Applicative Structures and Immediate Discourse in the Turkish Discourse Bank

İşin Demirşahin, Adnan Öztürel, Cem Bozşahin, and Deniz Zeyrek
Department of Cognitive Science, Informatics Institute
Middle East Technical University
{disin, ozturel, bozsahin, zeyrek}@metu.edu.tr

Abstract
Various discourse theories have argued for data structures ranging from the simplest trees to the most complex chain graphs. This paper investigates the structure represented by the explicit connectives annotated in the multiple-genre Turkish Discourse Bank (TDB). The dependencies that violate tree-constraints are analyzed. The effects of information structure in the surface form, which result in seemingly complex configurations with underlying simple dependencies, are introduced; and the structural implications are discussed. The results indicate that our current approach to local discourse structure needs to accommodate properly contained arguments and relations, and partially overlapping as well as shared arguments; deviating further from simple trees, but not as drastically as a chain graph structure would imply, since no genuine cases of structural crossing dependencies are attested in TDB.

1 Introduction
A variety of structures for discourse representation has been proposed, including successive trees of varying sizes connected and occasionally intertwined at the peripheries (Hobbs, 1985), a single tree structure (Rhetorical Structure Theory, RST, Mann & Thompson, 1988), entity chains (Knott et al., 2001), tree-adjoining grammars (Discourse-Lexicalized Tree Adjoining Grammar, D-LTAG, Webber, 2004), directed acyclic graphs (Lee et al., 2006, 2008) and chain graphs (Wolf & Gibson, 2005).

The simplest of these structures is a tree, which treats discourse structure simpler than sentence-level syntax. The most complex representation, chain graphs that allow for crossing dependencies and other tree-violations, treats discourse as more complex than sentence level. We know since the work of Shieber (1985) and Joshi (1985) that sentence-level structures require more than context-free power, but not to the extent of dealing with general graphs, or with strings that grow out of constant control. It is of general interest to discover whether such complexity occurs in natural discourses, because we would like to know how far discourse structures deviate from applicative semantics. (Applicative structures are binary operations on data; for example a connective’s meaning depending only on two arguments. A system is applicative if it only makes use of function application, but not e.g. graph reduction or general function composition. The concepts are distinct but related: function application can be linked to applicative structures by currying.) If more complex structures are found, we must go above applicative semantics, and we must worry about function compositions and graph reductions, which are known to require more computational power.

2 Turkish Discourse Bank
Turkish Discourse Bank (TDB) is the first large-scale publicly available language resource with discourse level annotations for Turkish built on a ~ 400,000-word sub-corpus of METU Turkish Corpus (MTC) (Say et al., 2002), annotated in the style of Penn Discourse Tree Bank (PDTB) (Prasad et al., 2008). The TDB Relations are annotated for explicit discourse connectives, which link two spans of text that can be interpreted as Abstract Objects (Asher, 1993). Connectives are annotated together with their modifiers and arguments, and with supplementary materials for the arguments (Zeyrek & Webber, 2008; Zeyrek et al., 2010). The first release of TDB is available at http://medid.ii.metu.edu.tr/.
As in PDTB, the connectives in TDB come from a variety of syntactic classes (Zeyrek & Webber, ibid). The coordinating and subordinating conjunctions such as ve ‘and’ and için ‘for’ and ‘in order to’, respectively, are considered structural connectives, meaning that they take both arguments structurally. Discourse adverbials and phrasal expressions that are built by combining a discourse-anaphoric element with a subordinating conjunction are considered to be anaphoric connectives, meaning that they only take the argument that is syntactically related, and the other argument is interpreted anaphorically. In PDTB and TDB style, the syntactically related argument is called the second argument (Arg2), and the other argument is called the first argument (Arg1), for both structural and anaphoric connectives. The syntactic class of the discourse connective will be included in the further releases of TDB along with the sense of the discourse relations, and some morphological features for the arguments of subordinating conjunctions (Demirşahin et al., 2012).

3 Discourse Relation Configurations in Turkish

Lee et al. (2006) identified independent relations and fully embedded relations as conforming to the tree structure, and shared arguments, properly contained arguments, pure crossing, and partially overlapping arguments as departures from the tree structure in PDTB. Although most departures from the tree structure can be accounted for by non-structural explanations, such as anaphora and attribution, Lee et al. (2006, 2008) state that shared arguments may have to be accepted in discourse structure.

