METU Turkish Discourse Bank Browser

Utku Şirin[1], Ruket Çakıcı[1], Deniz Zeyrek[2]

Computer Engineering Department[1], Informatics Institute[2]
Middle East Technical University, Ankara, Turkey[1,2]
utkusirin@gmail.com, ruken@ceng.metu.edu.tr, dezeyrek@metu.edu.tr

Abstract

In this paper, the METU Turkish Discourse Bank Browser, a tool developed for browsing the annotated annotated discourse relations in Middle East Technical University (METU) Turkish Discourse Bank (TDB) project is presented. The tool provides both a clear interface for browsing the annotated corpus and a wide range of search options to analyze the annotations.

Keywords: Browser, Turkish Discourse Bank (TDB), discourse annotation

1. Introduction

The METU TDB project extends the METU Turkish Corpus (MTC) from a sentence-level language resource to a discourse-level language resource (Zeyrek et al., 2008; Zeyrek et al., 2009; Zeyrek et al., 2010). For this purpose, discourse connectives (e.g., *ama* 'but', ve 'and', *ayrıca* 'in addition', *rağmen* 'despite') are annotated along with the text spans they relate. The TDB includes annotations of ˜8400 discourse relations created on the ˜400,000-word-subcorpus of the two-million-word MTC. The MTC itself has texts tagged with genre, author, publisher and publishing date (Say et al., 2002). The distribution of the genres is reflected in the TDB. The TDB aims to capture discourse relations to the extent that they are instantiated by explicit discourse connectives and it examines the structural and semantic aspects of discourse relations. The annotations are created by the DATT (Aktaş et al., 2010), which generates a layer of annotation data in XML format by means of character indexes. For the sentence in (1), for example,

(1) *Çay çok güzeldi* <u>ama</u> **kahve beş para etmezdi.**

The tea was very good <u>but</u> **the coffee was not worth a dime.**

The DATT keeps the beginning and end offsets of the connective and arguments of the relation with respect to their positions in the whole text file. The annotation tool has a user interface showing the connectives and arguments in different colours. Annotations are stored in well-formed XML files separate from the original text data. (See Section 2 for the description of the discourse-related terms.)

The METU TDB browser uses these annotation files and the indexes created by the DATT to serve as a clear interface for the annotations in the TDB to effectively identify and exploit various aspects of Turkish discourse.

In the rest of this paper, we provide information on the features and capabilities of the TDB browser. Section 2 presents a brief overview of Turkish discourse structure and how various discourse phenomena are reflected in the annotations in the TDB. Section 3 is an overview of the characteristics of the PDTB browser. It also provides a brief comparison of the TDB and the PDTB. Section 4 provides a comprehensive description of the METU TDB browser. Section 5 includes a summary and a brief mention of future work.

2. Turkish Discourse Structure and TDB Formalism

As in English, discourse connectives are basically identified from three grammatical classes in Turkish: coordinating conjunctions *(and, but)*, subordinating conjunctions *(although, before, after)* and discourse adverbials *(however, in addition)* (Zeyrek and Weber, 2008). These three types of connectives establish discourse relations mainly by taking two arguments, i.e., text spans that the connective relates. From a structural perspective, arguments are defined as tensed or untensed clauses. The argument-connective-argument structure is the basis of the TDB formalism enhanced with some other elements of Turkish discourse. The TDB follows the principles of the PDTB in the annotations (Prasad et al., 2008; Joshi et al., 2006). Therefore, the text span that is syntactically bound to the discourse connective is taken as the second argument. The text span that the connective relates to the second argument is taken as the first argument. Throughout this paper and in the TDB browser, we abbreviate the first argument of a discourse connective as ARG1 and the second one as ARG2. In all the examples in the paper, ARG1 is rendered in italics, ARG2 in bold. The connective is underlined. Discourse connectives are shown as CONN (Prasad et al., 2008). In addition to these basic categories, text spans that supplement ARG1 and ARG2 are also annotated, respectively called SUPP1 and SUPP 2. Finally, modifiers of the connectives, abbreviated as MOD, and grammatical elements that are shared by two arguments, abbreviated as SHARED, are annotated as well (Zeyrek et al., 2010).

