Pair Annotation as a Novel Annotation Procedure: The Case of Turkish Discourse Bank

Işın Demirşahin and Deniz Zeyrek

Abstract

In this chapter, we provide an overview of Turkish Discourse Bank, a resource of ~400,000 words built on a sub-corpus of the 2-million-word METU Turkish Corpus annotated following the principles of Penn Discourse Tree Bank. We first present the annotation framework we adopted, explaining how it differs from the annotation of the original language, English. Then we focus on a novel annotation procedure that we have devised and named pair annotation after pair programming. We discuss the advantages it has offered as well as its potential drawbacks.

1 An Overview of the 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 [21]). It is intended to be a resource for language teachers, linguists, and NLP researchers to help reveal aspects of Turkish discourse. The annotations it contains may be of particular use for various NLP applications such as discourse parsing, data mining, and auto summarization tasks.

I. Demirşahin · D. Zeyrek (✉)
Middle East Technical University, Ankara, Turkey
e-mail: e128500@metu.edu.tr

It may also allow researchers to examine the structures sanctioned by the annotations to reach generalizations about the structure of Turkish discourse.

TDB includes published texts from 1990 to 2000 covering different genres (novels, stories, research articles, essays, travel, interviews, diaries and memoires, news from several different newspapers) with at most two samples from one source. Each sample contains ∼2000 words. TDB uses the MTC files as source texts, keeping the original genre distribution of the texts. It creates annotations in the style of Penn Discourse Tree Bank (PDTB) [20], treating discourse connectives as discourse level predicates that take as argument two text spans that can be interpreted as abstract objects (facts, events, situations, propositions, etc., as in Asher [3]). In TDB 1.0, explicit discourse connectives and a set of phrasal expressions are annotated with their two arguments, modifiers, and supplementary materials as well as shared elements, amounting to 8483 annotations on 197 files.[1] Work on implicit connectives and senses have been started; annotation of attribution is left for future research.[2]

An important issue before starting to build the corpus was how to identify an initial set of discourse connectives. We observed that just like English and many other languages, in Turkish, discourse relations are signaled by discourse connectives belonging to major syntactic classes; therefore, an initial set of discourse connectives was determined by examining the following syntactic classes:

- Conjunctions

  - coordinating conjunctions, e.g. *ve* 'and', *ama* 'but, yet', *ya da* 'or'
  - other conjoining devices, e.g. *çünkü* 'because'[3]

- Subordinators

  - Complex subordinators: two-part subordinators (a postposition accompanied with suffixes on the nominalized verb):

---

[1]The project website is at http://medid.ii.metu.edu.tr/. The corpus is freely available to researchers. Due to copyright agreements with the publishers, the content of the texts from the MTC cannot be redistributed in any commercial products.

[2]One of our reviewers suggests that we speculate on what percentage of PDTB-style discourse relations are covered by annotating explicit connectives, their arguments and supplementary materials. However, without annotating a substantial portion of TDB for implicit connectives, it is very difficult to make a speculation. Also, the ratio might change according to the genre and this makes a speculation even more difficult.

[3]Csató and Johanson [6] classify *çünkü* 'because' as a conjoining device on the basis of examples as *Ali gelemiyor çünkü çalışıyor* 'Ali is not coming because he is working', which cannot be subordinated as the complement of verbs such as *bilmek* 'know': *[Ali gelemiyor çünkü çalıştığını] biliyorum*. 'I know [that Ali is not coming because he is working].' This is possible for coordinated structures, e.g. *[Ali'nin geldiğini ve çalıştığını] biliyorum*. 'I know that Ali came and worked'. This supports our categorization of *çünkü* and various coordinating conjunctions under a single category.

(1) –DIğ-I *için*
   - NOM- ACC için
   'since (causal)'

(2) –mA-sı-nA rağmen
   -NOM- AGR- DAT
   'despite'

- Simplex subordinators, e.g. the suffixes *–ken* 'while', *-cAğInA* 'rather than'[4]

• Discourse adverbials, e.g. *ayrıca* 'in addition', *tersin*e 'on the contrary'

Typically, the coordinating conjunctions as well as subordinators are intra-sentential. They show an affinity with their Arg2, evidenced in part through their ability to move to the end of Arg2 (example 3) and by the use of the comma (example 4) [29]. In the examples throughout this chapter, Arg1 is shown in italics, Arg2 is boldfaced. The connective is underlined and the supplementary material is rendered between square brackets.

(3) [Kitaplarımı yaktım, biliyor musun]? … *Buna şaşmayacaksınız,* **yeni bir şey değil** çünkü.
   [I burned my books, did you know]? ... *you won't be surprised,* because **it's not something new.**

While the arguments of coordinating conjunctions normally have the Arg1-Arg2 order, the usual order of arguments to subordinators is Arg2-Arg1. The second argument to a subordinator may be transposed, yielding a sentence-final subordinator, as in example (4).

(4) *Kimi zaman bir bitki gibi durmak gerekebilir*, **hayatın olanaklarını daha iyi fark edebilmek** için.
   Sometimes it might be necessary to live like a plant in order **to be able take better notice of the opportunities in life**.

The subordinator class, particularly the simplex subordinators, would be difficult to annotate without morphologically parsed data (which was unavailable at the time); therefore, we left them out of the scope of TDB 1.0 and formed a preliminary list of connectives on the basis of the remaining classes. Once a list was formed, annotation

---

[4]Simplex subordinators, and the dependent part of the complex subordinators have morphological variants due to the vowel and consonant harmony rules of Turkish. Briefly, vowel harmony works incrementally in a word, affecting all of the vowels in the root as well as the suffixes. Consonant harmony is an assimilatory process affecting, for example, the consonants at the boundary of a root and suffix. The capitalization we use represents a harmonizing vowel or consonant.

exercises were performed, where the connective, its two arguments and supplementing material were annotated (see Sects. 2.1 and 2.2 below). The annotation exercises led to more categories, e.g. phrasal expressions and the material shared by both arguments.

