AUTOMATIC SENSE PREDICTION OF IMPLICIT DISCOURSE RELATIONS IN TURKISH

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS
OF
MIDDLE EAST TECHNICAL UNIVERSITY

MURATHAN KURFALI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
COGNITIVE SCIENCE

SEPTEMBER 2016

Approval of the thesis:

AUTOMATIC SENSE PREDICTION OF IMPLICIT DISCOURSE RELATIONS IN TURKISH

submitted by **MURATHAN KURFALI** in partial fulfillment of the requirements for the degree of **Master of Science in Cognitive Science, Middle East Technical University** by,

Prof. Dr. Nazife Baykal
Director, **Graduate School of Informatics**

Assist. Prof. Dr. Cengiz Acartürk
Head of Department, **Cognitive Science, METU**

Prof. Dr. Deniz Zeyrek Bozşahin
Supervisor, **Cognitive Science**

Examining Committee Members:

Prof. Dr. Cem Bozşahin
Cognitive Science Department, METU

Prof. Dr. Deniz Zeyrek Bozşahin
Cognitive Science Department, METU

Assist. Prof. Dr. Burcu Can Buğlalılar
Department of Computer Engineering, Hacettepe University

Assoc. Prof. Dr. Annette Hohenberger
Cognitive Science Department, METU

Assist. Prof. Dr. Umut Özge
Cognitive Science Department, METU

Date:

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: MURATHAN KURFALI

Signature :

ABSTRACT

AUTOMATIC SENSE PREDICTION OF IMPLICIT DISCOURSE RELATIONS IN TURKISH

Kurfalı, Murathan
M.S., Department of Cognitive Science
Supervisor : Prof. Dr. Deniz Zeyrek Bozşahin

September 2016, 61 pages

In discourse parsing, the sense prediction of the Implicit discourse relations poses the most significant challenge. The thesis aims to develop a supervised system to predict the sense of implicit discourse relations in Turkish Discourse Bank (TDB). In order to accomplish that goal, the discourse level annotations obtained from TDB are used. TDB follows the PDTB-2's sense hierarchy and for all experiments within the current study, only CLASS senses are considered. As the primary experiment, the classifiers are trained on merely implicit discourse relations based on the several linguistically informed features, such as polarity and tense information, to detect the possible sentence structures characteristic to each CLASS level sense. In the secondary experiment, the effect of Explicit discourse relations on the sense prediction of Implicit relations is investigated. The motivation behind this experiment is to provide insight regarding the differences and similarities of these two type of discourse relations which is another challenging topic in the discourse research. The results indicate that implicit discourse relations manifest significant differences in terms of their sentence structure depending on their sense. It is also revealed that using Explicit discourse relations alters the performance of the classification radically which suggests that these two type of the discourse relations are structurally different from each other.

Keywords: discourse, implicit discourse relations, supervised learning, turkish discourse bank, automatic sense prediction

ÖZ

TÜRKÇE' DE ÖRTÜK BAĞLAÇLARIN OTOMATİK OLARAK BELİRLENMESİ

Kurfalı, Murathan
Yüksek Lisans, Bilişsel Bilimler Programı
Tez Yöneticisi : Prof. Dr. Deniz Zeyrek Bozşahin

Eylül 2016 , 61 sayfa

Söylem çözümlemesinde örtük söylem ilişkilerinin anlamının belirlenmesi önemli bir engel oluşturmaktadır. Bu tezin amacı Türkçe Söylem Bankası'nda (TDB) örtük bağlaçların anlamlarını belirleyebilecek bir gözetimli model geliştirmektir. Bu hedefi gerçekleştirmek için, TDB' den elde edilen söylem düzeyindeki işaretlemeler kullanıldı. TDB PDTB-2' nin anlam hiyerarşisini takip etmektedir ve yapılan bütün deneylerde, sadece SINIF düzeyindeki anlamlar dikkate alınmıştır. Birincil deney olarak, sınıflandırıcılar, çeşitli dilbilimsel özelliklere göre , zaman ve kutupluluk bilgisi gibi, SINIF düzeyindeki anlamlara özgü olası cümle yapılarını saptamak için sadece örtük söylem ilişkileri üzerinde eğitilmiştir. İkincil deneyde ise açık söylem ilişkilerinin örtük ilişkilerin anlamının belirlenmesindeki etkisi araştırılmıştır. Bu deneyin arkasındaki motivasyon, söylem çalışmalarındaki başka bir araştırma olan, bu iki tür söylem ilişkisinin farklılıkları ve benzerlikleri hakkında içgörü sağlamaktır. Sonuçlar, örtük söylem ilişkilerinin, anlamlarına göre, cümle kurgularında ciddi farklılıklar olduğunu göstermiştir. Ayrıca, açık söylem ilişkilerinin, sınıflandırma performansını ciddi şekilde değiştirmiştir ki bu da açık ve örtük söylem ilişkilerinin yapısal olarak birbirinden farklı olduğunu ortaya koymaktadır.

Anahtar Kelimeler: söylem, örtük söylem ilişkileri, gözetimli öğrenme, türkçe söylem bankası, otomatik anlam belirlenmesi

To my family

ACKNOWLEDGMENTS

Foremost, I would like to extend my sincere gratitude to my advisor Prof. Dr. Deniz Zeyrek Bozşahin for her continuous support, immense knowledge, patience and for all the invested time in the supervision of my thesis. She has always encouraged me to study on discourse and without her guidance the thesis would have never existed.

I thank Prof. Dr. Cem Bozşahin for his support, motivation and being one of the primary reasons for my interest in language and cognition.

I should express my gratitude to Dr. Ali Arifoğlu for his never ending support, understanding and valuable insights regarding academic and social life. I feel lucky to work with him in the E-Government Research and Development Center where I have had the chance to expand my knowledge.

I must acknowledge that I am fortunate enough to work with Assist. Prof. Dr. Burcu Can Buğlalılar. I have gained a lot from her vast knowledge which she never hesitates to share. In less than a year, she has taught me more than I could have learned in decades on my own.

Word would fail me, if I were to express my gratitude to Merve Akdemir. Instead of wasting time on such a fruitless effort, I can only thank you for being around and express my interest to remain thankful to you for the rest of my life.

I should express my gratitude to Savaş Çetin, Faruk Büyüktekin, Evren Aykaç and Ahmet Üstün for their support and priceless friendship. Directly or indirectly, they have contributed a lot and are one of the primal factors that have made this thesis possible.

I would love to thank Tunç Güven Kaya, Ege Saygıner, Ece Kamer Takmaz for their support and especially for the provocative discussions which taught me a lot and taught expanded my horizon in every aspect possible.

I would love to thank to Hakan Güler, Sibel Gülnar, Ali Kantar, Yunus Canatan and the rest of the informatics family for being always extremely supportive and friendly. Thanks to them, the informatics institute never felt much different than home.

I wish to express my deepest gratitude to my family for their love and courage throughout my life.

Last but not the least, I would like to thank Prof. Dr. Varol Akman for recommending me to apply for cognitive science. Without him, who knows where I would have been wandering right now.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Among his other accomplishments, William S. Burroughs (1914 – 1997) is widely recognized as the inventor of the *cut-up technique* [1] where the author rearranges the text by splitting the text in numerous pieces and glues them to each other in a random fashion. The following passage is taken from the Burroughs' *The Soft Machine (1961; p.11)* :

> He went to Madrid. . . Alarm clock ran for yesterday. . . "No me hágas casa." Dead on arrival. . . you might say at the Jew Hospital. . . blood spilled over the American. . . trailing lights and water. . . The Sailor went so wrong somewhere in that grey flesh. . . He just sit down on zero... I nodded on Niño Perdido his coffee over three hours late. . . They all went away and sent papers. . . The Dead Man write for you like a major, . . Enter vecinos. . .

In terms of Discourse analysis, what the cut-up technique disturbs is called the *coherence* of the text. That is to say, the incremental reading of the sentences does not offer more than what reading the sentences in isolation would. Coherence is accepted as the main distinction between a set of randomly uttered (or written) sentences and a well-connected piece of text, i.e discourse.

Discourse is one of the most attractive yet challenging topics both in linguistics and Natural Language Processing (NLP) due to the its hidden nature unlike other levels of the linguistics such as morphology and syntax. By the hidden nature, what is meant is the difficulty of recognizing/ describing the fundamental concepts of discourse such as the its minimal unit or its structure. That is, we know that it is coherence which makes a text a discourse but we are not sure what makes a discourse coherent. In dealing with that question, *discourse relations*, among others such as lexical relations, co-references etc., stand out as a relevant notion.

Although the definition of the discourse relations differ among the theories, there are basically two main type of discourse relations: Explicit and Implicit [2]. Briefly, the difference between Explicit and Implicit discourse relations is the existence of a discourse connective. In the simple Example 1, the relation that holds between the first pair of sentences are signaled by the discourse connective "because", therefore it is called an Explicit discourse relation; whereas,the second example lacks such a connective and the causal relation between the sentences is understood through inference.

[1] inspired by the dadaist movement

[2] The other types of the discourse relations can be regarded as a sub-category of either of these two types.

(1) (i) I did not order coffee because I like tea better.

(ii) I did not order coffee. I like tea.

Discourse relations are named after their *senses* [3]. The sense of a discourse relation can be described as the semantic relation between the arguments of the relation. For example, the sense of the discourse relation in the Example 1 is "reason" since the second clause provides the "reason" of the act described in the first sentence. The set of all possible senses are yet to be agreed on and differ non-trivially among theories.

1.1 Thesis

The thesis examines the automatic means to predict the sense of Implicit discourse relations. By doing so, it is aimed to reveal the syntactic features which may, even partly, give rise to the sense of the discourse relations. Therefore, to some extent, the thesis aims to investigate the relation between the syntactic configuration of the sentences and the meaning at the discourse level. In order to accomplish that goal, several linguistically rich features have been implemented. As the data, the output of the recent sense annotation effort on Turkish Discourse Bank (TDB) as well as the additional annotations done by the author and the supervisor of thesis are used. Therefore, all the data utilized in the thesis are obtained through human annotators which is an important aspect since the classifier built within the thesis learns from the human input.

In addition to the Implicit discourse relations, the Explicit discourse relations, sense of which are labeled manually is also added to the training data to see what kind effect of the they may produce on the sense prediction task of the Implicit relations.

Implicit and Explicit discourse relations have been annotated in a subcorpus of TDB which constitutes the 10% of the whole corpus, which is shortly called TDB-Subcorpus. Implicit discourse relations may hold between VPs, clauses, sentences, or even across paragraphs. TDB currently annotates only inter-sentential Implicit discourse relations which refer to the Implicit relations held between two adjacent sentences[4]. Therefore, the thesis is limited to inter-sentential Implicit discourse relations.

1.2 Motivation

The researches conducted by the TDB group constitutes the greater portion of the studies on Turkish discourse (Zeyrek & Webber, 2008; Aktaş et al., 2010; Zeyrek et al., 2013). Those researches include producing discourse level annotations to be used in computational studies such as discourse parsing. However, to date, automatic discourse parsing in Turkish has not received any attention. This thesis is aimed at dealing with this shortcoming. Among other tasks in discourse parsing[5], sense prediction of the Implicit discourse relations are chosen to

[3] There is only one exception which is Entity-based relation

[4] The opposite of the inter-sentential relation is the intra-sentential relations where the arguments of the discourse relations are located in the same sentence.

[5] These are detailed in Section 2.4

be studied within this thesis since that task constitutes the bottleneck of the discourse parsing procedure(Lin et al., 2014; Biran & McKeown, 2015; J. Wang & Lan, 2015).

Another challenging question in discourse studies concerns the usability of Explicit relations in predicting the sense of implicit relations. In order to provide more insight regarding the nature of these two types of discourse relations, a separate set of experiments is conducted. In those experiments, an original way to use Explicit discourse relations is proposed and the effect of Explicit discourse relations on sense prediction of the Implicit discourse relations is investigated. In other words, this experiment is undertaken to understand whether the linguistic features relevant for Explicit discourse relations are also relevant for implicit relations.

1.3 Outline

The thesis consists of five main chapters: Literature Review, Data, Methodology, Results and Discussion, Conclusion.

The Literature Review chapter has two main sections. The first section gives detailed information regarding discourse relations and how they are defined in the prevalent theories as well as the widely used discourse resources such Penn Discourse Tree Bank (PDTB). The second section deals with the computational aspect of the discourse, *discourse parsing*. In this section, the tasks a discourse parser should perform are explained and the previous studies on discourse parsing and sense prediction of the Implicit discourse relations are reviewed.

The third chapter details the annotation procedure in the recent sense annotation effort undertaken by the TDB research group. The fourth chapter is devoted to explain the methodology followed in building the classifiers.

The Results and Discussion chapters summarize the findings and discuss the possible implications of the results in terms of discourse research. The final chapter concludes the thesis and draws some generalizations.

CHAPTER 2

LITERATURE REVIEW

This section, firstly, gives a brief account of what discourse relations are, how they are realized in the discourse and how they are classified in various theories in order to understand the place of *discourse relations* in the study of discourse. Then, the Penn Discourse Tree Bank, one of the most commonly used discourse resources, and the other well known discourse sources which follows PDTB schema, such as Hindi Discourse Relational Treebank or Chinese Discourse Treebank, is summarized in order to better understand the Turkish Discourse Bank and its stance in the discourse studies. Lastly, discourse parsing and the recent developments in the automatic sense prediction of the discourse relations are reviewed.

2.1 Discourse Relations

Coherence is often attributed as the key element which differentiates the discourse from a random set of sentences Knott & Sanders (1998). In the domain of discourse, coherence can by identified as the "semantic property of discourse, based on the interpretation of each individual sentence relative to the interpretation of other sentences" (Van Dijk, 1980). For example, in the Example 2, the sequential reading of the sentence pair in (i) conveys an observation, a claim perhaps, as well as its justification; whereas, sentences in (ii) do not seem to be connected in any sense. Therefore, providing a systematic account in order to understand what is that makes a set of sentences coherent but not the others is one of the fundamental questions in discourse studies.

(2) (i) Tim must love that Belgian beer. The Crate in the hall is already half empty.
    (ii) Tim must love that Belgian Beer. He's six foot tall. (Knott & Sanders, 1998)

In dealing with such a question, both sentences' being about the same entity does not seem to be a valid answer, as we can see in the (ii); although both sentences are about *Tim*, there seems to be no connection among them. Those connections are argued to be "captured in the form of coherence relations" (Taboada, 2009). Such relations are also called *discourse relations* or *rhetorical relations* in the literature [1]. Three main questions regarding discourse relations can be asked in order to comprehend their nature better:

(i) *What do discourse relations relate?:* Discourse relations hold among two or more discourse units; however, the definition of discourse unit is one of the most challenging

[1] Throughout the thesis, the term *discourse relation* is preferred among others

issues in discourse studies. There is no universally accepted definition of the minimal discourse units; some authors settle with vague definitions as "typically clauses" whereas others define it in favor of automated purposes (Degand et al., 2005). In practice, phrases which denote events or propositions are regarded to be minimal discourse units, hence, arguments of the discourse relations (see Section 2.3.1).

(ii) *What is the possible set of discourse relations?:* The set of discourse relations may be even more problematic than the definition of the discourse unit. Here, the distinction among theories are sharper; the number of relations proposed varies in number, from 2 to 100, or in source, they are defined in terms of semantics or intentions (Knott & Sanders, 1998). In section 2.2, how various prevalent discourse theories regard discourse relations are explained to some detail [2].

(iii) *Is there a structure among discourse relations?:* Although, Hobbs takes it is a fact that discourse has a structure (1985), what kind of structure it has is controversial. Most of the theories assume that discourse has a tree structure, Theory of coherence relations (Hobbs, 1985), Rhetorical Structure Theory (RST) (Mann & Thompson, 1988), Discourse - Lexicalized Tree Adjoining Grammar (D-LTAG) (Webber, 2004) to name a few, there are others who do not confine discourse with the tree structure and offers graph-based structures such as Wolf& Gibson (2005).