3. Related Work - PDTB Browser

The METU TDB follows the PDTB principles in annotation. The PDTB is built on the Penn Tree Bank. The PDTB browser serves as a querying facility for searching specific relations. This is done through PDTBXPath, which is a simple top-level query interface for the PDTB (Yao et al., 2010). The search options are formed around relation type, connective type, the connective's semantic and structural

aspects, different features of ARG1 and ARG2 of the connective. The PDTB browser uses three different source directories. The RawRoot directory keeps all text files that are included in the annotated corpus. The PtbRoot directory keeps the syntactic annotation (parse trees) of the sentences in the text files. And finally, the PdtbRoot directory keeps the discourse annotation information in the PDTB.
The TDB is not built on a syntactically enriched tree bank. It is a resource for discourse relations only. The TDB browser's querying facility is done completely through the user interface. One can search strings or regular expressions on any element of any relation in a very simple and understandable way. There are three different source files in TDB browser. Two of them are the same as those of the PDTB browser, one is different. Instead of the PtbRoot directory storing the parse tree information, the TDB browser uses a tag file storing the information of tags of text files such as title, author, publisher and publishing date.

4. METU TDB Browser

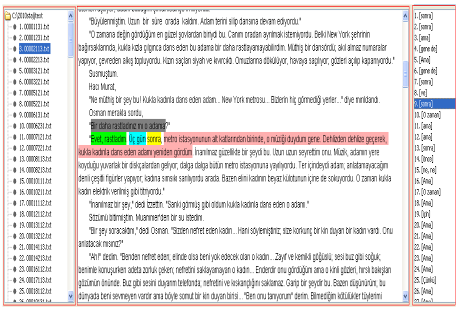
This section describes some highlighted features and some technical details of the METU TDB browser tool. The main feature in the TDB browser is searching through the relations. The browser has 4 search modes, which are quick search, general search, relation filter and advanced search. The details of how these components work are given in Section 4.2. The browser can also explore the structural aspects of the arguments and connectives such as discontinuity and adjacency.
When the browser is started, three file paths that are used by the browser are specified in the data path window, namely, the path of the text directory specifying the text files that will be browsed, the path of the annotation directory containing the XML files storing the annotation information, and the path of the tag file specifying a XLS or XLSX type file storing the tag information. The tag information includes title, author, publisher and publishing date information of the relevant documents. These path names are saved for future reference.

4.1. Browsing

The browsing window contains three basic parts. The text file list on the left of the main window is basically a directory tree whose root is the directory given by the first file path in the data path window (Figure 1). The middle window displays the text in the selected document/file. The rightmost window shows the relation list. It contains a list of all the annotated relations that are associated with the selected text file. Relations are listed by the name of the connective and are sorted according to their position indexes (relative positions) in the text file. Selected annotations are highlighted in the middle window. The main text area (the middle window) goes to the specific index that the relation starts and highlights it in accordance with the colors, as shown in Figure 2.

4.2. Search

The main contribution of the TDB browser is the extensive search functionalities provided. There are four different search modes in the browser, as described below.



C:\2010staj\text
1. 00001131.txt
2. 00001231.txt
3. 00002113.txt
4. 00002213.txt
5. 00003121.txt
6. 00003221.txt
7. 00005121.txt
8. 00005221.txt
9. 00006131.txt
10. 00006231.txt
11. 00007121.txt
12. 00007221.txt
13. 00008113.txt
14. 00008213.txt
15. 00010111.txt
16. 00010211.txt
17. 00011112.txt
18. 00012112.txt
19. 00013112.txt
20. 00013212.txt
21. 00014113.txt
22. 00014213.txt
23. 00016112.txt
24. 00017113.txt
25. 00018112.txt

"Büyülenmiştim. Uzun bir süre orada kaldım. Adam terini silip dansına devam ediyordu."

"O zamana değin gördüğüm en güzel şovlardan biriydi bu. Canım oradan ayrılmak istemiyordu. Belki New York şehrinin bağırsaklarında, kukla kızla çılgınca dans eden bu adama bir daha rastlayamayabilirdim. Müthiş bir dansördü; akıl almaz numaralar yapıyor, çevreden alkış topluyordu. Kızın saçları siyah ve kıvırcıktı. Omuzlarına dökülüyor, havaya saçılıyor; gözleri açılıp kapanıyordu."

