

TDB 1.1: Extensions on Turkish Discourse Bank

Deniz Zeyrek

Middle East Technical University
Ankara, Turkey
dezeyrek@metu.edu.tr

Murathan Kurfali

Middle East Technical University
Ankara, Turkey
kurfali@metu.edu.tr

Abstract

In this paper we present the recent developments on Turkish Discourse Bank (TDB). We first summarize the resource and present an evaluation. Then, we describe TDB 1.1, i.e. enrichments on 10% of the corpus (namely, added senses for explicit discourse connectives and new annotations for implicit relations, entity relations and alternative lexicalizations). We explain the method of annotation and evaluate the data.

1 Introduction

The annotation of linguistic corpora has recently extended its scope from morphological or syntactic tagging to discourse-level annotation. Discourse annotation, however, is known to be highly challenging due to the multiple factors that make up texts (anaphors, discourse relations, topics, etc.). The challenge may become even more heightened depending on the type of text to be annotated, e.g. spoken vs written, or texts belonging to different genres. Yet, discourse-level information is highly important for language technology and it is more so for languages such as Turkish that are relatively less resource-rich when compared to European languages.

Given that systematically and consistently annotated corpora would help advance state-of-the-art discourse-level annotation, this paper aims to describe the methodology of enriching Turkish Discourse Bank, a multi-genre, 400.000-word corpus of written texts containing annotations for discourse relations in the PDTB style. Thus, the motivation of this paper is to contribute to the empirical analysis of Turkish at the level of discourse relations and enable further LT applications on the corpus. The corpus can also be used by linguists, applied linguists and translators interested in Turkish or Turkic languages in general.

The rest of the paper proceeds as follows. §2 provides an overview of Turkish Discourse Bank, summarizes the linguistic decisions underlying the corpus and presents an evaluation of the corpus. §3 introduces TDB 1.1, explains the added annotations and how the data are evaluated. §4 shows the distribution of discourse relation types and presents a preliminary cross-linguistic comparison with similarly annotated corpora. Finally, §5 summarizes the study and draws some conclusions.

2 An Overview of Turkish Discourse Bank (TDB)

The current release of Turkish Discourse Bank, or TDB 1.0 annotates discourse relations, i.e. semantic relations that hold between text segments (expansion, contrast, contingency, etc.). Discourse relations (DRs) may be expressed by explicit devices or may be conveyed implicitly. Explicit discourse connecting devices (*but, because, however*) make a DR explicit. These will be referred to as discourse connectives in this paper. Even when a DR lacks an explicit connective, the sense can be inferred. In these cases, native speakers can add an explicit discourse connective to the text to support their inference. These have been known as implicit (discourse) relations. However, TDB 1.0 only annotates DRs with an explicit connective.

While sharing the goals and annotation principles of PDTB¹, TDB takes the linguistic characteristics of Turkish into account. Here we briefly review some of these characteristics, which have an impact on the annotation decisions (see §2.1 for more principles that guide the annotation procedure).

Turkish belongs to the Altaic language family with the SOV as the dominant word order, though it exhibits all other possible word orders. It is an agglutinating language with rich morphology.

¹ <https://www.seas.upenn.edu/pdtb/>

Two of its characteristics are particularly relevant for this paper. Firstly, it is characterized (a) by clause-final function words, such as postpositions that select a verb with nominalization and/or case suffixes; (b) by simple suffixes attached to the verb stem (termed as converbs). These are referred to as complex and simplex subordinators, respectively (Zeyrek and Webber, 2008). Both types of subordinators largely correspond to subordinating conjunctions in English (see Ex.1 for a complex subordinator *için* ‘for/in order to’ and the accompanying suffixes on the verb, and Ex.2 for a converb, *-yunca* ‘when’, underlined). Only the independent part of the complex subordinators have been annotated so far.

- (1) Gör-me-si için Ankara’ya gel-dik.
see-NOM-ACC to to-Ankara came-we
For him to see her, we came to Ankara.
- (2) Kuru-yunca fırçala-yacağ-ım.
dry-when brush-will-I
I will brush it when it dries.