The rest of the chapter is organized as follows: In Sect. 2, we introduce the annotation scheme and discuss the major divergences from PDTB. Section 3 explains the annotation process with information about the annotators and introduces the annotation environment. Section 4 presents the pair annotation procedure along with its observed benefits and possible drawbacks. In Sect. 5 we summarize the chapter and draw some conclusions.

2 Annotation Scheme of TDB: Major Differences from PDTB

In Table 1, we present the annotation categories used in TDB 1.0. In the rest of the chapter, the term annotation refers to the procedure of identifying the discourse use of connectives on the basis of the abstract object criterion and manually marking the categories in Table 1.

2.1 The Supp Tag

Turkish is a null subject language with word order variation, where all six orders are attested. For example, unlike English, in Turkish, only a deictic expression can be linked anaphorically to a clause (example 5). Neither the pro nor the third person pronoun has this potential (example 6) [25]. TDB aims to capture the anaphoric link between a deictic expression in a discourse relation and the clause outside the relation by means of the Supp tag (see example 7 below).

Table 1 Annotation scheme of TDB [32]

Conn	The connective head
Arg1	First argument of the connective
Arg2	Second argument of the connective
Supp1	Supplement to the first argument
Supp2	Supplement to the second argument
Shared	The subject, object or adverbial phrase *shared* by a relation
Shared supp	Supplement for the *shared* material
Mod	Modifier of the connective or the modifier of the relation

(5) Eğer geç kalırsan, bu/bu durum anneni endişelendirir.
     If you stay late, that/that situation will worry your mother.

                                                            (adapted from Turan, 1995:25)

(6) *[Eğer geç kalırsan]$_i$ o$_i$ /pro$_i$ anneni endişelendirir.   (Turan, 1995:25)
     If you stay late, it will worry your mother.

TDB also uses the Supp tag to specify the material that is needed to support the comprehension of a discourse relation though it is not minimally necessary for interpreting it (as in PDTB). Example (7) shows the use of the Supp1 tag to show where the deictic expression *bu* 'that' in Arg2 is resolved in the previous discourse.

(7) [Milliyet'i arayan Arınç, "Yanlışlara ortak olmam" haberindeki ifadenin rahatsızlık yarattığını, Erdoğan'ın Danimarka'dan arayarak tepki gösterdiğini belirtti]. … Deşifre metnini dinleyen Arınç, "*Bunu demiş olabilirim*, <u>ama</u> **kastım bu değildi**. ….." dedi.
     [Arınç, who called Milliyet, said that the sentence in the news report "I won't take part in the wrongdoings" had caused disturbance, and that Erdoğan had called them form Denmark to express his own reaction]. … Arınç, who listened to the recording, said "*I might have said that*, <u>but</u> **I didn't mean it**. ..."

2.2 The Shared Tag

While Turkish has SOV as the basic word order [5], it allows word order variations, which is largely sensitive to discourse-related facts [11,15,23,25]. This variability of word order often causes difficulties for the annotators in identifying the shortest text span as an argument to a discourse connective. We introduced the *shared* tag to mark the text pieces that belong to both arguments, e.g., the locative or temporal adverbial expressions (example 8) as well as subjects and objects (example 9). This tag mainly assists the annotators to produce annotations that are maximally free of span length errors, though further analysis of the shared tag is hoped to reveal new facts of Turkish discourse, e.g. the role of discourse-initial adverbs as in (9).[5] In the example, the shared element is shown between wavy brackets.

(8) {İnsanların da hayvanların da tok olduğu o zengin, bakımlı, temiz ülkelerde} *açlık yoktu*, <u>ama</u> **özgürlük de yoktu**.

---

[5]In example (9), the temporal adverbial is used discourse-initially and scopes over the whole relation. This is very similar to Asher et al. [4] who argue that locative sentence adverbials have a topic framing role due to their forward-looking character. Asher uses such examples from French to discuss a specific kind of backgrounding relations, i.e. Background$_{forward}$ within the framework of SDRT. Further research will identify the role of adverbials marked as *shared* material in TDB and their contribution to discourse interpretation.

{In those rich, well-kept, clean countries where both the people and the animals were well-fed}, *there was no hunger*, <u>but</u> **there was no freedom either.**

(9) "<u>Sonra</u> *birden kesilirdi* {*yağmur*} <u>ama</u> **kesilene dek filmin sesi duyulmazdı**"
Then *suddenly it would stop,* {*the rain*} **<u>but</u> the sound track of the movie wouldn't be heard until [the rain] ended.**

2.3 Phrasal Expressions

TDB annotates *phrasal expressions*, e.g. *buna rağmen* 'despite this', *bunun için* 'for this', etc. to the extent they constitute a postposition and a deictic expression. Our phrasal expressions correspond to a type of alternative lexicalizations (AltLex) in PDTB [19]. In creating TDB, we search explicit discourse connectives by what we call a search token and annotate the retrieved connectives in the whole corpus. A single search token, e.g. a postposition (i.e. a complex subordinator) such as *rağmen 'despite'* and *için* 'for' conveniently retrieves both the discourse and non-discourse uses as well as any phrasal expressions based on this postposition (cf. Sect. 3.2). Hence it is quite convenient to annotate phrasal expressions while annotating subordinator connectives. The deictic elements of phrasal expressions have a clausal antecedent and can be replaced with a nominalized clause (but never with a noun); the phrasal expression itself can be used both intra- and intersententially (examples 10 and 11, respectively); sentence-final uses are not attested in TDB.

(10) *Tabii, eroinin alındığı günlerin sayısı da artmaya başlar ve eroin kullanımı günlük hale gelir.* <u>Buna rağmen</u> **bağımlı hala eroine bağımlı olmadığını, istediği an bırakabileceğini sanır.**
Of course, the number of days when heroin is injected also increases, and heroin use becomes a daily habit. <u>Despite this</u>, **the addict still thinks he is not addicted to heroin and could quit anytime he wants.**

(11) *Komutanlar, savaş sırasında 250 bin kişinin yerlerinden olabileceğinin hesaplandığını,* <u>bunun için</u> **18 kampın 36-37. paralellerde kurulacağını** bildirdi.
The commanders announced that *during the war, 250 thousaund people could lose their homes,* <u>for this reason</u> **18 camps would be built between the 36**[th] **and 37**[th] **parallels**.

