

# Towards an Annotated Corpus of Discourse Relations in Hindi

Rashmi Prasad\*, Samar Husain†, Dipti Mishra Sharma† and Aravind Joshi\*

## Abstract

We describe our initial efforts towards developing a large-scale corpus of Hindi texts annotated with discourse relations. Adopting the lexically grounded approach of the Penn Discourse Treebank (PDTB), we present a preliminary analysis of discourse connectives in a small corpus. We describe how discourse connectives are represented in the sentence-level dependency annotation in Hindi, and discuss how the discourse annotation can enrich this level for research and applications. The ultimate goal of our work is to build a Hindi Discourse Relation Bank along the lines of the PDTB. Our work will also contribute to the cross-linguistic understanding of discourse connectives.

## 1 Introduction

An increasing interest in human language technologies such as textual summarization, question answering, natural language generation has recently led to the development of several discourse annotation projects aimed at creating large scale resources for natural language processing. One of these projects is the Penn Discourse Treebank (PDTB Group, 2006),<sup>1</sup> whose goal is to annotate the discourse relations holding between eventualities described in a text, for example causal and contrastive relations. The PDTB is unique in using a lexically grounded approach for annotation: discourse relations are anchored in lexical items (called “explicit discourse connectives”) whenever they are

explicitly realized in the text. For example, in (1), the causal relation between ‘the federal government suspending US savings bonds sales’ and ‘Congress not lifting the ceiling on government debt’ is expressed with the explicit connective ‘because’.<sup>2</sup> The two arguments of each connective are also annotated, and the annotations of both connectives and their arguments are recorded in terms of their text span offsets.<sup>3</sup>

(1) *The federal government suspended sales of U.S. savings bonds because Congress hasn’t lifted the ceiling on government debt.*

One of the questions that arises is how the PDTB style annotation can be carried over to languages other than English. It may prove to be a challenge cross-linguistically, as the guidelines and methodology appropriate for English may not apply as well or directly to other languages, especially when they differ greatly in syntax and morphology. To date, cross-linguistic investigations of connectives in this direction have been carried out for Chinese (Xue, 2005) and Turkish (Deniz and Webber, 2008). This paper explores discourse relation annotation in Hindi, a language with rich morphology and free word order. We describe our study of “explicit connectives” in a small corpus of Hindi texts, discussing them from two perspectives. First, we consider the type and distribution of Hindi connectives, proposing to annotate a wider range

---

<sup>2</sup> The PDTB also annotates implicit discourse relations, but only locally, between adjacent sentences. Annotation here consists of providing connectives (called “implicit discourse connectives”) to express the inferred relation. Implicit connectives are beyond the scope of this paper, but will be taken up in future work.

<sup>3</sup> The PDTB also records the senses of the connectives, and each connective and its arguments are also marked for their attribution. Sense annotation and attribution annotation are not discussed in this paper. We will, of course, pursue these aspects in our future work concerning the building of a Hindi Discourse Relation Bank.

---

\* University of Pennsylvania, Philadelphia, PA, USA, {rjprasad,joshi}@seas.upenn.edu

† Language Technologies Research Centre, IIIT, Hyderabad, India, samar@research.iiit.ac.in, dipti@iiit.ac.in

<sup>1</sup> <http://www.seas.upenn.edu/pdtb>

of connectives than the PDTB. Second, we consider how the connectives are represented in the Hindi sentence-level dependency annotation, in particular discussing how the discourse annotation can enrich the sentence-level structures. We also briefly discuss issues involved in aligning the discourse and sentence-level annotations.

Section 2 provides a brief description of Hindi word order and morphology. In Section 3, we present our study of the explicit connectives identified in our texts, discussing them in light of the PDTB. Section 4 describes how connectives are represented in the sentence-level dependency annotation in Hindi. Finally, Section 5 concludes with a summary and future work.

## 2 Brief Overview of Hindi Syntax and Morphology

Hindi is a free word order language with SOV as the default order. This can be seen in (2), where (2a) shows the constituents in the default order, and the remaining examples show some of the word order variants of (2a).

