

TCL - A Lexicon of Turkish Discourse Connectives

Deniz Zeyrek

Graduate School of Informatics
Middle East Technical University
Ankara, Turkey
dezeyrek@metu.edu.tr

Kezban Başbüyük

Department of Computer Engineering
Middle East Technical University
Ankara, Turkey
kezban.demirtas@metu.edu.tr

Abstract

It is known that discourse connectives are the most salient indicators of discourse relations. State-of-the-art parsers being developed to predict explicit discourse connectives exploit annotated discourse corpora but a lexicon of discourse connectives is also needed to enable further research in discourse structure and support the development of language technologies that use these structures for text understanding. This paper presents a lexicon of Turkish discourse connectives built by automatic means. The lexicon has the format of the German connective lexicon, DiMLex, where for each discourse connective, information about the connective's orthographic variants, syntactic category and senses are provided along with sample relations. In this paper, we describe the data sources we used and the development steps of the lexicon.

1 Introduction

Discourse connectives (alternatively labelled as cue phrases, discourse markers, discourse operators, etc.) are lexical anchors of coherence relations. Such relations (with semantic labels such as expansion, contingency, contrast, concession) can be signalled with discourse connectives, but languages vary in the way they express them. For example, while languages like English and German express discourse relations lexically (with conjunctions and adverbials), Turkish conveys discourse relations through morphological suffixes, as well as lexically. Languages also diverge in the number of connectives that express the same discourse relation. For example, French and Dutch differ in the number of connectives that convey causal relations (Zufferey and Degand, 2017). Finally, discourse connectives are polysemous, expressing several discourse relations. These issues are an obvious challenge for language technologies, translation studies and language learners. What is needed is a resource that goes beyond

traditional dictionaries. Our goal in this paper is to reveal the nature of Turkish discourse connectives through discourse-annotated corpora and describe the steps in constructing a discourse connective lexicon that hosts the connectives' various properties. The Turkish Lexicon will ultimately be part of the connective lexicon database (<http://connective-lex.info/>) that aims to synchronize the lexicons that exist.

The interest in discourse connectives goes hand in hand with the development of discourse-annotated corpora. There are three major approaches that have guided discourse research and inspired other languages to annotate discourse: RST (Mann and Thompson, 1988), SDRT (2012), and the PDTB (Prasad et al., 2014). Our focus in this paper will be the PDTB, one of the best known resources for English discourse. The PDTB takes discourse connectives (henceforth, DCs) as two-place predicates where argumenthood is based on abstract objects (eventualities, facts, propositions, etc.) as in Asher (1993). It annotates the DC together with its binary arguments, which are semantic representations of discourse parts (cf. (Danlos, 2009)). The PDTB-style annotation has been extended to various languages other than English, namely, Arabic (Al-Saif and Markert, 2010), Chinese (Zhou and Xue, 2015), Hindi (Kolachina et al., 2012), and Turkish (Demirşahin and Zeyrek, 2017) as well as a recent multilingual resource, TED-Multilingual Discourse Bank, or TED-MDB (Zeyrek et al., 2019).

In addition to these efforts, there has been an important initiative, namely DiMLex, the discourse connective lexicon first developed for German (Stede and Umbach, 1998; Scheffler and Stede, 2016), which has subsequently been extended to multiple languages, e.g. French (Roze et al., 2012), Italian (Feltracco et al., 2016), Portuguese (Mendes et al., 2018) and recently English (Das et al., 2018). Such lex-

icons are sure to complement the ongoing efforts of discourse-annotated corpora, support discourse research and various language technology applications such as discourse parsers.

In this paper, we describe the development of TCL, a lexicon for Turkish discourse connectives, which follows the format of DiMLex. To the best of our knowledge, there is no such resource for Turkish. Thus, our aim is to fill this gap with a resource that covers Turkish discourse connectives with their various properties and a representation of their meanings. This resource will not only benefit discourse studies in Turkish but will also form the basis of future multilingual studies on discourse connectives and their meanings.

In the rest of this paper, we describe the steps in creating the TCL. In Section 2, we provide information about the data sources we used and in Section 3, we discuss the criteria for selecting connectives as TCL entries. Section 4 presents the structure of TCL and Section 5 shows how the lexicon is populated. Section 6 brings the paper to an end and draws some conclusions.