Phrasal expressions will be categorized together with alternative lexicalizations by post-processing once other types of the AltLex class have been identified.

In Appendix 1, we present the search tokens, the number of files searched and the discourse connectives as well as phrasal expressions annotated. Table 2 provides the frequencies of explicit discourse connectives and phrasal expressions annotated in TDB 1.0.

Table 2 Absolute and relative frequencies of explicit discourse connectives and phrasal expressions in TDB

Syntactic Class	# of relations in TDB	% of relations in TDB (%)
Coordinators	4477	52.78
Subordinators	2287	26.96
Discourse adverbials	1225	14.44
Phrasal expressions	494	5.82
Total	**8483**	**100**

3 Annotation Process

The TDB 1.0 annotations were created manually by means of three different annotation procedures: independent annotation (IA), group annotation (GA) and pair annotation (PA). Regardless of the annotation procedure, the annotators are asked to obey the minimality principle, i.e. they have to select as arguments the minimal textual span necessary to interpret the discourse relation [18]. The minimality principle ensures that the annotators focus on the local text while annotating a particular discourse connective without having to consider the overall structure of the text.[6] All the annotations are adjudicated in periodical agreement meetings with the leadership of at least one senior researcher. The leader helps the annotators to resolve the differences (if any) and the team produces an agreed version of the annotations unanimously.

In the IA procedure, the data is triply-annotated blindly; i.e. three annotators annotate the data without seeing the others' annotations, and the other search tokens previously annotated on the file. In the GA procedure, the annotators gather to produce a single set of annotations for a search token, noting any disagreements to be discussed in a subsequent agreement meeting. The GA procedure was particularly used for annotating connectives that were too few in number. In the PA procedure, a pair of annotators produces a single set of annotations, which is blind to a third annotator's annotations. The PA process, inspired by Pair Programming, is a novel annotation approach developed during the TDB project. Section 3 below explains this procedure in more detail. Of the total 8483 annotations in TDB 1.0, 3804 (44.84%) discourse relations were annotated by the IA procedure, 3985 (46.98%) by PA, and 694 (8.18%) were annotated by GA [32].

---

[6]The minimality principle may sometimes lead to disagreements among annotators, as discussed in Zeyrek et al. [30].

3.1 Annotators and the Pilot Phase

Three graduate students (of Middle East Technical University Cognitive Science Department) were involved in the creation of TDB as annotators and researchers. In the pilot phase, the annotators were trained theoretically in reading groups. As the annotation tool was being developed (see Sect. 3.2), early annotation exercises were conducted on word processors. These exercises included multiple independent annotations by the annotators and the senior researchers involved in the project. The resulting annotations were compared manually, and disagreements were resolved in weekly discussions. The result was an initial set of annotation guidelines.

The method of annotation in the pilot phase and the later stages was as follows: the annotators were given a specific connective from the pre-determined list of connectives. They went through the files in the corpus, identifying and manually annotating the discourse uses of the connective, leaving the non-discourse uses unmarked. They were asked to follow the annotation guidelines but were also encouraged to reflect their native speaker intuitions on the annotations. With the annotators' constant feedback, the initial guidelines were updated through several iterations. The list of connectives was also updated as the annotators informed the research team about connectives not in the original list.

3.2 DATT: Discourse Annotation Tool for Turkish

TDB is annotated using DATT, the Discourse Annotation Tool for Turkish [1]. DATT is an XML-based infrastructure created specifically for the TDB project.

DATT takes a folder of text files and indexes the files for character offsets. The user interface lets the annotators search the tokens either by basic word search or regular expressions. The regular expression search is meant to facilitate finding the morphological variants of a discourse connective (e.g. *dolayı* 'owing to', *dolayısıyla* 'in consequence of' and *dolayısı ile* 'in consequence of')[7] and limit the search space for high frequency discourse connectives. For example, the postposition *gibi* 'as' occurs 1265 times in TDB. However, the majority of these occurrences should not be annotated. Since the source data was not POS tagged, regular expressions could not filter out the cases that accidentally matched the search pattern. Still, they allowed the annotators to sort out most of the irrelevant occurrences. Regarding *gibi* 'as', the regular expression search returns 455 instances, of which 228 were annotated as discourse connectives.

The regular expression search has a specific feature to accommodate the vowel and consonant harmony in Turkish (see footnote). For example, to capture the four variants of the simplex subordinator equivalent to 'because of' *(-den, -dan, -ten, -tan)*, the annotators can make a search with a simple *–DAn* instead of *–*[d|t][a|e]n.

---

[7]The connective devices *dolayısı-yla* and *dolayısı ile* are different forms with the same meaning. The first word contains the suffix *–yla*, which is semantically equivalent to the clitic *ile* 'with'.
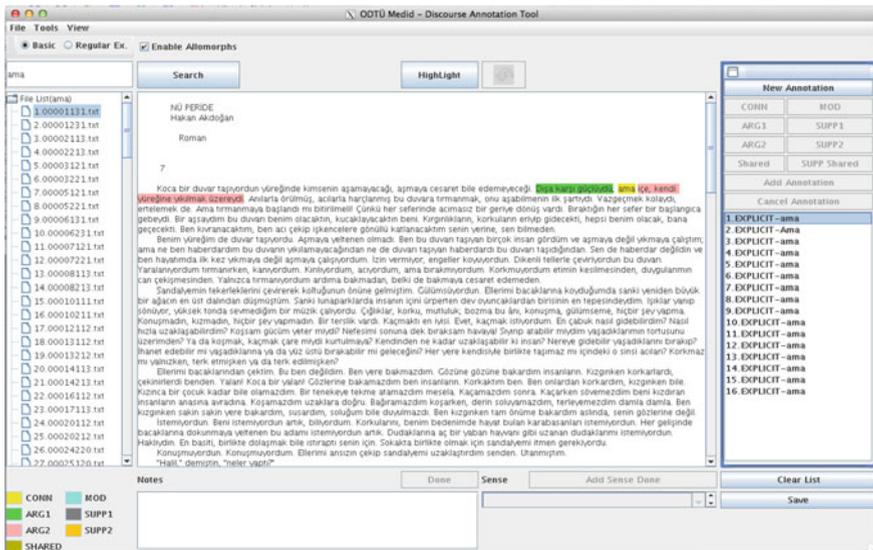
Fig. 1 A screenshot of DATT (Discourse annotation tool for Turkish)

