifications were driven by the desire to introduce greater conceptual clarity in the PDTB scheme and to facilitate better annotation quality, our findings indicate that overall, some of the changes render the annotation task much more difficult for the annotators, as also reflected in lower inter-annotator agreement for the relevant sub-tasks. Our study emphasizes the importance of best practices in annotation task design and guidelines, given that a major goal of an annotation effort should be to achieve maximally high agreement between annotators. Based on our study, we suggest modifications to the current version of the HDRB, to be incorporated in our future annotation work.

**Keywords:** discourse relation annotation, discourse structure, annotation evaluation

## 1. Introduction

Following the development and release of the Penn Discourse Treebank (PDTB) (Prasad et al., 2008), a large-scale annotated corpus of discourse relations, interest in cross-linguistic studies of discourse relations has led to the initiation of similar PDTB-based annotation projects in a number of languages, such as Arabic (Al-Saif and Markert, 2010), Chinese (Zhou and Xue, 2012), Czech (Mladová et al., 2008), Hindi (Oza et al., 2009b; Oza et al., 2009a), Italian (Tonelli et al., 2010), Tamil (Rachakonda and Sharma, 2011), and Turkish (Zeyrek and Webber, 2008). The Hindi discourse relation bank (HDRB) project is aimed at the creation of a large scale corpus of Hindi texts annotated with discourse relations (e.g., causal, contrastive, elaboration and temporal relations). The source corpus consists of 400K words, comprising multi-domain news articles in Hindi, taken from a leading daily newspaper, AMAR UJALA. There are three other levels of linguistic annotation currently being carried out on the same corpus: syntactic dependency relations (Begum et al., 2008), lexical predicate argument structure, and phrase structure, which, together with the HDRB, form a multi-level and multi-representational linguistically annotated resource (Xia et al., 2008; Bhatt et al., 2009; Palmer et al., 2009).

In adapting the PDTB annotation scheme and guidelines to annotate discourse relations in Hindi texts, our previous work (Oza et al., 2009b; Oza et al., 2009a) proposed extensions and modifications to the scheme, motivated by the need to introduce more conceptual clarity and semantic uniformity in the PDTB sense classification and to increase

annotation quality and specificity. In this paper, we investigate the impact of these modifications on the annotation quality, based on the results of an inter-annotator agreement study. Our results indicate that some aspects of the scheme lead to greater difficulty in annotation, as reflected in the lower inter-annotator agreement. Our study emphasizes the importance of best practices in annotation task design and guidelines, given that a major goal of an annotation effort should be to achieve maximally high agreement between annotators. In Section 2., we present an overview of the HDRB annotation scheme as proposed in Oza et al. (2009b; 2009a) and discuss the points of departure from the PDTB scheme (Prasad et al., 2008). In section 3., we present our experiments on applying the HDRB scheme on a subset of the corpus and the inter-annotator agreement evaluation. We discuss the evaluation in terms of its impact on designing the annotation task and guidelines, and accordingly propose revisions to the HDRB scheme in our concluding discussion in Section 4.

## 2. Current Guidelines and Task Design

Discourse relation annotation for Hindi follows the underlying principles of the PDTB framework, namely, that the annotation approach should be theory-neutral and lexically-grounded. Theory-neutrality implies that the annotation is unspecified for high-level discourse structure, motivated by a lack of agreement on the representational nature of such structure. By lexically-grounding the annotations of discourse relations on lexical items, for both explicit as well as implicit relations, the framework aims to effectively elicit

annotator judgements for this difficult task, thereby boosting annotation reliability.

As in the PDTB, the HDRB project annotates both explicit and implicit discourse relations, their abstract object (AO) arguments, called Arg1 and Arg2, and the senses of discourse relations.<sup>1</sup> Here, we first provide an overview of the HDRB annotation scheme and guidelines, and then discuss the main differences of the scheme with that of the PDTB. The complete annotation scheme and guidelines are described in the HDRB manual.<sup>2</sup> (In all examples in the paper, the relation is underlined, Arg1 is enclosed in square brackets, and Arg2 is enclosed in curly braces.) The annotation is done over the raw text, without reference to other layers of annotation, such as syntax, and the annotation representation format is fully stand-off. We used the PDTB annotation tool, which provides support for multiple languages, including Hindi.<sup>3</sup>