2.1.1 Explicit vs. Implicit Discourse Relations

Discourse relations can be realized in different forms. Generally, discourse relations are realized through *discourse connectives*. Discourse connectives may belong different syntactic groups[3] such as conjunctions (and, or, but) or adverbs (however, instead).

On the other hand, an inspection of the discourse relations in a natural discourse reveal that it is more than likely that (see Table 2.1) the relations are not signaled explicitly. The discourse relations where the explicit connective is missing between the arguments are called *Implicit discourse relations*. In these relations, the discourse relations is inferred by the sequential reading of the discourse segments. It is a common practice to annotate Implicit discourse relations by inserting a suitable discourse connective, which are called *implicit connectives*.

2.2 Discourse Relations in Various Theories

Discourse relations constitute a significant part in the discourse theories. Below, various prevalent theories' take on the discourse relations are provided [4].

2.2.1 Theory of Coherence Relations

Hobb's groups discourse relations under four main classes based on the following observation that in a discourse between a speaker and a listener: "(a) The speaker wants to convey a

[2] Also, (Hovy, 1990) can be referred for a detailed documentation of the proposed discourse relations
[3] which will be explained in a greater detail through the Section 2.2
[4] Although most of the theories also posit a discourse structure, those parts are omitted since discourse structure does not fall within the scope of the thesis

message. (b) The message is in service of some goal. (c)The speaker must link what he says to what the listener already knows. (d) The speaker should ease the listener's difficulties in comprehension" (Hobbs, 1985). From this analysis, Hobbs identifies the following categories:

1. *Occasion relations:* In these relations one of the segments depicts a change of state which can be in location, value, mental state etc.
2. *Evaluation relations:* These relations hold between the discourse units where one of the units allows us to infer the goal/plan set in the other segment as in *"Did you bring your car today? My car is at the garage."* (Hobbs, 1985)
3. *Background/ Explanation relations:* In these relations one of the segments relates to the interlocutor's prior knowledge.
4. *Expansion relations:* These relations expand the discourse in place instead of filling the background. This set corresponds to the largest set of relations among the four classes and involves the inferential relations which are *parallel, generalization, exemplification, contrast, violated expectation relations*.

2.2.2 Rhetorical Structure Theory (RST)

Rhetorical Structure Theory (RST) is proposed by Mann and Thompson and identifies the discourse structure as a single tree over the discourse relations(1988). RST requires discourse units to be adjacent in order to hold a relation and differentiates the discourse units in the discourse relations with respect to their contribution to the meaning: the central part of the relation, which can be interpreted independently, is called *the nucleus*, whereas the other part is *the satellite*. Depending on the types of the constituents, a discourse relation can be either *mononuclear*, where there is one nucleus and one satellite, or *multinuclear* where there are more units which bear equal importance.

In RST, a discourse relation is defined using four fields. Below, those fields and what they correspond in the *Background* relation are given for illustration purposes Mann & Thompson (1988):

(i) *Constraints on the Nucleus N:* Reader R won't comprehend N sufficiently before reading text of S.

(ii) *Constraints on the Satellite S:* (none)

(iii) *Constraints on both (Nucleus and Satellite):* S increases the ability off R to comprehend and element in N

(iv) *The Effect:* R's ability to comprehend N increases

In the original paper Mann and Thompson provides 23 different relations in two categories: Subject Matter and Presentational (Figure 2.1). Subject matter relations are those whose intended effect on the reader is that "reader recognizes the relation in questions" whereas the presentational relations seek to "increase some inclication in the reader such the desire to act" Mann & Thompson (1988).

Subject Matter	Presentational	
Elaboration	Motivation	(increases desire)
Circumstance	Antithesis	(increases positive regard)
Solutionhood	Background	(increases ability)
Volitional Cause	Enablement	(increases ability)
Volitional Result	Evidence	(increases belief)
Non-Volitional Cause	Justify	(increases acceptance)
Non-Volitional Result	Concession	(increases positive regard)
Purpose		
Condition		
Otherwise		
Interpretation		
Evaluation		
Restatement		
Summary		
Sequence		
Contrast		

Figure 2.1: Original Set of Rhetorical Relations in RST (Mann & Thompson, 1988)

2.2.3 Cognitive Theory of Discourse Representation

Cognitive Theory of Discourse Representation is proposed by Sanders et al. (1992) and takes a slightly different view on discourse relations. Rather than identifying the discourse relation itself, it tries to find out the fundamental components, called *primitives*, which give rise to the sense of the discourse relations. The discourse relations are generated by combining these primitives (see Figure 2.2). According to Cognitive Theory of Discourse Representation, there are four main primitives:

(i) *Basic Operation:* There are two operations; causality and addition. According to Cognitive Theory of Discourse Representation, an addition relation exists between the discourse segments P and Q if the conjunction relation $P \ \& \ Q$ can be deduced; whereas, the causal relation entails the existence of the implication relation, that is $P \rightarrow Q$ where P is antecedent and Q is consequent.

(ii) *Source of Coherence:* This primitive has two values; semantic and pragmatic. As is evident from the names, the discourse segments P and Q are related pragmatically if the relation holds between the illocutionary meaning of one or both of the segments. On the other hand, in semantic relations, the relation exists between the locutionary meanings.

(iii) *Order of the Segments:* A relation which is in basic order if the discourse segment P precedes the segment Q and they are in either $P \ \& \ Q$ or $P \rightarrow Q$ basic operation. Similarly, the relation is said to be in non-basic order if the segment Q precedes the segment P, yet they are still in either $P \ \& \ Q$ or $P \rightarrow Q$ basic operation. Therefore, the basic operation is the determinant for the order of a discourse relation.

(iv) *Polarity:* A relation is said to be *positive* if the discourse segments P and Q functions in the basic operation and is said to *negative* if either not-P or not-Q functions in the basic operation.

Basic Operation	Source of Coherence	Order	Polarity	Class	Relation
Causal	Semantic	Basic	Positive	1.	Cause–consequence
Causal	Semantic	Basic	Negative	2.	Contrastive cause–consequence
Causal	Semantic	Nonbasic	Positive	3.	Consequence–cause
Causal	Semantic	Nonbasic	Negative	4.	Contrastive consequence–cause
Causal	Pragmatic	Basic	Positive	5a.	Argument–claim
				5b.	Instrument–goal
				5c.	Condition–consequence
Causal	Pragmatic	Basic	Negative	6.	Contrastive argument–claim
Causal	Pragmatic	Nonbasic	Positive	7a.	Claim–argument
				7b.	Goal–instrument
				7c.	Consequence–condition
Causal	Pragmatic	Nonbasic	Negative	8.	Contrastive claim–argument
Additive	Semantic	—	Positive	9.	List
Additive	Semantic	—	Negative	10a.	Exception
				10b.	Opposition
Additive	Pragmatic	—	Positive	11.	Enumeration
Additive	Pragmatic	—	Negative	12.	Concession

Figure 2.2: Set of Discourse Relations in Cognitive Theory of Discourse Representation (Sanders et al., 1992)

2.3 Discourse Resources

2.3.1 Penn Discourse Tree Bank (PDTB)

Penn Discourse Tree Bank (henceforth PDTB) project started to meet the need of richer annotations required by NLP applications (Miltsakaki et al., 2004). PDTB follows the lexically grounded approach to annotate discourse relations. The discourse relations are searched between two Abstract Objects following the definition of (Asher, 1993) (Figure 2.3). PDTB consists of the manual annotation of the 1 million word Wall Street Journal (WSJ) Corpus for both Explicit and Implicit discourse relations with their two arguments and their senses as well as AltLex, EntRel and NoRel relations, which are used if a suitable Implicit connective cannot be provided.

The set of senses used in PDTB is organized hierarchically and given in Figure 2.4[5] (Prasad et al., 2008). In PDTB hierarchy, the senses which are at the highest level of the sense hierarchy are called "CLASS". The senses at the second level of the hierarchy are called "Type" and those who are located at the lowest level of the hierarchy are called "subtypes". The relation distribution of PDTB is given in Table 2.1. PDTB 2.0 is available through the Linguistic Data Consortium [6].

A sample annotation of an Explicit discourse relation is given in Example 3. The argument which is syntactically bound to the connective is marked as Arg2, whereas the other argument is identified as Arg1. The explicit connectives annotated in PDTB can be realized as one of the following four syntactic classes (Prasad et al., 2014):

- *Subordinating conjunctions:* because, although, when, if, as, etc.

[5] The sense hierarchy is elaborated in Section 3.1.2
[6] https://catalog.ldc.upenn.edu/LDC2008T05

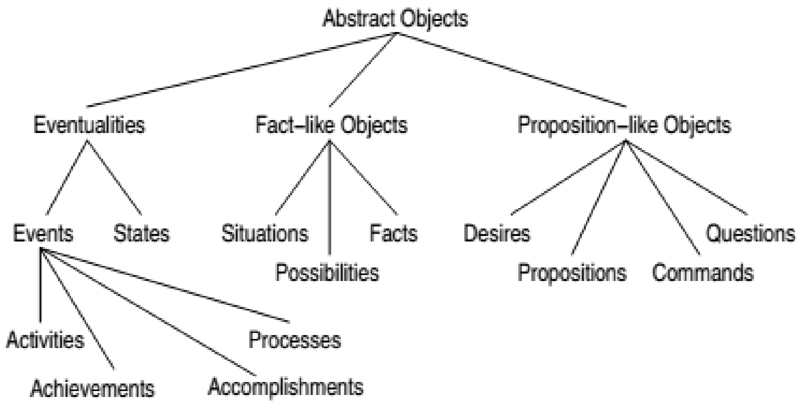

Figure 2.3: Hierarchy of Abstract Objects(Asher, 1993)

Table 2.1: Distribution of Relations in PDTB 2.0 (Prasad et al., 2008)

PDTB Relations	No. of tokens
Explicit	18459
Implicit	16224
AltLex	624
EntRel	5210
NoRel	254
TOTAL	**40600**

- *Coordinating conjunctions:* and, but, so, nor, or
- *Prepositional phrases*: as a result, in comparison, on the one hand
- *Adverbs:* then, however, instead, yet, likewise, subsequently, etc.

(3) Michelle lives in a hotel room, and although, **she drives a canary-colored Porsche,** *she hasn't time to clean or repair it*[7] (Prasad et al., 2008)

For the Implicit discourse relations, the annotators were asked to annotate the consecutive sentence pairs if they could infer a discourse relation by inserting a suitable connective[8]. In Example 4, when the sentences are read adjacently, a causal relation is inferred.

(4) But a few funds have taken other defensive steps. *Some have raised their cash positions to record levels.* IMPLICIT= BECAUSE **High cash positions help buffer a fund when the market falls.** (Prasad et al., 2008)

[7] Throughout the thesis, the following format is followed in line with the previous PDTB papers: the connective is underlined;Arg1 is italicized and Arg2 is written in bold

[8] which are called Implicit connective

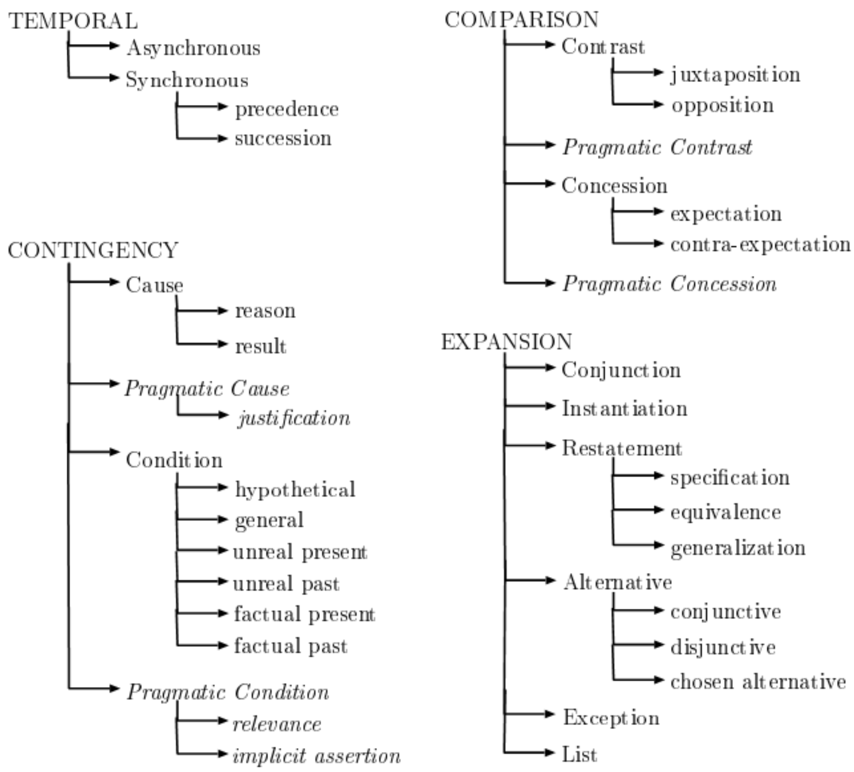

Figure 2.4: Hierarchy of Sense tags in PDTB 2.0(Prasad et al., 2008)

Besides annotating Explicit and Implicit discourse relations, PDTB also annotated expressions which assumed discursive roles such as *one reason is* in Example 5.

(5) *Now, GM appears to be stepping up the pace of its factory consolidation to get in shape for the 1990s.* **One reason is mounting competition from new Japanese car plants in the U.S. that are pouring out more than one million vehicles a year at costs lower than GM can match.** (Prasad et al., 2010a)

The need for annotating expressions which were not originally discourse connectives stemmed from the observation that during the annotations of the Implicit discourse relations, in certain cases addition of an implicit connective led redundancy (Prasad et al., 2010a). Based on this finding, PDTB decided to annotate those expressions as *Alternative Lexicalizations (AltLex).*

Although the underlying principles of PDTB can be traced back to D-LTAG (Webber, 2004), the PDTB is claimed to be theory-neutral. It only includes the annotation of local discourse without making any assumptions regarding the global structure of the discourse. That is to say, it does not take sides with any theoretical assumptions about "what kinds of high-level structures may be created from the low-level annotations of relations and their arguments" (Prasad et al., 2008).

2.3.2 Hindi Discourse Relational Treebank (HDRB)

Hindi Discourse Relational Treebank is a continuing annotation effort on the 200K-word corpus extracted from the 400K-word Hindi News data. It mainly follows the framework set by PDTB; yet, there are several main differences between HDRB and PDTB's annotation scheme. In HDRB, a discourse relation can be realized in one of the following three ways: as *explicit connectives*, as *alternative lexicalizations* or as *implicit connectives*. If none of the three relations can be inferred, then the given adjancent sentence pair is marked as either *entity-based coherence relation (EntRel)* or *NoRel* which indicates the absence of a discourse relation. One of the differences between HDRB and PDTB is that in addition to the subordinating conjunctions, coordinating conjunctions and adverbialsl, HDRB annotates *sentential relatives*, *subordinators* and *particles* as explicit connectives. As for Implicit discourse relations, HDRB annotates the relations across paragraph boundaries which is not allowed in PDTB (Oza et al., 2009).

In HDRB, wherever applicable, the senses of the relations are also annotated. However, HDRB introduced three main refinements to the PDTB's sense hiearchy, which are basically *eliminating the argument-specific labels*, *uniforming the pragmatic relations* and *introducing the "goal" sense*. (Oza et al., 2009). Likewise, HDRB's argument labeling is also different from that of PDTB's. In HDRB, labels of the both arguments are assigned semantically; that is, the sense of the relation is the determinant of the relation's arguments (Oza et al., 2009).

HDRB is still being annotated and as of 2013, 75k portion of the HDRB is annotated (Sharma et al., 2013).