To capture the eight variants of the complex subordinator –*DIğI için* 'since (causal)', they input –*DHğH için* instead of [d|t][ı|i|u|ü]ğ[ı|i|u|ü] *için*.

All the instances of the search token are highlighted in the text in the annotation tool. For each explicit relation, the discourse connective and its two argument spans must be annotated. In addition to these mandatory text spans, annotators can select modifiers, shared elements and supplementary materials where needed. DATT supports discontinuous text spans to be selected as part of the same argument. Each discourse relation can be further enriched with notes, which are free texts entered by annotators.

The annotations are represented as XML trees. A sample XML representation for the discourse relation in (12) is provided in (13).

(12) *İnsanlar tabiattan eşit doğarlar*. Dolayısıyla **özgür ve köle ayrılığı olma-malıdır**.
People are born equal by nature. Consequently, **there should be no such distinction as the freeman and the slave.**

(13)
```
<Relation note="" type="EXPLICIT">
  <Conn>
    <Span>
      <Text>dolayısıyla</Text>
      <BeginOffset>15624</BeginOffset>
```

```
            <EndOffset>15635</EndOffset>
          </Span>
      </Conn>
      <Arg1>
        <Span>
            <Text>İnsanlar tabiattan eşit doğarlar</Text>
            <BeginOffset>15590</BeginOffset>
            <EndOffset>15622</EndOffset>
        </Span>
      </Arg1>
      <Arg2>
        <Span>
            <Text>özgür ve köle ayrılığı olmamalıdır</Text>
            <BeginOffset>15636</BeginOffset>
            <EndOffset>15670</EndOffset>
        </Span>
      </Arg2>
    </Relation>
```

Whereas XML strictly enforces tree-structures in the data, stand-off annotations create a separate file for annotations and preserve the source data as is. However, stand-off annotations are highly vulnerable to changes in the source data, because if the changes in the source data are not reflected in the annotation files, the source and the annotations will be misaligned. As a precaution, the annotation files keep the content of the annotated spans as well as the start and end character offsets.

The annotations that belong to a raw text file are saved in an XML file with the same name as the raw text file; the annotations for the search token are saved in a folder named after the search token. This makes it easier to go over and edit all the annotations for a search token.[8] The physical appearance of the annotation tool is provided in Fig. 1.

4 Pair Annotation

When the inter-annotator reliability among three (independent) annotators stabilized, a new procedure was proposed, namely the use of a pair of annotators to carry out the task together. We call the procedure Pair Annotation after the pair programming (PP) procedure in software engineering [9]. In order to eliminate the risk of getting

---

[8]We are aware that this results in multiple annotation files for one raw text file. The next version of TDB is planned to include all the annotations for a raw text in the same XML file sorted by the character offset of the connective. This will result in fewer annotation files and allow easier processing [8].

high agreements too early in the process, we first carried out individual blind annotations on one thirds of the files of the high frequency connectives. During this phase we determined the connective-specific dynamics and updated the guidelines where necessary. Only then we proceeded to pair annotation for the remaining two thirds of the files containing that particular connective.

PP is a collaborative programming paradigm where two programmers work on an algorithm or a piece of code as a unit, assuming equal responsibility and credit for the work done [27, 28]. The unit is composed of two roles, the driver and the navigator. The driver is the one who is physically creating the code or algorithm, whereas the navigator is the one who monitors the driver. The monitoring is an active process: the navigator is expected to be involved in the creation of the code at all times by watching for errors, suggesting alternatives and supplementing the driver with additional resources when necessary. The pair periodically switches the roles of the driver and the navigator. Maintaining active involvement of the navigator and changing roles regularly ensures that the pieces of code created via PP not only belong to the programmer who was the driver at the time, but the pair as a unit; i.e. the result is a joint ownership.

The PA annotation procedure emerged out of the need to accelerate the annotation process. It was proposed by two of the annotators quite independently of PP, and its principles emerged in a short time on their own accord. In quite a spontaneous way, one of the annotators came to annotate the data while the other annotator checked, corrected or otherwise simply agreed with the first annotator's annotation. Therefore, the roles of the driver and the navigator used in the PP literature arose. The PA, then, is the procedure where one of the annotators assumes the driver role physically handling the keyboard and the mouse with the other annotator sitting next to her, looking at the screen and working together with her as a navigator as in PP (Fig. 1). The driver and navigator roles are occasionally switched between the annotators, as in PP. To assess the reliability of pair-annotations, we always compare them with the annotations produced by a third, independent annotator (Fig. 2).

4.1 Observed Benefits of Pair Annotation

We observed that in the PA procedure, physical errors, e.g. erroneously leaving a few letters of a word unmarked, or selecting spaces at the peripheries of the arguments are more easily noticed and corrected: the navigator readily sees such mistakes and warns the driver who then corrects them immediately. A related benefit is that the annotation of ambiguous cases can be handled more efficiently because the pair can easily resolve the ambiguity by discussing the options among them. The end result of this collaborative task is fewer disagreements in the annotations.