2 Data Sources

In building the TCL, we use three PDTB-inspired annotated corpora to compile explicit DCs, namely, Turkish Discourse Bank or TDB 1.0 (Zeyrek et al., 2013), TDB 1.1 (Zeyrek and Kurfalı, 2017), and the Turkish section of TED-MDB.

- TDB 1.0 is a 400,000-word resource of modern written Turkish containing annotations of explicit DCs and the discourse segments they relate. It also annotates “phrasal expressions” such as *bunun için* ‘for this (reason/purpose)’, which are linking devices compositionally derived from postpositions (*için* ‘since/in order to’) and a deictic term. They are a subset of the PDTB’s alternative lexicalizations and correspond to “secondary connectives” (Danlos et al., 2018). We used 8439 relations (explicit DCs and “phrasal expressions”) from this corpus.
- TDB 1.1 is a 40,000-word-subset of TDB 1.0, where all five relation types of the PDTB are annotated together with their binary arguments (i.e., explicit and implicit relations, alternative lexicalizations, entity relations and no relations). Based on the PDTB-3 relation hierarchy (Lee et al., 2016), the senses

of explicit and implicit connectives as well as alternative lexicalizations are annotated. We used 912 explicit relations from this corpus.

- TED-MDB is a corpus of TED talks transcripts in 6 languages (English, German, Polish, European Portuguese, Russian and Turkish). We used 276 explicit relations from the Turkish section of this corpus. TDB 1.0 and TED-MDB annotation files are in pipe-delimited format, the TDB 1.1 annotation files are in XML format.

By using different resources, we take advantage of the different coverage of the three corpora. As expected, while some of the connectives exist in all of the data sources, some connectives (and the information needed for the connective lexicon database) may exist in only one source (see Table 2). Moreover, resorting to different corpora is helpful as different corpora may spot new senses of a DC. For example, different senses of the postpositions *gibi* ‘as’ and *kadar* ‘until/as well as/as much as’ have been compiled from different corpora as indicated in Table 1.

3 The criteria for selection of connectives as TCL entries

Turkish is a morphologically rich, agglutinating language with suffixes added to the word root in the order licensed by the morphology and syntax of the language. In this section we describe the major syntactic categories we used to determine DCs, and how we represent suffixal connectives (converbs) in TCL. We also explain our method of determining the syntactic category of other DCs when different POS taggers provide different parses.

TCL only considers explicit discourse connectives annotated in the existing Turkish discourse-annotated corpora. Unlike other DC lexicons such as DimLex and the lexicon of Czech discourse connectives (Mírovský et al., 2017) it does not record non-connective usages.

3.1 Major syntactic categories

DCs are determined on the basis of the following syntactic categories:

- Conjunctions, comprising both the single type *ama* ‘but/yet’ and the paired or noncontinuous type such as *ne ... ne* ‘neither ... nor’.

TCL Entries	TDB 1.1	TED-MDB
gibi ‘as’	EXPANSION: Conjunction EXPANSION: Manner: Arg2-as-manner	COMPARISON: Similarity
kadar ‘until/as well/much as’	COMPARISON: Degree TEMPORAL: Asynchronous: Precedence	COMPARISON: Similarity CONTINGENCY: Purpose: Arg2-as-goal
artık ‘no longer’	EXPANSION: Level-of-detail: Arg2-as-detail TEMPORAL: Synchronous	CONTINGENCY: Cause: Result

Table 1: Different senses of connectives captured via discourse-annotated corpora in Turkish

TCL Entries	Data Sources		
	TDB 1.0	TDB 1.1	TED-MDB
ama ‘but/yet’	ama	ama	ama
çünkü ‘because’	Çünkü	Çünkü	çünkü
aksine ‘in contrast’	aksine		aksine
sadece ‘only’			sadece
sayesinde ‘thanks to’	sayesinde		
keza ‘as well’		Keza	Keza
dahası ‘furthermore’	Dahası	Dahası	