2.3.3 The Chinese Discourse Treebank (CDTB)

The Chinese Discourse Treebank (CDTB) annotates discourse relations with their two arguments on the 98 files of the Chinese newswire text obtained from Chinese Treebank. It follows the PDTB framework with several exceptions. One of the main difference in CDTB's annotation scheme is that their sense hierarchy has flat structure with only 12 sense categories instead of the original three-level hierarchy. Besides, like HDRB, CDTB also labels arguments in a semantically driven fashion. Explicit, Implicit and Altlex relations are annotated. The ratio of Implicit discourse relations is found to be fairly high in Chinese (82% vs 54.5% in PDTB). Therefore, the annotation of Explicit and Implicit discourse relations are performed simultaneously. Implicit discourse relations are divided into EntRel and NoRels as in PDTB (Y. Zhou & Xue, 2012, 2015).

2.3.4 The Turkish Discourse Bank

The Turkish Discourse Bank (henceforth TDB) is a 400.000-word subcorpus of METU Turkish Corpus (MTC) (Say et al., 2002). TDB is compiled in order to "produce a large-scale discourse level annotation resource for Turkish" (Aktaş et al., 2010). It consists of 197 texts from various genres written between 1990-2000.

TDB started with annotating Explicit discourse relations following the PDTB's lexical approach: Discourse relations are grounded to discourse connectives which are treated as discourse level predicates which take two arguments. Explicit discourse connectives belong to one of the three syntactic classes, which are listed below (all the examples are taken from (Zeyrek & Webber, 2008)):

(i) *coordinating conjunctions:* can be realized as a single lexical items such as *ama* 'but' or as paired coordinating conjunctions such as *hem ... hem* 'both and'.

(6) **Konuşmayı unuttum diyorum** da *gülüyorlar bana.*
I said I've forgotton to talk and *they laughed at me.*

(ii) *subordinating conjunctions:* can be realized as simplex (coverbs, such as *–(y)ArAk* 'by means of') or complex (connectives with two parts such as *rağmen* 'despite') subordinators.

(7) Kafiye Hanım beni kucakladı, **yanağını yanağıma sürt**erek *iyi yolculuklar diledi.*
Kafiye hugged me and **by rubbing her cheek against mine**, *she wished me a good trip.*

(iii) *Anaphoric Connectives:* are either discourse adverbials such as *y*oksa 'or else' or phrasal expressions e.g. *onun için* 'for this/that'.

(8) *Bu örgütlerin birleşerek Türkiye'yi etkilemesi ve Türkiye'ye özgü politikaları gündeme getirmesi lazım.* Yoksa **Tony Blair şöyle yaptı şimdi biz de şimdi böyle yapacağızla olmaz.**
These organizations must unite, have an impact on Turkey and introduce political strategies unique to Turkey. Or else **talking about what Tony Blair did and hoping to do what he did is outright wrong.**

The last category of *Anaphoric Connectives* actually fit the Alternative Lexicalizations of PDTB (Prasad et al., 2010a). The TDB group uses both terms in their publications (Zeyrek et al., 2013, 2015).

Similar to PDTB, the arguments of the discourse relations in TDB are required to be abstract objects (Asher, 1993). The first release of the TDB contains 8483 relations for 147 discourse connectives [9]. Upon request, TDB can freely be obtained at *http://medid.ii.metu.edu.tr* .

Annotation of Explicit discourse relations are succeeded by the sense annotations. In a subcorpus, which corresponds to the 10% of the whole TDB (henceforth TDB-Subcorpus), senses are added to previously annotated Explicit discourse relations. Then, the TDB-Subcorpus is annotated for Implicit, Entrel and Altlex discourse relations for their two arguments as well as their senses (Zeyrek et al., 2015) (see Section 3.1 for detailed information).

The detailed information regarding the recent annotation effort on TDB which includes sense annotations is provided in Section 3.1.

[9] Currently, these set includes the relations which are connected by both Explicit connectives and AltLexes which are to be separated the future

Table 2.2: Genre Distribution in the MTC and the TDB (Demirşahin et al., 2012)

GENRE	MTC		TDB	
	#	%	#	%
Novel	123	15.63	31	15.74
Story	114	14.49	28	14.21
Research/Survey	49	6.23	13	6.60
Article	38	4.83	9	4.57
Travel	19	2.41	5	2.54
Interview	7	0.89	1	1.02
Memoir	18	2.29	4	2.03
News	419	53.24	105	53.30
TOTAL	787	100	197	100

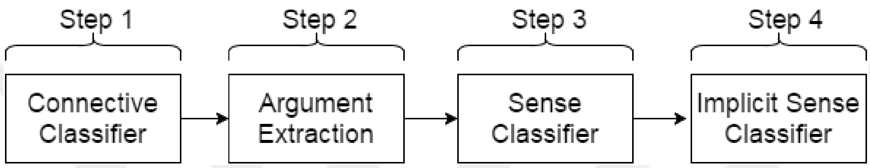

Figure 2.5: An overview of discourse parsing

2.4 Discourse Parsing

2.4.1 Overview

Discourse parsing aims to automatically reveal the internal structure of the discourse by determining how discourse units are related to each other. In NLP, it is one of the most challenging tasks and can be divided into four main sub-tasks. Although it may vary depending on the implementation, the Figure 2.5, inspired from (Lin et al., 2014), illustrates the tasks a typical discourse parser should perform.

Recognition of the discourse connectives is the first challenge to be tackled since discourse connectives, such as *however, but, because*, are ambiguous between discoursive and non-discoursive usage. For instance, in Example 9a, "and" is a discourse connective linking two clauses, whereas in Example 9b "and" does not bear any discoursive role.

(9) a. Selling picked up as previous buyers bailed out of their positions *and* aggressive short sellers– anticipating further declines–moved in

b. My favorite colors are blue *and* green (Pitler et al., 2009)

Resolving the ambiguity between discursive and non-discursive usage, inherently, brings the challenge of determining the arguments of the discourse connective because the phrases surrounding the connective determines whether of not it possesses a discursive role. For example, In PDTB the connectives which relate two abstract objects are regarded as discourse connectives as can be see in Example 9a. Additionally, the discourse connectives do not necessarily link the arguments which are found in the same sentence as the discourse connective. The arguments of the discourse relation can be located in the previous or the next sentences or even in a far location depending on the discourse theory. In PDTB, in almost 30% of the discourse relations, the first argument is located in the immediately previous sentence and in 9% of the relations it is found in a non-adjacent previous sentence (Prasad et al., 2008). Example 10 shows an annotation, in which the first argument is located in a non-adjacent sentence (Prasad et al., 2007).

(10) Mr. Robinson of Delta & Pine, the seed producer in Scott, Miss., said *Plant Genetic's success in creating genetically engineered male steriles doesn't automatically mean it would be simple to create hybrids in all crops.* [SUP1 That's because pollination, while easy in corn because the carrier is wind, is more complex and involves insects as carriers in crops such as cotton]. "It's one thing to say you can sterilize, and another to then successfully pollinate the plant," he said. <u>Nevertheless</u>, **he said, he is negotiating with Plant Genetic to acquire the technology to try breeding hybrid cotton.** (WSJ: 0209)

Moreover, aside from the position or the content of the arguments (e.g. abstract object or not), identification of the arguments' span also needs to be dealt with. In example 11, the connective "if" relates the clauses *the U.S. will default on Nov. 9* and *Congress doesn't act by then*. The clause " The Treasury said" is not included in the relation signaled by the "if" but it belongs to another discourse relation, *attribute relation*, with the rest of the sentences. Therefore, the second task of discourse parsing is the identification of the spans of the arguments.

(11) The Treasury said *the U.S. will default on Nov. 9* <u>if</u> **Congress doesn't act by then.** (Lin et al., 2014)

The next task of discourse parsing is resolving the ambiguity among the relations that a connective may convey. It is not uncommon that same connective signals different senses in different discourse relations, depending on the context it is found. The connective "but" may signal more than 8 different senses, including Contrast, opposition, Conjunction, according to PDTB (Prasad et al., 2008). The situation is the same in Turkish as "ama" (but) can convey 9 different senses. To illustrate this point, the most ambiguous 10 Explicit connectives in Turkish are give in Table B.4.

Typically, the last sub-task of a discourse parser is to classify the discourse relations which are not signaled by an explicit connective. To this end, the adjacent sentence pairs which are not labeled as holding an Explicit discourse relations are passed to the system. The task here is to predict the sense of the Implicit relation, if there is any, among the possible sense set. If the system cannot find a suitable sense, it may identify the relation as EntRel or NoRel. The task of sense classification for both Explicit and Implicit relations are elaborated in the Section 2.4.3.

2.4.2 Previous Studies on Discourse Parsing

The first attempts on discourse parsing were made on corpus annotated following the RST framework such as SPADE by Soricut & Marcu (2003), although SPADE was limited with only sentence-level parsing in the sense that it could not detect discourse relations beyond sentence level. HILDA, the first full discourse parser for RST, is implemented by (Hernault et al., 2010). Unlike (Soricut & Marcu, 2003), HILDA was able to perform document level parsing. HILDA consisted of two different classifiers, which are Support Vector Machines (SVM). The first classifier was responsible for discourse segmentation, that is, to find sentence pairs which were related to each other by means of a discourse relation. The second classifier was a multi-class classifier and chose the appropriate sense, among the 18 senses of the RST framework, for the discourse relation in hand. HILDA achieved F-score of 95.0% for discourse segmentation and 66.8% of sense classification.

As for PDTB, the first end-to-end discourse parser is implemented by Lin et al. (2014). The researchers divided discourse parsing into five components which were implemented in a pipeline architecture. The components of the system were (i) connective classifier, (ii) argument labeler, (iii) explicit classifier, (iv) non-explicit classifier, (v) attribution span labeler. That is, the researchers mainly followed the discourse parsing procedure explained in the previous section with the exception of the additional attribution span labeler. Lin et al. (2014) utilized the previous works for some of the subtasks, such as Pitler & Nenkova (2009)'s Explicit connective classifier and their own previous work on classifying Implicit discourse relations Lin et al. (2009). The sense classification were performed over 16 Type level senses of the PDTB hierarchy for Explicit discourse relations and over 11 Type level senses [10] (plus EntRel and NoRel in case none of the Type level senses were found suitable) for Implicit discourse relations. The overall performance of the system was 38.18%. They reported that the most of errors stemmed from Implicit classification (Lin et al., 2014).

In 2015, shallow discourse parsing was chosen as the shared task of Nineteenth Conference on Computational Natural Language Learning (henceforth CoNLL-2015) (Xue et al., 2015).The problem definition of the shared task was to determine and classify the discourse relations in the given text. The discourse relations were defined according to the PDTB framework, that is, a discourse relation took two abstract objects as its arguments and the presence of discourse connective was not obligatory. Therefore, it captured both Explicit and Implicit discourse relations.

The current state-of-the-art performance in discourse parsing belongs to J. Wang & Lan (2015) which is the winner of CoNLL-2015. In their work, J. Wang & Lan extends the framework developed by (Lin et al., 2014) with the introduction of more components. They approach the problem of argument extraction as two different problems by implementing separate classifiers to extract Arg1 and Arg2 in two cases: (i) if one of the arguments of the Explicit relation is located in one of the previous sentences of the discourse connective [11], (ii) if the type of the relation is not Explicit and is not EntRel. They also define *linked context* as the POS of the connective, its parent and its children in the parse tree so that they provide more syntactic information. Their parser achieves the increase of almost 4% in overall performance compared to (Lin et al., 2014).

[10] They excluded the *Condition, Pragmatic Condition, Pragmatic Contrast, Pragmatic Concession, and Exception* senses due to the lack of sufficient data (Lin et al., 2014)

[11] They disregard the cases where the argument of a discourse relation is not located in the immediate next sentence since such relations constitute only 0.1% of all Explicit relations, according to Prasad et al. (2008)

2.4.3 Automatic Sense Classification of Discourse Relations

Among the subtasks of discourse parsing, identifying the correct sense of the relations proves to be a harder challenge than extraction the sentence pairs between which a discourse relation is held. However, the difficulty of the task heavily relies on the presence, or absence, of a discourse connective. Therefore, sense classification of Explicit relations and of Implicit relations. Therefore, below the work on Explicit and Implicit discourse relations are reviewed separately.

2.4.3.1 Predicting of the Sense of The Explicit Discourse Relations

Labeling the sense of Explicit discourse relations can be formulated as a disambiguation problem on the ground that there is a closed set of senses a connective may convey. Therefore, the degree of ambiguity reduces dramatically by the presence of a discourse connective. In their 2008 work, Pitler et al. analyzed the PDTB for the distribution of Explicit and Implicit discourse relations as well as the degree of ambiguity created by discourse connectives. They revealed that the majority of discourse connectives were not very ambiguous: Within Comparison sense 93.43%, within Contingency sense 94.72% , within Temporal 84.10% and within Expansion 97.63% of the discourse connectives conveyed their predominant sense. Based on this finding, they prepared four binary classification settings to distinguish a Class level relation from each other (such as Comparison vs. others) and a four-way classifier to disambiguate the given relation among the the four main Class levels. They implemented a decision tree for each task and used solely the Explicit discourse connectives to distinguish relations from each other. Despite lack of any information other than discourse connective itself, the binary classifiers achieved F1 score over 90% for Comparison and Temporal, 84% for Contingency and 77% for Expansion relations when tested on a data comprised of both Explicit and Implicit relations.

(Pitler & Nenkova, 2009) focused on sense classification of Explicit discourse relations with syntactic features. They used the node in the parse tree which only covers the connective, which they called *self category*, and the parent, left sibling (which would be NONE if the connective is the left-most node) and right sibling (which would be NONE if the connective is the right-most node) nodes of the self category. They implemented a Naive Bayes classifier with those features and achieved the accuracy of 94.17% in four way classification.If the fact that the inter-annotator agreement in PDTB on Class level was 94% was taken into account, the result was remarkable. Therefore, classification of Explicit discourse relations can be seen as a solved problem, however, it should be noted that those experiments were conducted only on Class level senses.

2.4.3.2 Predicting the Sense of The Implicit Discourse Relations

As it has been pointed out, sense prediction of Implicit discourse relations constitutes the hardest subtask of discourse parsing (Lin et al., 2014; Biran & McKeown, 2015; J. Wang & Lan, 2015). One reason is that due to the lack of an explicit connective, the sense of the discourse relation becomes a matter of inference, which can also depend on the reader's world knowledge.

The work on Implicit discourse relations can be categorized in two categories: The first line of work depends on a large amount of unannotated data and some deterministic rules to extract discourse relations, whereas, the second line of work uses annotated data and linguistically informed features to predict the sense of Implicit discourse relations.

(Marcu et al., 2002) can be regarded as the first thorough study on recognizing the sense of Implicit discourse relations. Their work falls within the first category where their data for Implicit discourse relations were created by stripping the discourse connectives of the Explicit discourse relations in text. The researches termed such Implicit relations as *artificial Implicit discourse relations* since they did not occur in the natural discourse as Implicit relations but converted into by removing the Explicit connective. In order to create Implicit relations, they relied on extraction rules that they formed by analyzing the Explicit discourse relations. For example, they assigned the sense of "CONTRAST" to all sentence pairs where there was a "but" occurring at the beginning of the second sentence. This was based on the finding that out of 106 such occurrences, 89 instances were labeled as "CONTRAST" in the RST corpus built by (Carlson et al., 2003). They used a huge amount of data, ~42.000.000 sentences, and extracted the sentence pairs according to their extraction rules. Then they mainly used word pairs as the their features, which were defined as the Cartesian product of the words in both arguments, and built Naive Bayes Classifiers. Although their six-way classifier[12] achieved 49.7% accuracy, they tested their classifier also on artificial data, leaving the performance of the system on the natural Implicit relations open to question.

In 2008, Sporleder & Lascarides conducted a series of experiments to test the method proposed by Marcu et al. in more realistic settings. They prepared two classifiers, one of which was trained on artificial Implicit relations which were obtained by stripping the unambiguous connectives, such as *in short* which always signals SUMMARY, of the Explicit relations. They identified 55 such connectives and finally prepared a training set including 72.000 instances of artificial Implicit discourse relations. The second classifier, on the other hand, were trained on manually labeled data which included 1.051 discourse relations in total. Then, they tested both models on artificial and natural Implicit relations. The results showed that the model which was trained on artificial Implicit discourse relations achieved the F-score 59.60% on the test set which merely included artificial Implicit relations, however, the performance dropped dramatically to 24.50% when the model was tested on natural Implicit relations. On the other hand, the performance of the classifier which was trained on natural Implicit relations achieved 33.69% indicating that the Implicit discourse relations obtained via stripping the unambitious connective did not represent the natural Implicit relations (Sporleder & Lascarides, 2008).