We also noticed that the annotators have higher motivation during the PA procedure, as mentioned in the PP literature. During PA, the annotators are quite focused on the task and can easily resist being sidetracked since they do not want to waste each other's time. In our case, annotating numerous instances of the same connective is often monotonous. The pair of annotators uses the advantage of having a partner to
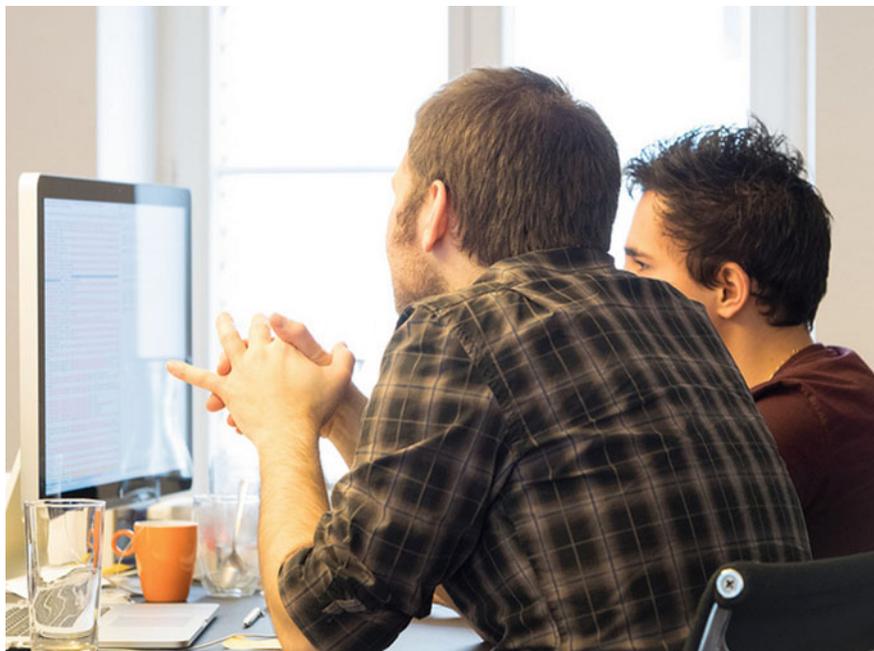
Fig. 2 Pair programming by Guido Gloor Modjib (https://www.flickr.com/photos/glodjib/ 16146549307/) is licensed under CC BY 2.0 (http://creativecommons.org/licenses/by/2.0/)

collaborate, discuss, and occasionally joke to lighten up the mood. Thus, the task that is tiresome when carried out alone becomes interactive and pleasant when carried out with a partner.

Thirdly, the PA can be timesaving because the pair is well prepared for the discussion of the hard cases in the agreement meetings. The pair annotators share the results of their discussions with the research team (through the notes field of the annotation tool) and offer their solution resulting from in-depth discussions and careful thinking. In hard cases, the pair annotators were particularly careful in recording their first intuitions and their reasoning process in producing the joint annotation; sometimes they even declared an unresolved difference of opinion. These comments were highly beneficial for the research team as they provided more insight about the reasoning behind the annotation itself, thus accelerating the agreement meetings (also see Sect. 4.2).

4.2   Possible Disadvantages of Pair Annotation

Just as PP is criticized, questions may arise against PA. One of the most prominent objections is the increased man-hours. In the IA procedure, three annotators produce three sets of annotations, whereas in the PA procedure, three annotators produce two

sets of annotations; it is as if PA increases the cost of a set of annotations by 50%. Yet, the benefits are high because the PA procedure increases the annotation pace of the pair and improves inter-annotator agreement.

Another concern is the possibility of losing the input of one of the annotators, most likely those of the navigator. This can take place in several ways. For example, the navigator may lose interest and watch passively as the driver annotates, or the driver may take control over the whole annotation and ignore the input from the navigator. The TDB team was an already well-established research group before the inception of PA, and the annotators had intrinsic and extrinsic motivations to produce a high quality corpus in a limited time; hence these issues did not arise. In other projects where annotators are not a part of the research team or their involvement is limited to annotations only, they might be inclined to overlook the principles of PA. If such cases arise, it would be advisable to incorporate peer evaluation to get periodic feedback and ensure that the procedure is working as intended.

These concerns are common to PP and PA, but issues specific to annotation projects may also arise. In annotation projects it may be desirable to involve several annotators to annotate the same text files so as to capture the intuitions of many native speakers. PA may appear as if a limited range of native speaker intuitions is captured. It may also be argued that the constant interaction between the pair may contaminate their own intuitions.

To avoid both criticisms, we used the notes field in DATT to record the pair annotators' initial intuitions, particularly in cases where one of the members of the pair felt that the pair annotation did not reflect her own intuitions. The discussions that occurred during PA as well as other procedures are retained. Table 3 provides the number of relations, notes, and the number of notes per 100 relations for all the procedures, which reveals that the majority of the notes have been recorded during pair annotation. According to Table 3, a total of 1398 notes were recorded. Only 15 of these notes were produced during the GA procedure for 697 relations. A total of 512 notes were recorded by the 3 annotators involved in the IA procedure for 3018 relations, and 871 notes were recorded by the pair and the independent annotator for 4145 relations during the PA procedure. The pair recorded 705 notes. The high number of notes per relation in the PA procedure indicates that the individual opinions of the members of the pair (as well as the pair's common opinion) did not go unnoticed; any disagreements were recorded so that they are discussed in agreement meetings.

We do not claim that PA is the solution to all problems in annotation, or that it offers the perfect annotation procedure. That is why we suggest keeping an independent individual annotator in the process. As such, this procedure is akin to having two independent annotators, where one of the annotators is like a composite consisting of two individuals thinking independently but producing a single set of annotations collaboratively. Similar to the joint ownership of PP, neither annotator claims the annotation as her own. It is treated as a single set of annotations both during the agreement meetings and in calculating the agreement statistics.