Aktas et al. (2010) identified similar structures in TDB, adding nested relations that do not violate tree structure constraints, as well as properly contained relations that introduce further deviations from trees. Following their terminology, we will reserve the word relation to discourse relations (or coherence relations), and use the term configuration to refer to relations between discourse relations.

1.1 Independent, Fully Embedded and Nested Relations

The first release of TDB consists of 8,484 explicit relations. The argument spans of some discourse connectives do not overlap with those of any other connectives in the corpus. We call them independent relations. All others are called non-independent relations. We have identified 2,548 non-independent configurations consisting of 3,474 unique relations, meaning that 5,010 relations (59.05%) are independent. Table 1 shows the distribution of 2,548 non-independent configurations.

| Configuration            | #   | %    |
|-------------------------|-----|------|
| Full Embedding          | 695 | 27.28|
| Nested Relations        | 138 | 5.42 |
| Total Non-violating Configurations | 833 | 32.69|
| Shared Argument         | 489 | 19.19|
| Prop. Cont. Argument    | 194 | 7.61 |
| Prop. Cont. Relation    | 1018| 39.95|
| Pure Crossing           | 2   | 0.08 |
| Partial Overlap         | 12  | 0.47 |
| Total Violating Configurations | 1715| 67.31|
| Total                   | 2548| 100.00|

Table 1: Distribution of non-independent configurations

Since full embedding and nested relations conform to tree structure, these configurations will not be discussed further. The following subsections discuss the suitability of explanations involving anaphora and attribution to tree-violating configurations. Those that cannot be completely explained away must be accommodated by the discourse structure.

1.2 Shared Arguments

Lee et al. (2006, 2008) state that shared argument is one of the configurations that cannot be explained away, and should be accommodated by discourse structure. Similarly, Egg & Redeker (2008) admit that even in a corpus annotated within RST Framework, which enforces tree structure by annotation guidelines, there is a genre-specific structure that is similar to the shared arguments in Lee et al. (2006).
Of the 489 shared arguments in TDB, 331 belong to anaphoric discourse relations (i.e. relations in which at least one of the connectives involved is either a discourse adverbial or a phrasal expression) (67.69%). In the remaining 158 relations (32.31%), arguments are shared by structural connectives. (1) is an example of a shared argument.

(1) 00001131-2&3
(a) Vazgeçmek kolaydı, ertelemek de. Ama tırmanmaya başlandı mı bitirilmeli! Çünkü her seferinde acımasız bir geriye dönüş vardı. “It was easy to give up, so was to postpone. But once you start climbing you have to go all the way! Because there was a cruel comeback every time.”
(b) Vazgeçmek kolaydı, ertelemek de. Ama tırmanmaya başlamış bir bitirilmeli! Çünkü her seferinde acımasız bir geriye dönüş vardı. “It was easy to give up, so was to postpone. But once you start climbing you have to go all the way! Because there was a cruel comeback everytime.”

All examples are from TDB; the first line indicates the file name (00077211 in (1)), and the browser index of the connectives involved in the configuration (2 & 3 in (1)). The first arguments (Arg1) of the connectives are in italic, the second arguments (Arg2) are in bold. The connectives themselves are underlined. For the sake of simplicity, the modifiers of the connectives are displayed as part of the connective, and the shared tags are omitted when they are immaterial to the configuration being discussed.

In (1), the first argument of but (relation 2) annotated in (a) completely overlaps with the first argument of because (relation 3), annotated in (b) on the same text for comparison. The result is a shared argument configuration.

1.3 Properly Contained Relations and Arguments

In TDB there are 1,018 properly contained relations, almost half of which (471 relations; 46.27%) are caused by anaphoric relations.

Properly contained relations where anaphoric connectives are not involved can be caused by attribution, complement clauses, and relative clauses. (2) is a relation within a relative clause (a), which is part of another relation in the matrix clause (b). The result is a properly contained relation.