## 2.1. Types of Discourse Relations

Discourse relations between two AOs can be realized in one of three ways:

**Explicit connectives.** These are ‘closed-class’ expressions drawn from four well-defined grammatical classes: subordinating conjunctions (Ex. 1), coordinating conjunctions (Ex. 2), sentential relatives (Ex. 3), discourse adverbials (Ex. 4). The task of identifying explicit connectives involves functional disambiguation, since words and phrases that function as discourse connectives can have other non-discourse functions as well.<sup>4</sup>

- (1) [आज दीया जलाया गया है] क्योंकि {मेरी वर्षगांठ है} [Today the lamp has been lit] because {it is my birthday}.
- (2) [संघ के संगठन अनेक हैं] किंतु {विचारधारा एक ही है} [There are many groups in the Sangh] but {there is just one ideology}.
- (3) [सारा काम छोड़कर वह उस चिड़िया को उठाकर दवा घर की ओर भागा] जिससे {उसका सही इलाज किया जा सके} [Dropping all his work, he picked up the bird and ran towards the dispensary] so that {it could be given proper treatment}.
- (4) [दानवी लहरों के कारण अंडमान के पश्चिमी तट पर तटीय वनस्पति पूरी तरह बर्बाद हो गयी] इसके अलावा {मूंगे की चट्टानों को भी नुकसान हुआ है} [The coastal vegetation on the west coast of the

<sup>1</sup>The distinct relation of attribution, annotated in the PDTB for discourse relations and each of their two arguments (Prasad et al., 2007), is currently not in the scope of HDRB, although it is planned for future work.

<sup>2</sup><http://researchweb.iiit.ac.in/~sudheer.kpg08/hdrb-guidelines-0.1.pdf>.

<sup>3</sup>The tool is available from: <http://www.seas.upenn.edu/~pdtb/PDTBAPI>.

<sup>4</sup>The class of subordinators and particles are also included within the scope of connectives. However, because they are relatively much harder to identify and distinguish from their non-discourse roles, their annotation has been postponed to a later stage of the project.

Andaman has been completely destroyed due to wild waves]. In addition, {the coral reefs have also been damaged}.

**Implicit connectives.** In many cases, although a discourse relation can be inferred between abstract objects, no explicit connective occurs to express the relation. In such cases, the annotator must “insert a connective” that best expresses the inferred relation. Thus, in Example (5), there is a clear consequence relation inferred between the highlighted AOs although no explicit connective is present to express this relation, and the implicit connective inserted by the annotator is इसलिये. Implicit discourse relations are taken to be triggered in adjacent sentence contexts, that is, they are annotated between every adjacent pair of sentences (otherwise unrelated by an explicit connective).<sup>5</sup>

- (5) {इस गेम के सारे खिलाड़ी सचिन तेन्दुल्कर से भी महान हैं} Implicit=इसलिये [इनको क्लीन बॉल्ड करना किसी के बस की बात नहीं] [All players in this game are greater than even Sachin Tendulkar] Implicit=therefore [it is not possible for anyone to get them clean bowled.]

**Alternative Lexicalization (AltLex).** If a discourse relation is inferred but insertion of an implicit connective leads to a perception of “redundancy” in the expression of the relation, it suggests that the second sentence of the pair contains an alternatively lexicalized non-connective expression. The relation is tagged as “AltLex” (Alternative Lexicalization), and the AltLex expression is identified and marked. An AltLex expression is any expression that doesn’t belong to any of the grammatical classes identified for explicit connectives. Example (6) illustrates an AltLex annotation.

- (6) {बांग्लादेश में कानून-व्यवस्था की हालत में सुधार हुआ है} AltLex [इसी वजह से भारत ने सम्मेलन में शामिल होने का फैसला किया है] [Bangladesh’s judiciary has seen an improvement]. That is why [India has decided to participate in the conference.]

**When adjacent sentences are not related by a discourse relation.** If no discourse relation is inferred between two adjacent sentences, an attempt is made to identify an entity-based relation across the two sentences. In an entity-based relation, the only relational inference made by the reader is that the second sentence identifies one or more entities from the previous sentence, and provides a further description about this entity (or entities). Such a relation is tagged as “EntRel” (Entity Relation). Example (7) illustrates an EntRel annotation, where the only purpose of the second sentence is to provide the reader with some additional information about “Jha’s second film”. No discourse relation, such as a causal or contrastive relation, is inferred between the sentences. If neither a discourse relation nor an EntRel is found, a ‘NoRel’ (no relation) tag is assigned.

