The Annotation Scheme of the Turkish Discourse Bank and An Evaluation of Inconsistent Annotations

Deniz Zeyrek, Işın Demirşahin, Ayışığı Sevdik-Çallı, Hale Ögel Balaban, İhsan Yalçınkaya
Middle East Technical University, Ankara, Turkey
and
Ümit Deniz Turan
Anadolu University, Eskişehir, Turkey

Corresponding author: dezeyrek@metu.edu.tr

Abstract

In this paper, we report on the annotation procedures we developed for annotating the Turkish Discourse Bank (TDB), an effort that extends the Penn Discourse Tree Bank (PDTB) annotation style by using it for annotating Turkish discourse. After a brief introduction to the TDB, we describe the annotation cycle and the annotation scheme we developed, defining which parts of the scheme are an extension of the PDTB and which parts are different. We provide inter-coder reliability calculations on the first and second arguments of some connectives and discuss the most important sources of disagreement among annotators.

1 A brief introduction to the Turkish Discourse Bank

1.1 The Data

The Turkish Discourse Bank (TDB) project aims to annotate the 500,000-word-subcorpus of the two-million-word METU Turkish Corpus (MTC) (Say et al, 2002). The subcorpus includes a wide range of texts, e.g. fiction, interviews, memoirs, news articles, etc. reflecting the distribution of the genres in the MTC (Zeyrek et al, 2009). The main objective of the project is to annotate discourse connectives with their two arguments, modifiers and supplementary text spans. Following the Penn Discourse Tree Bank (PDTB), we take discourse connectives as discourse-level predicates taking two (and only two) arguments, called Arg1 and Arg2, which may span one or more clauses and sentences that are adjacent or nonadjacent to the connective (Prasad et al, 2007, Webber, 2004). Discourse relations can certainly be expressed without connectives but we have chosen to annotate discourse relations encoded by connectives since they are more specific about their semantics.

Discourse connectives are identifiable from three syntactic classes, namely, coordinating conjunctions, subordinating conjunctions, and discourse adverbials. As in the PDTB, we take elements belonging to these syntactic classes as discourse connectives when they semantically relate syntactic entitities such as clauses, sentences, sequences of sentences, and nominalizations having an abstract object interpretation i.e., eventualities, possibilities, situations, facts, and propositions (as in Asher, 1993, cf. Webber, et al, 2005). Major departures from the PDTB are, attribution is not annotated, only overt connectives are being annotated, and the nominal arguments of connectives are being annotated where they denote an abstract object. Annotation of implicit connectives is further work.

1.2 The annotation cycle

Before the annotation process started, the annotators studied the guidelines, which defined some general principles and illustrated difficult cases. The guidelines were written in a way to allow the annotators enough freedom to reflect their intuitions on the annotations. The annotators were also told to observe the minimality principle (MP) of the PDTB guidelines, which expects them to mark as argument parts of a clause or sentence that are

minimally sufficient and necessary for the discourse relation encoded by the connective.

The annotation cycle includes three steps. First, the annotators go through the whole subcorpus to annotate a given connective at a time. Any disagreements are discussed and resolved by the project team. In the second step, the definitions in the annotation guidelines are revised with the new issues that emerged in annotating the connective. Finally, the agreed annotations are checked to ensure they obeyed the annotation guidelines fully. The annotations were created by a tool designed by Aktaş (2008).

The connectives are being annotated for the categories given in the next section by three annotators, who have been in the project since the annotation effort started. The three-step annotation process and the number of annotators we use slow down the task considerably but given the complexity of discourse annotation and the need for annotation efforts in Turkish, we were compelled to target maximum reliability achieved by three annotators.

The inter-coder reliability has recently stabilized and to speed the annotation effort, two annotators have started to carry out their task as a pair, while the other annotator works independently. This annotation style involves two annotators working side-by-side at one computer, continuously collaborating on one connective type at a time to code all its tokens in the subcorpus. One of the annotators carries out the task on the annotation tool, while the other observes her continuously for any defects and problems and suggests alternative solutions. This style of annotation, created by our group independently of pair programming, corresponds to the practice explained in Williams, et al (2000) and Williams and Kessler (2000). It is quite a beneficial and reliable method that also speeds up the process (Demirşahin, et al ms). [1] We give the preliminary results of this procedure in section 2.1.4.