Secondly, Turkish is a null-subject language; the subject of a tensed clause is null as long as the text continues to talk about the same topic (Ex.3).

- (3) Ali her gün koşar. Sağlıklı yiyecekler yer.
Ali jogs everyday. (He) maintains a healthy diet.

We take postpositions (and converbs) as potential explicit discourse connectives and consider the null subject property of the language as a signal for possible entity relations.

TDB adopts PDTB’s lexical approach to discourse as an annotation principle, which means that all discourse relations are grounded on a lexical element (Prasad et al., 2014). The lexically grounded approach applies not only to explicitly marked discourse relations but also to implicit ones; i.e., it necessitates annotating implicit DRs by supplying an explicit connective that would make the sense of the DR explicit, as in Ex.4.

- (4) ... bu çocuğun sınırsız bir düş gücü var.
[IMP=bu yüzden] Sen bunu okulundan mahrum etme.
... the child has a vivid imagination.
[IMP=for this reason] Don’t stop him from going to school.

2.1 Principles that Guide Annotation

In TDB 1.0, explicit discourse connectives (DCs) are selected from three major lexical classes. This is motivated by the need to start from well-defined syntactic classes known to function as discourse connectives: (a) complex subordinators (postpositions, e.g. *rağmen* ‘despite’, and similar clause final elements, such as *yerine* ‘instead of’), (b) coordinating conjunctions (*ve* ‘and’, *ama* ‘but’), and (c) adverbials (*ayrca* ‘in addition’). TDB 1.0 also annotates phrasal expressions; these are devices that contain a postposition or a similar clause final element taking a deictic item as an argument, e.g. *buna rağmen* ‘despite this’, as in Ex.5 below. This group of connectives are morphologically and syntactically well-formed but not lexically frozen. Moreover, due to the presence of the deictic element in their composition, they are processed anaphorically. Because of these reasons, phrasal expressions, which are annotated separately in TDB 1.0, are merged with alternative lexicalizations in TDB 1.1 (see §3).

It is important to note that connectives may have a DC use as well as a non-DC use. The criterion to distinguish the DC/non-DC use is Asher’s (2012) notion of abstract objects (AO) (events, activities, states, etc.). We take a lexical signal as a DC to the extent it relates text segments with an AO interpretation. The DC is referred to as the head of a DR, the text segments it relates are termed as the arguments. We also adhere to the minimality principle of PDTB (MP), a principle that applies to the length of text spans related by a DC. It means that annotators are required to choose an argument span that is minimally necessary for the sense of the relation (Prasad et al., 2014).

With the MP and the AO criterion in mind, the annotators went through the whole corpus searching for predetermined connectives one by one in each file, determining and annotating their DC use, leaving the non-DC use unannotated. Here, to annotate means that (explicit) DCs and phrasal expressions are tagged mainly for their predicate-argument structure; i.e. for their head (Conn) and two arguments (Arg1, Arg2) as well as the material that supplements them (Supp1, Supp2)².

²Following the PDTB principles, Arg2 is taken as the text segment that syntactically hosts the discourse connective; the other text segment is Arg1. The clause order of sentences with complex subordinators is Arg2-Arg1 while the other relations have the Arg1-Arg2 order. Supp1 and Supp2 stand for text segments that support the interpretation of an argument.

In the examples in the rest of the paper, Arg2 is shown in bold, Arg1 is rendered in italics; the DC itself is underlined. Any null subjects are shown by parenthesized pronouns in the glosses.

- (5) *Çalışması gerekiyordu.* Buna rağmen,
üniversiteyi bırakmadı.
(She) had to work. Despite this, **(she) did not quit university.**

2.2 Evaluation of TDB 1.0

TDB 1.0 has a total of 8483 annotations on 77 Conn types and 147 tokens including coordinating conjunctions, complex subordinators, and discourse adverbials. However, it does not contain sense annotations; it does not annotate implicit DRs or entity relations; neither does it annotate alternative lexicalizations as conceived by the PDTB. The addition of these relations and their senses would enhance the quality of the corpus. Thus, this study describes an effort that involves the addition of new annotations to TDB 1.0, part of which involves sense-tagging of pre-annotated explicit DCs.