<sup>5</sup>In contrast to the current version of the PDTB, the HDRB also annotates implicit relations between adjacent sentences separated by a paragraph boundary.

- (7) [फ़िल्म महोत्सव में प्रकाश झा की नई फ़िल्म अपहरण का भी प्रीमियर होना है ] EntRel {गंगाजल के बाद झा की यह किसी अलग विषय पर बनी दूसरी फ़िल्म है }  
 ‘[Prakash Jha’s latest film Apaharan will be premiered at the film festival.] {This is Jha’s second film on a different subject after Gangajal.}’

## 2.2. Arguments of Relations

The two arguments of a relation are labeled Arg1 and Arg2, and the label assignment is semantically driven, in that it is based on the *sense* of the relation to which the arguments belong. Thus, each sense definition for a relation specifies the sense-specific semantic role of each of its arguments, and stipulates one of the two roles to be Arg1, and the other, Arg2. For example, the ‘cause’ sense definition, which involves a causal relation between two eventualities, specifies that one of its arguments is the cause, while the other is the effect, and further stipulates that the cause will be assigned the label Arg2, while the effect will be assigned the label Arg1. The effect of this convention can be seen with Examples (8) and (9). The connectives in both the sentences have the ‘cause’ sense, but while the cause appears after the effect in Example (8), it appears before the effect in Example (9). With the sense-specific labelling convention for ‘cause’, these argument label orderings are Arg1-Arg2 (cause after effect) for Example (8), and Arg2-Arg1 (cause before effect) for Example (9).

- (8) प्रतियोगिता के बाद सोनल ने बताया कि [विजेता के रूप में जब उसका नाम पुकारा गया तो उसे कुछ देर तक खुद पर विश्वास नहीं हुआ], क्योंकि {वह मान कर चल रही थी कि यह प्रतियोगिता फ़िक्स है }  
 ‘After the competition, Sonal said that [when her name was announced as the winner, she could not believe herself for some time], because {she was thinking that the competition was fixed}.’
- (9) फ़ैशन डिजाइनरों का कहना है कि सबसे ज्यादा नकल या चोरी मोनोपोली डिजाइन की होती है। {डिजाइनर इन बातों को बखूबी जानते हैं} इसलिये [कई बार ध्यान नहीं देते हैं ]  
 ‘Fashion designers say that the most prevalent thefts or copies are of monopoly designs. {Designers know this fact very well} so [it does not matter to them many times].’

Arguments of explicit connectives can appear anywhere in the text, which means that they can be non-adjacent in the text. On the other hand, arguments of all the other relation types must be adjacent. Annotation of spans for the arguments of discourse relations (i.e., those expressed by explicit connectives, implicit connectives, and AltLex) follows the ‘minimality principle’ which, while adding no maximal constraint on the extent of the argument, stipulates that only as much is selected as the argument text span as is minimally necessary to interpret the relation. Any additional text deemed as relevant to the arguments’ interpretation is annotated as supplementary material: Sup1 for material supplementary to Arg1, and Sup2 for material supplementary to Arg2. Arguments of EntRel must contain complete sentences but can include multiple sentences, while

### **Class: Comparison**

Contrast  
 Concession  
 Similarity  
 Pragmatic Contrast  
*Epistemic, Speech-Act, Propositional*  
 Pragmatic Concession  
*Epistemic, Speech-Act, Propositional*

### **Class: Contingency**

Cause  
 Goal  
 Condition  
 Pragmatic Cause  
*Epistemic, Speech-Act, Propositional*  
 Pragmatic Condition  
*Epistemic, Speech-Act, Propositional*

### **Class: Temporal**

Synchronous  
 Asynchronous

### **Class: Expansion**

Conjunction  
 Instantiation  
 Exception  
 Restatement  
*Specification, Generalization, Equivalence*  
 Alternative  
*Conjunctive, Disjunctive, Chosen Alternative*

Figure 1: HDRB Sense Classification

arguments of NoRel must contain all and only the two adjacent sentences.