The release of large annotated corpus, such as PDTB (Prasad et al., 2008), made possible supervised methods to be employed in classifying Implicit discourse relations, thus led to another line of work. The very first work in this category belonged to Pitler et al. (2009). Unlike (Marcu et al., 2002), Pitler et al. exploited linguistically informed features and human annotated data to predict the sense of Implicit relations. Some of the features they used are given below since it greatly inspired and were frequently used in later works:

(i) *Word pairs*: Cross product of the words in the arguments.

(ii) *Modality*: Existence of modal verbs in the arguments. The authors argued that existence

[12] The senses they were concerned with were CONTRAST-EVIDENCE-CONDITION-ELABORATION and two types of sentences pairs where there were no discourse relation held.

of modality signals Contingency relation.

(iii) *Context*: Whether or not the previous discourse relations is Explicit.

(iv) *Verbs*: The number of verb pairs which belong to the same verb class according to Levin verb classes (Levin, 1993).

(v) *First-last and the First3*: The first, last and the first three words of the arguments. This feature was added in order to capture phrases such as AltLex.

The authors built binary Naive Bayes classifiers for each class in PDTB hierarchy as well as a four-way classifier. Except Comparison, some subset of the features they implemented led to an increase for all classes, however to various degrees: they achieved the F-score 76.42% for Expansion, yet only 16.76% for Temporal relations. In the same year, (Lin et al., 2009) also built a classifier to recognize Implicit discourse relations. Their work differed from that of Pitler et al. in the sense that Lin et al. also took Type level senses of the PDTB hierarchy into account and preferred Maximum entropy over Naive Bayes classifier. In terms of features, Lin et al. (2009) relied on more syntactically informed features, such as constituent parse features and dependency parse features, which was different from (Pitler et al., 2009) that relied on mostly semantic features of the words. (Lin et al., 2009) achieved accuracy of 40.2% which was remarkable given that they were concerned with more fine-grained senses.

(Z.-M. Zhou, Xu, et al., 2010; Z. M. Zhou, Lan, et al., 2010) approached the problem more indirectly. Using the finding of (Pitler & Nenkova, 2009) that the existence of discourse connective greatly reduced the ambiguity of the relation, (Z.-M. Zhou, Xu, et al., 2010; Z. M. Zhou, Lan, et al., 2010) set their first task to find an appropriate connective for the Implicit relation, in hand. They trained a language model using the Implicit connectives that were provided by PDTB to predict an Implicit connective for unseen instances. Then, the predicted Implicit connective was used in two ways: firstly, they compiled a feature set which involved the features used in the previous works of Pitler et al. (2009); Lin et al. (2009) as well as the predicted Implicit connective. Secondly, they converted the Implicit discourse relations into Explicit by inserting the predicted Implicit connective and performed the classification using merely the predicted discourse connective, following the work of (Pitler et al., 2008) on Explicit discourse relations. (Z.-M. Zhou, Xu, et al., 2010) reported that using the predicted connective led the improvement of 1.07% to 4.16% improvement in F-Score over the (Pitler et al., 2009) which was the previous state-of-the-art. (Xu et al., 2012) extended the same methodology by using more linguistically informed model to predict the Implicit connective. Park & Cardie (2012) studied the previously proposed features and managed to optimize the results by selecting the best feature subset for each Class through a greedy feature selection.

Sparsity is one of the most serious problems, classifiers suffer to recognize Implicit relations (Biran & McKeown, 2013). (Rutherford & Xue, 2014) used Brown clusters to represent the words in order to deal with the sparsity problem [13]. They replaced the words with one of the 3200 Brown cluster assignments so the overall feature space was reduced to $O(3200^2)$ instead of the original $O(V^2)$ where V corresponded to the size of the vocabulary. Consequently, the classifier trained on Brown clusters outperformed the previous works. More recently, as a result of the current neural network wave, (Ji et al., 2016) have implemented recurrent neural networks. Their system currently holds the state-of-the-art performance for multi-class sense identification with the F-Score of 42.3%.

[13] Brown clusters are explained in Section 4.2 as they are also used within this study

There are limited number of works on Implicit relations for other languages. For Japanese, (Saito et al., 2006) extended the work of (Marcu et al., 2002) by using phrasal patterns in addition to word pairs. Researchers defined phrasal patterns as having at least three phrases, with at least one from each argument. In those phrasal patterns functions words were mandatory whereas content words were optional. The motivation behind constructing such phrasal patterns was the observation that discourse relations could be identified from the fragments of the arguments. For example, CONTRAST relations are more likely to hold between a pair of sentences which involved *".. should have done .."* and *".. did .."*. Their phrasal patterns achieved 12% increase in the accuracy (Saito et al., 2006).

Huang & Chen (2011) built four one-way and a multi-class SVM classifier for Implicit discourse relation recognition in Chinese. The feature set they used included length of the arguments, punctuations that ended both arguments, whether there was a shared word in the arguments, bag of words and POS of the words in the arguments. Overall, they achieved the F-Score of 63.69% in the four-way classification and of 93.57% in the recognition of the Expansion relation.

CHAPTER 3

DATA

This section describes the recent extensions on Turkish Discourse Bank. The output of the recent annotation effort were used as the training/ test data of the classifier built to predict the sense of the Implicit discourse relations. The recent annotation effort covers[1]:

- Sense annotation of previously annotated Explicit discourse relations;
- Annotation of Implicit relations along with sense tags;
- Annotation of Alternative Lexicalizations (AltLex) along with sense tags;
- Annotation of Entity-based (Entrel) relations.

The rest of the chapter summarizes the principles, according to which the annotations were performed.

3.1 Extensions on Turkish Discourse Bank

3.1.1 Annotation Cycle

Four annotators participated in the annotation process[2]. All of the annotators were graduate students in Cognitive Science Department. Before the annotations started, a meeting was held in which the annotators and the supervisor went over the PDTB Annotation Manual 2.0 Prasad et al. (2007) to get familiarized with the PDTB's sense hierarchy. During this period, which lasted approximately 2 weeks, a first draft of annotation guidelines for Turkish was prepared. The first draft of the guideline contained the Turkish connectives which conveyed the senses in the PDTB 2's hierarchy. Later, those connectives were used as "implicit connectives" during the annotations.

Annotators worked in pairs and each pair annotated the 50% of the TDB-Subcorpus. The annotation system adopted was as follows:

---

[1] This annotation effort was realized as part of the project no. BAP-07-04-2015-004 supported by METU, Informatics Institute.

[2] Savaş Çetin, Murathan Kurfalı, Serkan Kumyol, Tuğçe Nur Bozkurt

- Annotators performed the annotation blindly. That is to say, none of the annotators saw the annotations performed by the other annotations.

- Before the weekly held meeting, the annotations from the pairs were collected and the inter-annotator agreement between the members of each pair calculated. The inter-annotator agreement scores are provided in the Table 3.1.The exact match method [3] was adopted to calculate the inter-annotator agreements for each level (Miltsakaki et al., 2004).

- Only after, the inter-annotator agreement were gathered, annotators in the pairs were allowed to see the annotations of their partner. Firstly, Members in the pairs discussed the disagreements between them. If annotators were able to eliminate any disagreements, those cases were not brought up in the general meeting where the team leader [4] were also present.

- During general meetings, the disagreements which could not be resolved discussed by all annotators and the team leader in order to produce the 'gold standard' version of the annotations. In short, all the disagreements are discussed by the annotators and the research leader, all disagreements are eliminated and a final version is obtained.

- During the meetings, if the discussion yielded a modification in the annotation guideline, past annotations which were related to the new decision were reexamined so that the inconsistencies among annotations could be avoided.

- Annotations were performed using the annotation tool developed for TDB (DATT) Aktaş et al. (2010).

Table 3.1: Inter-annotator Agreement for sense tags for Explicit and Implicit DRs

LEVEL	IMPLICIT	EXPLICIT
CLASS	0.52	0.84
Type	0.43	0.71

Table 3.2: Genre Distribution of the TDB-Subcorpus

GENRE	DATA	
	#	%
Novel	7	35
Research/Survey	2	10
Article	2	10
Interview	1	5
Memoir	2	10
News	6	30
TOTAL	20	100

[3] In exact match criterion 1 is assigned for the annotations where the annotators are exactly agree and 0 is assigned if annotators partially agree or totally disagree.

[4] Prof. Dr. Deniz Zeyrek Bozşahin

Regardless of the discourse relation type, annotators were allowed to annotate multiple senses to a relation whenever they inferred multiple interpretations. For example, the relation in Example 12 was labelled as having both the Conjunction and the result senses.

(12) *...toplumsal etiği olduğu gibi sarsacak* ve **toplumsal yıkıntılar yaratacak**
... will tremble the social ethic and **lead to social devastation**
(*"EXPANSION: Conjunction"; "CONTINGENCY: Cause: result"*, fileNo: 20630000 in TDB)

For Implicit discourse relations, the annotators were provided a sample Turkish Explicit connective for each sense of the PDTB's sense hierarchy. Yet, during the annotations, the annotators were allowed to choose other Explicit discourse connectives wherever they saw fit. The annotators were only asked not to provide an Explicit discourse connective which would require paraphrasing the sentences or sound unnatural. The Implicit DRs were sought between adjacent sentences which were delimited with a full stop, colon, semicolon or question mark. An example of an Implicit DR is given in Example 13.

(13) *Kazandığı ünden pek bir şey kaybetmedi.* IMPLICIT: ŞÖYLE Kİ **Sürekli saraya ve zengin konaklara davet ediliyor, en iyi biçimde ağırlanıyordu.**
He did not lose much of his fame. IMPLICIT: IN FACT, **He was frequently invited to the castle and residences, being hosted in the best way**.
(*"EXPANSION: Restatement: specification"*, fileNo: 00001231 in TDB)

Implicit relations were further divided into two types. Those types were used when annotators could not provide a suitable connective which conveyed the Implicit relation. These types[5];

- *Alternative Lexicalizations (AltLexs):* is used for the cases when the insertion of the Implicit connective leads to redundancy since the discourse relation is conveyed by an expression which is not inherently connective. These were either *phrasal expressions* (connective devices with a deictic anaphora, such as *buna göre* 'accordingly') or any other expression with a discourse connective role as in the Example 14.'

  (14) *Birçok endüstride kullanılan kloroflorokarbon gazlarıyla, Halon türü gazların, çok yüksekten uçan jetlerden çıkan azot oksitlerin ve nükleer denemelerin ozonu tüketen unsurlar olduğu belirlendi.* **Bu olaylarda, özellikle klorun ozonu parçalamasının önüne geçilmeyişi, ozon incelmesinin geri dönüşümsüz, tamiri imkansız ve devamlı seyrini** AltLex: beraberinde getirdi.
  It has been identified that along with chlorofluorocarbon gases which were used in various industries, halon gases, nitric oxides coming from jet planes which fly at high altitudes and nuclear tests are the components which consume the ozone. **On those occasions, especially the failure of avoidance of chlorine's decomposition of ozone** brought along **the irreversible, irreparable and continuous progress of the ozone depletion** with itself.
  (*"CONTINGENCY: Cause: result"*, fileNo: 00011112 in TDB)

[5] In PDTB, there is also the third label called *NoRel* which emphasize that there is neither discourse nor entity-based relation between the given pair of sentences. However, NoRels have not been annotated in TDB, yet.

- *Entity Relations (EntRels):* is used when there is only an entity-based relation between the pair of sentences. That is to say, the only relation of the sentences is that they describe the same entity, as in 15

(15) *O gün öğleden sonra okula dönmedim.* EntRel **Ertesi sabah Fındıklı'da tramvaydan indim.**
I did not get back to school. EntRel **Next morning, I got off the train in Fındıklı.**
(fileNo: 00050220 in TDB)

During annotation, annotators firstly looked for Implicit or AltLex relations. If none of the types were suitable for the given discourse relation, EntRel was annotated as the last resort. Therefore, it can be said that while a discourse relation can have multiple senses, it can belong to only one type (e.g Explicit, Implicit, Altlex or Entrel).

Table 3.3: Sense Distribution among Discourse Relations in TDB: CLASS Level

CLASS	Explicit	Implicit	Altlex
TEMPORAL	116	29	22
CONTINGENCY	188	136	44
COMPARISON	207	48	6
EXPANSION	289	189	39
SUM	800	402	111

Table 3.4: Sense Distribution among Discourse Relations in TDB: Type Level

TYPE	Explicit	Implicit	Altlex
Synchronous	18	20	14
Asynchronous	98	9	8
Contrast	81	30	3
Concession	77	14	3
Purpose	79	2	1
Cause	99	129	30
Condition	8	1	5
Manner	13	0	0
Conjunction	226	32	6
Instantiation	5	13	9
Restatement	8	130	17
Alternative	20	3	0
Exception	7	3	0
List	1	0	0
SUM	740	386	96

3.1.2 Sense Hierarchy in Turkish Discourse Bank

The sense hierarchy of the Turkish Discourse Bank is mainly based on the PDTB 2's sense hierarchy (which was provided in Figure 2.4). Yet, a number of modifications have been per-

Table 3.5: Sense Distribution among Discourse Relations in TDB: subtype Level

SUBTYPE	Explicit	Implicit	Altlex
precedence	52	10	4
succession	53	5	8
juxtaposition	9	7	1
opposition	31	22	0
expectation	19	0	0
contra-expectation	26	14	3
reason	62	58	3
result	37	69	27
justification	1	2	2
specification	8	106	8
equivalence	0	13	1
generalization	0	3	2
conjunctive	10	1	0
disjunctive	9	0	0
SUM	317	310	59

formed on the original hierarchy so that it captures new relations that have been encountered in Turkish (see Section 3.1.2.1).

The discourse relations are grouped under four main semantic categories, called "*CLASS*", which are at the highest level of the sense hierarchy. Each class is further divided into other senses which are at the second level of the hierarchy and are called "*Type*". Lastly, some of the *Type* level senses are further divided into "*subtypes*". Below, CLASS level senses and several Subtype and type senses are described and exemplified.

- **"TEMPORAL"**: This tag is used whenever the discourse segments are related temporally. In the second level, whether or not those situations occur concurrently ("Synchronous") or not ("Asynchronous") is defined (Example 16).

  (16) **Orada aynı yerde durur** <u>ve</u> *onu bekler.*
  He stands there on the same spot <u>and</u> *waits for him/her.*
  (*"TEMPORAL: Synchronous"*, fileNo: 00002213 in TDB)

- **"CONTINGENCY":** This tag indicates that one of the situations in the relation causally influences the other. Directionality of the causal relation is specified in the subtype level; the tag "reason" specifies that Arg2 is the cause and "result' specifies that the Arg2 is the effect of the cause described in Arg1 (Example 17).

  (17) **Gözlerine bakamazdım insanların.** <u>IMPLICIT: Çünkü</u> *Korkaktım ben.*
  I could not look people in the eye. <u>IMPLICIT: Because</u> *I was a coward.*
  (*"CONTINGENCY: Cause: reason"*, fileNo: 00001131 in TDB)

- **"COMPARISON":** This tag covers the relations which indicate a difference among the situations or entities described in the arguments. If the arguments are about the same entity or situation, but mark different aspects of it, they are tagged as "Contrast" or its subtypes. If the difference highlighted is about an expectation which is raised by one argument and denied by the other, then it is a "Concession" relation (Example 18).