Table 2: TCL entries obtained from various discourse-annotated corpora

- Subordinators:
 - Converbs (simplex subordinators), e.g. –sA, ‘if’, -(y)ArAk ‘by means of/ and’.
 - Postpositions (complex subordinators), which involve an accompanying suffix on the (non-finite) verb of the subordinate clause, *gibi* ‘as’.
- Adverbs, involving single tokens such as *ayrıca* ‘in addition’ as well as phrasal tokens, e.g. *ne var ki* ‘even so’.¹

3.2 Representing suffixal connectives

In Turkish, suffixal connectives are essentially converbs forming non-finite adverbial clauses. Converbs have complex allomorphy based on vowel harmony as well as consonant harmony (Zeyrek and Webber, 2008). We decided that such variation has to be represented in TCL. To illustrate, -(y)ArAk ‘by means of/and’ is a converb shown in the standard morphological notation, where the capital letters indicate alternation (-erek, -arak) and the parentheses show that y is needed if the verb root ends in a vowel (see examples (1), (2), (3)). Other converbs may additionally carry dedicated nominalization markers or person agreement markers, which have different morphological realizations. To identify all occurrences of a converb, the allomorphs need to be specified in

¹We note that the TDB’s term “phrasal expression” is different from the DimLex term “phrasal connective”, which refers to discourse connectives that involve more than one words without specifying the type of words involved in the composition of the connective.

the lexicon. By means of the TCL search tool (see 5.1 below), we specified 15 converbs and their allomorphs to be added to TCL. If any allomorph of a specific converb was missing in the corpora we used, those allomorphs were added manually.

- (1) **Ali okula gid-erek öğretmenle görüştü.**²
‘Ali went to school and talked with the teacher.’
- (2) **Ali sıkı çalış-arak başarı kazandı.**
‘Ali gained success by working hard.’
- (3) **Ali şarkı söyle-yerek başarı kazandı.**
‘Ali gained success by singing.’

3.3 Noun-based connectives

Turkish has a group of connectives which are the lexicalized forms of nominal roots, e.g. *dahası* ‘furthermore’, *amacıyla* ‘with the aim of’, *sonuçta* ‘eventually.’ For this group of connectives, the available POS taggers sometimes provide incomplete information. Table 3 shows different parses provided by different POS parsers for these connectives.

Connectives	UDPipe	TRmorph
dahası ‘furthermore’	Noun	Cnj:adv, Adv
amacıyla ‘for the purpose of’	Noun	Noun
sonuçta ‘eventually’	Adv	Adv, Noun

Table 3: Different parses for three noun-based DCs

In such cases, we compare different sources to determine the connective’s syntactic category for TCL. For example, for the connectives in Table 3, we settled on the syntactic categories provided in Table 4.

²As in the PDTB, Arg2 is the discourse part that hosts the connective and in the examples, it is shown in bold fonts. Arg1 is the other argument and it is rendered in italics. The discourse connective is underlined.

Connectives	TCL
dahası ‘furthermore’	Adv
amacıyla ‘for the purpose of’	Other
sonuçta ‘eventually’	Adv

Table 4: Syntactic categories of three noun-based DCs in TCL

4 The structure of TCL

The TCL structure is based on the structure of the connective lexicon database. Thus, it contains the following components.

- *Orthographical variants*: This criterion specifies whether the connective is a single token (part=single) or a phrasal token (part=phrasal); continuous (orth=cont) or discontinuous (orth=discont). For example, the phrasal connective *ne...ne* ‘neither ... nor’ is annotated as “discont” while the connective *öte yandan* ‘on the other hand’ is annotated as “cont”. An entry illustrating the orthographical variants of the single connective *ama* ‘but’ is provided in Figure 1.

```
<orths>
<orth canonical="0" orth_id="16o1" type="cont">
  <part type="single">ama</part>
</orth>
<orth canonical="1" orth_id="16o1" type="cont">
  <part type="single">Ama</part>
</orth>
<orth canonical="2" orth_id="16o1" type="cont">
  <part type="single">AMA</part>
</orth>
</orths>
```

Figure 1: Variants of *ama* ‘but/yet’

In addition to these, we added the type “suffixal” to TCL to indicate converbs.