- (2) a. मलय ने समीर को किताब दी ।  
malay ERG sameer DAT book gave  
“Malay gave the book to Sameer” (S-IO-DO-V)<sup>4</sup>  
b. मलय ने किताब समीर को दी. (S-DO-IO-V)  
c. समीर को मलय ने किताब दी. (IO-S-DO-V)  
d. समीर को किताब मलय ने दी. (IO-DO-S-V)  
e. किताब मलय ने समीर को दी. (DO-S-IO-V)  
f. किताब समीर को मलय ने दी. (DO-IO-S-V)

Hindi also has a rich case marking system, although case marking is not obligatory. For example, in (2), while the subject and indirect object are explicitly for the ergative (ERG) and dative (DAT) cases, the direct object is unmarked for the accusative.

## 3 Discourse Connectives in Hindi

Given the lexically grounded approach adopted for discourse annotation, the first question that arises is how to identify discourse connectives in Hindi. Unlike the case of the English connectives in the PDTB, there are no resources that alone or together provide an exhaustive list of connectives in the

language. We did try to create a list from our own knowledge of the language and grammar, and also by translating the list of English connectives in the PDTB. However, when we started looking at real data, this list proved to be incomplete. For example, we discovered that the form of the complementizer ‘कि’ also functions as a temporal subordinator, as in (3).

- (3) [ वह बाल्टी के गंदे पानी से अपनी चॉकलेट  
[he bucket of dirty water from his chocolates  
निकालने ही वाला था] कि {उसकी मम्मी ने  
taking-out just doing was] that {his mother ERG  
उसे रोक दिया }  
him stop did}

“He was just going to take out the chocolates from the dirty water in the bucket when his mother stopped him.”

The method of collecting connectives will therefore necessarily involve “discovery during annotation”. However, we wanted to get some initial ideas about what kinds of connectives were likely to occur in real text, and to this end, we looked at 9 short stories with approximately 8000 words. Our goal here is to develop an initial set of guidelines for annotation, which will be done on the same corpus on which the sentence-level dependency annotation is being carried out (see Section 4). Table 1 provides the full set of connectives we found in our texts, grouped by syntactic type. The first four columns give the syntactic grouping, the Hindi connective expressions, the English gloss, and the English equivalent expressions, respectively. The last column gives the number of occurrences we found of each expression. In the rest of this section, we describe the function and distribution of discourse connectives in Hindi based on our texts. In the discussion, we have noted our points of departure from the PDTB where applicable, both with respect to the types of relations being annotated as well as with respect to terminology. For argument naming, we use the PDTB convention: **the clause with which the connective is syntactically associated is called Arg2 and the other clause is called Arg1**. Two special conventions are followed for paired connectives, which we describe below. **In all Hindi examples in this paper, Arg1 is enclosed in square brackets and Arg2 is in braces.**

<sup>4</sup> S=Subject; IO=Indirect Object; DO=Direct Object; V=Verb; ERG=Ergative; DAT=Dative

Connective Type	Hindi	Gloss	English	Num
Sub. Conj.	क्योंकि	why-that	because	2
	(क्यों)कि..इसलिए	(why)-that..this-for	because	3
	(अगर यदी)..तब तो	(if)..then	if..(then)	15
	(जब).. तब तो	(when)..then	when	50
	जब तक.. तब तक (के लिए)	when till..then till (of for)	until	2
	जैसे ही..(तो)	as just..(then)	as soon as	5
	इतना ऐसा..की	so such..that	so that	12
	ताकि	so-that	so that	1
	कि	that	when	5
Sentential Relatives	जिससे	which-with	because of which	5
	जो	which	because of which	1
	जिसके कारण	which-of reason	because of which	1
Subordinator	पर	upon	upon	9
	(-कर -के करके)	(do)	after while	111
	समय	time	while	1
	हुए	happening	while	28
	के बाद	of later	after	3
	से	with	due to	1
	के पहले	of before	before	1
	के लिए	of for	in order to	4
	में	in	while	1
	के कारण	of reason	because of	3
Coord. Conj.	लेकिन पर परन्तु	but	but	51
	और तथा	and	and	117
	या	or	or	2
	यों तो..पर	such TOP..but	but	2
	ना केवल..बल्कि	not only..but	not only..but	1
Adverbial	तब	then	then	2
	बाद में	later in	later	5
	फिर	then	then	4
	इसीलिए	this-for	that is why	7
	नहीं तो	not then	otherwise	5
	तभी तो	then-only TOP	that is why	1
	सो	so	so	10
	वही यही नहीं	that this-only not	not only that	1
<b>TOTAL</b>				<b>472</b>