Susmuştum.

Hacı Murat,

"Ne müthiş bir şey bu! Kukla kadınla dans eden adam... New York metrosu... Bizlerin hiç görmediği yerler..." diye mırıldandı.

Osman merakla sordu,

"Bir daha rastladınız mı o adama?"

"Evet, rastladım. Üç gün sonra, metro istasyonunun alt katlarından birinde, o müziği duydum gene. Dehlizden dehlize geçerek, kukla kadınla dans eden adamı yeniden gördüm. İnanılmaz güzellikte bir şeydi bu. Uzun uzun seyrettim onu. Müzik, adamın yere koyduğu yuvarlak bir diskçalardan geliyor; dalga dalga bütün metro istasyonuna yayılıyordu. Ter içindeydi adam; anlatamayacağım denli çeşitli figürler yapıyor; kadına sımsıkı sarılıyordu arada. Bazen elini kadının beyaz külotunun içine de sokuyordu. O zaman kukla kadın elektrik verilmiş gibi titriyordu."

"İnanılmaz bir şey," dedi İzzettin. "Sanki görmüş gibi oldum kukla kadınla dans eden o adamı."

Sözümü bitirmiştim. Muammer'den bir su istedim.

"Bir şey soracaktım," dedi Osman. "Sizden nefret eden kadın... Hani söylemiştiniz; size korkunç bir kin duyan bir kadın vardı. Onu anlatacak mısınız?"

"Ah!" dedim. "Benden nefret eden; elinde olsa beni yok edecek olan o kadın... Zayıf ve kemikli göğüslü; sesi buz gibi soğuk; benimle konuşurken adeta zorluk çeken; nefretini saklayamayan o kadın... Enderdir onu gördüğüm ama o kinli gözleri, hırslı bakışları gözümün önünde. Buz gibi sesini duyarım telefonda; nefretini ve kıskançlığını saklamaz. Garip bir şeydir bu. Bazen düşünürüm; bu dünyada beni sevmeyen vardır ama böyle somut bir kin duyan birisi... "Ben onu tanıyorum" derim. Bilmediğim kötülükler tüylerimi

1. [sonra]
2. [sonra]
3. [ama]
4. [gene de]
5. [Ama]
6. [gene de]
7. [sonra]
8. [ve]
9. [sonra]
10. [O zaman]
11. [ama]
12. [ama]
13. [sonra]
14. [önce]
15. [ne, ne]
16. [Ama]
17. [O zaman]
18. [Ama]
19. [için]
20. [Ama]
21. [Ama]
22. [Ama]
23. [Ama]
24. [Ama]
25. [Çünkü]
26. [Ama]

Figure 1: Browsing

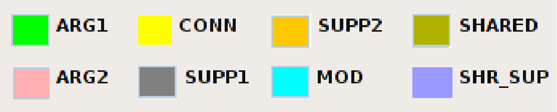
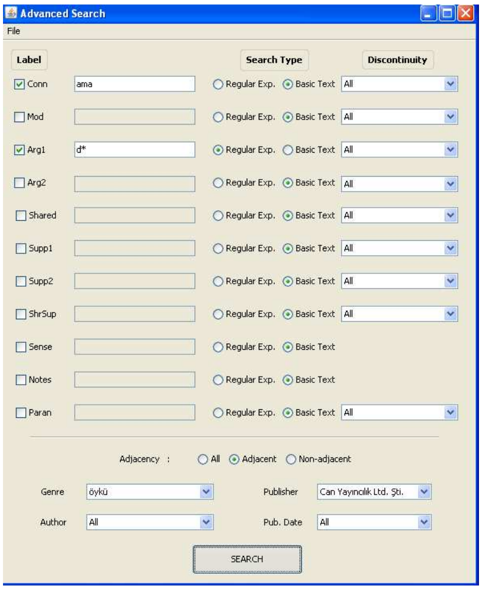