(2) 00001131-27&28
(a) Sabah çok erken saatte bir önceki akşam gün batmadan hemen önce astığı çamaşırları toplamaya çıkmış, ve doğal olarak da gün batmadan o günkü çamaşırlarını asmak için geliyordu. “She used to go out to gather the clean laundry she had hung to dry right before the sun went down the previous evening, and naturally she came before sunset to hang the laundry of the day.”
(b) Sabah çok erken saatte bir önceki akşam gün batmadan hemen önce astığı çamaşırları toplamaya çıkmış ve doğal olarak da gün batmadan o günkü çamaşırları asmak için geliyordu. “She used to go out to gather the clean laundry she had hung to dry the previous evening right before the sun went down, and naturally she came before sunset to hang the laundry of the day.”
relation on the discourse level. Therefore, those stranded verbs of attribution should not be regarded as tree-structure violations. In (3) the properly contained relations occur in a quote, but the intervening materials are more than just verbs of attribution. Because the intervening materials in (3) are whole sentences that participate in complex discourse structures, we believe that (3) is different than the case proposed by Lee et al. (2006) and should be considered a genuine case of properly contained relation.

(3) 00003121-10, 11&13
(a) "Evet, küçük amcamdı o, nur içinde yatsın, yetmişlik bir rakıyi devirip ıpi sek sek geçmeye kalkmış; kaptan olan amcam ise kocaman bir gemiye sulara gömülü. Ayıldan kasımdı, ben çocukum, çok iyi anımsıyorum, firtınalı bir gecede, Karadeniz’in batısında batmışlardı. Kaptandı, ama yüze bilmezdi amcam. Bir namaz tahtası sarımlı olarak kıyıya vurduğu kolları zor açımlar, yarınıyı dönmuş. Belki de o anda Tanrı’ya yakarıp yardımı istiyordu, çünkü çok dindar bir adamdı. Ama artık değil; küp gibi içip meyhanelerde keman çalıyor." Sonra da Nesli’nin ilgisiyle çatılmış alına bakıp güliyor: "Çok istavritsiyin!"
   “Yes, he was my younger uncle, may he rest in peace, he tried to hop on the tightrope after quaffing down a bottle of raki, my other uncle who was a captain, on the other hand, sank a whole ship. It was October, I was a child, I remember it vividly, in a stormy night, they sank by the west of the Black Sea. He was a captain, but he couldn’t swim, my uncle. When he washed ashore holding onto a piece of driftwood, they tried open his arms with great difficulty, he was half frozen. Maybe at that moment he was begging God for help, because he was a very religious man. But not anymore, now he hits the bottle and plays the violin in taverns.” Then he sees Nesli’s interested frown and laughs: “You’re so gullible!”
(b) “Evet, [...] Ama artık değil; küp gibi içip meyhanelerde keman çalıyor.” Sonra da Nesli’nin ilgisiyle çatılmış alına bakıp güliyor: "Çok istavritsiyin!"
   “Yes, [...] But not anymore, now he hits the bottle and plays the violin in taverns.” Then he sees Nesli’s interested frown and laughs: “You’re so gullible!”

Whereas attribution can be discarded as a non-discourse relation, a discourse model based on discourse connectives should be able to accom-

modate partially contained relations resulting from relations within complements of verbs and relative clauses.

Figure 3 - Properly Contained Argument
As in properly contained relations, properly contained arguments may arise when an abstract object that is external to a quote is in a relation with an abstract object in a quote. Likewise, a discourse relation within the complement of a verb or a relative clause can cause properly contained arguments. Anaphoric connectives account for the 129 (66.49%) of the 194 properly contained arguments in TDB.

1.4 Partial Overlap

There are only 12 partial overlaps in TDB, and 3 of them involve anaphoric relations.

Figure 4 - Partial Overlap
In (4), the argument span of in order to partially overlaps with the argument span of to. This is a partial overlap of the arguments of two structural connectives.

(4) 20630000-44&45
(a) Hükümetin, 1998’de kapattığı kumarhaneleri, kaynak sorununa çözüm bulmak amacıyla yeniden açmak için harekete geçmesi, tartışma yarattı.
   “The fact that the government took action in order to reopen the casinos that were closed down in 1998 in order to come up with a solution to the resource problem caused arguments.”
(b) Hükümetin, 1998’de kapattığı kumarhaneleri, kaynak sorununa çözüm bulmak amacıyla yeniden açmak için harekete geçmesi, tartışma yarattı.
“The fact that the government took action in order to reopen the casinos that were closed down in 1998 in order to come up with a solution to the resource problem caused arguments.”