1.3 An outline of Turkish connectives and the annotation scheme

We annotate discourse connectives belonging to the syntactic classes listed below, leaving out converbs that may function as discourse connectives.

- Coordinating conjunctions (*ve* 'and', *ya da* 'or', *ama* 'but')
- Complex subordinators (*için* 'for', *rağmen* 'although, despite'), converbs/simplex subordinators (*-Ince* 'when,' *–ken* 'while, now that')[2]
- Anophoric connectives (*bundan başka* 'in addition to/separate from these', *bunun sonucunda* 'as a result of this,' *bunun için* 'due to/for this reason', *buna rağmen* 'despite this') and discourse adverbials (*oysa* 'however', *öte yandan* 'on the other hand', *then* 'sonradan')

In Turkish, coordinators are typically s(entence)-medial, they may also be found s-initially, or s-finally. Coordinators show an affinity with the second clause, as evidenced by punctuation and their ability to move to the end of the second clause. Subordinators take as their second argument a nonfinite clause that contains a genitive marked subject that agrees with the subordinate verb in terms of person and number. The subordinate clause may also be assigned case by the postposition that functions as the connective. The subordinator and its host clause are always adjacent and the subordinate clause may appear s-initially or s-finally. Anaphoric connectives are characterized by an anaphoric element in the phrase and hence they have the ability to access the inference in the prior discourse (Webber, et al 2003). Furthermore, they may take as their first argument text spans that are nonadjacent to the sentence containing the connective (Zeyrek and Webber, 2008).[3] As example (1) illustrates, discourse adverbials can be used with connectives from other syntactic classes, e.g. a coordinating conjunction, *fakat* 'but' may be used with *sonradan* 'then', and in accessing its first argument, the discourse adverbial may cross one or more clauses. In the examples, Arg1 is italicized, Arg2 set in bold, and the connective head is underlined.

[1] Except for *ve* 'and', the statistics reported in this paper reflect the agreement among 3 independent annotators.

[2] The capital letters are used to capture the cases where a vowel agrees with the vowel harmony rules of the language. The vowel rendered by the capital letter I may be resolved as any of the high vowels in the language, i.e., i, ü, ı, u.

[3] In the PDTB, expressions like *after that* are coded as "alternative lexicalization" while coding implicit connectives, i.e., as a "nonconnective expression" (Prasad et al, 2007:22). In Turkish, such phrasal expressions are abundant. They are two-part expressions with one part referring to the relation, the other anaphorically to Arg1 as in English. We decided to take these expressions as connectives because otherwise, we would be missing an important fact about Turkish. Therefore they are being annotated as connectives in the TDB project.

(1)

a. *Bunları açıkladığımız vakit yöneticiler evvela şaşırdılar*
When (we) explained these, the administrators were first surprised.

b. Böyle bir şeyi asla beklemiyorlardı.
(They) were never expecting such a thing.

c. Fakat **sonradan kendilerini toparladılar.**
But **then they gained their composure**.

Largely following the annotation style of the PDTB, we determined the categories that form the annotation of a relation as follows:

Conn: This is the connective head of an explicit connective.

Arg2: This tag refers to the argument that forms a syntactic unit with the connective.

Arg1: This tag is for the other argument that the connective relates semantically to Arg2.

Sup1/Sup2: This attribute specifies either the material that makes the semantic contribution of the argument more specific (as in the PDTB), or the clause/sentence where an anaphoric element expressed in the argument is resolved. The Sup tag is not specifically used for anaphor resolution in the PDTB.

Mod: This tag specifies the following features: (a) the adverbs that are used along with connective heads, e.g. *tam aksine* 'just to the contrary', (b) the focus particle *dE* used together with the connective head (e.g., *ve de* 'and-focus particle 'and'), (c) adverbs showing the determinacy of the relation, e.g. *belki* 'perhaps', *sadece* 'only' etc., (d) polarity of postpositional phrases (e.g. *için değil* 'not for'). In the PDTB, the Mod category is utilized only for adverbs used together with connective heads. The other categories are used to capture aspects of attribution and verbs of attribution.