Figure 2: Color Specifications

The **quick search facility** filters the text file list by only the connective and the genre. Connective and genre are used as quick filter parameters as they are the most commonly used filtering parameters. The quick search facility also provides the numbers of annotated relations, whose connective is specified by the quick search facility in the whole data and in a specific text file. These numbers are shown just above the main text area.
A **general search** is performed within a selected text file. After specifying a string in the general search text box, the user can see all of the matching (sub)strings in the text file with red highlighting.
The **relation filter** option filters the annotated relations that are listed in the relation list. Filtering is done by a prefix of the connective entered in a box just above the relation list. For example, one can easily filter the connectives containing the letters *an* by means of this feature, which would retrieve the connective *ancak* 'but' as well as the phrasal expression *o zaman* 'at that time/then'.
The **advanced search** mode serves via a separate window and provides a wider range of search options. One can perform a string search in any element of a relation, which can be done either by regular expression or basic text search. Moreover, the discontinuity of any element of a relation and the adjacency information for arguments can be retrieved through advanced search. See Section 4.2.1. for detailed explanations of discontinuity and adjacency notions.
In addition to the search facility on the discourse relations, one can also specify the genre, the author, the publisher and the publishing date of text. For instance, a discontinuous connective with the substring *ard* in the modifier, with a discontinuous ARG1 in a text whose author is *Atilla Atalay* and genre is novel would be one of the example queries in advanced search.
Advanced search queries can be saved permanently or temporarily. Figure 3 shows the advanced search window of the METU TDB browser.
The advanced search query results are shown in a separate



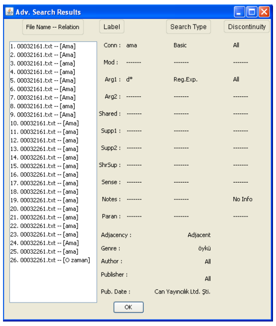
Label		Search Type	Discontinuity
☑ Conn	ama	○ Regular Exp. ● Basic Text	All
□ Mod		○ Regular Exp. ● Basic Text	All
☑ Arg1	d*	● Regular Exp. ○ Basic Text	All
□ Arg2		○ Regular Exp. ● Basic Text	All
□ Shared		○ Regular Exp. ● Basic Text	All
□ Supp1		○ Regular Exp. ● Basic Text	All
□ Supp2		○ Regular Exp. ● Basic Text	All
□ ShrSup		○ Regular Exp. ● Basic Text	All
□ Sense		○ Regular Exp. ● Basic Text	
□ Notes		○ Regular Exp. ● Basic Text	
□ Paran		○ Regular Exp. ● Basic Text	All

Adjacency : ○ All ● Adjacent ○ Non-adjacent

Genre	öykü	Publisher	Can Yayıncılık Ltd. Şti.
Author	All	Pub. Date	All

Figure 3: Advanced Search

window. One can perform different queries with the advanced search facility, get multiple search result windows and use them concurrently. The results window shows each annotated relation compatible with the query. It also shows the details of the query. Figure 4 shows the advanced search result window in the browser.



File Name -- Relation

1. 00032161.txt -- [Ama]
2. 00032161.txt -- [Ama]
3. 00032161.txt -- [ama]
4. 00032161.txt -- [ama]
5. 00032161.txt -- [Ama]
6. 00032161.txt -- [Ama]
7. 00032161.txt -- [Ama]
8. 00032161.txt -- [ama]
9. 00032161.txt -- [Ama]
10. 00032161.txt -- [Ama]
11. 00032161.txt -- [ama]
12. 00032161.txt -- [ama]
13. 00032261.txt -- [ama]
14. 00032261.txt -- [ama]
15. 00032261.txt -- [ama]
16. 00032261.txt -- [ama]
17. 00032261.txt -- [ama]
18. 00032261.txt -- [ama]
19. 00032261.txt -- [ama]
20. 00032261.txt -- [ama]
21. 00032261.txt -- [ama]
22. 00032261.txt -- [ama]
23. 00032261.txt -- [ama]
24. 00032261.txt -- [Ama]
25. 00032261.txt -- [ama]
26. 00032261.txt -- [O zaman]

Label		Search Type	Discontinuity
Conn :	ama	Basic	All
Mod :	-------	-------	-------
Arg1 :	d*	Reg.Exp.	All
Arg2 :	-------	-------	-------
Shared :	-------	-------	-------
Supp1 :	-------	-------	-------
Supp2 :	-------	-------	-------
ShrSup :	-------	-------	-------
Sense :	-------	-------	-------
Notes :	-------	-------	No Info
Paran :	-------	-------	-------

Adjacency :	Adjacent
Genre :	öykü
Author :	All
Publisher :	All
Pub. Date :	Can Yayıncılık Ltd. Şti.