The first argument of relation 44 (a) properly contains the first argument of 45 (b), whereas the second argument of (b) properly contains the second argument of (a). This double containment results in a complicated structure that will be analyzed in detail in §3.5.

In (5) the second argument of but (relation 42 (a)) contains only one of the two conjoined clauses, whereas the first argument of after (relation 43 (b)) contains both of them. The most probable cause for this difference in annotations is the combination of “blind annotation” with the “minimality principle.” This principle guides the participants to annotate the minimum text span required to interpret the relation. Since the annotators cannot see previous annotations, they have to assess the minimum span of an argument again when they annotate the second relation. Sometimes the minimal span for one relation is annotated differently than the minimal span required for the other, resulting in partial overlaps.

(5) 00001131-42&43
(a) Yine istediği kişi, bir türlü görememişti, ama aylarca sabrettikten sonra gizlediği bir kadın suçunu daralttı, tüyleri diken diken oldu.
   “Once again he couldn’t see the person he wanted to see, but after waiting patiently for months, a woman he peeped at took his breath away, gave him goose bumps.”
(b) Yine istediği kişi, bir türlü görememişti, ama aylarca sabrettikten sonra gizlediği bir kadın suçunu daralttı, tüyleri diken diken oldu.
   “Once again he couldn’t see the person he wanted to see, but after waiting patiently for months, a woman he peeped at took his breath away, gave him goose bumps.”

1.5 Pure Crossing

There are only 2 pure crossing examples in the current release of TDB, a number so small that it is tempting to treat them as negligible. However, the inclusion of pure crossing would result in the most dramatic change in discourse structure, raising the complexity level to chain graph and making discourse structure markedly more complex than sentence level grammar. Therefore, we would like to discuss both examples in detail.

(6) 00010111-54&55
(a) Sonra ansızın sesler gelir. Ayak sesleri. Birilerinin ya işi vardır, acelyle yürürlar, ya koşarlар. O zaman kız katlasırmı ansızın. Oğlan da katlasır ve her koşunun gizli bir isteği var.
   “And then suddenly there is a sound. Footsteps. Someone has an errand to run, they walk hurriedly or run. Then the girl stiffens suddenly. The boy stiffens, too; and every run has a hidden wish.”
(b) Sonra ansızın sesler gelir. Ayak sesleri. Birilerinin ya işi vardır, acelyle yürürlar, ya koşarlar. O zaman kız katlasırmı ansızın. Oğlan da katlasır ve her koşunun gizli bir isteği var.
   “And then suddenly there is a sound. Footsteps. Someone has an errand to run, they walk hurriedly or run. Then the girl stiffens suddenly. The boy stiffens, too; and every run has a hidden wish.”

In (6), the discourse relation encoded by then is not only anaphoric -and therefore not determinant in terms of discourse structure- but also the crossing annotation does not necessarily arise from the coherence relation of the connective’s arguments. It is more likely imposed by lexical cohesive elements (Halliday & Hasan, 1976), as the annotators apparently made use of the repetitions of ansızın ‘suddenly’ and koş ‘run’ in the text when they could not interpret the intended meaning.

(7) 20510000-31,32&34
(a) Ceza, Telekom’un iki farklı internet alt yapısı pazarında tekel konumunu körükle kovaladığı için ve uydu istasyonu işletmeciliği pazarında artık tekel hakkı kalmadığı halde räkplerinin faaliyetlerini zorlaştırdığı için verildi.
"The penalty was given because Telekom abused its monopoly status in the two different internet infrastructure markets and because it caused difficulties with its rivals’ activities although it did not have a monopoly status in the satellite management market anymore."

(b) Ceza, Telekom’un iki farklı internet alt yapısı pazarında tekel konumunu kötüye kullanıldığı için ve uydudur istasyonu işletmeciliği pazarında artık tekel hakkı kalmadığı halde rakiplerinin faaliyetlerini zorlaştırdığı için verildi.