Before explaining the details about the enrichment of the corpus, we provide an evaluation of TDB 1.0. In earlier work, we reported the annotation procedure and the annotation scheme (Zeyrek et al., 2010) and provided inter-annotator agreement for complex subordinators and phrasal expressions (Zeyrek et al., 2013), but a complete evaluation of the corpus has not been provided. Table 1 presents inter-annotator agreement (IAA) of the connectives by syntactic type. We measured IAA by Fleiss’ Kappa (Fleiss, 1971) using words as the boundaries of the text spans selected by the annotators, as explained in Zeyrek et al. (2013).

The agreement statistics for argument spans are important because they show how much the annotators agreed on the AO interpretation of a text span. Table 1 shows that overall, IAA of both arguments is ≥ 0.7 . Although this is below the commonly accepted threshold of 0.8, we take it satisfactory for discourse-level annotation, which is highly challenging due to the ambiguity of coherence relations (Spooren and Degand, 2010).

³Some phrasal expressions are retrieved by the same search token as subordinators; thus, ‘Subord’ indicates IAA for subordinators and phrasal expressions calculated jointly.

⁴‘Subtotal’ represents the total of connectives for which IAA could be calculated; ‘IAA not avl.’ (available) means IAA could not be calculated.

Conn. Syn. Type	DC	Non-DC	Arg1	Arg2
Coord.	3609	6947	0.78	0.83
Subord. ³	3439	5154	0.75	0.80
Disc. Adv.	698	223	0.74	0.83
Subtotal ⁴	7746	12324	0.76	0.82
IAA not avl.	737	903	-	-
TOTAL	8483	13227		

Table 1: DC/Non-DC counts of connective types in TDB 1.0 (coordinators, complex subordinators, adverbials) and Fleiss’ Kappa IAA results for argument spans (Sevdik-Çalli, 2015)

3 Creating TDB 1.1

Due to lack of resources, we built TDB 1.1 on 10% of TDB (40.000 words). We used PDTB 2.0 annotation guidelines and the sense hierarchy therein (see fn 1).

Four part-time working graduate students annotated the corpus in pairs. We trained them by going over the PDTB guidelines and the linguistic principles provided in §2.1. Each pair annotated 50% of the corpus using an annotation tool developed by Aktaş et al. (2010). The annotation task took approximately three months, including adjudication meetings where we discussed the annotations, revised and/or corrected them where necessary.

3.1 Annotation Procedure

The PDTB sense hierarchy is based on four top level (or level-1) senses (TEMPORAL, CONTINGENCY, COMPARISON, EXPANSION) and their second and third level senses. The annotation procedure involved two rounds. First, we asked the annotators to add senses to the pre-annotated explicit DCs and phrasal expressions. The annotators implemented this task by going through each file. In this way, they fully familiarized themselves with the predicate-argument structure of DCs in TDB 1.0, as well as the PDTB 2.0 sense hierarchy.

In the second round, the annotators first tagged alternative lexicalizations (AltLexs) independently of all other DRs in each file. Given that phrasal expressions could be considered as a subset of PDTB-style AltLexs, this step ensured that TDB 1.1 not only includes phrasal expressions but various subtypes of Altlexs as well. Finally, the annotators identified and annotated implicit DRs and entity relations (EntRels) simultaneously in each file by searching them within paragraphs and between adjacent sentences delimited by a full stop, a colon, a semicolon or a question mark.

Alternative Lexicalizations: This refers to cases which could be taken as evidence for the lexicalization of a relation. The evidence may be a phrasal expression (Ex. 5), or a verb phrase, as in Ex. 6:

- (6) ... genç Marx, Paris’de Avrupa’nın en devrimci işçi sınıfı ile tanışır. **Bu, onun düşüncesinin oluşmasında en önemli kilometre taşlarından birini teşkil eder.**
... in Paris, young Marx meets Europe’s the most revolutionary working class. This constitutes one of the most important milestones that shapes his thoughts.

Entity Relations: In entity relations, the inferred relation between two text segments is based on an entity, where Arg1 mentions an entity and Arg2 describes it further. As mentioned in §2, a null subject in Arg2 (or in both Arg1 and Arg2) is often a sign of an EntRel (Ex. 7).