- *Canonical Form*: The canonical form of a connective is the most commonly used variant of that connective. For example, the canonical form of *çünkü* ‘because’ is the sentence-initial *Çünkü*, a property which is determined by the TCL search tool.
- *Frequency*: The frequency of the connective shows both how often it occurs in the corpora and the frequency of each of its sense tags.
- *Syntactic category*: The syntactic category of connectives is assigned using several sources as described in Section 5.1, namely the Turkish section of UDPipe³, the search tool pro-

³<http://lindat.mff.cuni.cz/services/udpipe/>

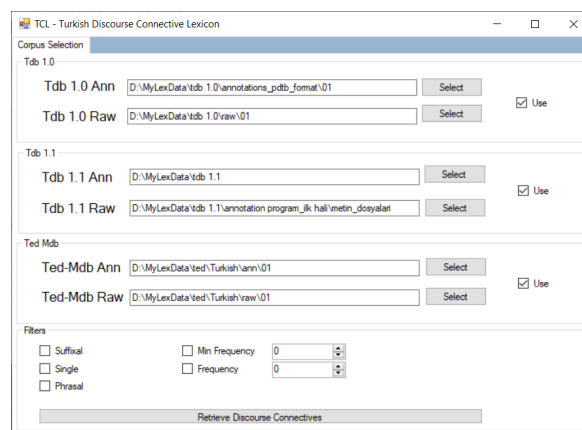


Figure 2: Main window of the TCL search tool

vided in the Turkish Linguistic Society website or TDK⁴ and TRmorph⁵.

5 Populating the lexicon

5.1 The TCL search tool

We developed a search tool to populate TCL. The search tool was developed using the C# programming language. It is also extendable with new features. The main feature of the tool is that it searches different corpora to retrieve DCs (see Figure 2 for a snapshot of the main window of the tool). In addition, it uses filters for DC types, such as suffixal, single, and phrasal. When the search tool is started, file paths used by the tool are specified in the data path window, namely, the path of the text directory specifying the raw text files that will be searched, and the path of the annotation directory containing the XML or pipe-delimited files storing the annotation information.

5.2 The workflow

Using the search tool, we populated the TCL entries. Our work flow involves several steps, as described below and summarized in Figure 3.

- Firstly, the annotation files of the three corpora are parsed and the relations encoded by explicit connectives are retrieved. For this purpose, an XML parser and a pipe-delimited file parser have been developed.
- *Relation Builder*: The Relation Builder module reads the connective and its sense(s) in each relation directly from the annotation files while it reads the respective relation