"The penalty was given because Telekom abused its monopoly status in the two different internet infrastructure markets and because it caused difficulties with its rivals’ activities although it did not have a monopoly status in the satellite management market anymore."

The first reason is the repetition of the subordinator için ‘because’. Had there been only the rightmost subordinator, the relation would be a simple case of Full Embedding, where ve ‘and’ in (b) connects the two reasons for the penalty, while the rightmost subordinator connects the combined reasons to the matrix clause (see Figure 6). However, since both subordinators were present, they were annotated separately. They share their first arguments, and take different spans as their second arguments, which are also connected by ve ‘and’, resulting in an apparent pure crossing.

Our alternative analysis is that ve ‘and’ actually takes the subordinators için ‘because’ in its scope, and it should be analyzed similar to an assumed single-subordinator case. This kind of annotation was not available in TDB because the annotation guidelines state that the discourse connectives at the peripheries of the arguments should be left out. Machine Learning can help us spot these instances.

The second reason for crossing is the wrapping of the first arguments of (a) and (c) around the subordinate clause. This crossing is in fact not a configuration-level dependency, but a relation-level surface phenomenon confined within the relation anchored by için ‘because’, without underlying complex discourse semantics. Example (8) is a simpler case where the surface crossing within the relation can be observed.
In (9) ve ‘and’ links two relative clauses, one of which seems to be embedded in the other. It should be noted that the first part of Arg1 (Bezirci-nin) has an ambiguous suffix. The suffix could be the agreement marker of the relative clause, as reflected in the annotation, or it could be the genitive marked complement of the genitive-possessive construction Bezirci’nin antoloji hazırlığı ‘Bezirci’s anthology preparation’. The latter analysis does not cause wrapping.

(10) 00003121-26
Biz yaşalar karşısında evli sayılacak, ama gerçekten evli iki insan gibi değil de (evlilikler sıradanlaşmıştır çünkü, teknüze ve sıkıcıdır; biz farklı olacaktı, aynı evi paylaşan iki öğrencisi gibi yaşayacaktı).

“We would be married under the law, but in reality we would live like two students sharing the same house rather than two married people (because marriages were getting ordinary, they were) monotonous and boring; we would be different.”

(11) 00008113-10
Masa ya da duvar saatleri bulunmuyan, ezan seslerini her zaman duymayıp zamanı öğrenmek için erkeklerin (evde oldukları zaman, tabii) cep saatiley doğanış işık saatine ve kendi içgüdüleriyle tahminlerine bel bağlayan birçok aile, yaşamlarını bu top sesine göre ayarlarlardı.

“Lots of families who didn’t have a table clock or a wall clock and couldn’t always hear the prayer calls, who relied upon the men’s pocket watch (when they were home, of course) and their instincts and guesses to learn the time adjusted their lives according to this cannon shot.”

Both (10) and (11) are parentheticals, resulting in a double-wrapping-like construction (Figure 8). However, parentheticals move freely in the clause and occupy various positions, so we believe that this construction should be taken as a peculiarity of the parenthetical, rather than the structural connectives involved in the relation.

(8) 10380000-3
1882’de İstanbul Ticaret Odası, bir zahire ve ticaret borsası kurulması için girişimde bulunuyor ama sonuç alamıyor.

“In 1882, Istanbul Chamber of Commerce makes an attempt for founding a Provisions and Commodity Exchange Market but cannot obtain a result.”

Subordinators in Turkish form adverbial clauses (Kornfilt, 1997), so they can occupy any position that is legitimate for a sentential adverb.

Wrapping in discourse seems to be motivated information-structurally. In the unmarked position, the subordinate clause comes before the matrix clause and introduces a theme. However, the discourse constituents can occupy different positions or carry non-neutral prosodic features to express different information structures (Demirşahin, 2008). In (7), wrapping takes ceza ‘penalty’ away from the rheme and makes it part of the theme, at the same time bringing the causal discourse relation into the rheme.

As is clear from the gloss in (7) and its stringset, this is function application, where ceza verildi ‘penalty was given’ wraps in the first argument as a whole. Double occurrence of the “connective” within the wrapped-in argument is causing the apparent crossing, but there is in fact one discourse relation.