Table 1: A Partial List of Discourse Connectives in Hindi. Parentheses are used for optional elements; “|” is used for alternating elements; TOP = topic marker.

### 3.1 Types of Discourse Connectives

#### 3.1.1 Subordinating Conjunctions

Finite adverbial subordinate clauses are introduced by independent lexical items called “subordinating conjunctions”, such as *क्योंकि* (“because”), as in (4), and they typically occur as right or left attached to the main clause.

- (4) [मैं इस सभी धन को राज्य के बादशाह को दे देता], *क्योंकि* {वही समस्त

DAT give would], *why-that* {he-EMPH all धरती की सम्पदा का स्वामी है} earth of wealth of lord is}

“I would give all this wealth to the king, because he alone is the lord of this whole world’s wealth.”

As the first group in Table 1 shows, subordinating conjunctions in Hindi often come paired, with one element in the main clause and the other in the subordinate clause (Ex.5). One of these elements can also be implicit (Ex.6),

and in our texts, this was most often the subordinate clause element.

- (5) क्योंकि {यह तुम्हारी ज़मीन पर मिला है}, इसलिए  
because {this your land on found has}, this-for  
[इस धन पर तुम्हारा अधिकार है]  
[this treasure on your right is]

“Because this was found on your land, you have the right to this treasure.”

- (6) [उसका वश चलता] तो {वह उसे घर से  
[her power walk] then {she it home from  
बाहर निकाल देती}  
out take would}

“Had it been in her power, she would have banished it from the house.”

When both elements of the paired connective are explicit, their text spans must be selected discontinuously. The main clause argument is called Arg1 and the subordinate clause argument, Arg2.

Subordinating conjunctions, whether single or paired, can occur in non-initial positions in their clause. However, this word order variability is not completely unconstrained. First, not all conjunctions display this freedom. For example, while ‘जब’ (‘when’) can be clause-medial (Ex. 7), ‘क्योंकि’ (‘because’) cannot. Second, when the main clause precedes the subordinate clause, the main clause element, if explicit, cannot appear clause-initially at all. Consider the causal ‘क्योंकि.. इसलिये’ (Ex.5), which represents the subordinate-main clause order. In the reverse order, the explicit main clause ‘इसलिये’ (Ex.8) appears clause medially. Placing this element in clause-initial position is not possible.

- (7) {लकड़हारे की पत्नी को} जब {यह  
{woodcutter of wife DAT} when {this  
मालूम पड़ा कि इस चिड़िया के कारण  
knowledge put that this bird of reason  
काम छोड़कर घर आ गया है} तो [वह  
work leaving home come went is} then [she  
उस पर बरस पड़ी].  
him on anger-rain put}

“When the woodcutter’s wife found out that he had left his work and come home to care for the bird, she raged at him.”

- (8) [. . . पर चिराग की बत्ती उसका या दोहरी  
[. . .but lamp of light light or another  
बत्ती लगाना] शायद इसलिए [उचित नहीं

light putting] perhaps this-for [appropriate not समझते थे] कि {तेल का अपव्यय होगा}.

Consider did] that {oil of waste be-FUT}.

“... but he did not consider it appropriate to light the lamp repeatedly or light another lamp, perhaps because it would be a waste of oil.”