Figure 4: Advanced Search Results

4.2.1. Discontinuity and Adjacency

Discontinuity concerns either a connective or an argument. A discontinuous connective would be *either ... or* and its equivalent *ya ... ya da* in Turkish. Discontinuity of an argument means that there is intervening material (one or more words, phrases, sentences, etc.) inside an argument. Adjacency concerns the positioning of the arguments with respect to the corresponding connective. The second argument of a connective is often adjacent to the connective but there is no restriction as to how far away the first argument may be from its corresponding connective. In other words, the first argument may be adjacent or nonadjacent to the corresponding connective. Both discontinuity and nonadjacency are captured by the TDB browser by means of the corresponding features in the search facility.

A discontinuous first argument is exemplified below, in (2). This example shows that ARG1 consists of two full sentences with another intervening between them.

(2) *Belediye Başkanlığı, şehirde 6 saat boyunca su kesintisi olacağını ilan etti.* Bu karara karşı şehrin çeşitli yerlerinde protestolar yapıldı. *Bu süre içinde hastanelerdeki suyun kesilmeyeceği belirtildi.* Ancak **Belediye Başkanlığı daha sonra su kesintisinin 6 yerine 4 saat olacağını belirtti.**

The municipality announced that there would be a water cutoff in the city for 6 hours. There were many protests against this in various parts of the city. *It was indicated that there would be no water cutoff in the hospitals.* However, **the Municipality later announced that the water cutoff would last 4 hours instead of 6.**

In (3), the connective *aslında* divides its ARG2 into two parts, making ARG2 a discontinous argument.

(3) *Bakan açıklama yapmadı.* Daha önce dediğimiz gibi, **bu durum** aslında **beklenmedik değil.**

The minister has not made an explanation. Indeed, as we explained earlier, **this is not something we do not expect**.

Adjacency is a relationship between the ARG1 of a connective and its ARG2. In other words, when ARG1 and ARG2 spans are consecutive with only punctuation marks, the corresponding connective, its modifiers, or shared arguments intervening, then ARG1 is defined as adjacent to ARG2. For all other conditions, ARG1 is nonadjacent to ARG2. In (3), ARG1 is nonadjacent to ARG2 due to the fact that there is a phrase *as we explained earlier* between ARG1 and ARG2. In (4), on the other hand, ARG1 is adjacent to ARG2 since there is only a punctuation mark between ARG1 and ARG2.

(4) *Bakan açıklama yapmadı.* **Bu durum** aslında **beklenmedik değil.**

The minister has not made an explanation. **This is not something we do not expect**, indeed.

4.3. Miscellaneous

This section presents some minor but important facilities that the METU TDB browser provides.

The **sense and note representation** feature is to represent sense and note attributes of the annotation. These two attributes are represented in two separate text areas on the main window.

The **numbers** feature gives the statistics of the annotations. There are four different numbers. The first one is the total number of annotations in the TDB. This number is displayed after a connective quick search and denotes the total number of annotations returned by that query. The second number is the number of annotations in a particular file. It is set after selecting a text file from the text file list and shows the number of annotations in the selected file. Note that the user needs to do a quick search first to make this second number set. The third and fourth numbers are ”Number of Annotations - Caps” and ”Number of Annotations - Sml”. These two numbers show the number of annotations in a selected text file. Caps shows the number of connectives starting with uppercase and Sml shows the number of connectives starting with lower case.

The **highlight all** feature is used for highlighting all of the annotations listed in the annotation list. There is also a **remove highlights** feature which removes any highlighting from the main text area.

The font size can be adjusted with the use of **font size** feature.

4.4. Technical Characteristics

4.4.1. Programming Language, Platform and Speed

The METU TDB browser is written in java using Java SE 6. We have implemented three different TDB browser versions for three different platforms, Mac, Ubuntu and Windows. The METU TDB browser is licenced by LGPL. One can reach the source code of all the versions of the browser from https://sourceforge.net/projects/tdbbrowser/.