Table 3 Number of relations annotated, number of notes recorded, and number of notes per 100 relations for GA, IA and PA

Annotation procedure	# of relations annotated	# of notes recorded	# of notes/100 relations
GA	697	15	2.15
IA			
Ann1	3018	172	5.7
Ann2	3018	184	6.1
Ann3	3018	156	5.17
Average		170.67	5.65
PA			
Pair	4145	705	**17.01**
Individual	4145	166	4.00

Table 4 Pair-wise averaged inter-annotator agreement (K) for 3 individual annotators in the IA phase

Connective	Arg1	Arg2
ama 'but'	0.832	0.901
sonra 'after'	0.820	0.902
ve 'and'	0.692	0.791
ya da 'or'	0.843	0.974

4.3 Evaluation Exercise

We carried out an evaluation exercise on four connectives annotated both by the IA and PA procedures and six connectives annotated only by the PA procedure [9]. The four discourse connectives annotated by means of two annotation procedures were: *ama* 'but', *sonra* 'after', *ve* 'and' and *ya da* 'or'. The first 1/3 of all files in the data were annotated via the IA procedure, the rest of the files were annotated via the PA procedure. The six connectives annotated only by the PA procedure were: *aslında* 'actually', *halde* 'in spite of', *nedeniyle* 'for the reason that', *nedenle* 'for this reason', *ötürü* 'due to' and *yüzden* 'since (causal)'.

Table 4 provides the averaged pair-wise Fleiss' Kappa (K) [12] agreement coefficient values of the IA phase for the first group of connectives.

Table 5 shows the K values of the PA phase for the same group of connectives.

In Tables 4 and 5, all the cells but one indicate good agreement ($0.80 < K < 1.00$). Only the first argument of *ve* 'and' in the IA phase shows some agreement ($0.60 < K < 0.80$).

The inter-annotator agreement statistics of the annotations of two phases show that the K values for both arguments have increased after the transition from the IA

Table 5 Inter-annotator agreement (K) for pair versus individual in the PA phase (of the connectives in Table 1)

Connective	Arg1	Arg2
ama 'but'	0.956	0.969
sonra 'after'	0.889	0.953
ve 'and'	0.945	0.964
ya da 'or'	0.939	0.973

Table 6 Individual annotator versus agreed agreement (K) in PA

Connective	Arg1	Arg2
aslında 'actually'	0.766	0.889
halde 'in spite of'	0.834	0.898
nedeniyle 'for the reason that'	0.905	0.984
nedenle 'for this reason'	0.952	0.987
ötürü 'due to'	1.000	0.907
yüzden 'since (causal)'	0.916	0.983

Table 7 Pair annotator versus agreed agreement (K) in PA

Connective	Arg1	Arg2
aslında 'actually'	0.937	0.984
halde 'in spite of'	0.973	1.000
nedeniyle 'for the reason that'	0.937	0.984
nedenle 'for this reason'	1.000	1.000
ötürü 'due to'	1.000	0.953
yüzden 'since (causal)'	0.992	1.000

procedure to the PA procedure. A repeated measures test shows that the increase is significant ($p < 0.01$).

Tables 6 and 7 show the agreement statistics for the second group of connectives, where only the PA annotation was conducted. Each set of annotations is compared to the agreed annotations that were produced after the final agreement meeting for that particular connective. In Table 6, the K values show the agreement between the individual annotations and the agreed annotations, and in Table 7, they indicate the agreement between the pair's annotations and the agreed annotations.

Except for the 0.766 value for Arg1 of *aslında* 'actually' in Table 6, all K values indicate good agreement.[9] A repeated measures test shows that the agreement of the annotator pair and the agreed annotations are significantly higher than the agreement of the individual annotator and the agreed annotations ($p < 0.001$). *Aslında* is a discourse adverbial, whereas the rest of the connectives in Tables 6 and 7 are complex subordinators. Unsurprisingly, identifying the Arg1 to discourse adverbials creates problems for annotators. We attribute this to the fact that discourse adverbials take their Arg1 anaphorically, a problem also noted by the PDTB group [17]. The difficulty of reaching a perfect agreement on Arg1 of *aslında* withstanding, our evaluation exercise shows that PA yields both higher inter-annotator agreement and annotator-agreed agreement.

## 5   A Study on Turkish Discourse Structure and Conclusions

We conclude this chapter by summarizing a study on TDB 1.0 which investigate the structures in discourse. The TDB research group assumes that discourse structure is hierarchical but it is constructed and processed incrementally, an idea borrowed from Grosz and Sidner [14]. As in PDTB, rather than imposing a hierarchy on discourse structure, we ask the annotators to annotate discourse connectives together with their modifiers, arguments and supplementary material locally. Annotations created in this way can shed light on the structural aspects of discourse in later analysis and show the interaction of discourse structures with other phenomena, such as information structure.

Lee et al. [16] analyze PDTB for the cases where the shared discourse pieces are subordinate clauses introduced by explicit subordinating conjunctions (e.g. *although*). The study reveals the existence of tree-conforming structures (e.g. fully embedded relations) as well as tree-violating structures such as shared arguments, properly contained arguments, pure crossing, and partially overlapping arguments. Lee et al. argue that all tree violations but the shared arguments can be explained away through non-structural elements of discourse such as anaphora and attribution. Aktaş et al. [1] analyze Turkish with respect to the shared discourse structures without limiting them to particular syntactic constructions. They find that Turkish discourse displays all these configurations; in addition, they discover nested relations (which conform with the tree structure) and properly contained relations (which are tree-violating). Demirşahin et al. [10] expand on Aktaş's study and reveal that one of the crossing examples between relations in Turkish discourse is surface crossing which results from wrapping. In Turkish, wrapping is an operation motivated by information structure where adverbial clauses introduced by complex subordinators

---

[9]Artstein and Poesio [2] suggest 0.8 as a good cut-off point for reasonable quality. On the other hand, Spooren and Degan [22] suggest reaching a minimal value of 0.7 in annotating discourse coherence. In this paper, we take 0.8 as the cut-off point.