### 3.1.2 Sentential Relative Pronouns

Since discourse relations are defined as holding between eventualities, we have also identified relations that are expressed syntactically as relative pronouns in sentential relative clauses, which modify the main clause verb denoting an eventuality, rather than some entity denoting noun phrase. For example, in (9), a result/purpose relation is conveyed between ‘the man’s rushing home’ and ‘the bird being taken care of’, and we believe that this relation between the eventualities should be captured despite its syntactic realization as the relative pronoun ‘जिससे’ (‘because of which/so that’). (10) gives an example of a modified relative pronoun.

- (9) [सारा काम चोड़कर वह उस बीमार चिड़िया  
[all work leaving he that sick bird  
को उठाकर दवा घर की ओर भागा],  
ACC picking-up fast home of direction ran],  
जिससे {उसका सही इलाज किया जा सके}  
from-which {her proper care do go able}

“Leaving all his work, he picked up the bird and ran home very fast, so that the bird could be given proper care.”

- (10) [ऊँटों के हर वार कदम रखने पर  
[camels of every time step keeping upon  
चिड़ियों के सिर आपस में तथा ऊँट की  
birds of head each-other in and camels of  
गरदन से टकरा रहे थे] जिसके कारण  
neck with hit-against be had] of-which reason  
{उन पक्षियों की दरदभरी चीखें निकल  
{those birds of painful screams come-out  
रही थीं}.  
be had}

“With each step of the camels, the birds heads were hitting against each other as well as with the camels’ necks because of which the birds were screaming painfully.”

### 3.1.3 Subordinators

In contrast to the subordinating conjunctions, elements introducing non-finite subordinate clauses are called “subordinators”. Unlike

English, where certain non-finite subordinate clauses, called “free adjuncts”, appear without any overt marking so that their relationship with the main clause is unspecified, Hindi non-finite subordinate clauses almost always appear with overt marking. However, also unlike English, where the same elements may introduce both finite and non-finite clauses (cf. *After leaving, she caught the bus* vs. *After she left, she caught the bus*), different sets of elements are used in Hindi. In fact, as can be seen in the subordinator group in Table 1, the non-finite clause markers are either postpositions (Ex.11), particles following verbal participles (Ex.12), or suffixes marking serial verbs (Ex.13).

- (11) {मम्मी के मना करने} के कारण [रामू  
 {mummy of warning doing} of reason [Ramu  
 थोड़ी थोड़ी चॉकलेट बड़े अनंद के साथ  
 little little chocolate big pleasure of with  
 खा रहा था].  
 eat being be]

“Because of his mother’s warning, Ramu was eating bits of chocolate with a lot of pleasure.”

- (12) . . . और {खेलते} हुए [यह भूल जाता है  
 . . . and {playing} happening [this forget go is  
 कि यदि उसका मित्र भी अपने खिलौने को  
 that if his friends also their toys to  
 उसे हाथ नहीं लगाने देता, तो उसे  
 him hand not touching did, then he  
 कितना बुरा लगता]  
 how-much bad feel]

“. . . and while playing, he forgets that if his friends too didn’t let him touch their toys, then how bad he would feel.”

- (13) {अपनी पत्नी से यह सुन}कर [लकड़हारा  
 {self wife from this listen}-do [woodcutter  
 बहुत दुखी हुआ]  
 much sad became]

“Upon hearing this from his wife, the woodcutter became very sad.”

While subordinators constitute a frequently-used way to mark discourse relations, their annotation raises at least two difficult problems, both of which have implications for the reliability of annotation. **The first is that these markers are used for marking both argument clauses and adjunct clauses, so that annotators would be required to make difficult decisions for distinguishing them: in the former case, the**

marker would not be regarded as a connective, while in the latter case, it would. Second, the clauses marked by these connectives often seem to be semantically weak. This is especially true of verbal participles, which are nonfinite verb appearing in a modifying relation with another finite verb. Whereas in some cases (Ex.12-13) the two verbs are perceived as each projecting “two distinct events” between which some discourse relation can be said to exist, in other cases (Ex.14), the two verbs seem to project two distinct actions but as part of a “single complex event” (Verma, 1993). These judgments can be very subtle, however, and our final decision on whether to annotate such constructions will be made after some initial annotation and evaluation.