Shared: This attribute identifies the subjects, objects, or any temporal adverbs shared by the arguments of the discourse relation. This category was required for Turkish, which is a pro-drop and free word-order language. In Turkish, subjects, objects or adverbs can appear s-initially, s-medially or s-finally. Subjects and objects are dropped if they are salient in the discourse. This category allows us to capture the variable position of subjects, objects and adverbs shared by the arguments of a discourse relation. The PDTB does not have this feature.

In what follows, we will report on the inter-coder reliability statistics on Arg1 and Arg2 of a set of connectives for which we obtained low inter-coder reliability results and discuss the most common inconsistencies. The remaining categories mentioned above are under use but inter-coder reliability statistics have not been calculated for them.

2 A quantitative and qualitative evaluation of the inconsistent annotations in the TDB

So far, 60 types of discourse connectives amounting to 6873 relations have been annotated in the TDB project. We computed the reliability of the coders' agreement for Arg1 and Arg2 of these connectives by means of the Kappa statistic (Carletta, 1996). A value of K agreement coefficient (henceforth K values) between 0.80 and 1.00 shows a good agreement, and a value between 0.60 and 0.80 indicates some agreement (Poesio, 2000). The K values we obtained for Arg1 and Arg2 of most connectives annotated so far range between 0.80 and 1.00 but for the connectives that are in focus in this paper, the K values for Arg1 are less than 0.80 (see Appendix B). It is these connectives that we now turn to.

These connectives are listed below again, along with the K values obtained for their Arg1 and Arg2. Before calculating the K values, all annotated text spans were re-processed in order to express the annotations in (pseudo) categories. During re-processing, for each annotator the annotated text span boundary characters (i.e., the beginning and end characters) were coded as 1 and the remaining text was coded as 0, so that an agreement table could be constructed (Artstein, and Poesio, 2008; Di Eugenio and Glass, 2004). It is on the basis of this table which we measured inter-coder reliability.

Connective	K value	
	Arg1	Arg2
yandan 'on the other hand'	0.523	0.645
ayrıca 'in addition, separately'	0.545	0.760
rağmen 'despite, despite this'	0.688	0.742
fakat 'but'	0.719	0.855
tersine 'on the contrary'	0.741	1.000
dolayısıyla 'as a result'	0.759	0.930
oysa 'however'	0.767	0.913
amaçla 'for this purpose'	0.785	0.876

Table 1. Eight connectives with K values less than 0.80 (total number of annotations: 554)

For comparison, we provide the K values for two discontinuous connectives in Table 2:

Connective	K value	
	Arg1	Arg2
ne .. ne 'neither nor'	0.820	0.930
hem .. hem 'both .. and'	1.000	0.982

Table 2. Two discontinous connectives with K values higher than 0.80 (total number of annotations: 126)

As seen in Table 2, inter-coder agreement in discontinuous connectives is high. We argue that discontinuous connectives are maximally different from anaphoric connectives (and discourse adverbials) since they unambiguously draw the boundaries of their arguments. As a result, the inter-coder reliability tests yield good agreements, with K values > 0.80. Anaphoric connectives relate their second argument with another argument adjacent or nonadjacent to the connective, in a way much similar to how definite NPs find their antecedents in the previous discourse. Depending on the relation encoded by the connective, the previous discourse is likely to contain clauses that elaborate and expand a generalization, refute an assertion, list the components of a statement, explain the cause of an eventuality, etc. It may not be an easy task to decide whether one should take all or part of these clauses as Arg1; therefore inconsistencies are expected in drawing the Arg1's boundaries. Arg2, on the other hand, is relatively easier to determine since it is syntactically related to the connective and hence its domain is determined.

Example (2), which shows a relation encoded by the connective *tersine* 'on the contrary', presents one of the most common cases of inconsistency in determining the Arg1 span.

(2)

a. Eyleme değil, karaktere ağırlık veren modern romanda biliyoruz ki roman kişilerinin psikolojisi, iç dünyası, bilinci ve bilinçaltı yazarın dikkatle çözmeye çalıştığı ilginç sorunları içerir.
(We) know that in the modern novel, which emphasizes character rather than action, the novel contains the interesting problems that the writer wants to solve and the characters' psychology, their inner world, consciousness, and the subconscious.

b. Bundan ötürü önemli bir yönünü oluşturur romanın.
For this reason, (it) constitutes an important aspect of the novel.

c. Ama gene biliyoruz ki *halk edebiyatı ürünlerinde önemli olan kişinin iç dünyası değil,*
But we also know that *in folk literature what is important is not the person's inner world*,

d. **tersine**, **eylemidir.**
on the contrary, (**it) is his action**.