(18) **Sözlerini onaylamanı bekliyor** <u>IMPLICIT: ama</u> *Gülümsüyor, susuyorsun.*
S/he expects you to confirm his/her words <u>IMPLICIT: but</u> *You are smiling, keeping quiet.*
(*"COMPARISON: Concession: contra-expectation"*, fileNo: 00001131 in TDB)

- **"EXPANSION":** This tag refers to the relations which expand the discourse. For example, the subtype "specification" is one of the most common senses encountered in Implicit discourse relations and refers to the relations where one of the arguments describes the situation mentioned in the other argument in a greater detail (Example 19). "Conjunction", on the other hand, is used to tag the relations where both arguments introduce new information about the same situation but do not fall within the scope of any other "EXPANSION" relations (Example 20).

(19) **Bilginin kaynağı toplumsal pratiktir.** <u>IMPLICIT: ŞÖYLE Kİ</u> *Büyük bilimsel ve düşünsel devrimlerin kökenine indiğimiz zaman büyük toplumsal pratikleri ve insan ihtiyaçlarını görürüz.*
The source of the information is the social practice. <u>IMPLICIT: IN FACT</u> *When we trace the great scientific and intellectual revolutions to their roots, we see great social practices and human needs*
(*"EXPANSION: Restatement: specification"*, fileNo: 00001131 in TDB)

(20) **Gün ağrana dek uğraşıyor;** <u>IMPLICIT: ve</u> *Kadın terasa çıkmadan önce kaçıyordu.*
s/he strives until the dawn; <u>IMPLICIT: and</u> *runs away before the woman appears on the terrace*
(*'EXPANSION: Conjunction"*, fileNo: 00001131 in TDB)

3.1.2.1 Differences from PDTB's hierarchy

As stated earlier, TDB followed the PDTB framework to annotate the discourse relations. However, on certain instances, a suitable sense could not be found in the PDTB sense hierarchy. In order to deal with those relations, the following senses are introduced to the sense hierarchy [6].

- *COMPARISON: Degree* : This sense is introduced to capture meaning of the Explicit connective *kadar* "so much that". In total, 11 tokens have been encountered bearing the *Degree* sense.

(21) **Tanınmayacak** <u>kadar</u> *değişmişti.*
He changed <u>so much that</u> *he cannot be recognized.*
(*'COMPARISON: Degree"*, fileNo: 00001231 in TDB)

[6] Adding this sense to the hierarchy should be regarded as a claiming that other languages missing these senses ?

- *EXPANSION: Manner* : The sense manner is introduced the sense hierarchy in order to express the meaning of *gibi* "as". In total, 13 such tokens have been encountered.

(22) **Kimsenin anlatamayacağı** gibi *anlattın.*
You have told [the story] as **no one else could.**
(*'EXPANSION: Manner"*, fileNo: 00002213 in TDB)

- *EXPANSION: Correction* : In Turkish, correction is conveyed through a negative particle *("yok", "değil")* or the negative verbal morpheme ("-ma") (Zeyrek et al., 2015).

(23) *Ben yere bakmazdım.* IMPLICIT: ama **Gözüne bakardım insanların.**
I wouldn't look down IMPLICIT: but **I would look into people's eyes.**
(*'EXPANSION: Correction"*, fileNo: 00001131 in TDB)

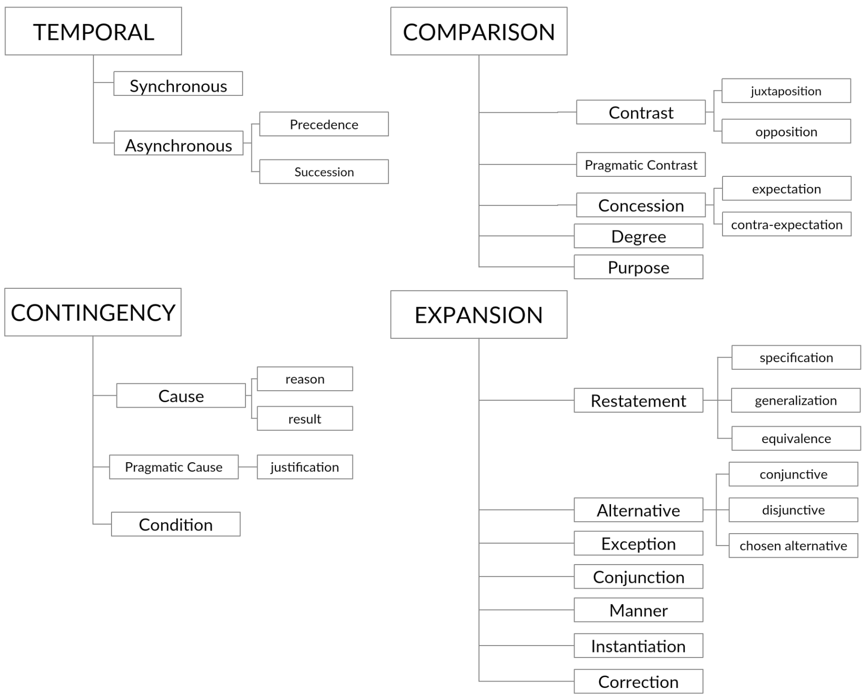

Figure 3.1: Sense Hierarchy in TDB-Subcorpus (developed on the basis of PDTB-2)

3.2 Enriching the Dataset

In addition to the annotations obtained from the latest extensions to TDB, the dataset of the Implicit discourse relations are, also, enriched by the author and the supervisor of the thesis;

1. *Additional annotations on TDB:* In order to increase data, 12 more files are annotated in TDB. However, since there were not any other annotators, a different annotation procedure was applied. The annotations were only confined to Implicit discourse relations

and were performed by the author and checked by the supervisor of the thesis[7]. The disagreements between the author and the supervisor were resolved before finalizing the annotations.

The distribution of the senses in the overall annotations, which are used as the training/test data during classification, as well as the source of those annotations are given in Table 3.6. In the table, the column *Original Annotations on TDB* denotes the annotations created as the output of the recent annotation effort on TDB. The column *Additional Annotations*, on the other hand, refers to the additional annotations which were created by the author.

Table 3.6: Sense Distribution of the all Implicit Annotations: CLASS Level

CLASS	Original Annotations	Additional Annotations	OVERALL
TEMPORAL	29	29	58
CONTINGENCY	136	70	206
COMPARISON	48	24	72
EXPANSION	189	98	287
OVERALL	402	221	**623**

[7] I should thank Prof. Dr. Deniz Zeyrek Bozşahin for accepting the second annotator and providing her insightful comments regarding each and every annotation.

CHAPTER 4

EXPERIMENTAL SETUP

4.1 Classifiers

4.1.1 Support Vector Machines (SVM)

Support Vector Machines (SVM) is a supervised classification method proposed by (Cortes & Vapnik, 1995). SVM is one of the most popular techniques frequently used in various NLP applications including classification of Implicit discourse relations (Z.-M. Zhou, Xu, et al., 2010; Louis et al., 2010; Huang & Chen, 2011) among others. SVM performs the classification by mapping the inputs into a high dimensional feature space and aims to find the optimal hyperplane which separates the data points. The optimal hyperplane is found by maximizing the margin from the closest data points of each class and the support vectors are the points which are located on the boundaries (see Figure 4.1).

If the dataset is not linearly separable, the points can be projected to higher dimensional space in order to achieve linearly separableness.

Basically, SVM can perform binary classification. Given the set of pairs (x_i, y_i) where $x_i \in R$ and y is the label such that $y_i \in \{-1, +1\}$ and $i = 1...N$, in the linearly separable case the support vectors will be of form:

$$w * x_i + b = 1 \; for \; all \; y_i = +1$$

$$w * x_i + b = -1 \; for \; all \; y_i = -1$$

where the parameters w and b are learned in the learning procedure. Recall that SVM tries to find the optimal hyperplane which is the one at the maximum distance from both of the support vectors. The distance between those support vectors becomes

$$d = \frac{2}{||w||}$$

where the ||w|| denotes the euclidean distance of w and is to be minimized.

4.1.2 Maximum Entropy (MaxEnt)

Maximum entropy is the most unbiased machine learning technique in the sense that it looks for the most uniform distribution objected to given constraints (Berger et al., 1996). MaxEnt

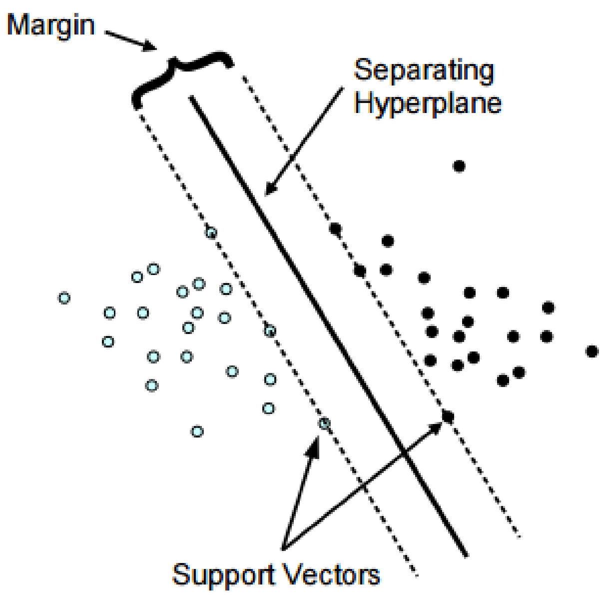

Figure 4.1: Margin, separating planes and support vectors in SVM, taken from (Meyer & Wien, 2015)

estimates the conditional probability of $P(c|d)$ as follows:

$$P(c|d) = \frac{exp[\sum_i \lambda_i f_i(c,d)]}{\sum_{c'}[exp[\sum_i \lambda_i f_i(d,c)]]}$$

where, in the current context, λ is the weight vector, c is the CLASS sense , d is the discourse relation. The $f_i(d,c)$ denotes a feature function which takes the form:

$$f_i(d,c) = \begin{cases} 1 & if\ c = c_i\ and\ d\ contains\ wp \\ 0 & otherwise \end{cases}$$

For example, the feature function defined above fires if the current discourse relation contains a certain word pair *wp*. The advantage MaxEnt models is that they do not assume any relation among these feature functions so even if the features are dependent to each other, MaxEnt models can be used.

4.2 Feature Set

The following feature set is devised, following the previous research, mainly (Marcu et al., 2002; Sporleder & Lascarides, 2008; Pitler et al., 2009; Lin et al., 2009; Huang & Chen, 2011; Rutherford & Xue, 2014).

- *Word Pairs (WP):* The cross product of the words in both arguments. E.g, if the first argument of the relation consists of the words (x, y) and second argument (a, b), the following word pairs are generated: (x_a, x_b, y_a, y_b). In producing word pairs, the punctuations are ignored and the words are lemmatized in order to reduce the sparsity. This feature is expected to be useful to detect the word pairs where a semantic relationship, such as synonymy or antonymy, holds. Such word pairs are hypothesized to be a good indicator of causal or contrast discourse relations (Pitler et al., 2009).

- *Brown Cluster Pairs:* Brown clustering is proposed by (Brown et al., 1992) in order to deal with the sparsity problem. Brown clustering works on a large amount of unannotated data to build word hierarchies based on the word co-occurrences within the specified distance. The pair of Brown clusters of the words in the relation were used in (Rutherford & Xue, 2014) and were reported to yield a better performance than the traditional word pair feature which is described above. Brown Cluster pairs are trained on the rest of the TDB which contains approximately $\sim$ 330K words.

- *Length:* This feature includes the length of the first argument and the length of the second argument and the total length in terms of number of words. Since EXPANSION relations give detailed information regarding the event or the situation described in one of their arguments, EXPANSION relations are expected to be longer than the others.

- *Genre:* Following (Webber, 2009), the genre of the text which the discourse relation is located at is provided to the classifier as a feature. Overall, there are 8 different type of genres tagged in TDB which are *novel, travel, news, story, memoir, interview, research, article*.

- *Polarity:* This feature indicates whether or not the main verb of the first and the second argument is negated. Additionally, the combination of the polarity of the both verbs are provided to the classifier. For example, the corresponding polarity feature of the artificial example of *Oraya gidemedim. Bazı işlerim vardı.* ("I could not go there. I had some errands to do"), would be "arg1Neg", "arg2pos" and "arg1Neg_arg2pos". It is expected that polarity feature will be useful in order to catch COMPARISON or CONTINGENCY relations on the ground that in those relations one of the arguments are likely to refute or oppose the proposition of other argument.

- *Tense:* This feature refers to the tense of the main verbs in both arguments. Also, whether or not those tenses are the same is also provided to the classifier as a feature. It is predicted that the discourse relations behave differently in terms of the tense of their main verb. For example, EXPANSION relations can be expected to have arguments with the same tense, whereas CONTINGENCY or COMPARISON relations link sentences with different tenses.

4.3 Evaluation of the Effect of Explicit Discourse Relations

As discussed in Section 2.4.3.2, Explicit discourse relations are regarded as a source to obtain Implicit discourse relations in an effortless way. The methodology offered by (Marcu et al., 2002) was to convert the Explicit discourse relations into Implicit relations by stripping the connective providing that the connective was unambiguous. However, that methodology has been shown to be ineffective to classify Implicit discourse relations (Sporleder & Lascarides, 2008; Webber, 2009).

Within this study, a slightly different approach is adopted. The Explicit discourse relations are converted into Implicit relations by stripping the Explicit connective, but unlike (Marcu et al., 2002) where the researchers merely used the Explicit relations which are signaled by an unambiguous connectives, all Explicit relations, signaled by both ambiguous and unambiguous connectives, extracted from human annotated data (TDB-Subcorpus) are used . The senses of the Explicit discourse relations with ambiguous connectives are created via human effort. In the rest of thesis, the term *manually-labeled Explicit relations*[1] will be used.

The main motivation behind using manually-labeled Explicit relations is the observation that unambiguous connectives constitute a small portion of all discourse connectives (see Table 4.1 for their distribution in TDB) and the relations signaled by unambiguous connectives are more likely to belong to certain senses than others (see Table 4.2). Moreover, Explicit relations realized with only unambiguous connectives may not reflect true nature of Explicit relations. Therefore, the relations obtained according to (Marcu et al., 2002) may result in incomplete representations of the senses and may be the reason why they do not generalize to Implicit discourse relations. In the light of these findings, it is hypothesized that manually-labeled Explicit relations can provide a more compact representation of the Explicit discourse relations and enable a more valid comparison between Implicit and Explicit discourse relations in terms of their sentence structure.

Although manual annotation of Explicit discourse relations demand human effort and is a more tedious way to obtain Implicit relations than the method of (Marcu et al., 2002), if it can be shown to generalize to Implicit relations, manually-labeled Explicit relations are still a faster and more reliable source [2] to increase Implicit discourse relations.

Therefore, in order to test the hypothesis formulated above, a separate set of training data where *manually-labeled Explicit relations* are added are prepared.

Table 4.1: List of the Unambiguous Discourse Connectives vs. List of the human-annotated Discourse Connectives in TDB-Subcorpus

Unambiguous Discourse Connectives (20 search tokens)	All Discourse Connectives (44 search tokens)
amacıyla, ayrıca, beraber, birlikte, çünkü, dolayı, dolayısıyla, gene de, halbuki, içindir, mesela, ne ki, ne var ki, nedeniyle, örneğin, rağmen, tersine, veya, ya, yine de	ama, amacıyla, ancak, ardından, aslında, ayrıca beraber, bir yandan birlikte, böylece, çünkü, dahası dolayı, dolayısıyla, fakat, gene de, gibi, halbuki halde, hem, için, içindir, iken, kadar, karşın mesela, ne ki, ne var ki, ne, nedeniyle, önce örneğin, oysa, rağmen, sonra, tersine, ve veya, ya da, ya, yalnız, yine de, yoksa, zaman

[1] This term is selected in order to highlight the difference between the current approach and (Marcu et al., 2002) where the researchers used the term "automatically labeled Explicit relations".

[2] The reliability is used in the sense that inter-agreement results on Explicit discourse relations are much higher than that of Implicit discourse relations (Zeyrek et al., 2015).