## 2.3. Senses of Discourse Relations

Each discourse relation is assigned a sense label which characterizes the semantics of the relation. The sense classification is hierarchical with three levels, where the lower level labels specify meaning refinements of the higher level labels. Thus, the notion of inheritance is inherent in the sense scheme. Since multiple senses are possible for a relation, annotators were allowed to enter upto two senses for a relation. The complete HDRB sense classification is shown in Figure 1. The top “class” level is shown in bold, the second “type” level is in normal font, while the third “sub-type” level appears in italics. As can be seen in the figure, while all sense classes have second-level sense types, some sense types are not further specified at the subtype level. This does not reflect a strict belief that no further meaning refinements can be described for these sense types. Rather, it was felt that further refined distinctions in the meaning of the relations would be hard to annotate.

In all cases, the annotators were encouraged to provide the most specific sense available from the classification, but were allowed to back-off to the type-level sense when they were not able to distinguish the meaning at the subtype-level. Back-off to class-level sense was disallowed.

## 2.4. Annotation Task Workflow

Annotation of all the different relation types was carried out simultaneously during a sequential reading of the text. Thus, for each new sentence encountered during the reading, annotators identified the sentence-internal relations as well as the relation of the sentence to its prior context. No particular order was imposed on which type of relation to annotate first (e.g., explicit before implicit, or intra-sentential before inter-sentential). For identifying explicit connectives, annotators were provided with an initial list as a guide for detecting connectives, but the basic instruction was to “discover” explicit connectives. This was primarily due to lack of resources with comprehensive cataloguing of discourse connectives for the language. For each relation (of any type), annotators also marked at the same time its arguments and its sense.

## 2.5. Differences with PDTB

The current version of the HDRB scheme differs in some significant ways from the PDTB scheme (Oza et al., 2009a), as described below:

**Changes in task workflow.** In contrast to the HDRB, where the different relation types were annotated simultaneously, the PDTB task workflow consisted of multiple independent subtasks. First, explicit connectives were collected in a list from various resources and provided to annotators, who were instructed to annotate one connective at a time over the entire corpus before moving on to the next connective. Annotation of implicit connectives was done after the annotation of explicit connectives, and here, annotators read the text sequentially while annotating implicit relations between adjacent sentences where needed. Because the PDTB scheme was to some extent developed incrementally, the categories of EntRel and AltLex were devised after the annotation of implicit connectives, when the annotation revealed that annotators were not able to insert a connective in many cases. A subsequent annotation phase looked specifically at such cases, and labeled them as one of EntRel, AltLex, or NoRel. In all cases, arguments and senses were annotated at the same time as the relation type.

The simultaneous annotation of all the relation types in HDRB had a dual motivation. The first was to abstract away from the incremental development of the PDTB scheme and to design the workflow based on the final scheme. The second was the hypothesis that simultaneous annotation was necessary to ensure that the annotators were taking the full context of the discourse into account during the annotation. This was especially so for explicit connectives which need not occur in every sentence in the text, and there was no clear way to ensure that annotators would take the complete prior context into account when annotating any particular connective.

**Modification to the sense classification and semantic labeling convention for arguments.** In PDTB, the assignment of the Arg1 and Arg2 labels to a discourse relation’s arguments is syntactically driven, in that the Arg2 label is assigned to the argument with which the connective is syntactically associated, while the Arg1 label is assigned to the

‘other’ argument. In the HDRB, on the other hand, a semantic criterion is used for argument naming, as described in Section 2.2. This modification was based on the observation that many of the subtype-level senses of the PDTB did not in fact specify further refinements of meaning, but rather reflected differences in the order of the arguments. The modification in HDRB aimed to specify only meaning distinctions in the sense classification, and accordingly, eliminated the argument ordering labels from the PDTB sense hierarchy. All levels in the HDRB sense hierarchy thus have the purpose of specifying the semantics of the relation to different degrees of granularity. The relative ordering of the arguments is instead specified in the definition of the type-level senses, which then forms the basis for the argument naming convention, as described for Examples (8) and (9).