Two annotators selected as Arg1 the italicized part in (2c) while the third one selected as Arg1 the clauses in (2a), and (2c). In fact, a careful analysis of the discourse connective *tersine* 'on the contrary' reveals that in all tokens in the corpus, the speaker introduces an assertion, then refutes some aspect of it with an overt negation and then rectifies it in Arg2. (Turan and Zeyrek, 2010). The third annotator's selection of Arg1 is compatible with this observation, while the other annotators' selection of Arg1 appears to be guided by the MP.

2.1 Common Sources of Disagreement

An examination of the 8 connectives for which K values were below 0.80 for Arg1 or Arg2 showed that there were 6 main sources for the inconsistencies. These were (a) no overlapping annotations for Arg1, (b) partially overlapping annotations for Arg1 or (c) Arg2, (d) lack of adequate definitions in the guidelines, (e) annotators' errors in following the linguistic definitions in the guidelines, (f) other inconsistencies, e.g., errors in selecting spaces, leaving characters out, etc. (Appendix A).[4] Among these, partially overlapping Arg1 annotations is the major source of discrepancy observed in 63.98% of the inconsistent cases, followed by partially overlapping Arg2 annotations observed in 10.17% of the cases, and no overlapping annotations in 9.74% of the cases. While errors grouped under the 'other' category is 9.74%, annotators' errors in following the linguistic definitions in the guidelines is negligible (2.97%). The percentage of lacking definitions in the guidelines is also low (3.39%), showing that the coverage is good in the updated guidelines. Let us now turn to the

[4] There were also missing annotations but since we did not calculate inter-coder statistics for them, they are not mentioned in this work.

common sources of disagreement among annotators.

2.1.1 Interpretations of the minimality principle

A frequent reason for inconsistent annotations was lack of agreement in determining the exact boundaries of argument spans, which is ultimately related to how the MP is interpreted. For example, in (3), the connective *fakat* 'but' may be taken as linking clauses (3a) and (3b). Yet, the scope of the predicative morpheme (i.e. *–tır* in (3c)) that determines finiteness is shared by the verbs of (3b) and (3c), i.e. this morpheme takes into its scope two consecutive clauses. While two annotators coded only clause (3b) as Arg2, the third annotator tended to interpret Arg2 as the clauses within the scope of the shared predicative morpheme (*-tır*), coding (3b) and (3c) as Arg2. It appears that the disagreement in example (3) stems from different interpretations of the MP coupled with a structural property of Turkish.

(3)

a. Onlara sunulan kurbanlar, başlangıçta insanlardı
At the beginning, it was humans that were sacrified for them.

b. Fakat bu âdet sonraları hafifletilerek, insan yerine hayvanlar kurban edilmeğe başlanmış,
But later on, loosening this tradition, (they) started to sacrifice animals instead of humans,

c. sonunda da bu hayvanları temsil eden bazı şeylerin (…) kâğıt hayvan figürlerinin (…) yahut da bir taşın suya atılmasının yeterli olacağına inanılmıştır.
finally, it was believed that it would be sufficient to throw a stone or paper animal figures to the water, as well as other objects that represent these animals.

The text given in (4) further illustrates a case where the annotators disagreed on the final boundary of Arg2. One annotator selected as Arg2 the span "this is .. noted" (4b), while the other annotators selected the span "this is…. a lost place" ((4b)-(4c)); i.e., they included as Arg2 not only the clause adjacent to the connective, but they also selected the clause that followed where the cataphor is resolved. Faced with such inconsistencies, we decided to annotate the material that is needed for pronoun resolution as supplementary text. In this case, the clause in (4c) is marked as Sup2.

(4)

a. … ikincisindeki ayrıntı bolluğu Recaizade Ekrem'in gerçekçiliğine atfedilmiştir.
.. the richness of details in the second (novel) was attributed to Recaizade Ekrem's realism.

b. **Oysa asıl dikkat çekmesi gereken şudur**:
However, this is what should be noted:

c. Araba Sevdasının Çamlıca'sı yitik bir Çamlıca'dır.
The Çamlıca described in Araba Sevdası is a lost place.