Table 4.2: Number of Relations Signaled by Unambiguous Connectives vs. Number of Relations Signaled by All Explicit Connectives in TDB-Subcorpus

CLASS	# of DRs realized with Unambiguous Discourse Connectives	# of DRs realized with All Discourse Connectives
COMPARISON	19	207
TEMPORAL	0	116
CONTINGENCY	60	188
EXPANSION	28	289
TOTAL	107	800

4.4 Methodology

DATT converts the annotations into a pre-defined XML format, where beginning and end offsets of the arguments and the connectives, their content as well as the senses are kept. A sample annotation in XML format is provided in Appendix A.

Those XML files are converted to text format where the sense and the arguments are kept. The necessary morphological and syntactic information is obtained by the morphological analyzer developed by (Sak et al., 2007).

As it was explained in Section 3.1.1, annotators were allowed to annotate multiple senses for a discourse relation if they could infer more than one sense. Those relations with multiple senses are added to the data as if there are two relations, each of which signal the inferred sense.

For each Class sense in the hierarchy, a Support Vector Machine and Maximum Entropy classifiers with the features detailed in the Section 4.2 are built. For the SVM classifier the LIBSVM (Chang & Lin, 2011) implementation and for the MaxEnt classifier the MALLET (McCallum, 2002) are used. In order to provide negative instances to the classifier, the instances which do not belong to the target CLASS are chosen randomly. That is, the negative instances for the sense COMPARISON are obtained from the instances of EXPANSION, TEMPORAL and CONTINGENCY.

Overall, the experiments are divided into two categories based on the instances in the training data. The motivation behind creating two sets of different training data is to assess the effect of manually-labeled Explicit discourse relations. For all experiments, 30% of the Implicit discourse relations are used for testing. The two sets of training data used in the experiments are:

1. *TrainingSet-Imp:* In this set of experiments, the training data only involves natural Implicit discourse relation instances. The size of the training and test data is given in Table 4.3.

2. *TrainingSet-ImpExp:* In this set of experiments, the training data is obtained by adding the manually-labeled Explicit relations to the TrainingSet-Imp. The size of training and test data is given in Table 4.4.

The idea behind using merely manually-labeled Explicit discourse relations is to be able to compare them directly with the natural Implicit discourse relations. The motivation behind TrainingSet-ImpExp is the hypothesis that manually-labeled Explicit discourse relations can be helpful to increase the Implicit relations' data as supplementary material. Therefore,these two experiments are aimed to reveal how close Explicit discourse relations are to the Implicit relations.

Table 4.3: Number of Annotations in Training and Test data

CLASS	Training Size (Positive/Negative)	Test Size (Positive/Negative)
COMPARISON	100 (50/ 50)	44 (22/ 22)
TEMPORAL	80 (40/ 40)	36 (18/ 18)
CONTINGENCY	288 (144/ 144)	124 (62/ 62)
EXPANSION	400 (200/ 200)	172 (86/ 86)

Table 4.4: Number of Annotations in Training and Test data when both Implicit and manually-labeled Explicit discourse relations are used

CLASS	Training Size (Positive/Negative)	Test Size (Positive/Negative)
COMPARISON	514 (257/ 257)	44 (22/ 22)
TEMPORAL	312 (156/ 156)	36 (18/ 18)
CONTINGENCY	664 (332/ 332)	124 (62/ 62)
EXPANSION	978 (489/ 489)	172 (86/ 86)

CHAPTER 5

RESULTS

This section is devoted to provide the results of the classifiers built following the methodology explained in the previous chapter. For each experiment, the accuracy, precision, recall and F1 scores which are obtained by the SVM and the MaxEnt classifiers are reported. The definitions of those metrics are given below:

(i) Accuracy: The percentage of the predictions which are correct.

(ii) Precision: The percentage of the positive predictions the model catches:

$$Precision = \frac{TP}{TP + FP}$$

where TP = true positive and FP = false positive.

(iii) Recall: The percentage of the positive predictions which are correct:

$$Recall = \frac{TP}{TP + FN}$$

where FN = false negative

(iv) F1 Score: F1 score is the harmonic mean of precision and recall. It is calculated as follows:

$$F_1 = 2 * \frac{precision * recall}{precision + recall}$$

5.1 Results of the Sense Prediction of the Implicit Discourse Relations

This section covers the experiments conducted by using TrainingSet-Imp data which merely contains Implicit discourse relations. Sense predictions depended on each sense in the CLASS level of the PDTB-2 sense classification. The results concerning each CLASS sense are provided separately. For each experiment, the last lines of the tables contain the best feature set, that is, the set of the features which yield the best performance.

5.1.1 Temporal

Among the four CLASS level senses, Temporal relations are the ones with the lowest number of instances. The results of the classification task of Temporal vs. Others for both SVM and MaxEnt classifiers are provided in the Table 5.1. The best feature set which achieves the best F-score is Polarity, Tense and Word Pairs. On the other hand, the MaxEnt model achieves the same performance with merely Polarity and Tense features. In the MaxEnt model, the Word Pair to Polarity and Tense feature reduces the performance by 4%.

Table 5.1: Results of Temporal vs. Other

SVM				
Feature	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
1. BCpairs	0.472	0.486	0.944	0.642
2. Genre	0.583	0.571	0.667	0.615
3. Length	0.472	0.455	0.278	0.345
4. Polarity	0.694	0.769	0.556	0.645
5. Tense	0.528	0.529	0.500	0.514
6. Word Pairs	0.472	0.484	0.833	0.612
Polarity & Tense & Word Pairs	0.639	0.609	0.778	**0.683**
MaxEnt				
Feature	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
1. BC Pairs	0.472	0.484	0.833	0.612
2. Genre	0.583	0.571	0.667	0.615
3. Length	0.583	0.588	0.556	0.571
4. Polarity	0.694	0.769	0.556	0.645
5. Tense	0.444	0.450	0.500	0.474
6. Word Pair	0.556	0.750	0.167	0.273
Polarity & Tense	0.639	0.609	0.778	**0.683**

5.1.2 Comparison

Similar to Temporal relations, the best feature set which differentiates Comparison relations from the others is the combination of Polarity and Tense for the SVM classifier. However, the MaxEnt model achieves a higher F-score of 64.2% by using only Polarity feature. In the MaxEnt model, the combination of Length and Polarity features yields 56.5% F-score. The results are provided in Table 5.2.

5.1.3 Contingency

For CONTINGENCY relations, in terms of the individual features, the most helpful ones are the BC pairs, Word Pairs and the Polarity for both classifiers. However, the combination of those features behave differently. The highest F-score is achieved by the combination of Polarity, Tense and Word Pairs for the SVM model; however, BC pairs achieved better on its own in the MaxEnt classifier. The results are provided in the Table 5.3.

Table 5.2: Results of Comparison vs. Other

SVM				
Feature	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
1. BCpairs	0.500	0.500	0.091	0.154
2. Genre	0.409	0.409	0.409	0.409
3. Length	0.477	0.485	0.727	0.582
4. Polarity	0.523	0.515	0.773	0.618
5. Tense	0.432	0.435	0.455	0.444
6. Word Pairs	0.455	0.250	0.045	0.077
Polarity & Tense	0.568	0.552	0.727	**0.627**
MaxEnt				
Feature	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
1. BC Pairs	0.500	0.333	0.045	0.080
2. Genre	0.409	0.400	0.364	0.381
3. Length	0.636	0.636	0.636	0.636
4. Polarity	0.568	0.548	0.773	0.642
5. Tense	0.455	0.464	0.591	0.520
6. Word Pair	0.500	0.500	0.091	0.154
Polarity	0.568	0.548	0.773	**0.642**

5.1.4 Expansion

EXPANSION relations constitutes the largest part of the all Implicit relations (as these relations constitute the 46% of all Implicit relations). Among the features implemented, the Length feature yields the best performance on its own in the SVM model. The best performance achieved by the MaxEnt model is obtained via the combination of Genre and Length features, yet, MaxEnt performs worse than SVM model by almost 5%. The overall results are provided in the Table 5.4.

5.2 The Effect of the Manually Labeled Explicit Discourse Relations

The difference between this set of experiments from the previous one is the addition of the *manually-labeled Explicit discourse relations* to the training data. The test data is not altered in any way.

5.2.1 Temporal

For TEMPORAL relations, the best score achieved by using only Implicit relations was 68.3%. The addition of *manually-labeled Explicit discourse relations* seems to increase the performance. However, the feature set which achieves the best performance alters significantly. Here, MaxEnt model needs all features together to achieve the F-score of 77.8% whereas the Polarity feature achieves 72.2% in SVM classifier. The results are provided in Table 5.5.

Table 5.3: Results of Contingency vs. Other

SVM				
Feature	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
1. BCpairs	0.484	0.490	0.806	0.610
2. Genre	0.516	0.519	0.435	0.474
3. Length	0.492	0.488	0.323	0.388
4. Polarity	0.460	0.455	0.403	0.427
5. Tense	0.589	0.585	0.613	0.598
6. Word Pairs	0.532	0.519	0.871	0.651
Polarity& Tense & Word Pairs	0.637	0.620	0.710	**0.662**
MaxEnt				
Feature	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
1. BC Pairs	0.524	0.513	0.984	0.674
2. Genre	0.508	0.509	0.435	0.470
3. Length	0.476	0.475	0.452	0.463
4. Polarity	0.460	0.455	0.403	0.427
5. Tense	0.589	0.593	0.565	0.579
6. Word Pair	0.500	0.500	0.919	0.648
BC Pairs	0.524	0.513	0.984	**0.674**

5.2.2 Comparison

The highest F-score achieved by using merely Implicit discourse relations was 64.2% using MaxEnt classifier and the Polarity feature. *Manually-labeled Explicit discourse relations* increase the performance by almost 10% in the SVM classifier when all features are provided. On the other hand, MaxEnt classifier achieves only 1% increase. The results are given in the Table 5.6.

5.2.3 Contingency

In Contingency relations, the addition of *manually-labeled Explicit discourse relations* causes a decrease in both classifiers by almost 4%. Moreover, the set of the features which achieves the best performance is also changed. When *manually-labeled Explicit discourse relations* are present in the training data, SVM classifier with the Tense features yield the best performance of 63.7%. The results are given in the Table 5.7.

5.2.4 Expansion

Similar to Contingency relations, the *manually-labeled Explicit discourse relations* decresases the performance of the classifiers. When *manually-labeled Explicit discourse relations* are added to the training data, the SVM classifier with the Genre feature achieves the F-score of 64%, whereas the highest score was 65% when the model was trained merely on the Implicit discourse relations. The results are provided in the Table 5.8.

Table 5.4: Results of Expansion vs. Other

SVM				
Feature	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
1. BCpairs	0.500	0.500	0.105	0.173
2. Genre	0.669	0.754	0.500	0.601
3. Length	0.674	0.703	0.605	0.650
4. Polarity	0.494	0.494	0.442	0.466
5. Tense	0.430	0.439	0.500	0.467
6. Word Pairs	0.494	0.483	0.163	0.243
Length	0.674	0.703	0.605	**0.650**

MaxEnt				
Feature	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
1. BC Pairs	0.471	0.308	0.047	0.081
2. Genre	0.669	0.754	0.500	0.601
3. Length	0.634	0.677	0.512	0.583
4. Polarity	0.483	0.480	0.419	0.447
5. Tense	0.413	0.407	0.384	0.395
6. Word Pair	0.465	0.350	0.081	0.132
Genre & Length	0.651	0.697	0.535	**0.605**

Table 5.5: Results of Temporal vs. Other when the classifier is trained on manually-labeled Explicit relations + Implicit discourse relations

SVM				
Feature	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
1. BCpairs	0.528	0.515	0.944	0.667
2. Genre	0.444	0.438	0.389	0.412
3. Length	0.389	0.400	0.444	0.421
4. Polarity	0.722	0.722	0.722	0.722
5. Tense	0.444	0.462	0.667	0.545
6. Word Pairs	0.556	0.533	0.889	0.667
Polarity	0.722	0.722	0.722	**0.722**

MaxEnt				
Feature	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
1. BC Pairs	0.553	0.545	0.900	0.679
2. Genre	0.556	0.542	0.722	0.619
3. Length	0.444	0.429	0.333	0.375
4. Polarity	0.722	0.722	0.722	0.722
5. Tense	0.528	0.524	0.611	0.564
6. Word Pair	0.667	0.615	0.889	0.727
All Features	0.778	0.778	0.778	**0.778**

Table 5.6: Results of Comparison vs. Other when the classifier is trained on manually-labeled Explicit relations + Implicit discourse relations

SVM				
Feature	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
1. BCpairs	0.500	0.500	0.091	0.154
2. Genre	0.477	0.485	0.727	0.582
3. Length	0.568	0.579	0.500	0.537
4. Polarity	0.591	0.667	0.364	0.471
5. Tense	0.386	0.414	0.545	0.471
6. Word Pairs	0.500	0.500	0.273	0.353
All Features	0.727	0.692	0.818	**0.750**
MaxEnt				
Feature	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
1. BC Pairs	0.523	0.556	0.227	0.323
2. Genre	0.477	0.485	0.727	0.582
3. Length	0.545	0.545	0.545	0.545
4. Polarity	0.568	0.571	0.545	0.558
5. Tense	0.409	0.438	0.636	0.519
6. Word Pair	0.591	0.700	0.318	0.438
All Features	0.636	0.625	0.682	**0.652**

Table 5.7: Results of Contingency vs. Other when the classifier is trained on manually-labeled Explicit relations + Implicit discourse relations

SVM				
Feature	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
1. BCpairs	0.468	0.481	0.823	0.607
2. Genre	0.484	0.489	0.742	0.590
3. Length	0.500	0.500	0.452	0.475
4. Polarity	0.484	0.476	0.323	0.385
5. Tense	0.605	0.589	0.694	0.637
6. Word Pairs	0.516	0.510	0.790	0.620
Tense	0.605	0.589	0.694	**0.637**
MaxEnt				
Feature	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
1. BC Pairs	0.508	0.505	0.839	0.630
2. Genre	0.484	0.489	0.742	0.590
3. Length	0.452	0.446	0.403	0.424
4. Polarity	0.484	0.476	0.323	0.385
5. Tense	0.581	0.576	0.613	0.594
6. Word Pair	0.444	0.455	0.565	0.504
BC Pairs	0.508	0.505	0.839	**0.630**

Table 5.8: Results of Expansion vs. Other when the classifier is trained on manually-labeled Explicit relations + Implicit discourse relations

SVM				
Feature	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
1. BCpairs	0.535	0.650	0.151	0.245
2. Genre	0.634	0.629	0.651	0.640
3. Length	0.599	0.605	0.570	0.587
4. Polarity	0.529	0.523	0.663	0.585
5. Tense	0.459	0.434	0.267	0.331
6. Word Pairs	0.552	0.636	0.244	0.353
Genre	0.634	0.629	0.651	**0.640**

MaxEnt				
Feature	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
1. BC Pairs	0.547	0.595	0.291	0.391
2. Genre	0.663	0.700	0.570	0.628
3. Length	0.570	0.577	0.523	0.549
4. Polarity	0.552	0.544	0.651	0.593
5. Tense	0.419	0.405	0.349	0.375
6. Word Pair	0.616	0.616	0.616	0.616
Genre	0.663	0.700	0.570	**0.628**

CHAPTER 6

DISCUSSION

This chapter provides the analysis of the findings reported in Chapter 5. The possible implications of the results regarding the sentence configuration of the arguments of a discourse relation is discussed. The second section concentrates on the effect of Explicit discourse relations in the prediction task and aims to compare Explicit and Implicit discourse relations in terms of the features built. In the last section, the performance of the classifiers, Support Vector Machines and Maximum Entropy, are compared.

6.1 Sense Prediction of Implicit Discourse Relations

Below, the results presented in Chapter 5 are discussed for each CLASS level sense. It should be noted that the classification task for each sense is set up to be a binary classification. That is to say, the results of each experiment points out the different aspects of the relations from each other. In terms of the classifiers, the SVM classifier is found to be more informative regarding the sentence configuration of discourse relations. On most of the cases, MaxEnt classifier achieves maximum F-score by using either a single feature or all features together. Therefore, the discussion below is mostly based on the findings obtained by the SVM classifier because to meet the objectives of the current thesis, the syntactic features revealed by the classifiers are as important as the maximum F-scores achieved.