Wrapping in discourse is almost exclusive to subordinating conjunctions, possibly due to their adverbial freedom in sentence-level syntax. The subordinators make up 468 of the total of 479 wrapping cases identified in TDB. However, there are also four cases of coordinating conjunctions with wrapping. Two of them result in surface crossing as in (9), and the other two build a nested-like structure, as in (10) and (11). The latter two are both parentheticals.

(9) 10690000-32
Bezirci’nin sonradan elimize geçen ve 1985’lerde yaptığı antoloji hazırlığında, […]

“In the preparation for an anthology which Bezirci made during 1985’s and which came into our possession later[…]”

In (9) ve ‘and’ links two relative clauses, one of which seems to be embedded in the other. It should be noted that the first part of Arg1 (Bezirci-nin) has an ambiguous suffix. The suffix could be the agreement marker of the relative clause, as reflected in the annotation, or it could be the genitive marked complement of the genitive-possessive construction Bezirci’nin antoloji hazırlığı ‘Bezirci’s anthology preparation’. The latter analysis does not cause wrapping.
4 Conclusion

In this paper we presented possible deviations from the tree structure in the first release of TDB. Following Lee et al. (2006, 2008) and Aktaş et al. (2010). We have scanned the corpus for shared arguments, properly contained relations and arguments, partial overlaps, and pure crossings. Overall, about half of these configurations can be accounted for by anaphoric relations, i.e. they are not applicative structures (see Table 2). Note that if one of the relations in a configuration is anaphoric, we treat the configuration as anaphoric.

| Configuration     | Structural | Anaphoric | Total |
|-------------------|------------|-----------|-------|
| Shared Argument   | 158        | 331       | 489   |
|                   | 32.31%     | 67.69%    | 100.00% |
| Prop. Cont. Arg.  | 65         | 129       | 194   |
|                   | 33.51%     | 66.49%    | 100.00% |
| Prop. Cont. Rel.  | 547        | 471       | 1018  |
|                   | 53.73%     | 46.27%    | 100.00% |
| Pure Crossing     | 1          | 1         | 2     |
|                   | 50.00%     | 50.00%    | 100.00% |
| Partial Overlap   | 9          | 3         | 12    |
|                   | 75.00%     | 25.00%    | 100.00% |
| Total             | 780        | 935       | 1715  |
|                   | 45.48%     | 54.52%    | 100.00% |

Table 2: Distribution of anaphoric relations among tree-violating configurations

In addition to the shared arguments that were accepted in discourse structure by Lee et al., we have also come up with partially contained relations arising from verbal complements and relative clauses. These structures can be treated differently in other frameworks; for instance in RST, they are treated as discourse constituents taking part in coherence relations. However, for the connective-based approach adopted in this study, they need to be accommodated as deviations from tree structure.

The few partial overlaps we have encountered could mostly be explained away by wrapping and by different interpretations of annotation guidelines by the annotators, especially the minimality principle. Recall that wrap has applicative semantics. Of the two pure crossing examples we have found, one was also anaphoric, whereas the other could be explained in terms of information-structurally motivated relation-level surface crossing, rather than configuration-level crossing dependency. In other words, if we leave the processing of information structure to other processes, the need for more elaborate annotation disappears. In Joshi’s (2011) terminology, immediate discourse in the TDB appears to be an applicative structure, which, unlike syntax, seems to be in no need of currying.

As a result, we can state that pure crossing (i.e. crossing of the arguments of structural connectives) is not genuinely attested in the current release of TDB. The annotation scheme need not be enriched to allow more complex algorithms to deal with unlimited use of crossing. There seems to be a reason in every contested case to go back to the annotation, and revise it in ways to keep the applicative semantics, without losing much of the connective’s meaning.

In summary, our preliminary analysis shows that discourse structure may have to accommodate partial containment and wrap in addition to shared arguments. TDB has an applicative structure.

Taking into account that independent relations, fully embedded relations and nested relations are frequent in discourse structure, and that the discourse structure should accommodate shared arguments and partial containments; we are currently inclined to think of discourse structure as Hobbs (1985) does: local trees of various sizes connected and occasionally intertwined at the edges. Further complications within trees are an open field for further studies.