⁴<http://www.tdk.gov.tr>

⁵<http://coltekin.net/cagri/trmorph/>

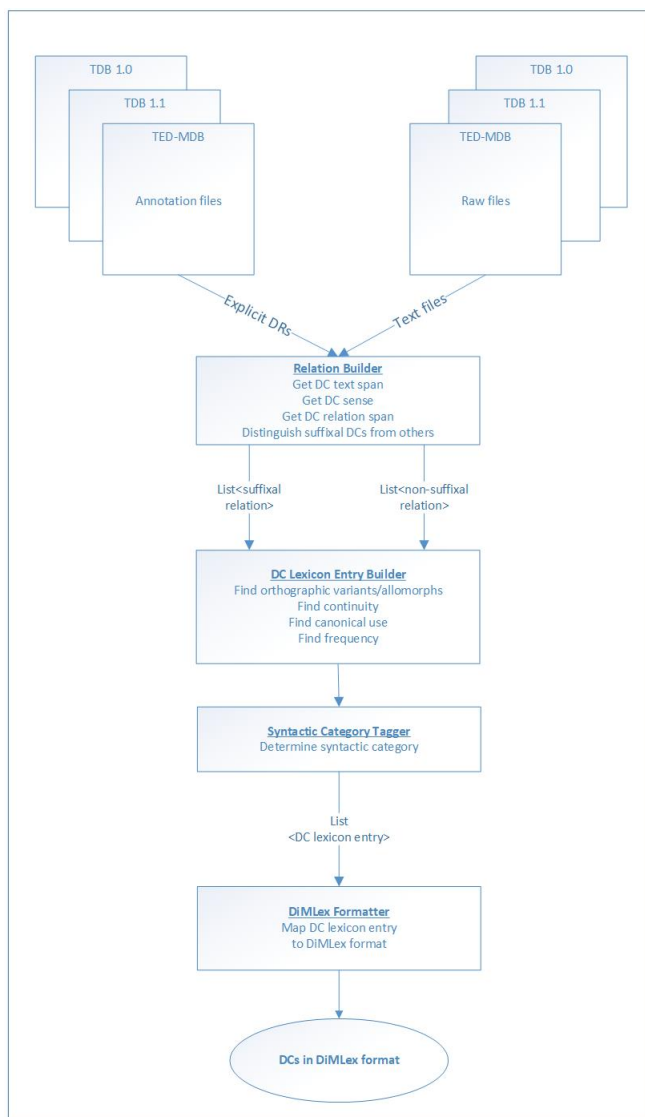


Figure 3: Flow of populating TCL entries

spans from the text files; then it distinguishes suffixal DCs from non-suffixal DCs, i.e. all other types of DCs. To characterize a DC as suffixal, the Relation Builder simply checks the character preceding the DC token. If the previous character is white space, a new line, or a separator (. , ; ! ? () ‘ “ -), the connective is labeled as *Not Suffixal*; in other cases, it is labeled as *Suffixal*. Hence, two lists of relations are formed, a list of relations containing suffixal connectives and a list of relations containing non-suffixal connectives.

- *DC Lexicon Entry Builder*: Using the two lists of relations from the Relation Builder, this module computes the following properties for each DC to be represented in the lexicon:

- *Orthographical Variants*: Since suffixal connectives are uniform in terms of orthography, continuity, etc. they do not have variants. Therefore, the DC Lexicon Entry Builder only searches the list of non-suffixal relations to determine the variants of all connectives. Suffixal connectives are simply tagged as “suffixal”.
- *Allomorphs*: To handle the allomorphs of converbs, the list of suffixal relations is used. The entries of this list are analyzed to find out which entries are allomorphs of a suffixal DC.
- *Continuity*: Phrasal DCs can be continuous or discontinuous. This property is specified automatically by going through the multi-word connectives in the list of non-suffixal relations, and whether there are any words between the two parts of the connective is determined.
- *Canonical Use*: The DC Lexicon Entry Builder counts the times each variant of a DC occurs in our relation lists and labels the most frequently used variant as canonical.
- *Frequency*: To set the frequency property, the DC Lexicon Entry Builder uses both lists of relations and computes the number of occurrences of a DC as well as the number of occurrences of each sense of the DC.

- *Syntactic Category Tagger*: This module assigns a syntactic category to each DC. If a connective is suffixal, it is assigned the converb category. The syntactic category of non-suffixal connectives is determined on the basis of the available POS taggers to the extent possible, otherwise by comparing various parses as described in Section 4 above.

Some of the connectives may belong to more than one syntactic category (Zeyrek and Kurfalı, 2018). For such connectives, we provide both of the syntactic categories. E.g. the DC *önce* ‘before’ is both an adverb (4) and a post-position (5).

- (4) *Ali matematiği iyice anladı. Ama daha önce bir problemi bile*

Connective-Lex	Syntactic Category of Turkish DCs
coc (Coordinating conjunction)	CCONJ
csu (Subordinating conjunction)	Converb, ADP
adv (Adverb of adverbial)	ADV
other	Secondary connective, Other

Table 5: Mapping of TCL syntactic categories onto Connective-Lex categories

yardımsız çözememdi.

Ali has now grasped math fully. But before he could not solve even one math problem without help.'

- (5) **Bu filmi görmeden önce romanını okumalısın.**
'Before seeing this movie, you should read the novel.'

The syntactic categories we assign to the DCs are; CCONJ, Converb (Simplex Subordinator), ADP (Postposition, Complex Subordinator), ADV, Secondary Connective ("phrasal expressions") and Other categories, such as noun-based connectives. These syntactic categories are mapped onto the categories which the Connective-Lex website offers (cf. Table 5).

- *DiMLex Formatter*: After creating the list of DC lexicon entries with all the properties described so far, the entries are mapped onto the DiMLex XML format. Firstly, an XML document is created and for each entry of the lexicon, an XML node is created. The XML elements and attributes are filled with the properties of lexicon entries following the DimLex format.
- *DCs in the DiMLex format*: At the end of these steps, we have a list of 180 DCs with their respective syntactic categories and other properties.