(e.g. *için* 'for') can move freely in the sentence and can land right before the matrix verb, which is an information structurally prominent position [13]. In TDB 1.0, wrapping occurs 479 times in total. An example is provided below in (14) followed by a diagram representing the associated discourse structure.

(14) *1882'de İstanbul Ticaret Odası*, **bir zahire ve ticaret borsası kurulması** için *girişimde bulunuyor* ama sonuç alamıyor.
I*n 1882, Istanbul Chamber of Commerce makes an attempt* for **founding a Provisions and Commodity Exchange Market** but cannot obtain a result.
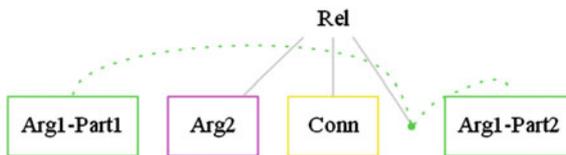
Wrapping structures have applicative semantics, which utilizes function application but not function composition. Although they result in surface-crossing at the discourse level, computationally they are not more complex than tree-structures, as they are not the product of function composition. (Function application is the only operation required to derive the semantics of wrapping.) Demirşahin et al.'s [10] finding draws attention to the interaction of an information structure-motivated syntactic phenomenon with discourse connectives (particularly the complex subordinator connectives) and it is a promising result for further research on aspects of Turkish interacting with discourse structure.

To conclude, in this chapter we presented Turkish Discourse Bank 1.0, a discourse resource annotated with the principles of PDTB, where discourse connectives are taken as predicates with two arguments. We explained the core differences of TDB from PDTB and introduced the discourse annotation tool specifically designed for this project. We then offered a novel annotation procedure we named pair annotation after pair programming. This is the procedure where two annotators team up to create a single set of annotations. The pair's annotations are treated as a single set of annotations and compared with the annotations of an independent annotator to assess reliability. We presented the observed benefits and possible drawbacks of the PA procedure as well as an evaluation exercise that compares the PA procedure with the IA procedure. We concluded the chapter with a study on TDB 1.0 investigating possible discourse structures allowed by the annotations.

Discourse presents many challenges for linguists as well as language technology; in the future, we plan to enrich TDB with more annotations to allow the use of this resource more effectively. Ultimately, analyses of the annotations on TDB could lead to cross-linguistic comparisons and a better understanding of discourse-level properties.

Appendix 1: The Number of Files and Annotations in TDB 1.0

See Table 8.

Table 8 List of search tokens for TDB sorted by their occurrence as discourse connectives

Search token	# of files	# of annotations
ve 'and'	185	2112
için 'for/because'	177	1103
ama 'but'	170	1024
sonra 'after/then'	179	713
ancak 'however'	114	419
çünkü 'because'	125	300
gibi 'as'	106	228
kadar 'until'	94	159
zaman 'when'	90	159
ya da 'or'	81	139
oysa 'conversely'	74	136
önce 'before'	84	134
nedenle 'for (this) reason'	73	117
ayrıca 'besides'	66	108
böylece 'thus'	54	85
hem 'as well as'	57	82
aslında 'actually'	52	81
fakat 'however'	49	80
rağmen 'despite'	51	77
yoksa 'otherwise'	52	75
karşın 'regardless of'	57	71
ardından 'after'	49	71
yandan 'on (one) hand'	53	70
yüzden 'since (causal)'	47	66
dolayısıyla 'consequently'	46	66
yine de 'still'	40	65
amacıyla 'with the intention of'	48	64

(continued)

Table 8 (continued)

Search token	# of files	# of annotations
örneğin 'for example'	40	64
halde 'in spite of'	48	61
ne 'neither… nor'	35	44
nedeniyle 'for the reason that'	34	42
zamanda 'at (that) time'	27	39
veya 'or'	28	38
birlikte 'together/though'	30	33
ne var ki 'nevertheless'	26	32
karşılık 'although'	21	28
gene de 'still'	13	26
iken 'while'	16	22
dolayı 'owing to'	16	21
halbuki 'whereas'	13	17
ne ki 'nonetheless'	7	14
aksine 'on the contrary'	12	13
mesela 'for instance'	11	13
yalnız 'however/only that'	12	12
sonucunda 'as a result of'	10	12
amaçla 'for (this) purpose'	11	11
tersine 'inversely'	10	11
ötürü 'due to'	4	11
bu yana 'since (temporal)'	10	10
sonuçta 'as a result'	10	10
dahası 'moreover'	7	10
ya '(either)… or'	7	8
beraber 'nontwithstanding'	6	6
ister 'whether..or'	5	6
sözgelimi 'for instance'	3	6
sonuç olarak 'as a result'	5	5
yüzünden 'since (causal)'	5	5
sayede 'thanks to'	4	5
beri 'since (temporal)'	4	4
içindir 'because of'	4	4
nedenlerle 'for these reasons'	4	4
veyahut 'or'	1	4
nedeni ile 'for the reason that'	3	3
sayesinde 'thanks to'	3	3

Table 8 (continued)

Search token	# of files	# of annotations
taraftan 'on (one) hand'	3	3
yahut 'or'	2	3
fekat 'however'	1	3
gerek 'whether… or'	2	2
ha 'whether… or'	2	2
örnek olarak 'for example'	2	2
amacı ile 'with the aim of'	1	1
dolayısı ile 'consequently'	1	1
ek olarak 'in addition'	1	1
neticede 'as a result'	1	1
neticesinde 'as a result of'	1	1
sebeple 'therefore'	1	1
söz gelimi 'for instance'	1	1
Total	**2797**	**8483**