- (14) {देखते ही देखते सब बैल भागते }  
 {looking EMPH looking all buffalos running}  
 हुए [गोशाला पहुँच गए]  
 happening [shed reach did]

“Within seconds all the buffalos came running to the shed.”

The naming convention for the arguments of subordinators is the same as for the subordinating conjunctions: the clause associated with the subordinator is called Arg2 while its matrix clause is called Arg1.

**Unlike subordinating conjunctions, subordinators do not come paired and they can only appear clause-finally.** Clause order, while not fixed, is restricted in that the nonfinite subordinate clause can appear either before the main clause or embedded in it, but never after the main clause.

### 3.1.4 Coordinating Conjunctions

Coordinating conjunctions in Hindi are found in both inter-sentential (Ex.15) and intra-sentential (Ex.16) contexts, **they always appear as independent elements, and they almost always appear clause-initially.**<sup>5</sup> For these connectives,

<sup>5</sup> While the contrastive connectives ‘पर’, ‘परन्तू’ appear only clause-initially, it seems possible for the contrastive ‘लेकिन’ to appear clause-medially, suggesting that these two types may correspond to the English ‘but’ and ‘however’, respectively. However, we did not find any examples of clause-medial ‘लेकिन’ in our texts, and this behavior will have to be verified with further annotation.

the first clause is called Arg1 and the second, Arg2.

(15) [जब वह लौटता तो गा-गाकर उसका मन  
[when he return then sing-singing his mind  
खुश कर देती]. लेकिन {उसकी पत्नी को वह  
happy do gave}. But {his wife DAT the  
चिड़िया फूटी आँख नहीं सुहाती थी}.  
bird torn eye not bear did}

“Upon his return, she would make him happy by singing. But his wife could not tolerate the bird even a little bit.”

(16) [ तभी दरवाज़ा खुला] और {मालकिन आ  
[then-only door opened] and {wife come  
गई }.  
went}

“Just then the door opened and the wife came in.”

We also recognize paired coordinating conjunctions, such as ‘ना केवल..बल्कि’ (See Table 1). The argument naming convention for these is the same as for the single conjunctions.

### 3.1.5 Discourse Adverbials

Discourse adverbials in Hindi modify their clauses as independent elements, and some of these are free to appear in non-initial positions in the clause. Example (17) gives an example of the consequence adverb, ‘सो’. The Arg2 of discourse adverbials is the clause they modify, whereas Arg1 is the other argument.

(17) [चिड़िया जबान कट जाने और मालकिन के ऐसे  
[bird tongue cut going and wife of this  
व्यवहार से डर गई थी]. सो {वह किसी  
behavior with fear go had}. So {she some  
तरह उड़कर चली गई}.  
manner flying walk went}.

“The bird was scared due to her tongue being cut and because of the wife’s behavior. So she somehow flew away.”

As with the PDTB, one of our goals with the Hindi discourse annotation is to explore the structural distance of Arg1 from the discourse adverbial. If the Arg1 clause is found to be non-adjacent to the connective and the Arg2 clause, it may suggest that adverbials in Hindi behave anaphorically. In the texts we looked at, we did not find any instances of non-adjacent Arg1s.

Additional annotation will provide further evidence in this regard.

## 4 Hindi Sentence-level Annotation and Discourse Connectives

The sentence-level annotation task in Hindi is an ongoing effort which aims to come up with a dependency annotated treebank for the NLP/CL community working on Indian languages. Presently a million word Hindi corpus is being manually annotated (Begum et al., 2008). The dependency annotation is being done on top of the corpus which has already been marked for POS tag and chunk information. The scheme has 28 tags which capture various dependency relations. These relations are largely inspired by the Paninian grammatical framework. Given below are some relations, reflecting the argument structure of the verb.