In Figures 4, 5 and 6, we provide how DCs are represented in the DiMLex format. Figure 4 presents the entry of a suffixal connective, Figure 5 shows the entry for a single connective belonging to the postposition category and Figure 6 illustrates the entry for a phrasal discontinuous connective.

We computed the sense distribution of Turkish explicit DCs by using our corpora and compared the results with the sense distribution of explicit

in the PDTB 2.0 (Prasad et al., 2014). Table 6 displays the distribution of top-level classes comparatively and shows that the PDTB 2.0 displays an order of Expansion (33%), Comparison (28.8%), Contingency (19.2%) and Temporal (19%). This distribution is preserved in Turkish to a great extent in the order of Expansion (36%), Contingency (24.4%), Comparison (22.3%), and Temporal (17.3%).

Sense Class	Turkish corpora	PDTB
TEMPORAL	360	3696
CONTINGENCY	507	3741
COMPARISON	463	5589
EXPANSION	748	6423
TOTAL	2078	19449

Table 6: Distribution of top-level sense classes among explicit in the PDTB 2.0 and discourse-annotated corpora of Turkish

Table 7 provides the most frequent 15 discourse connectives and their second-level senses compiled from all data sources.

6 Conclusion

In sum, the major contributions of this paper have been:

- to characterize various properties of Turkish discourse connectives including their syntactic categories and the senses they convey via discourse-annotated corpora,
- to develop a DimLex-style lexicon of discourse connectives to host Turkish discourse connectives together with their various properties and sample relations retrieved from annotated corpora.

TCL is populated by DCs gleaned from texts belonging to different genres. Given that DCs are sensitive to genre (Webber, 2009), in future work, we will compute the distribution of senses in different genres and work on incorporating this information into DiMLex. This aim goes in parallel with our plan of extending the DC search tool with new facilities.

DC	Gloss	Senses	Total
ve	and	Conjunction (395), Cause (39), Cause+Belief (2), Asynchronous (24), Synchronous (8), Level-of-detail (3), Conjunction Level-of-detail (3), Conjunction Contrast (1), Conjunction Synchronous (1), Conjunction Cause (3), Conjunction Instantiation (1)	480
ama	but/yet	Contrast (92), Concession (135), Exception (8), Concession+SpeechAct (8), Correction (6), Cause+SpeechAct (2), Conjunction (3), Concession Synchronous (1), Concession Conjunction (1)	256
için	to/since	Purpose (167), Cause (39), Cause+Belief (3), Degree (2), Level-of-detail (1)	212
sonra	then	Asynchronous (142)	142
çünkü	because	Cause+Belief (17), Cause (76)	85
ancak	however	Concession (36), Exception (4), Contrast (27), Conjunction (1), Exception (1)	69
ayrıca	in addition	Conjunction (41)	41
-ken	while	Synchronous (33), Conjunction (2), Concession+SpeechAct (1), Contrast (1)	37
gibi	as	Conjunction (6), Manner (30), Similarity (1)	37
-(y)HncA	when	Synchronous (19), Cause (6), Asynchronous (10), Level-of-detail (1)	36
-(y)Hp	and	Conjunction (33), Manner (2), Synchronous Conjunction (1)	36
yani	that is	Equivalence (17), Level-of-detail (4), Cause+Belief (10), Substitution (3), Cause+SpeechAct (1)	35
-sA	if	Condition (23), Concession (2), Negative-condition (4), Condition Purpose (1), Condition+SpeechAct (3), Substitution (1)	34
-dA	when	Synchronous (29), Condition (1)	30
önce	before	Asynchronous (30)	30