6.1.1 Temporal Relations

The results of the classification of TEMPORAL discourse relations task indicate that the most important aspects of the arguments of the TEMPORAL relations are regarding their Tense and Polarity. Whether or not the both arguments of the given relation has positive polarity and at least one of the arguments' tense is past have been found as the most important features by the SVM classifier. The inspection of the data reveals that 60% of the Temporal Implicit relations have positive verbs on their both arguments since Temporal relations often describe the situation. Moreover, since Temporal relations are mostly used to describe a situation, their arguments are in past tense. The ratio of TEMPORAL relations with at least one of the argument has a verb in past tense is found to be 74.1% in the data. A typical Temporal relation Example from the TDB-Subcorpus is given in the Example 24.

(24) *Sokağın başına kadar yürüdü.* <u>IMPLICIT: Sonra</u> **Birden koşmaya başladı.**

(S/he) walked to the end of the street IMPLICIT: Then **(S/he) suddenly started to run**.
(*"TEMPORAL: Asynchronous: succession"*, fileNo: 00001131 in TDB)

6.1.2 Comparison Relations

The results regarding discourse relations with COMPARISON sense also highlight the importance of the Tense and Polarity of the main verbs of its arguments. However, unlike TEMPORAL relations, the arguments of the COMPARISON relations differ in polarity. The most distinguishing feature of the COMPARISON relations turns out to be the configuration when the first argument has positive polarity whereas the second argument has a negated verb such as in the Example 25. In terms of tense, the COMPARISON relations behave similar to the TEMPORAL relations in the sense that only 35% of the COMPARISON relations do not have argument with the verb in the past tense.

(25) *Seçimin üstünden neredeyse bir ay geçti.* IMPLICIT: Buna rağmen **Barajın altında kalan siyasi partiler şoku atlatamadı.**
It has been almost a month after the elections. IMPLICIT: Despite this **The political parties which failed to pass the election threshold could not recover from the shock**. (*"COMPARISON: Concession: contra-expectation"*, fileNo: 20580000 in TDB)

6.1.3 Contingency Relations

Regarding the CONTINGENCY relations, the results indicate that one of the most characteristic features of these relations is that most of the time the main verbs of the arguments are in different tenses (see Example 26). The inspection of the data shows that, in 66% of the CONTINGENCY relations, the tenses of the main verbs of its arguments are different from each other. The results are, partly, in line with (Miltsakaki et al., 2005) where the authors suggest that in Causal relations, the first argument is more likely to be prior to the second argument. With the regard to polarity, the CONTINGENCY relations are more likely to bear a negative verb in the second argument than the TEMPORAL and EXPANSION relations. Therefore, similiar to COMPARISON relations, having a negated verb in the second argument differentiates CONTINGENCY, as well as COMPARISON relations, from the others (see Table 6.1).

(26) *Yıllar geçti üstünden.* IMPLICIT: Bu yüzden **Artık üzmez.**
[The incident] happened ages ago. IMPLICIT: For this reason, **It does not upset me anymore**.
(*"CONTINGENCY: Cause: result"*, fileNo: 00065111 in TDB)

6.1.4 Expansion Relations

EXPANSION relations, by constituting the 45% of the all Implicit relations, are the most frequently occurring sense among the CLASS level sense. However, unlike other three CLASS

Table 6.1: The Percentage of the Discourse Relations which contains a second argument with a negated main verb

CLASS	%
TEMPORAL	13.79
COMPARISON	44.44
EXPANSION	23.77
CONTINGENCY	33.49

level senses, for the EXPANSION sense the most helpful feature is found to be the length of the discourse relation while the genre is the second. On general, EXPANSION relations tend to be longer as shown in the Table 6.2. There is not any characteristic manifestation of the arguments of EXPANSION relations in terms of polarity or tense. Among the 8 type of genres, 31.8% of the EXPANSION relations are found in news and 30.7% are found in novels.

Table 6.2: Average Length of Discourse Relations in words

CLASS	Average Length
TEMPORAL	12.735
COMPARISON	17.578
EXPANSION	20.262
CONTINGENCY	13.867

6.2 Analysis of the Implicit and Explicit Discourse Relations

The idea of using Explicit discourse relations as Implicit relations have received many attention. However, the fast way of converting Explicit discourse relations with an unambiguous connectives into Implicit relations by stripping the connective, proposed by (Marcu et al., 2002), has been shown to generalize poorly to natural Implicit discourse relations (Sporleder & Lascarides, 2008; Webber, 2009). Instead, the current study adopted a different approach, where the human annotators labeled all Explicit connectives manually in order to be able to utilize all Explicit discourse relations without being limited to those with unambiguous connectives. Those Explicit connectives are called *manually-labeled Explicit discourse relations*.

The results given in Tables 5.5 - 5.8 indicate the effect of *manually-labeled Explicit discourse relations*. However, the results are hard to explain since *manually-labeled Explicit discourse relations* behave very differently depending on the CLASS sense. The overall F-scores seem to increase for TEMPORAL and COMPARISON relations in both classifiers and decrease for CONTINGENCY and EXPANSION relations. What is more interesting is the fact that for all class level senses the set of features which yields the best performance differ when *manually-labeled Explicit discourse relations* are added to training data. This finding indicates that, from the linguistic point of view, the sentence configuration of Explicit and Implicit discourse relations, in terms of the features implemented, differ from each other.

The methodology adopted within the current study can also be regarded as the continuation of (Sporleder & Lascarides, 2008). (Sporleder & Lascarides, 2008) has suggested several reasons why the methodology adopted by (Marcu et al., 2002) did not work well on Implicit discourse relations. One of the reasons, they claimed, could be the fact that automatically labeling the discourse relations merely based on they were signaled by an unambiguous connective could cause labeling errors. Their inspection of randomly selected 50 annotations which were labeled automatically revealed that 15% of the examples for the SUMMARY[1] sense were invalid. The other reason brought up by (Sporleder & Lascarides, 2008) was that Explicit discourse relations with unambiguous connectives do not represent the all Explicit discourse relations [2]. However, the methodology adopted within the current study eliminates these two reasons since in the experiments all Explicit discourse relations, instead of a small portion of them which were signaled by unambiguous connectives, were used and also all Explicit discourse relations were annotated by human annotators. Therefore, using *manually-labeled Explicit relations* provided a solid ground to compare Explicit and Implicit of discourse relations, without allowing to concerns raised by (Sporleder & Lascarides, 2008).

6.3 Comparison of the Classifiers

In sense prediction of Implicit discourse relations Support Vector Machines and Maximum Entropy models are the most popular classifiers (Lin et al., 2009; Z.-M. Zhou, Xu, et al., 2010; Louis et al., 2010; Park & Cardie, 2012; Rutherford & Xue, 2014). In terms of single feature, that is to say when only one feature is present, the both classifiers perform similarly. Their performance differs in the feature set which achieves the highest performance. In general, MaxEnt model performs slightly better than SVM. When only Implicit relations are present in the training data, MaxEnt model seems to perform better than SVM for TEMPORAL, COMPARISON and CONTINGENCY relations. On the other hand, when *manually-labeled Explicit discourse relations* are added to training data, the SVM model performs better than the MaxEnt except for the sense TEMPORAL. However, the feature set which achieves the maximum performance seems to be less informative with the MaxEnt classifier. SVM results are found to reveal the differences in sentence structures among CLASS senses in a more explanatory way.

[1] SUMMARY sense belongs to the RST framework but can be thought of a Type relation of the EXPANSION sense of the PDTB 2.

[2] This issue was mentioned in Section 4.3; see Table 4.1 4.2 .

CHAPTER 7

CONCLUSION

The current thesis, to our best knowledge, is the first study about automatic sense prediction of Implicit discourse relations in Turkish. Using the annotations performed on TDB-Subcorpus, which are partly the output of the recent annotation effort by TDB team and partly produced by the author and the supervisor of the thesis, fully supervised classifiers are built.

The implementations for the classifiers are from libSVM for the SVM model and from the MALLET for the MaxEnt model. For each CLASS level sense, a binary classification task, such as TEMPORAL vs. others, is set. Several linguistically rich features such as polarity, tense information regarding the main verb of the arguments of the discourse relations are used during classification.

The results reveal several syntactic patterns that are characteristic to certain senses. For example, TEMPORAL discourse relations are found, on general, to be composed of the sentences which do not have negated verbs in their arguments. That is because those relations often are used to describe a series of events. On the other hand, COMPARISON and CONTINGENCY relations have an argument with a negated verb. In terms of tense, COMPARISON and CONTINGENCY relations behave similarly, in that the tenses of their arguments are likely to be the same whereas in CONTINGENCY relations the tenses of the arguments differ more frequently. EXPANSION relations differ from the rest by having a longer arguments and being more likely to be located in news.

As the second part of the thesis, the effect of the Explicit discourse relations on the sense prediction task is analyzed. In discourse research, how close Explicit discourse relations to Implicit relations is a challenging question. To this end, in order to provide insight regarding the nature of Explicit and Implicit discourse relations, the Explicit discourse relations are added to training data. In literature, the Explicit relations are tried to be converted to Implicit discourse relations, however, the methods used are proved to be fruitless (Sporleder & Lascarides, 2008; Webber, 2009). In the current thesis, instead of using automatically extracted Explicit discourse relations based on the unambiguity of their connective, *manually-labeled Explicit discourse relations* (the Explicit discourse relations senses of which are labeled by human annotators) are added to training data. By doing so, it is aimed to capture the Explicit discourse relations completely, rather than a small subset of entire Explicit relations. Extending the natural Implicit relations by adding the *manually-labeled Explicit discourse relations* leads to an increase in TEMPORAL and COMPARISON relations whereas they decrease the performance for CONTINGENCY and EXPANSION relations. The crucial point that must be noted here is that for all senses the feature set which yield the best performance when only Implicit discourse relations are present differed when *manually-labeled Explicit discourse re-*

lations are added to training data. This finding seems to support the position which denies the hypothesis that Implicit discourse relations are merely Explicit discourse relations without a discourse connective.

In terms of Cognitive Science perspective, firstly, the features implemented are "cognitively plausible" in the sense that they are known to exist in human languages. The thesis has demonstrated, to some extent, the relation between syntax and discourse. The sense of a Implicit discourse relation is actually the meaning inferred by the adjacent reading of the sentences and the results have revealed that certain meanings require certain syntactic configurations.

7.1 Limitations

The similar studies conducted for English have utilized various features including the class to which the main verbs of the arguments belong in Levin Verb Classes or the lexical relations between the word pairs using Wordnet. However, since Turkish lack such linguistic resources, the set of features used in this study is limited. Secondly, thanks to the existence of PDTB, the studies on English have access to thousands of annotations with their sense tags whereas in TDB sense annotation task has recently began. Although extra annotations have been created for this study, annotating discourse relations is a tedious task since it requires at least two experienced annotators. Therefore, in the future, the current study can be reconducted hopefully with more features and larger amount of annotations.

Appendix A

A SAMPLE ANNOTATION IN XML FORMAT

(from 65111)


```
Relation note="result: sonuc olarak"
      sense="Contingency: Cause: result" type="IMPLICIT"
         genre="novel"
       Conn
           Span
               Text Birinci /Text
               BeginOffset 8413 /BeginOffset
               EndOffset 8420 /EndOffset
           /Span
       /Conn
       Mod/
       Arg1 f
           Span
               Text Beni okula gondermeye bir turlu
                 kiyamayan babam
                  Lutfullah Bey ozel hocalar tuttu /Text
               BeginOffset 8331 /BeginOffset
               EndOffset 8411 /EndOffset
           /Span
       /Arg1
       Arg2
           Span
               Text Birinci sinifi evde, cam bir fanus
                 icinde okudum /Text
               BeginOffset 8413 /BeginOffset
               EndOffset 8461 /EndOffset
           /Span
       /Arg2
       Supp1/
       Supp2/
       Shared/
       SuppShared/
   /Relation
```

Appendix B

AMBIGUITY OF EXPLICIT CONNECTIVES IN TURKISH

The statistics provided below are extracted from the TDB-Subcorpus.

Table B.1 provides a distribution and counts of the types of explicit connectives along with their sense types. There are **44 distinct types** of Explicit connectives.

Table B.2 provides a distribution of all the distinct senses annotated for explicit connectives. **29 distinct senses** are recorded for Explicit connectives.

Table B.3 provides the list of unambigous Explicit discourse connectives in Turkish along with their senses.

Table B.4 provides the list of unambigous Explicit discourse connectives in Turkish along with their senses.

EXPLICIT CONNECTIVE	SENSES
ama (118)	Opposition (11), Pragmatic Contrast (34), Concession (31), Contrast (29), Juxtaposition (4), Contra-Expectation (3), Conjunction (1), Exception (4), Pragmatic Concession (1)
amacıyla (6)	Purpose (6)
ancak (28)	Opposition (7), Contrast (2), Juxtaposition (3), Contra-Expectation (13), Exception (2), Condition (1)
ardından (5)	Precedence (3), Succession (2)
aslında (11)	Pragmatic Contrast (1), Contra-Expectation (2), Specification (5), Conjunction (1), Exception (1)
ayrıca (20)	Conjunction (20)
beraber (1)	Conjunction (1)
bir yandan (3)	Conjunction (2), Synchronous (1)
birlikte (1)	Expectation (1)
böylece (9)	Result (7), Justification (1), Pragmatic Cause (1)
çünkü (37)	Reason (37)
dahası (3)	Specification (1), Conjunction (2)
dolayı (3)	Reason (3)
dolayısıyla (9)	Result (9)
fakat (11)	Pragmatic Contrast (2), Concession (1), Contrast (8)
gene de (2)	Contra-Expectation (2)
gibi (18)	Conjunction (1), Expansion (6), Manner (11)
halbuki (4)	Opposition (4)
halde (5)	Expectation (2), Condition (1), Manner (2)
hem (5)	Specification (1), Conjunction (3), Conjunctive (1)
için (91)	Purpose (72), Reason (18)
içindir (1)	Purpose (1)
iken (2)	Opposition (1), Condition (1)
kadar (13)	Precedence (2), Expansion (11)
karşın (9)	Contra-Expectation (1), Expectation (8)
mesela (2)	Instantiation (2)
ne ki (1)	Opposition (1)
ne var ki (1)	Juxtaposition (1)
ne (4)	Conjunction (1), Conjunctive (3)
nedeniyle (4)	Reason (4)
önce (21)	Precedence (9), Succession (7), Synchronous (1)
örneğin (3)	Instantiation (3)
oysa (14)	Opposition (5), Pragmatic Contrast (1), Contrast (2), Juxtaposition (1), Contra-Expectation (4), Comparison (1)
rağmen (8)	Expectation (8)
sonra (72)	Precedence (26), Succession (44)
tersine (1)	Opposition (1)
ve (234)	Specification (1), Conjunction (194), Precedence (3), Result (21), Asynchronous (1), Synchronous (4), Precedence (9), List (1)
veya (1)	Disjunctive (1)
ya da (12)	Conjunctive (6), Disjunctive (6)
ya (1)	Disjunctive (1)

yalnız (2)	Opposition (1), Comparison (1)
yine de (1)	Contra-Expectation (1)
yoksa (4)	Disjunctive (2), Condition (2)
zaman (7)	Asynchronous (1), Condition (3), Synchronous (3)