- (7) Kerem ter içindeydi. **“Kurtulamamışım demek,” diye mırıldandı.**
 Kerem was all sweaty. “So I was not set free” (he) muttered.

Implicit DRs: For the annotation of implicit DRs, we provided the annotators with an example explicit DC or a phrasal expression (in Turkish) for each level of the PDTB 2.0 sense hierarchy. We told the annotators to insert the example connective (or another connective of their choice if needed) between two sentences where they inferred an implicit DR (Ex. 5 above). While EntRels were only annotated for their arguments, Altlexs and implicit DRs required senses as well. While annotating the senses, the annotators were free to chose multiple senses where necessary.

3.2 Additional Sense Tags

To capture some senses we came across in Turkish, we added three level-2 senses to the top-level senses, COMPARISON and EXPANSION.

COMPARISON: Degree. This sense tag captures the cases where one eventuality is compared to the other in terms of the degree it is similar to or different from the other eventuality. The label seemed necessary particularly to capture the sense conveyed by the complex subordinator *kadar*, which can be translated to English as, ‘as ADJ/ADV as’ or ‘so AJD/ADV that’. When *kadar* is used to compare two eventualities in terms of

how they differ, Arg2 is a negative clause (Ex. 8). So far, this label has only been used to annotate explicit DRs.

- (8) **Tanınmayacak kadar değişmişti.**
(He) changed so much that (he) could not be recognized.

EXPANSION: Manner. This tag indicates the manner by which an eventuality takes place.⁵ It was particularly needed to capture the sense of the pre-annotated complex subordinator *gibi* ‘as’, and the simplex subordinator *-erek* ‘by’, which we aim to annotate. So far, the Manner tag has only been used to annotate explicit DRs.

- (9) **Dediği gibi yaptı.**
(S/he) did as (S/he) said (s/he) would

EXPANSION: Correction. The Correction tag is meant to capture the relations where an incorrect judgement or opinion gets corrected or rectified in the other clause. So far, the majority of Correction relations in TDB 1.1 are implicit. There are polysemous tokens (Ex. 10), as well as single-sense tokens (Ex. 11). These do not convey the PDTB chosen alternative sense (the sense where one of the alternatives replaces the other). For example, to insert *onun yerine* ‘instead of this’ in Ex. 11 would be odd (though this connective would fit Ex. 10). Although further research is needed, we predict that Correction relations are characterized by the negative marker of nominal constituents, *değil* (underlined) in Arg1.

- (10) *Ben yere bakmazdım. (IMP=ama ‘but’) Gözüne bakardım insanların. (Chosen alternative; Correction)*
I wouldn’t look down. (I) would look into peoples eyes.
- (11) *O olayları yaşayan ben değilim. (IMP= bilakis ‘to the contrary’) Benim yaşamım bambaşka. (Correction)*
I am not the one who went through those events. My life is completely different.

⁵PDTB-3 sense hierarchy (Webber et al., 2016) introduces Expansion:Manner and Comparison:Similarity, among other sense tags. The PDTB Manner label conveys the same sense we wanted to capture. On the other hand, the PDTB label ‘Similarity’ is similar to Degree only to the extent it conveys how two eventualities are similar. To the best of our knowledge, the Similarity label does not indicate comparison on the basis of how two things differ. Finally, we became aware of the revised PDTB sense hierarchy after we have started our annotation effort. We decided to continue with PDTB 2.0 labels (plus our new labels) for consistency.

3.3 Annotation Evaluation

TDB 1.1 was doubly-annotated by annotators who were blind to each other’s annotations. To determine the disagreements, we calculated IAA regularly by the exact match method (Miltsakaki et al., 2004). At regular adjudication meetings involving all the annotators and the project leader, we discussed the disagreements and created an agreed set of annotations with a unanimous decision.