Table 7: 15 most frequent discourse connectives and their second-level sense distribution in discourse-annotated corpora

```

<entry id="6" word="- (y)Hp">
  <orths>
    <orth canonical="0" orth_id="6o1" type="cont">
      <part type="suffixal">ip</part>
    </orth>
    <orth canonical="1" orth_id="6o1" type="cont">
      <part type="suffixal">ip</part>
    </orth>
    <orth canonical="2" orth_id="6o1" type="cont">
      <part type="suffixal">up</part>
    </orth>
    <orth canonical="3" orth_id="6o1" type="cont">
      <part type="suffixal">üp</part>
    </orth>
    <orth canonical="4" orth_id="6o1" type="cont">
      <part type="suffixal">yip</part>
    </orth>
    <orth canonical="5" orth_id="6o1" type="cont">
      <part type="suffixal">yip</part>
    </orth>
    <orth canonical="6" orth_id="6o1" type="cont">
      <part type="suffixal">yup</part>
    </orth>
    <orth canonical="7" orth_id="6o1" type="cont">
      <part type="suffixal">yüp</part>
    </orth>
  </orths>
  <syn>
    <cat>Converb</cat>
    <sem>
      <pdtb2_relation anno_N="36" freq="33" sense="EXPANSION:Conjunction" />
      <example>hazırlanıp arabaya bindi.</example>
    </sem>
    <sem>
      <pdtb2_relation anno_N="36" freq="2" sense="EXPANSION:Manner:Arg2-as-manner" />
      <example>Oğlum oturup yesene,</example>
    </sem>
    <sem>
      <pdtb2_relation anno_N="36" freq="1" sense="TEMPORAL:Synchronous" />
      <example>bu tarafa çevirip, 6 milyar kilometre öteden Dünya'nın fotoğrafını çektirdiler.</example>
    </sem>
  </syn>
</entry>

```

Figure 4: A suffixal connective -(y)Hp ‘and’, the senses it conveys and representative examples

```

<entry id="34" word="gibi">
  <orths>
    <orth canonical="0" orth_id="34o1" type="cont">
      <part type="single">gibi</part>
    </orth>
  </orths>
  <syn>
    <cat>ADP</cat>
    <sem>
      <pdtb2_relation anno_N="37" freq="6" sense="EXPANSION:Conjunction" />
      <example>kahve deđirmenininde nerede olduđunu bilmediđi gibi bulacađını da sanmıyordu.</example>
    </sem>
    <sem>
      <pdtb2_relation anno_N="37" freq="30" sense="EXPANSION:Manner:Arg2-as-manner" />
      <example>beni bylenmiř gibi dinliyorlardı.</example>
    </sem>
    <sem>
      <pdtb2_relation anno_N="37" freq="1" sense="COMPARISON:Similarity" />
      <example>adeta ok kk bir kasabada yařıyormuřuz gibi gzkyor.</example>
    </sem>
  </syn>
</entry>

```

Figure 5: A single connective *gibi* ‘as’, the senses it conveys and representative examples

```

<entry id="50" word="ya ... ya da">
  <orths>
    <orth canonical="0" orth_id="50o1" type="discont">
      <part type="phrasal">ya ya da</part>
    </orth>
    <orth canonical="1" orth_id="50o1" type="discont">
      <part type="phrasal">Ya ya da</part>
    </orth>
  </orths>
  <syn>
    <cat>CCONJ</cat>
    <sem>
      <pdtb2_relation anno_N="2" freq="2" sense="EXPANSION:Disjunction" />
      <example>ya tamamen Yok olmuřlar, ya da cins ve tr dzeyinde nemli lde azalmıřlardır.</example>
    </sem>
  </syn>
</entry>

```

Figure 6: A phrasal connective *ya ... ya da* ‘either ... or’, its sense and a representative example

References

- Amal Al-Saif and Katja Markert. 2010. The Leeds Arabic Discourse Treebank: Annotating Discourse Connectives for Arabic. In *LREC*, pages 2046–2053.
- Nicholas Asher. 1993. Reference to abstract objects in English.
- Nicholas Asher. 2012. *Reference to abstract objects in discourse*, volume 50. Springer Science & Business Media.
- Laurence Danlos. 2009. D-STAG: a formalism for discourse analysis based on SDRT and using synchronous TAG. In *International Conference on Formal Grammar*, pages 64–84. Springer.
- Laurence Danlos, Katerina Rysova, Magdalena Rysova, and Manfred Stede. 2018. Primary and secondary discourse connectives: Definitions and lexicons. *Dialogue and Discourse*, 9:50–78.
- Debopam Das, Tatjana Scheffler, Peter Bourgonje, and Manfred Stede. 2018. Constructing a lexicon of English discourse connectives. In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*, pages 360–365.
- Iřın Demirřahin and Deniz Zeyrek. 2017. Pair annotation as a novel annotation procedure: The case of Turkish Discourse Bank. In *Handbook of Linguistic Annotation*, pages 1219–1240. Springer.
- Anna Feltracco, Elisabetta Jezek, Bernardo Magnini, and Manfred Stede. 2016. Lico: A lexicon of Italian connectives. *CLiC it*, page 141.
- Sudheer Kolachina, Rashmi Prasad, Dipti Misra Sharma, and Aravind K Joshi. 2012. Evaluation of Discourse Relation Annotation in the Hindi Discourse Relation Bank. In *LREC*, pages 823–828.
- Alan Lee, Rashmi Prasad, Bonnie Webber, and Aravind K Joshi. 2016. Annotating discourse relations with the PDTB annotator. In *Proceedings of COLING 2016, the 26th International Conference*