Starting the browser and initiating the Main Window takes ˜4000ms on a Intel Core 2 Duo machine. The most processor-intensive operation is querying by advanced search options. Each Advanced Search query results in traversing all the annotations and collecting the matching ones. Annotations in the TDB are stored in directories named after the connectives. These directories contain XML files storing the annotations of only one text file with that connective. Hence, the queries specifying the connective name traverse only one directory, whereas the queries that do not specify the connective check all directories separately. As a result, the query response times may vary depending on whether the connective type is known or not.

A simple query specifying only the connective type as *ama* ‘but’, takes ˜1700ms. When we make the query detailed by specifying the connective type *ama* ‘but’, having a continous connective, containing the *da* substring in the modifier and *ması* substring in ARG2, it takes ˜790ms. When we make the query even more detailed by adding constraints such as the inclusion of the *ki* substring in the SUPP1, nonadjacent ARG1 in a text whose author is “İnci Aral” and the genre of which is novel, it takes ˜40ms. These results are very much related with the number of returning values. As we make detailed queries, the possible number of annotations to be checked decreases, which also decreases the time that is required to answer the query.

On the other hand, a query that does not specify the connective type but that only contains the *da* substring in the modifier and the *ma* substring in ARG1 takes ˜5100ms. When we make the query even more complicated by checking the inclusion of the *ma* substring in ARG2 and choosing only nonadjacent ARG1s, it takes ˜4000ms. When we further detail the query by adding author and publishing date constraints, the querying time decreases to ˜400ms. This decrease in querying time is related with the number of files processed.

Although queries having connective type information is much faster than the queries that do not have connective type information, the METU TDB browser can answer both types of queries in reasonable times.

4.4.2. Software Architecture

Figure 5 shows the general software architecture of the METU TDB browser. The program starts with the Opening Window component. This is where the user defines the paths of the required files/directories. These paths are used to have connections to the data. For the annotation directory and the text file directory, no memory allocation is done. That is, at each time information of an annotation or a text file is required, the XML Parser and File Reader run again. This is due to the fact that the text files and annotations are too large to be kept in the memory. For the tag file, however, memory allocation is done and the tag information is kept in the memory since it is small.

After initialization, the browser steps into the Main Window component. Through the Main Window, users can get into the Advanced Search and the Advanced Search Results components. These are responsible for handling advanced search queries. The Advanced Search Results component has a connection back to the Main Window component so that the user can see the results of the advanced search queries in the Main Window. Here, the important thing is that the user can have more than one Advanced Search Results window at a time. Hence, users may see the results of multiple queries together, which appear without affecting each other. The miscellaneous component is for the miscellaneous facilities that the browser provides. There is also the Highlighter component which is responsible for highlighting the selected annotations according to the colors in Figure 2.

It is worth noting that the Opening Window, Main Window, Advanced Search and Advanced Search Results components have different user interfaces. Hence, they are all multi-threaded classes efficiently handling both the user interface and the background computations.

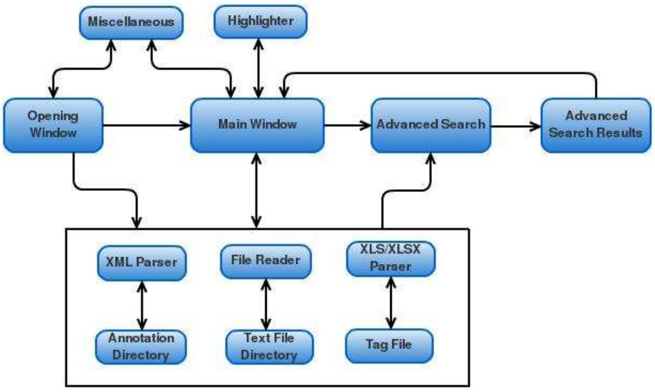


Figure 5: Architecture of the Browser

4.5. User Evaluations

This section represents some user evaluations about the METU TDB browser and how it is useful for understanding some facts about discourse. Firstly, we can say that we are able to search and retrieve the connectives whose first or second argument carries nominalization suffixes, e.g. *-mA (-mesi/-ması), -DHK (-dığı, -duğu, -düğü, -diği)*. One of the connectives, i.e. *için* ‘because/to’ has different senses depending on the suffix its ARG2 takes. For example in (5) *için* has a purposive sense, while in (6) it has a casual sense. With the browser, we are able to find the two types of *için*s.