References

Berfin Aktaş, Cem Bozşahin, Deniz Zeyrek. 2010. Discourse Relation Configurations in Turkish and an Annotation Environment. Proc. LAW IV - The Fourth Linguistic Annotation Workshop.

Nicholas Asher. 1993. Reference to Abstract Objects in Discourse. Kluwer Academic Publishers.

İşin Demirşahin. 2008. Connective Position, Argument Order and Information Structure of Discourse Connectives in Written Turkish Texts. Unpublished MS Thesis, Cognitive Science Program, Middle East Technical University.

İşin Demirşahin, Ayuğü Sevdiç-Çallı, Hale Ölç Balaban, Ruket Çakıcı and Deniz Zeyrek. 2012. Turkish Discourse Bank: Ongoing Developments. Proc. LREC 2012. The First Turkic Languages Workshop.

Markus Egg, Gisela Redeker. 2010. How Complex is Discourse Structure? Proc. 7th International Conference on Language Resources and Evaluation (LREC 2010) pp. 1619–23.
Michael A. K. Halliday, Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman

Jerry R. Hobbs. 1985. On the Coherence and Structure of Discourse. *Report CSLI-85-37, Center for Study of Language and Information*.

Aravind K. Joshi. 1985. How Much Context-sensitivity is Necessary for Characterizing Structural Descriptions: Tree Adjoining Grammars. In David Dowty, Lauri Karttunen and Arnold Zwicky (eds.) *Natural Language Parsing*, Cambridge University Press.

Aravind K. Joshi. 2011. Some Aspects of Transition from Sentence to Discourse. Keynote address, Informatics Science Festival, Middle East Technical University, Ankara, June 9.

Alistair Knott, Jon Oberlander, Michael O’Donnel, Chris Mellish. 2001. Beyond elaboration: The interaction of relations and focus in coherent text. In Ted Sanders, Joost Schilperoord and Wilbert Spooren (Eds.), *Text Representation: Linguistic and psycholinguistic aspects* (181-196): John Benjamins Publishing.

Jacqueline Kornfilt. 1997. *Turkish*. New York: Routledge.

Alan Lee, Rashmi Prasad, Aravind K. Joshi, Nikhil Dinesh, Bonnie Webber. 2006. Complexity of dependencies in discourse: are dependencies in discourse more complex than in syntax? *Proc. 5th Workshop on Treebanks and Linguistic Theory (TLT’06)*.

Alan Lee, Rashmi Prasad, Aravind K. Joshi, Bonnie Webber. 2008. Departures from tree structures in discourse. *Proc. Workshop on Constraints in Discourse III*.

William C. Mann, Sandra A. Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text, 8*(3), 243-281.

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind K. Joshi, Bonnie Webber. 2008. The Penn Discourse Treebank 2.0. *Proc. LREC’08 - The sixth international conference on Language Resources and Evaluation*.

Bilge Say, Deniz Zeyrek, Kemal Oflazer, and Umut Özge. 2002. Development of a Corpus and a Treebank for Present-day Written Turkish. *Proc. Eleventh International Conference on Turkish Linguistics (ICTL 2002)*.

Stuart Shieber. 1985. Evidence against the Context-Freeness of Natural Language. *Linguistics and Philosophy*: 8:333-343.

Bonnie Webber. 2004. D-LTAG: Extending Lexicalized TAG to Discourse. *Cognitive Science, 28*(5), 751-779.

Florian Wolf, Edward Gibson. 2005. Representing discourse coherence: a corpus-based study. *Computational Linguistics 31*: 249–87.

Deniz Zeyrek, Bonnie Webber. 2008. A Discourse Resource for Turkish: Annotating Discourse Connectives in the METU Turkish Corpus. *Proc. 6th Workshop on Asian Language Resources, The Third International Joint Conference on Natural Language Processing (IJNLP)*.

Deniz Zeyrek, Işın Demirşahin, Ayşığı Sevdik-Çallı, Hale Ögel Balaban, İhsan Yağmukaya, Ümit Deniz Turan. 2010. The annotation scheme of Turkish discourse bank and an evaluation of inconsistent annotations. *Proc. 4th Linguistic Annotation Workshop (LAW IV)*.