- a) कर्ता (agent) (k1)
- b) कर्म (theme) (k2)
- c) करण (instrument) (k3)
- d) सम्प्रदान sampradaan (recipient) (k4)
- e) अपादान (source) (k5)
- f) अधिकरण (location) (k7)

Figure 1 shows how Examples (2a-f) are represented in the framework. Note that agent and theme are rough translations for ‘कर्ता’ and ‘कर्म’ respectively. Unlike thematic roles, these relations are not purely semantic, and are motivated not only through verbal semantics but also through vibhaktis (postpositions) and TAM (Tense, aspect and modality) markers (Bharati et al., 1995). The relations are therefore syntactico-semantic, and unlike thematic roles there is a greater binding between these relations and the syntactic cues.

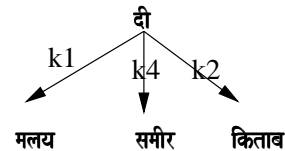


Figure 1: Dependency Diagram for Example (2) Some discourse relations that we have identified are already clearly represented in the sentence-level annotation. But for those that aren’t, the

discourse level annotations will enrich the sentence-level. In the rest of this section, we discuss the representation of the different types of connectives at the sentence level, and discuss how the discourse annotation will add to the information present in the dependency structures.

**Subordinating Conjunctions** Subordinating conjunctions are lexically represented in the dependency tree, taking the subordinating clause as their dependents while themselves attaching to the main verb (the root of the tree). Figure 2 shows the dependency tree for Example (4) containing the subordinating conjunction 'क्योंकि'. Note that the edge between the connective and the main verb gives us the causal relation between the two clauses, the relation label being 'rh' (relation hetu 'cause'). Thus, the discourse level can be taken to be completely represented at the sentence-level.

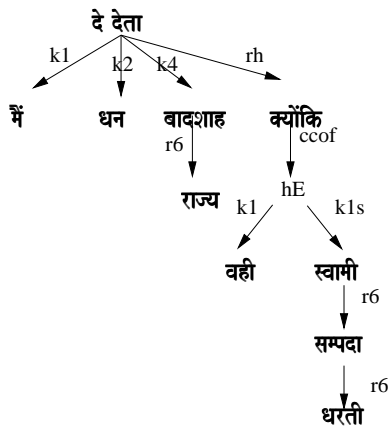


Figure 2: Dependency Tree for Subordinating Conjunction in Example (4)

**Paired Subordinating Conjunctions** Unlike Example (4), however, the analysis for the paired connective in Example (5), given in Figure 3, is insufficient. Despite the lexical representation of the connective in the tree, the correct interpretation of the paired conjunction and the clauses which it relates is only possible at the discourse level. In particular, the dependencies don't show that 'क्योंकि' and 'इसलिए' are two parts of the same connective, expressing a single relation and taking the same two arguments. Thus, the discourse annotation will

be able to provide the appropriate argument structure and semantics for these paired connectives.

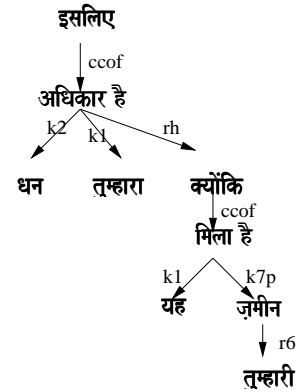


Figure 3: Dependency Tree for Paired Subordinating Conjunction in Example (5)

**Subordinators** As mentioned earlier, Hindi nonfinite subordinate clauses almost always appear with overt marking. But unlike the subordinating conjunctions, subordinators are not lexically represented in the dependency trees. Figure 4 gives the dependency representation for Example (11) containing a postposition subordinator 'के कारण', which relates the main and subordinate clauses causally. As the figure shows, while the causal relation label ('rh') appears on the edge between the main and subordinate verbs, the subordinator itself is not lexically represented as the mediator of this relation. The lexically grounded annotation at the discourse level will thus provide the textual anchors of such relations, enriching the dependency representation. Furthermore, while many of the subordinators in Table 1 are fully specified in the dependency trees for the semantic relation they denote (e.g., 'पर' and 'में' marked as the 'k7t' (location in time) relation, and 'के कारण' and 'से' marked as the 'rh' (cause/reason) relation), others, like the particle 'हुए' are underspecified for their semantics, being marked only as 'vmod' (verbal modifier). The discourse-level annotation will thus be the source for the semantics of these subordinators.