Table B.1: Explicit Connective – Sense Table

Table B.2: Sense – Explicit Connective Table

SENSE	EXPLICIT CONNECTIVES
Asynchronous (2)	ve (1), zaman (1)
Comparison (2)	oysa (1), yalnız (1)
Concession (32)	ama (31), fakat (1)
Condition (8)	ancak (1), yoksa (2), iken (1), zaman (3), halde (1)
Conjunction (226)	ama (1), aslında (1), gibi (1), ne (1), ve (194), ayrıca (20), hem (3), dahası (2), bir yandan (2), beraber (1)
Conjunctive (10)	ne (3), ya da (6), hem (1)
Contra-Expectation (26)	ama (3), ancak (13), aslında (2), gene de (2), oysa (4), karşın (1), yine de (1)
Contrast (41)	ama (29), ancak (2), oysa (2), fakat (8)
Disjunctive (12)	ya da (6), yoksa (2), ya (1), veya (1)
Exception (7)	ama (4), ancak (2), aslında (1)
Expansion (17)	gibi (6), kadar (11)
Expectation (19)	rağmen (8), karşın (8), halde (2), birlikte (1)
Instantiation (5)	mesela (2), örneğin (3)
Justification (1)	böylece (1)
Juxtaposition (9)	ama (4), ancak (3), ne var ki (1), oysa (1)
List (1)	ve (1)
Manner (13)	gibi (11), halde (2)
Opposition (31)	ama (11), ancak (7), oysa (5), ne ki (1), iken (1), yalnız (1), halbuki (4), tersine (1)
Pragmatic Cause (1)	böylece (1)
Pragmatic Concession (1)	ama (1)
Pragmatic Contrast (38)	ama (34), aslında (1), oysa (1), fakat (2)
Precedence (43)	önce (9), sonra (26), ve (3), kadar (2), ardından (3)
Precedence (9)	ve (9)
Purpose (79)	amacıyla (6), için (72), içindir (1)
Reason (62)	çünkü (37), için (18), dolayı (3), nedeniyle (4)
Result (37)	ve (21), böylece (7), dolayısıyla (9)
Specification (8)	aslında (5), ve (1), hem (1), dahası (1)
Succession (53)	önce (7), sonra (44), ardından (2)
Synchronous (9)	önce (1), ve (4), zaman (3), bir yandan (1)

Table B.3: List of Unambiguous Explicit Connectives in Turkish

EXPLICIT CONNECTIVE	SENSE
amacıyla	Purpose
ayrıca	Conjunction
beraber	Conjunction
birlikte	Expectation
çünkü	Reason
dolayı	Reason
dolayısıyla	Result
gene de	Contra-Expectation
halbuki	Opposition
içindir	Purpose
mesela	Instantiation
ne ki	Opposition
ne var ki	Juxtaposition
nedeniyle	Reason
örneğin	Instantiation
rağmen	Expectation
tersine	Opposition
veya	Disjunctive
ya	Disjunctive
yine de	Contra-Expectation

Table B.4: The 10 Most Ambiguous Explicit Connectives in Turkish

EXPLICIT CONNECTIVE	SENSE
ama (9)	Opposition, Pragmatic Contrast, Concession, Contrast, Juxtaposition, Contra-Expectation, Conjunction, Exception, Pragmatic Concession
ve (8)	Specification, Conjunction, Precedence, Result, Asynchronous, Synchronous, Precedence, List
ancak (6)	Opposition, Contrast, Juxtaposition, Contra-Expectation, Exception, Condition
oysa (6)	Opposition, Pragmatic Contrast, Contrast, Juxtaposition, Contra-Expectation, Comparison
aslında (5)	Pragmatic Contrast, Contra-Expectation, Specification, Conjunction, Exception
böylece (3)	Result, Justification, Pragmatic Cause
fakat (3)	Pragmatic Contrast, Concession, Contrast
gibi (3)	Conjunction, Expansion, Manner
halde (3)	Expectation, Condition, Manner
hem (3)	Specification, Conjunction, Conjunctive

Bibliography

Aktaş, B., Bozsahin, C., & Zeyrek, D. (2010). Discourse relation configurations in turkish and an annotation environment. In *Proceedings of the fourth linguistic annotation workshop* (pp. 202–206).

Asher, N. (1993). *Reference to abstract objects in discourse* (Vol. 50). Springer Science & Business Media.

Banko, M., & Brill, E. (1998). Scaling to Very Very Large Corpora for Natural Language Disambiguation. , 26–33.

Berger, A. L., Pietra, V. J. D., & Pietra, S. A. D. (1996). A maximum entropy approach to natural language processing. *Computational linguistics*, *22*(1), 39–71.

Biran, O., & McKeown, K. (2013). Aggregated word pair features for implicit discourse relation disambiguation..

Biran, O., & McKeown, K. (2015). Pdtb discourse parsing as a tagging task: The two taggers approach. In *16th annual meeting of the special interest group on discourse and dialogue* (p. 96).

Blair-goldensohn, S., Mckeown, K. R., & Rambow, O. C. (2002). Building and Refining Rhetorical-Semantic Relation Models.

Brown, P. F., Desouza, P. V., Mercer, R. L., Pietra, V. J. D., & Lai, J. C. (1992). Class-based n-gram models of natural language. *Computational linguistics*, *18*(4), 467–479.

Carlson, L., Marcu, D., & Okurowski, M. E. (2003). Building a discourse-tagged corpus in the framework of rhetorical structure theory. In *Current and new directions in discourse and dialogue* (pp. 85–112). Springer.

Chang, C.-C., & Lin, C.-J. (2011). Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *2*(3), 27.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*(3), 273–297.

Degand, L., Simon, A. C., et al. (2005). Minimal discourse units: Can we define them, and why should we. *Proceedings of SEM-05. Connectors, discourse framing and discourse structure: from corpus-based and experimental analyses to discourse theories, Biarritz*, 14–15.

Demirşahin, I., Sevdik-Çallı, A., Balaban, H. Ö., Çakıcı, R., & Zeyrek, D. (2012). Turkish discourse bank: Ongoing developments. In *First workshop on language resources and technologies for turkic languages* (p. 15).

Denis, P. (2014). Combining Natural and Artificial Examples to Improve Implicit Discourse Relation Identification Implicit Discourse Relation Identification. (August).

Hernault, H., Bollegala, D., & Ishizuka, M. (2011). Semi-supervised Discourse Relation Classification with Structural Learning. , 340–352.

Hernault, H., Prendinger, H., DuVerle, D. A., Ishizuka, M., & Paek, T. (2010). Hilda: a discourse parser using support vector machine classification. *Dialogue and Discourse*, *1*(3), 1–33.

Hobbs, J. R. (1985). *On the coherence and structure of discourse*. CSLI.

Hovy, E. H. (1990). Parsimonious and profligate approaches to the question of discourse structure relations. In *Proceedings of the fifth international workshop on natural language generation* (pp. 128–136).

Huang, H.-h., & Chen, H.-h. (2011). Chinese Discourse Relation Recognition. (2002), 1442–1446.

Ji, Y., Haffari, G., & Eisenstein, J. (2016). A latent variable recurrent neural network for discourse relation language models. *arXiv preprint arXiv:1603.01913*.

Knott, A., & Sanders, T. (1998). The classification of coherence relations and their linguistic markers: An exploration of two languages. *Journal of pragmatics*, *30*(2), 135–175.

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. University of Chicago press.

Lin, Z., Kan, M.-y., & Ng, H. T. (2009). Recognizing Implicit Discourse Relations in the Penn Discourse Treebank. (August), 343–351.

Lin, Z., Ng, H. T., & Kan, M.-Y. (2014). A pdtb-styled end-to-end discourse parser. *Natural Language Engineering*, *20*(02), 151–184.

Louis, A., Joshi, A., Prasad, R., & Nenkova, A. (2010). Using entity features to classify implicit discourse relations. , 59–62.

Mann, W. C., & Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text-Interdisciplinary Journal for the Study of Discourse*, *8*(3), 243–281.

Marcu, D., Echihabi, A., & Rey, M. (2002). An Unsupervised Approach to Recognizing Discourse Relations. (July), 368–375.

McCallum, A. K. (2002). Mallet: A machine learning for language toolkit.

Meyer, D., & Wien, F. T. (2015). Support vector machines. *The Interface to libsvm in package e1071*.

Miltsakaki, E., Dinesh, N., Prasad, R., Joshi, A., & Webber, B. (2005). Experiments on sense annotations and sense disambiguation of discourse connectives. In *Proceedings of the fourth workshop on treebanks and linguistic theories (tlt2005), barcelona, spain, december.*

Miltsakaki, E., Prasad, R., Joshi, A., & Webber, B. (2004). The Penn Discourse Treebank. , 1–4.

Oepen, S., Read, J., Scheffler, T., Sidarenka, U., Stede, M., Velldal, E., & Øvrelid, L. (2016). Opt: Oslo—potsdam—teesside pipelining rules, rankers, and classifier ensembles for shallow discourse parsing..

Oza, U., Prasad, R., Kolachina, S., Sharma, D. M., & Joshi, A. (2009). The hindi discourse relation bank. In *Proceedings of the third linguistic annotation workshop* (pp. 158–161).

Park, J., & Cardie, C. (2012). Improving Implicit Discourse Relation Recognition Through Feature Set Optimization. (July), 108–112.

Patterson, G., & Kehler, A. (2013). Predicting the Presence of Discourse Connectives. (October), 914–923.

Pettibone, J., & Pon-barry, H. (2003). A Maximum Entropy Approach to Recognizing Discourse Relations in Spoken Lanugage. (2002), 1–10.

Pitler, E., Louis, A., & Nenkova, A. (2009). Automatic sense prediction for implicit discourse relations in text. (August), 683–691.

Pitler, E., & Nenkova, A. (2009). Using Syntax to Disambiguate Explicit Discourse Connectives in Text. (August), 13–16.

Pitler, E., Raghupathy, M., Mehta, H., Nenkova, A., Lee, A., & Joshi, A. K. (2008). Easily identifiable discourse relations. *Coling 2008: Companion volume: Posters and Demonstrations*, 85–88.

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., & Webber, B. (2008). The Penn Discourse TreeBank 2.0. *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, 1–4. Retrieved from `http://www.seas.upenn.edu/{~}pdtb/papers/lrec04.pdf` doi: 10.1.1.165.9566

Prasad, R., Joshi, A., & Webber, B. (2010a). Realization of discourse relations by other means: alternative lexicalizations. In *Proceedings of the 23rd international conference on computational linguistics: Posters* (pp. 1023–1031).

Prasad, R., Joshi, A., & Webber, B. (2010b). Realization of Discourse Relations by Other Means : Alternative Lexicalizations. (1), 1023–1031.

Prasad, R., Miltsakaki, E., Dinesh, N., Lee, A., Joshi, A., Robaldo, L., & Webber, B. L. (2007). The penn discourse treebank 2.0 annotation manual.

Prasad, R., Webber, B., & Joshi, A. (2014). Reflections on the penn discourse treebank, comparable corpora, and complementary annotation. *Computational Linguistics*.

Ramesh, B. P., Prasad, R., Miller, T., Harrington, B., & Yu, H. (2012). Automatic discourse connective detection in biomedical text. , 800–808. doi: 10.1136/amiajnl-2011-000775

Redeker, G., & Bouma, G. (2009). Computational Analysis of Coherence Relations in Dutch. , 1–6.

Rutherford, A., & Xue, N. (2014). Discovering implicit discourse relations through brown cluster pair representation and coreference patterns. In *Eacl* (Vol. 645, p. 2014).

Rutherford, A., & Xue, N. (2015). Improving the Inference of Implicit Discourse Relations via Classifying Explicit Discourse Connectives. *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 799–808. Retrieved from `http://www.aclweb.org/anthology/N15-1081`

Saito, M. (2006). Using Phrasal Patterns to Identify Discourse Relations. (June), 133–136.

Saito, M., Yamamoto, K., & Sekine, S. (2006). Using phrasal patterns to identify discourse relations. In *Proceedings of the human language technology conference of the naacl, companion volume: Short papers* (pp. 133–136).

Sak, H., Güngör, T., & Saraçlar, M. (2007). Morphological disambiguation of turkish text with perceptron algorithm. In *International conference on intelligent text processing and computational linguistics* (pp. 107–118).

Sanders, T. J., Spooren, W. P., & Noordman, L. G. (1992). Toward a taxonomy of coherence relations. *Discourse processes*, *15*(1), 1–35.

Say, B., Zeyrek, D., Oflazer, K., & Özge, U. (2002). Development of a corpus and a treebank for present-day written turkish. In *Proceedings of the eleventh international conference of turkish linguistics* (pp. 183–192).

Sharma, H., Dakwale, P., Sharma, D. M., Prasad, R., & Joshi, A. (2013). Assessment of different workflow strategies for annotating discourse relations: A case study with hdrb. In *International conference on intelligent text processing and computational linguistics* (pp. 523–532).

Soricut, R., & Marcu, D. (2003). Sentence level discourse parsing using syntactic and lexical information. In *Proceedings of the 2003 conference of the north american chapter of the association for computational linguistics on human language technology-volume 1* (pp. 149–156).

Sporleder, C., & Lascarides, A. (2008). Using automatically labelled examples to classify rhetorical relations: An assessment. *Natural Language Engineering*, *14*(3), 369–416.

Taboada, M. (2009). Implicit and explicit coherence relations. *Discourse, of course. Amsterdam: John Benjamins*, 127–140.

The PDTB Research Group. (2007). The Penn Discourse Treebank 2.0 Annotation Manual.

Van Dijk, T. A. (1980). Text and context explorations in the semantics and pragmatics of discourse.

Vliet, N. V. D., Bouma, G., & Redeker, G. (2013). The automatic identification of discourse units in Dutch text. (Nodalida), 411–421.

Wang, J., & Lan, M. (2015). A refined end-to-end discourse parser. *CoNLL 2015*, 17.

Wang, X., Li, S., Li, J., & Li, W. (2012). Implicit Discourse Relation Recognition by Selecting Typical Training Examples. , *2*(December 2012), 2757–2772.

Webber, B. (2004). D-ltag: extending lexicalized tag to discourse. *Cognitive Science*, *28*(5), 751–779.

Webber, B. (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* (pp. 674–682).

Wellner, B., & Pustejovsky, J. (2006). Classification of Discourse Coherence Relations : An Exploratory Study using Multiple Knowledge Sources. (July), 117–125.

Wolf, F., & Gibson, E. (2005). Representing discourse coherence: A corpus-based study. *Computational Linguistics*, *31*(2), 249–287.

Xu, Y., Lan, M., Lu, Y., Niu, Z. Y., & Tan, C. L. (2012). Connective prediction using machine learning for implicit discourse relation classification. In *The 2012 international joint conference on neural networks (ijcnn)* (pp. 1–8).

Xue, N., Ng, H. T., Pradhan, S., Bryant, R. P. C., & Rutherford, A. T. (2015). The conll-2015 shared task on shallow discourse parsing. In *Proceedings of conll* (p. 2).

Zeyrek, D., Demirşahin, I., Sevdik-Çallı, A., & Çakıcı, R. (2013). Turkish discourse bank: Porting a discourse annotation style to a morphologically rich language. *Dialogue and Discourse*, *4*(2), 174–184.

Zeyrek, D., Demirşahin, I., Sevdik-Çallı, A., & Kurfalı, M. (2015, July). Annotating implicit discourse connectives in turkish: The challenge of corrective discourse relations. In *Oral presentation at the 14th ipra conference, july 26-31, 2015, antwerp, belgium.*

Zeyrek, D., & Webber, B. L. (2008). A discourse resource for turkish: Annotating discourse connectives in the metu corpus. In *Ijcnlp* (pp. 65–72).

Zhou, Y., & Xue, N. (2012). Pdtb-style discourse annotation of chinese text. In *Proceedings of the 50th annual meeting of the association for computational linguistics: Long papers-volume 1* (pp. 69–77).

Zhou, Y., & Xue, N. (2015). The chinese discourse treebank: a chinese corpus annotated with discourse relations. *Language Resources and Evaluation*, *49*(2), 397–431.

Zhou, Z. M., Lan, M., Niu, Z. Y., Xu, Y., & Su, J. (2010). The Effects of Discourse Connectives Prediction on Implicit Discourse Relation Recognition. , 139–146.

Zhou, Z.-M., Xu, Y., Niu, Z.-Y., Lan, M., Su, J., & Tan, C. L. (2010). Predicting discourse connectives for implicit discourse relation recognition. In *Proceedings of the 23rd international conference on computational linguistics: Posters* (pp. 1507–1514).