The inconsistencies that derive from different interpretations of the minimality requirement is particularly interesting from a theoretical perspective. It appears that this principle may be interpreted as syntactic minimality as illustrated in example (3), and as a factor that goes against basic insights of discourse interpretation such as anaphor/cataphor resolution as in example (4). In the former case, the MP pulls the annotators in one direction, and the need to reflect their understanding of the discourse in the annotations pulls them in the opposite direction, especially when there is morphological/syntactic evidence for them to choose more than one clause. In the latter case, the annotators seem to feel they would lose the anaphoric/coreference chains in the discourse if they left out the text span where the anaphor was resolved. After using the Sup label for anaphor resolution/coreference chains, disagreements of the latter sort diminished considerably but this was a methodological approach with a bias towards the MP rather than the desired solution of the role of anaphoric/coreference chains in argument spans. We aim to tackle this issue in further research.

A parenthetical or evaluative clause in the argument span also led to inconsistencies in determining argument boundaries. For example, the annotators gave conflicting decisions as to whether or not they should select parenthetical clauses, especially when they are s-medial, as in (5):

(5)

Kemal, **bir yandan** *askeri bir savaş verirken* **öte yandan yerli işbirlikçilerle** –ki bunların başında da basın- **savaşmak zorunda kalmıştır.**
Kemal, while **on the one hand** *fighting a military war*, **on the other hand (he) had to fight with local accomplices** –which mainly included the media.

Disagreements that arise from parenthetical clauses have diminished after we added a new principle to the guidelines, asking annotators to select the parenthetical together with the argument if it contributed to the meaning of the argument.

2.1.2 Ambiguity

Another reason for inconsistent annotations was ambiguity in meaning. Consider example (1) once again, where the contrast relation can be interpreted in three ways: it is not clear whether the contrast is between *to be surprised* in (1a) and *to regain composure* in (1c) or whether it is between *not expecting such a thing* in (1b) and *to regain composure* in (1c). Alternatively, the contrast can be interpreted between (1a) and (1b) on the one hand, and between (1a) and (1c) on the other. We observed that some of the disagreements concerning the span of Arg 1 stemmed from such cases.

2.1.3 Type of discourse relation

Yet another type of inconsistency appears to be associated with the type of discourse relation. Sanders and Noordman (2000) and Pitler, et. al. (2008) state that causal (contingency) relations are among the most salient coherence relations. They suggest that the connectives that signal comparison and contingency are mostly unambiguous. Being cognitively salient, causal and contingency relations are more tightly organized than the additive list relation. This is because in causal relations, one target sentence is more important than the other; while in a list relation there is more than one sentence contributing to the discourse. Sanders and Noordman (2000:53) argue that causal relations are more strongly connecting than additive relations. This salience in discourse relations can be universal. In fact, we found that the inter-coder agreement of the causal connective *çünkü* 'because' was high (0.888 for Arg1 and 0.941 for Arg2). However, for the connective *ayrıca 'in addition to',* which encodes the list relation, the inter-coder K value was 0.545 for Arg1, 0.765 for Arg 2. An example of the list reading interpretation of this connective is illustrated in (6) below.

(6)

a. Babanın yaşamı artık derli toplu olmuştu.
The father's life now became orderly.

b. Evde kavgalar da azalmıştı.
The fights at home have diminished

c. **Ayrıca** yeni bir çevrede de bulunuyorlardı.
Besides, (they) are now in a new neighborhood.

In the extract given in (6), the topic under discussion seems to be the list of the family's diminishing problems: Father's having an orderly life, reduced fights at home, etc. While two annotators preferred to select (6b) as Arg 1, the third one preferred to select (6a) and (6b) together. The connective *ayrıca,* marking a weaker relation between its two arguments, is among the connectives that yielded such instances of disagreement.

2.1.4 Nominalized arguments

In this section, we will report on some preliminary results about a common inconsistency that occurred while annotating the connective *ve* 'and,' namely the problem of teasing apart nominalized arguments that have an abstract object interpretation and those that do not. We also explain the pair annotation process.