**Coordinating Conjunctions** Coordinating conjunctions at the sentence level anchor the root of the dependency tree. Figure 5 shows the

dependency representation of Example (16) containing a coordinating conjunction.

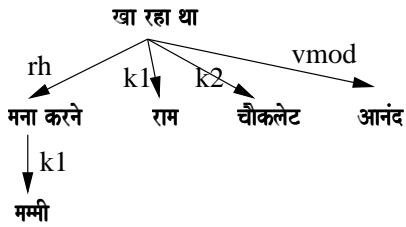


Figure 4: Dependency Tree for Subordinator in Example (11)

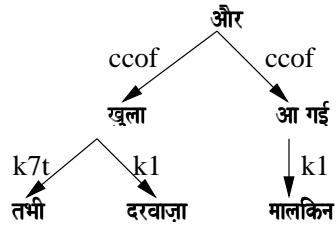


Figure 5: Dependency Tree for Coordinating Conjunction in Example (16)

While the sentence-level dependency analysis here is similar to the one we get at the discourse level, the semantics of these conjunctions are again underspecified, being all marked as ‘ccof’, and can be obtained from the discourse level.

**Discourse Adverbials** Like subordinating conjunctions, discourse adverbials are represented lexically in the dependency tree. They are attached to the verb of their clause as its child node and their denoted semantic relation is specified clearly. This can be seen with the temporal adverb ‘तभी’ (‘then-only’) and its semantic label ‘k7t’ in Figure 5. At the same time, since the Arg1 discourse argument of adverbials is most often in the prior context, the discourse annotation will enrich the semantics of these connectives by providing the Arg1 argument.

## 5 Summary and Future Work

In this paper, we have described our study of discourse connectives in a small corpus of Hindi texts in an effort towards developing an annotated corpus of discourse relations in Hindi. Adopting the lexically grounded approach of the Penn Discourse Treebank, we have identified a

wide range of connectives, analyzing their types and distributions, and discussing some of the issues involved in the annotation. We also described the representation of the connectives in the sentence-level dependency annotation being carried out independently for Hindi, and discussed how the discourse annotations can enrich the information provided at the sentence level. While we focused on explicit connectives in this paper, future work will investigate the annotation of implicit connectives, the semantic classification of connectives, and the attribution of connectives and their arguments.

## References

- Rafiya Begum, Samar Husain, Arun Dhvaj, Dipti Misra Sharma, Lakshmi Bai, and Rajeev Sangal. 2008. Dependency annotation scheme for Indian languages. In *Proceedings of IJCNLP-2008*. Hyderabad, India.
- Akshar Bharati, Vineet Chaitanya, and Rajeev Sangal. 1995. *Natural Language Processing: A Paninian Perspective*. Prentice Hall of India. <http://ltrc.iiit.ac.in/downloads/nlpbook/nlppanini.pdf>.
- Manindra K. Verma (ed.). 1993. *Complex Predicates in South Asian Languages*. New Delhi: Manohar.
- The PDTB-Group. 2006. The Penn Discourse TreeBank 1.0 Annotation Manual. Technical Report IRCS-06-01, IRCS, University of Pennsylvania.
- Bonnie Webber, Aravind Joshi, Matthew Stone, and Alistair Knott. 2003. Anaphora and discourse structure. *Computational Linguistics*, 29(4):545–587.
- Nianwen Xue. 2005. Annotating Discourse Connectives in the Chinese Treebank. In *Proceedings of the ACL Workshop on Frontiers in Corpus Annotation II: Pie in the Sky*. Ann Arbor, Michigan.
- Deniz Zeyrek and Bonnie Webber. 2008. A Discourse Resource for Turkish: Annotating Discourse Connectives in the METU Corpus. In *Proceedings of IJCNLP-2008*. Hyderabad, India.