References

1. Aktaş, B., Bozsahin, C., Zeyrek, D.: Discourse relation configurations in Turkish and an annotation environment. In: Proceedings of the Fourth Linguistic Annotation Workshop, pp. 202–206 (2010)
2. Artstein, R., Poesio, M.: Inter-coder agreement for computational linguistics. Comput. Linguist. **34**(4), 555–596 (2008)
3. Asher, N.: Reference to Abstract Objects in Discourse. Springer, New York (1993)
4. Asher, N., Prévot, L., Vieu, L.: Setting the background in discourse. Discours. Revue de linguistique, psycholinguistique et informatique (2007). http://discours.revues.org/301, 15 February 2015
5. Bozşahin, C.: Word order as projection. Dilbilim Araştırmaları Dergisi/J. Linguist. Res. 1–23 (2014)
6. Csató, É.Á., Lars, J.: Turkish. In: Csató, É.Á., Lars, J. (eds.) The Turkic Languages, pp. 203–235. Routledge, London (1998)
7. Cresswell, C., Forbes, K., Miltsakaki, E., Prasad, R., Joshi, A., Webber, B.: The discourse anaphoric properties of connectives. In: Proceedings of the 4th Discourse Anaphora and Anaphor Resolution Colloquium (DAARC), Lisbon, Portugal, pp. 45–50 (2002)
8. Demirşahin, I., Sevdik-Çallı, A., Ögel Balaban, H., Çakıcı, R., Zeyrek, D.: Turkish discourse bank: ongoing developments. In: Proceedings of LREC 2012 The First Turkic Languages Workshop (2012)
9. Demirşahin, I., Yalçınkaya, I., Zeyrek, D.: Pair annotation: adaption of pair programming to corpus annotation. In: Proceedings of the Sixth Linguistic Annotation Workshop, pp. 31–39 (2012)
10. Demirşahin, I., Öztürel, A., Bozşahin, C., Zeyrek, D.: Applicative structures and immediate discourse in the Turkish discourse bank. In: Proceedings of the Seventh Linguistic Annotation Workshop and Interoperability with Discourse, pp. 122–130 (2013)

11. Enç, M.: The semantics of specificity. Linguist. Inq. **22**, 1–25 (1991)
12. Fleiss, J.L.: Measuring nominal scale agreement among many raters. Psychol. Bull. **76**(5), 378–382 (1971)
13. Göksel, A., Keslake, C.: Turkish: A Comprehensive Grammar. Routledge, London (2005)
14. Grosz, B.J., Sidner, C.L.: Attention, intention and the structure of discourse. Comput. Linguist. **12**(3), 175–204 (1986)
15. Hoffman, B.: The computational analysis of the syntax and interpretation of " free" word order in Turkish. IRCS Technical Reports Series, 130 (1995)
16. Lee, A., Prasad, R., Joshi, A.K., Webber, B.: Departures from tree structures in discourse. In: Proceedings of the Workshop on Constrains in Discourse III (2008)
17. Miltsakaki, E., Prasad, R., Joshi, A., Webber, B.: Annotating discourse connectives and their arguments. In: Proceedings of the HLT/NAACL Workshop on Frontiers in Corpus Annotation, pp. 9–16 (2004)
18. Prasad, R., Webber, B., Joshi, A.: The Penn discourse treebank 2.0. In: Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008), p. 2961 (2008)
19. Prasad, R., Joshi, A., Webber, B.: Realization of discourse relations by other means: alternative lexicalizations. In: Proceedings of the 23rd International Conference on Computational Linguistics: Posters, pp. 1023–1031 (2010)
20. Prasad, R., Webber, B., Joshi, A.: Reflections on the penn discourse treeBank, comparable corpora and complementary annotation. Comput. Linguist. **40**(4), 921–950 (2014)
21. Say, B., Zeyrek, D., Oflazer, K., Özge, U.: Development of a corpus and a treebank for present-day written Turkish. In: Proceedings of the Eleventh International Conference of Turkish Linguistics, pp. 183–192 (2002)
22. Spooren, W., Degand, L.: Coding coherence relations: reliability and validity. Corpus Linguist. Linguist. Theory **6**(2), 241–266 (2010)
23. Taylan, E.E.: The Function of Word Order in Turkish Grammar, vol. 106. University of California Press, Berkeley (1984)
24. Traugott, E.C.: The role of the development of discourse markers in a theory of grammaticalization. In: ICHL XII, Manchester, pp. 1–23 (1995)
25. Turan, Ü.D.: Subject and object in Turkish discourse: a centering analysis. Doctoral dissertation, Ph.D dissertation, University of Pennsylvania (1995)
26. Webber, B.: D-LTAG: extending lexicalized TAG to discourse. Cogn. Sci. **28**(5), 751–779 (2004)
27. Williams, L.A., Kessler, R.R.: All I ever needed to know about pair programming I learned in kindergarten. Commun. ACM, **43**(5), 108–114 (2000)
28. Williams, L.A., Kessler, R.R., Cunningham, W., Jeffries, R.: Strengthening the case for pair programming. IEEE Software **17**(4), 19–25 (2000)
29. Zeyrek, D., Webber, B.: A discourse resource for Turkish: annotating discourse connectives in the METU Turkish corpus. In: Proceedings of the Sixth Workshop on Asian Language Resources, Hyderabad, India, pp. 65–72 (2008)
30. Zeyrek, D., Demirşahin, I., Sevdik-Çallı, A., Ögel, B.H., Yalçınkaya, İ.: The evaluation scheme of the Turkish discourse bank and an evaluation of inconsistent annotations. In: Proceedings of the Fourth Linguistic Annotation Workshop, ACL 2010, Uppsala, Sweden, pp. 282–289 (2010), 15–16 July 2010

31. Zeyrek, D., Turan, Ü.D., Demirşahin, I., Çakıcı, R.: Differential properties of three discourse connectives in Turkish: a corpus-based analysis of Fakat, Yoksa, Ayrıca. In: Benz, A., Kuehlnlein, P., Stede, M. (eds.) Constraints in Discourse III. John Benjamins, Amsterdam (2012)
32. Zeyrek, D., Demirşahin, I., Sevdik-Çallı, A., Çakıcı, R.: Turkish discourse bank: porting a discourse annotation style to a morphologically rich language. Dialogue Discourse **4**(2), 174–184 (2013)