- on *Computational Linguistics: System Demonstrations*, pages 121–125.
- William C Mann and Sandra A Thompson. 1988. Rhetorical Structure Theory: Toward a functional theory of text organization. *Text-Interdisciplinary Journal for the Study of Discourse*, 8(3):243–281.
- Amália Mendes, Iria del Rio, Manfred Stede, and Felix Dombek. 2018. A lexicon of discourse markers for Portuguese–ldm-pt. In *11th International Conference on Language Resources and Evaluation*, pages 4379–4384.
- Jiří Mírovský, Pavlína Synková, Magdaléna Rysová, and Lucie Poláková. 2017. CzeDLex—a lexicon of Czech discourse connectives. *The Prague Bulletin of Mathematical Linguistics*, 109(1):61–91.
- Rashmi Prasad, Bonnie Webber, and Aravind Joshi. 2014. Reflections on the Penn Discourse Treebank, comparable corpora, and complementary annotation. *Computational Linguistics*, 40(4):921–950.
- Charlotte Roze, Laurence Danlos, and Philippe Muller. 2012. Lexconn: a French lexicon of discourse connectives. *Discours. Revue de linguistique, psycholinguistique et informatique. A journal of linguistics, psycholinguistics and computational linguistics*, (10).
- Tatjana Scheffler and Manfred Stede. 2016. Adding Semantic Relations to a Large-Coverage Connective Lexicon of German. In *LREC*.
- Manfred Stede and Carla Umbach. 1998. Dimlex: A lexicon of discourse markers for text generation and understanding. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 2*, pages 1238–1242. Association for Computational Linguistics.
- Bonnie Webber. 2009. Genre distinctions for discourse in the Penn TreeBank. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*, pages 674–682. Association for Computational Linguistics.
- Deniz Zeyrek, Isin Demirsahin, A Sevdik-Çallı, and Ruket Çakıcı. 2013. Turkish Discourse Bank: Porting a discourse annotation style to a morphologically rich language. *D&D*, 4(2):174–184.
- Deniz Zeyrek and Murathan Kurfalı. 2017. Tdb 1.1: Extensions on Turkish discourse bank. In *Proceedings of the 11th Linguistic Annotation Workshop*, pages 76–81.
- Deniz Zeyrek and Murathan Kurfalı. 2018. An assessment of explicit inter-and intra-sentential discourse connectives in Turkish Discourse Bank. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*.
- Deniz Zeyrek, Amália Mendes, Yulia Grishina, Murathan Kurfalı, Samuel Gibbon, and Maciej Ogrodniczuk. 2019. TED Multilingual Discourse Bank (TED-MDB): A parallel corpus annotated in the PDTB style. *Language Resources and Evaluation*, pages 1–27.
- Deniz Zeyrek and Bonnie Webber. 2008. A discourse resource for Turkish: Annotating discourse connectives in the METU corpus. In *Proceedings of the 6th workshop on Asian language resources*.
- Yuping Zhou and Nianwen Xue. 2015. The Chinese Discourse Treebank: a Chinese corpus annotated with discourse relations. *Language Resources and Evaluation*, 49(2):397–431.
- Sandrine Zufferey and Liesbeth Degand. 2017. Annotating the meaning of discourse connectives in multilingual corpora. *Corpus Linguistics and Linguistic Theory*, 13(2):399–422.