We measured two types of IAA: type agreement (the extent at which annotators agree over a certain DR type), and sense agreement (agreement/disagreement on sense identity for each token). For the senses added to the pre-annotated explicit DCs and phrasal expressions, we only calculated sense agreement. For the new relations, we measured both type agreement and sense agreement. This was done in two steps. Following Forbes-Riley et al. (2016), in the first step, we measured type agreement. Type agreement is defined as the number of common DRs over the number of unique relations, where all discourse relations are of the same type. For example, assume annotator1 produced 12 implicit discourse relations for a certain text whereas annotator2 produced 13, where the total number of unique discourse relations were 15 and the common annotations 11. In this case, type agreement is 73.3%. Then, we calculated sense agreement among the common annotations using the exact match method ⁶ (see Table 2 and Table 3 below).

Relation Type	Agreement
Implicit	33.4%
AltLex	72.6%
EntRel	79.5%

Table 2: IAA results for type agreement in TDB 1.1

Sense	Explicit	Implicit	AltLex
Level-1	88.4%	85.7%	93.9%
Level-2	79.8%	78.8%	79.5%
Level-3	75.9%	73.1%	73.4%

Table 3: IAA results for sense agreement in TDB 1.1

According to Table 2, the type agreement for AltLexs and EntRels is satisfactory (≥ 0.7) but implicit DRs display too low a type agreement. Due to this low score, we evaluated the reliability of the gold standard implicit relations: one year after TDB 1.1 was created, we asked one of our

⁶Since no sense tag is assigned to Entrels, for them only type agreement is calculated.

Type	TDB 1.1	PDTB 2.0	Hindi DRB
Explicit	800 (43.1%)	18459 (45.4%)	189 (31.4%)
Implicit	407 (21.9%)	16224 (39.9%)	185 (30.7%)
Altlex	108 (5.8%)	624 (1.5%)	37 (6.15%)
Entrel	541 (29.1%)	5210 (12.8%)	140 (23.2%)
NoRel	-	254 (0.6%)	51 (8.4%)
TOTAL	1,856	40,600	602

Table 4: Cross linguistic comparison of DR types. The numbers within the parenthesis indicate the ratio of DR tokens.

four annotators to annotate the implicit DRs (both for type and sense) by going through 50% of the corpus he had not annotated before. He searched and annotated implicit DRs between adjacent sentences within paragraphs, skipping other kinds of relations. This procedure is different from the earlier one where we asked the annotators to annotate EntRels and implicit DRs simultaneously in each file. We also told the annotator to pay attention to the easily confused implicit EXPANSION:Restatement:specification relations and EntRels. (We stressed that in the former, one should detect an eventuality being further talked about rather than an entity as in the latter.)

Then, we assessed intra-rater agreement between the annotator’s annotations and the gold standard data. In this way, we reached the score of 72.9% for type agreement on implicit DRs.⁷ This result shows that implicit DRs have been consistently detected in the corpus; in addition, it suggests that annotating implicit DRs independently of EntRels is a helpful annotation procedure.

Table 3 shows that for explicit DCs, the IAA results for all the sense levels is ≥ 0.7 , indicating that the senses were detected consistently. Similarly, the sense agreement results for implicit DRs and AltLexs for all the sense levels are ≥ 0.7 , corroborating the reliability of the guidelines.

4 Distribution of Discourse Relation Types

This section offers a preliminary cross-linguistic comparison. It presents the distribution of discourse relation types in TDB 1.1 and compares them with PDTB 2.0 (Prasad et al., 2014) and Hindi Discourse Relation Bank (Oza et al., 2009), which also follows the PDTB principles (Table 4).

⁷Intra-rater agreement between the implicit relation sense annotations of the annotator and the gold standard data is also satisfactory, i.e. ≥ 0.7 for all sense levels (Level-1: 87.5%, Level-2: 79.3%, Level-3: 74.6%). We calculated sense agreement in the same way explained throughout the current section.