In Turkish, a nominalizing process realized by various inflectional suffixes forms nonfinite clauses. The clauses formed by some of these suffixes are abstract enough to be easily specified as an argument of a discourse relation, e.g. –mAk. On the other hand, some of the suffixes (e.g. –mA, -Iş) are very productive in deriving ordinary nouns referring to actual instances or things. It is these cases where disagreement among the annotators increases. Example (7) illustrates the use of –mAk, where the clauses it forms were easily determined as arguments with abstract object interpretations.

(7)

18. yüzyılın yaptığı, 17. Yüzyılın yarattıklarını *çoğaltmak* ve **yaymaktır**.
What the 18th century did was *to increase* and **to extend** what the 17th century created.

Example (8) shows a difficult case where the annotators were inconsistent in deciding whether the connective's arguments have abstract object interpretations or not. This is because the morphological form of the words *gelişme* (improve-mA) 'improvement' and *yapılaşma* (construct-mA) '(re)construction' are very much the same as the words *bekleme* (wait-mA) 'waiting' and *arama* (search-mA) 'search, searching' shown in (9). The final decision was to annotate (9) only.

(8)

Deprem bölgesinde yeniden [gelişme] ve [yapılaşmanın] planlanması gibi ciddi bir sorun bulunmaktadır.
There is the important issue of planning the [improvement] and [re-construction] of the areas affected by the earthquake.

(9)

Artık onu *beklemenin* **ve** **aramanın** boşuna olduğunu anlamıştır.
He has already figured out that it was futile *to wait for her* **and** **to search her**.

We noticed such inconsistencies in annotating 1/3 of the files for *and*. When we shifted to the pair annotation procedure, we obtained high agreement on Arg1 and Arg2 annotations of *and* because we observed that when done in pairs, resolving any disagreements between the annotations was faster since the members of the pair discussed difficult cases between them and sometimes determined a preferred annotation before presenting the results to the group. Table 3 shows the results for *and* annotations. A repeated measures test shows that the increase in K values is significant ($p < 0.01$).

Annotators	K value	
	Arg1	Arg2
3 annotators	0.692	0.791
A pair of annotators and an independent annotator	0.945	0.964

Table 3 K values for *ve* 'and' of 3 independent annotators, and a pair and an independent annotator

3 Summary

In this paper we presented common sources of disagreement we observed in annotating the arguments of discourse connectives in the TDB, a project of discourse-level annotation on written Turkish. We defined our annotation scheme and annotation cycle. We achieved high agreement on argument annotations of discontinuous connectives but agreement on some other connectives was low, particularly for Arg1. Since these connectives belong to different syntactic classes, the inconsistencies cannot be easily explained by the properties of the syntactic class of connectives. We discussed various potential factors affecting inter-coder agreement, including the minimality principle coupled with language specific properties, the structure of discourse (as in the case of our example in *tersine*), cognitive salience of discourse relations, and ambiguity. We discussed inconsistencies resulting form the difficulty of distinguishing the non-abstract object interpretation of a nominalized clause from its abstract object interpretation. We argued that once inter-coder reliability stabilizes, it is beneficial to shift to the procedure where a pair of annotators works together to annotate a specific connective while the third works independently.

Acknowledgments

We acknowledge the grant form The Scientific and Technological Research Council of Turkey and thank our anonymous referees for comments.

References

Berfin Aktaş. 2008. Computational Aspects of Discourse Annotation. Unpublished MS Thesis, Cognitive Science Program, Middle East Technical University

Ron, Artstein and Massimo Poesio (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics*, *34*(4). pp. 555-596.

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

Işın Demirşahin, İhsan Yalçınkaya, Deniz Zeyrek (ms). Pair Annotation: Adaption of Pair Programming to Corpus Annotation.

Barbara Di Eugenio, & Michael Glass (2004). The Kappa statistic: a second look. *Computational Linguistics*, *30*(1). pp. 95-101.

Eleni Miltsakaki, Rashmi Prasad, Aravind Joshi, Bonnie Webber. 2004. Annotating Discourse Connectives and Their Arguments. In *Proceedings of the HLT/NAACL Workshop on Frontiers in Corpus Annotation*. Boston, MA. 2004.

Emily Pitler, Mridhula Raghupathy, Hena Mehta, Ani Nenkova, Alan Lee, and Aravind Joshi. 2008. Easily Identifiable Discourse Relations *Proceedings of COLING*, 2008. Poster paper.