**Restricted back-offs:** In the PDTB, annotators were allowed to back off to higher levels in the hierarchy when they found it difficult to identify the more refined senses at the lower levels. Thus, for example, they could select “Comparison” at the class level instead of “Comparison.Contrast” or “Comparison.Concession” if they were unable to disambiguate between “Contrast” and “Concession”. In HDRB, however, such back-offs are allowed only up to the type level. This constraint was enforced based on the belief that class-level senses are too coarse-grained to be useful.

## 3. Annotation and Evaluation

In this section, we discuss an inter-annotator agreement study conducted to evaluate the quality of the annotation and also, the performance of annotators on different aspects of the annotation task. The study was conducted using a sample set of 11 texts (5K words). These texts were annotated by two annotators following the same set of guidelines.

In our annotation experiment, we follow the method described by Miltsakaki et al. (2004) to estimate inter-annotator agreement.

For agreement on the spans of arguments of connectives, Miltsakaki et al. (2004) use two diagnostics, both of which are based on the exact match criterion<sup>6</sup> which is a very conservative measure of inter-annotator agreement.

*Diagnostic 1: Arg1 and Arg2 are treated as independent tokens.* The total number of tokens is therefore, twice the number of connective tokens. For any Arg1 or Arg2 token, agreement was recorded as 1 when both annotators made identical textual selections for the annotation and 0 otherwise (exact match).

*Diagnostic 2: Arg1 and Arg2 are treated together as part of the connective token.* The total number of tokens is the number of connective tokens. For any connective token, agreement is 1 only when both annotators made identical textual selections for the annotation of both arguments and 0 otherwise. This is a more conservative measure of the agreement.

---

<sup>6</sup>binary-valued agreement: either 1 or 0

|                        | Explicit | Implicit |
|------------------------|----------|----------|
| # connectives by Ann1  | 73       | 119      |
| # connectives by Ann2  | 113      | 79       |
| # matching connectives | 54       | 57       |
| Diagnostic 1           | 0.47     | 0.49     |
| Diagnostic 2           | 0.31     | 0.24     |
| Sense agreement        | 0.96     | 0.91     |

Table 1: Inter-annotator agreement statistics for explicit and implicit connectives in HDRB

Inter-annotator agreement was also estimated with respect to the sense type of the discourse relation inferred. In our experiment, we use the four class-level labels (comparison, contingency, temporal and expansion) as the basic types to estimate agreement in annotation of the sense of implicit connectives. Agreement in sense type is estimated as the Dice coefficient of the two sets of senses of implicit connectives annotated by the two annotators.

The inter-annotator agreement statistics estimated using the method described above are shown in Table 1. The first three rows in the table contain information about the number of connective tokens annotated by each of the annotators (Ann1 and Ann2) and the number of matching tokens for explicit as well as implicit connectives. Inter-annotator agreement estimates based on the above diagnostics are shown in the next three rows.

One consequence of annotating discourse connectives over continuous text seems to be that the number of connectives annotated is not the same across annotators (cf. first two rows in Table 1). This was not the case in PDTB (Mitsakaki et al., 2004), where annotation was carried out connective-wise. This suggests that the notion of a connective itself was quite challenging for the annotators, probably due to the substantial annotation load in simultaneously annotating different kinds of connectives over continuous text. In particular, the distinction between connectives and AltLex expressions in Hindi seems to be a difficult one. In the next section, we discuss a strategy to break down the annotation task into multiple phases in order to handle this issue.

Inter-annotator agreement for annotation of argument spans was low regardless of connective type. Furthermore, agreement estimated using diagnostic 2 was much lower than that of diagnostic 1 for both explicit and implicit connectives. Such low values suggest two possible issues:

1. While annotating the spans of arguments of connectives, the minimality principle was not followed strictly. And since the diagnostics used to estimate agreement were based on the exact match criterion, the values are particularly low. If this is true, it also means that training annotators based on the minimality principle alone can bring immediate improvements in annotation quality.
2. The second and the more serious possibility is that annotators seem to prefer the syntactic (linear-order) convention for argument naming (followed in PDTB)

|               | Explicit | Implicit |
|---------------|----------|----------|
| matching Arg1 | 0.38     | 0.37     |
| matching Arg2 | 0.56     | 0.61     |

Table 2: Inter-annotator agreement: Arg1 versus Arg2

rather than the semantic one proposed for HDRB. This is evident from the relatively higher agreement values when matching Arg2 alone as compared to Arg1 alone, shown in Table 2. The main reason for this preference could be due to the relatively lower complexity of a syntactic naming convention as opposed to a semantic one. While the semantic convention requires that the annotator understand the semantics of the discourse relation to be able to assign argument names, the syntactic naming convention is invariant (Arg2 always the one in which the connective occurs and Arg1 is the ‘other’ argument), and therefore, relatively simpler to adhere to. This then suggests that we may need to reconsider the semantic naming convention proposed in earlier work.