It is known that implicit relations abound in texts; thus, it is important to reveal the extent of implicitation in discourse-annotated corpora. Table 4 indicates that in TDB 1.1, explicit DRs are highest in number, followed by EntRels and implicit DRs. The ratio of explicit DRs to implicit DRs is 1.96. This ratio is 1.13 for PDTB 2.0, and 1.02 for Hindi DRB. That is, among the corpora represented in the table, TDB displays the largest difference in terms of the explicit-implicit split. However, it is not possible at this stage to generalize the results of this cross-linguistic comparison to tendencies at the discourse level. TDB 1.1 does not annotate simplex subordinators and leaves implicit VP conjunctions out of scope. Thus, when these are annotated, the ratio of explicit DRs to implicit DRs would change. Issues related to the distribution of explicit and implicit relations across genres are also necessary to reveal. We leave these matters for further research.

5 Conclusion

We presented an annotation effort on 10% of Turkish Discourse Bank 1.0 resulting in an enriched corpus called TDB 1.1. We described how PDTB principles were implemented or adapted, and presented a complete evaluation of TDB 1.1 as well as TDB 1.0, which has not been provided before. The evaluation procedure of TDB 1.1 involved measuring inter-annotator agreement for all relations and assessing intra-annotator agreement for implicit relations. The agreement statistics are overall satisfactory. While inter-annotator agreement measurements show reliability of annotations (and hence the re-usability of the annotation guidelines), intra-rater agreement results indicate the reproducibility of gold standard annotations by an experienced annotator. Using the same methodology, we aim to annotate a larger part of the TDB including attribution and no relations in the future.

Acknowledgements We would like to thank our anonymous reviewers for their useful comments. We also thank METU Project Funds (BAP-07-04-2015-004) for their support.

References

- Berfin Aktaş, Cem Bozsahin, and Deniz Zeyrek. 2010. Discourse relation configurations in Turkish and an annotation environment. In *Proc. of the 4th Linguistic Annotation Workshop*, pages 202–206. ACL.
- Nicholas Asher. 2012. *Reference to abstract objects in discourse*, volume 50. Springer Science & Business Media.
- Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378.
- Kate Forbes-Riley, Fan Zhang, and Diane Litman. 2016. Extracting PDTB discourse relations from student essays. In *Proc. of the SIGDIAL*, pages 117–127.
- Eleni Miltsakaki, Rashmi Prasad, Aravind K Joshi, and Bonnie L Webber. 2004. The Penn Discourse Treebank. In *LREC*.
- Umangi Oza, Rashmi Prasad, Sudheer Kolachina, Dipti Misra Sharma, and Aravind Joshi. 2009. The Hindi Discourse Relation Bank. In *Proc. of the 3rd Linguistic Annotation Workshop*, pages 158–161. Association for Computational Linguistics.
- Rashmi Prasad, Bonnie Webber, and Aravind Joshi. 2014. Reflections on the Penn Discourse Treebank, comparable corpora, and complementary annotation. *Computational Linguistics*.
- Ayışığı Sevdik-Çallı. 2015. *Assessment of the Turkish Discourse Bank and a Cascaded Model to Automatically Identify Discursive Phrasal Expressions in Turkish*. Ph.D. thesis, Middle East Technical University.
- Wilbert Spooren and Liesbeth Degand. 2010. Coding coherence relations: Reliability and validity. *Corpus Linguistics and Linguistic Theory*, 6(2):241–266.
- Bonnie Webber, Rashmi Prasad, Alan Lee, and Aravind Joshi. 2016. A discourse-annotated corpus of conjoined VPs. In *Proc. of the 10th Linguistics Annotation Workshop*, pages 22–31.
- Deniz Zeyrek and Bonnie L Webber. 2008. A discourse resource for Turkish: Annotating discourse connectives in the METU corpus. In *IJCNLP*, pages 65–72.
- Deniz Zeyrek, Işın Demirşahin, Ayışığı Sevdik-Çallı, Hale Ögel Balaban, İhsan Yalçınkaya, and Ümit Deniz Turan. 2010. The annotation scheme of the Turkish Discourse Bank and an evaluation of inconsistent annotations. In *Proc. of the 4th Linguistics Annotation Workshop*, pages 282–289. Association for Computational Linguistics.
- Deniz Zeyrek, Işın Demirşahin, Ayışığı Sevdik-Çallı, and Ruket Çakıcı. 2013. Turkish Discourse Bank: Porting a discourse annotation style to a morphologically rich language. *Dialogue and Discourse*, 4(2):174–184.