(5) **...üyelik müzakerelerinin başlaması** için *tarih verilebilir.*

...the date could be arranged to **start to membership negotiations with Turkey.**

(6) **Borcu ödenmediği** için *satıcı firma tarafından haczedilmiştir.*

It is seised by the seller company since **its debt was not paid.**

Secondly, we are able to search and retrieve the discontinuous and nonadjacent arguments of a specific connective. This information is particularly important in understanding what connectives do. For example, Zeyrek et al. (in print) find that *ayrıca* ‘besides’ has more nonadjacent Arg1s than *oysa* ‘whereas’ and *fakat* ‘but’. The authors were able to use this information to prove the hypothesis that *ayrıca* is a discourse adverbial (i.e. a connective whose meaning is not necessarily derived by the adjacency of its arguments as in (Forbes-Riley et al., 2006)).

5. Conclusion and Future Work

In this paper, we have introduced the METU TDB browser. As future work, we will add some more parameters to the quick search option. Moreover, we want to use the METU TDB browser for some statistical analyses over the annotated MTC corpus and to extract useful information about Turkish and Turkish discourse structure.

6. Acknowledgements

We want to thank all METU Turkish Discourse Bank Project members for their invaluable comments and advice. We also thank METU Scientific Research Funds project no. BAP-07-04-2012-001 for the partial support provided.

7. References

Berfin Aktaş, Cem Bozşahin, and Deniz Zeyrek. 2010. Discourse Relation Configurations in Turkish and an Annotation Environment. In *Proceedings of the Fourth Linguistic Annotation Workshop, ACL 2010*, pages 202–206.

Katherine Forbes-Riley, Bonnie Webber, and Aravind Joshi. 2006. Computing Discourse Semantics: The Predicate-Argument Semantics of Discourse Connectives in D-LTAG. *Journal of Semantics*, 23(1):55–106.

Aravind Joshi, Rashmi Prasad, and Bonnie Webber. 2006. Discourse Annotation: Discourse Connectives and Discourse Relations. In *Tutorial at the Association for Computational Linguistics*. Sydney.

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Weber. 2008. The Penn Discourse Treebank 2.0. In *Proceedings of the 6th International Conference on Languages Resources and Evaluation (LREC).*

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 11th International Conference on Turkish Linguistics*, pages 183–192.

Xuchen Yao, Irina Borisova, and Mehwish Alam. 2010. PDTB XML: the XMLization of the Penn Discourse Treebank 2.0. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC’10).*

Deniz Zeyrek and Bonnie Weber. 2008. A Discourse Resource for Turkish: Annotating Discourse Connectives in the METU Corpus. In *Proceedings of the 6th Workshop on Asian Language Resources, The Third International Joint Conference on Natural Language Processing(IJCNLP).*

Deniz Zeyrek, Ümit Deniz Turan, Işın Demirşahin, and Ruket Çakıcı. (in print). Differential Properties of Three Discourse Connectives in Turkish: A Corpus-based Analysis of Fakat, Yoksa, Ayrıca. In *Anton Benz, Peter Khlein, Manfred Stede (Eds.) Constraints in Discourse III.*

Deniz Zeyrek, Umit Deniz Turan, and Cem Bozşahin. 2008. The Role of Annotation in Understanding Discourse. In *Proceedings of 14th International Conference on Turkish Linguistics*, pages 303–313.

Deniz Zeyrek, Ümit Turan, Cem Bozşahin, Ruket Çakıcı, Ayışığı Sevdik-Çallı, Işın Demirşahin, Berfin Aktaş, İhsan Yalçınkaya, and Hale Ögel. 2009. Annotating Subordinators in the Turkish Discourse Bank. In *Proceedings of the Third Linguistic Annotation Workshop, ACL-IJCNLP 2009*, pages 44–47.

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 *Proceedings of the Fourth Linguistic Annotation Workshop, ACL 2010*, pages 282–289.