The most encouraging outcome of our evaluation study is the high degree of inter-annotator agreement achieved with respect to the sense type of the connective both for explicit and implicit connectives (cf. last row in Table 1). Such high values confirm the efficacy of the hierarchical sense classification scheme for annotating senses of discourse connectives. However, it must be noted that annotators were allowed to back-off to the coarser type level sense labels (cf. section 2.5.) since argument naming was assumed to be based on a semantic naming convention. However, in case of a rollback to a syntactic convention for argument naming (which seems necessary), annotators will have to assign sub-type level sense labels to the connectives. Another oft-noticed phenomenon in Hindi discourse is implicit connectives between non-adjacent sentences, an example of which is as follows.

- (10) [वरिष्ठ माकपा नेता नीलोत्पल बसु के मुताबिक आम बजट में न्यूनतम साझा कार्यक्रम की झलक साफ नजर आती है।] वामदल लगातार सरकार पर शिक्षा, स्वास्थ्य व रोजगार क्षेत्रों में बजट प्रावधान बढ़ाने के लिए दबाव बनाए हुए थे। बजट में इन सभी क्षेत्रों पर ध्यान दिया गया है। ..... Implicit=दूसरी ओर {आरएसपी के अवनी राय ने कहा कि बजट में छोटे किसानों के लिए कुछ नहीं किया गया है।} लघु उद्योग के लिए नई योजना नहीं है और कस्टम ड्यूटी में कटौती का सीधा असर घरेलू उद्योग पर पड़ेगा। .....
- ‘[According to senior CPI(M) leader Mr.Nilotpal Basu, the general budget clearly reflects aspects of the Common Minimum Programme.] The left parties had been continuously exerting pressure on the government to increase allocation to education, health and employment sectors. And the budget has focused on all these areas..... Implicit=On the other hand, { RSP’s Mr.Avani Rai said that the budget has nothing to offer to small-scale farmers.} There is no planning for small-scale

industry and the drop in custom duty will have a direct adverse impact on cottage industries.....’

In the above example, a discourse relation of contrast type (juxtaposition) can be inferred between the highlighted text spans. An implicit discourse connective needs to be inserted between these non-adjacent sentences. Such instances seem to be quite common in Hindi discourse. This could possibly be due to the high incidence in Hindi of connective classes that take distant arguments (such as discourse adverbials). Annotation of implicit connectives was done only between adjacent sentences in PDTB and the same practice was adopted in HDRB. However, based on our observations in the annotation experiment, we plan to revise this earlier decision and allow for annotation of implicit connectives between non-adjacent sentences.

#### 4. Discussion and Conclusion

In an attempt to introduce more uniformity in the PDTB sense classification and task workflow, the HDRB annotation scheme of Oza et al. (2009a) proposed some modifications to the PDTB scheme. In this paper, we gave an overview of the HDRB scheme, discussing the modifications to the PDTB, and described our application of the scheme to a subset of the HDRB source corpus. Our results have shown that annotation quality is negatively affected by some of the changes. Moving forward for future annotation in the HDRB, we propose the following revisions to the current version of the HDRB scheme.

First, in order to annotate connectives over continuous text and yet keep the annotation load on annotators manageable, the annotation workflow should be split into two steps. This two-step annotation strategy is currently being tried out.

1. Annotate all inter-sentential relations, which include explicit and implicit connectives and AltLex expressions as well.
2. Annotate all intra-sentential relations, which mostly include explicit connective classes such as subordinators, sentential relatives and possibly AltLex expressions.

Second, annotators must be trained specifically to follow the Minimality principle when selecting the spans of a connective’s arguments. This is expected to boost the agreement on argument span selection.