Massimo Poesio. 2000. Annotating a Corpus to Develop and Evaluate Discourse Entity Realization Algorithms: Issues and Preliminary Results. *Proceedings of LREC-2000*. Athens, May 2000.

Rashmi Prasad, Eleni Miltsakaki, Nikhil Dinesh, Alan Lee, Aravind Joshi, Livio Robaldo and Bonnie Webber. 2007. The Penn Discourse Tree Bank 2.0 Annotation Manual. December 17, 2007.

Ted J. M. Sanders and Leo G. M. Noordman. 2000. The Role of Coherence Relations and Their Linguistic Markers in Text Processing. *Discourse Processes 29*(1): 37–60.

Bilge Say, Deniz Zeyrek, Kemal Oflazer, and Umut Özge. 2002. Development of a Corpus and a Treebank for Present-day Written Turkish. In *Proceedings of the Eleventh International Conference of Turkish Linguistics*.

Ümit Deniz Turan and Deniz Zeyrek. 2010. Context, Contrast, and the Structure of Discourse in Turkish. ms.

Bonnie Webber, Aravind Joshi, Eleni Miltsakaki, Rashmi Prasad, Nikhil Dinesh, Alan Lee, and Kate Forbes. 2005. A Short Introduction to the Penn Discourse Treebank. Copenhagen *Working Papers in Language and Speech Processing*.

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

Bonnie Webber, Aravind Joshi, Matthew Stone, and Alistair Knott. 2003. Anaphora and Discourse Structure. Computational Linguistics 29(4). pp. 545-587. 2003

Laurie Williams, Robert R. Kessler, Ward Cunningham, Ron Jeffries. 2000. Strengthening the Case for Pair Programming. *IEEE Software*, July/August 2000, pp. 19-25.

Laurie Williams and Kessler, Robert R. 2000. All I Really Need to Know about Pair Programming I Learned In Kindergarten*, Communications of the ACM*, Vol. 43, No., 5, pp. 108-114, May 2000.

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

Deniz Zeyrek, Ümit Turan, Cem Bozşahin, Ruket Çakıcı, Ayışığı Sevdik-Çallı, Işın Demirşahin, Berfin Aktaş, İhsan Yalçınkaya, Hale Ögel. 2009. Annotating Subordinators in the Turkish Discourse Bank.*ACL-IJCNLP,* In *Proceedings of LAW III Annotation Workshop III*. Singapore, August 6-7, 2009, pp. 44-48.

Appendix A. Sources of disagreement in 8 connectives (Turkish equivalents of 'but', 'however', 'for this reason', 'despite', 'on the other hand', 'for this reason', 'on the contrary', 'in addition')

Source of disagreement	**No.**	**%**
Partial Arg1 overlap	151	63.98
Partial Arg2 overlap	24	10.17
No overlap of Arg1	23	9.74
Other	23	9.75
Lack of guidelines	8	3.39
Guidelines not followed	7	2.97
Total	236	100

Appendix B. K values of connective types annotated in the TDB project[5]

		K Value	
Connective (type)	**English equivalent**	**Arg1**	**Arg2**
ne ..ne	neither .. nor	1.000	0.982
veya	or	0.942	0.980
dolayı	since	0.892	0.957
çünkü	because	0.888	0.941
örneğin	for example	0.870	0.898
ya da	or	0.843	0.974
yoksa	otherwise	0.837	0.938
ama	but	0.832	0.901
karşın	despite, despite this	0.824	0.893
hem .. hem	both .. and	0.820	0.930
dahası	moreover	0.785	0.908
amaçla	for the purpose of	0.785	0.876
için	for, for this reason	0.776	0.915
oysa	however	0.767	0.913
dolayı-sıyla	for this reason	0.759	0.930
tersine	on the contrary	0.741	1.000
fakat	but	0.719	0.855
amacıyla	for the purpose of	0.700	0.912
ve	and	0.692	0.791
rağmen	despite, despite this	0.688	0.742
ayrıca	in addition; separately	0.545	0.760
yandan	on the one hand	0.523	0.645

[5] The results about the connective types for which 10 or more relations have been annotated by three annotators are included in the appendices.