Third, a rollback to the older syntactic argument naming convention followed in PDTB will be implemented. This is based on our observations about the annotators’ preference to adhere to a linear order-based convention rather than a semantic one. Relatedly, if the argument naming convention is syntactic, annotation of senses of connectives must be done at the fine-grained subtype level.

Finally, implicit connectives between non-adjacent sentences must be annotated and guidelines for their annotation developed.

A revision of the HDRB guidelines based on these conclusions is currently underway. We expect that the resultant new version will help smoothen the process of training annotators and also increase the annotation quality.

#### Acknowledgements

This work was partially supported by NSF grant IIS-07-05671.

#### 5. References

- A. Al-Saif and K. Markert. 2010. The Leeds Arabic Discourse Treebank: Annotating Discourse Connectives for Arabic. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC-2010)*, Valletta, Malta.
- R. Begum, S. Husain, A. Dhvaj, D.M. Sharma, L. Bai, and R. Sangal. 2008. Dependency Annotation Scheme for Indian Languages. *Proceedings of IJCNLP-2008*.
- R. Bhatt, B. Narasimhan, M. Palmer, O. Rambow, D.M. Sharma, and F. Xia. 2009. A Multi-Representational and Multi-Layered Treebank for Hindi/Urdu. In *Proceedings of the Third Linguistic Annotation Workshop*, pages 186–189. Association for Computational Linguistics.
- E. Miltsakaki, R. Prasad, A. Joshi, and B. Webber. 2004. Annotating Discourse Connectives And Their Arguments. In *Proceedings of the HLT/NAACL Workshop on Frontiers in Corpus Annotation*, pages 9–16.
- L. Mladová, Š. Zikánová, and E. Hajičová. 2008. From Sentence to Discourse: Building an Annotation Scheme for Discourse Based on Prague Dependency Treebank. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*.
- U. Oza, R. Prasad, S. Kolachina, S. Meena, D.M. Sharma, and A. Joshi. 2009a. Experiments with Annotating Discourse Relations in the Hindi Discourse Relation Bank. In *Proceedings of the 7th International Conference on Natural Language Processing (ICON-2009)*, Hyderabad, India.
- U. Oza, R. Prasad, S. Kolachina, D.M. Sharma, and A. Joshi. 2009b. The Hindi Discourse Relation Bank. In *Proceedings of the Third Linguistic Annotation Workshop*, pages 158–161. Association for Computational Linguistics.
- M. Palmer, R. Bhatt, B. Narasimhan, O. Rambow, D.M. Sharma, and F. Xia. 2009. Hindi Syntax: Annotating Dependency, Lexical Predicate-Argument Structure, and Phrase Structure. In *The 7th International Conference on Natural Language Processing*, pages 14–17.
- R. Prasad, N. Dinesh, A. Lee, A. Joshi, and B. Webber. 2007. Attribution and its annotation in the Penn Discourse TreeBank. *Traitement Automatique des Langues, Special Issue on Computational Approaches to Document and Discourse*, 47(2).
- R. Prasad, N. Dinesh, A. Lee, E. Miltsakaki, L. Robaldo, A. Joshi, and B. Webber. 2008. The Penn Discourse Treebank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*, pages 2961–2968. Citeseer.
- R. T. Rachakonda and D. M. Sharma. 2011. Creating an Annotated Tamil Corpus as a Discourse Resource. In *Proceedings of the 5th Linguistic Annotation Workshop*, pages 119–123, Portland, OR.
- S. Tonelli, G. Riccardi, R. Prasad, and A. Joshi. 2010. Annotation of Discourse Relations for Conversational Spoken Dialogs. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010)*, Valletta, Malta.
- F. Xia, O. Rambow, R. Bhatt, M. Palmer, and D.M. Sharma. 2008. Towards a Multi-Representational Treebank. *Proceedings of the 7th International Workshop on Treebanks and Linguistic Theories (TLT-7)*, pages 207–238.
- D. Zeyrek and B. Webber. 2008. A Discourse Resource for Turkish: Annotating Discourse Connectives in the METU Corpus. *Proceedings of IJCNLP-2008*. Hyderabad, India.
- Y. Zhou and N. Xue. 2012. PDTB-style Discourse Annotation of Chinese Text. In *Proc. 50<sup>th</sup> Annual Meeting of the ACL*, Jeju Island, Korea.