THE ROLE OF MODIFIERS IN TURKISH DISCOURSE BANK

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

BY

DENİZ HANDE ÇAKMAK

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

OF

MASTER OF SCIENCE

IN

THE DEPARTMENT OF COGNITIVE SCIENCE

SEPTEMBER 2015

THE ROLE OF MODIFIERS IN TURKISH DISCOURSE BANK

Submitted by DENİZ HANDE ÇAKMAK in partial fulfillment of the requirements for the degree of **Master of Science in Cognitive Science Department, Middle East Technical University** by,

Prof. Dr. Nazife Baykal
Director, **Informatics Institute** _____

Prof. Dr. Cem Bozşahin
Head of Department, **Cognitive Science** _____

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

Examining Commitee Members:

Prof. Dr. Yeşim Aksan
English Linguistics, Mersin University _____

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

Assisst. Prof. Dr. Cengiz Acartürk
Cognitive Science, METU _____

Prof. Dr. Hüseyin Cem Bozşahin
Cognitive Science, METU _____

Assist. Prof. Dr. Murat Perit Çakır
Cognitive Science, 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: Deniz Hande Çakmak

Signature : _____

ABSTRACT

THE ROLE OF MODIFIERS IN TURKISH DISCOURSE BANK

Çakmak, Deniz Hande

MS, Department of Cognitive Science

Supervisor: Prof. Dr. Deniz Zeyrek Bozşahin

September 2015, 68 pages

This thesis focuses on the role of modifiers used with discourse connectives and investigates modifiers in terms of affecting the modality of the discourse relations. Modifiers are originally adverbs used for different semantic purposes. The already annotated data in Turkish Discourse Bank, or TDB are used for identifying modifiers and classifying them. In the light of previous studies, the discourse connectives occuring with modifiers are analyzed and classified. The semantic distribution of modifiers is determined and their effect on the modality of discourse relations are examined. The data are put into a decision-tree algorithm and 10-fold cross-validation was applied. The results confirm that modifiers of discourse connectives have a role in discourse relations in terms of the modality of the relation.

Keyword: Discourse Connectives, Discourse Relations, Modifiers, Turkish Discourse Bank, Decision Tree

ÖZ

TAMLAYICI SÖZCÜKLERİN TÜRKÇE SÖYLEM BANKASI'NDAKİ ROLÜ

Çakmak, Deniz Hande

Yüksek lisans, Bilişsel Bilimler Bölümü

Tez Yöneticisi: Prof. Dr. Deniz Zeyrek Bozşahin

Eylül 2015, 68 sayfa

Bu tez, esasında tamlayıcı sözcüklerin söylem bağlaçlarındaki rolü üzerine odaklıdır ve tamlayıcı sözcüklerin söylem ilişkileri üzerinde kiplik açısından etkisini araştırmaktadır. Tamlayıcı sözcükler kökeninde, farklı anlamsal amaçlar için kullanılan zarflardır. Daha önceden açımlanan Türkçe Söylem Bankası, veya TDB verileri bu tamlayıcı sözcükleri tanımlamak ve sınıflandırmak için kullanılmıştır. Önceki çalışmaların ışığında, tamlayıcı sözcüklerle bulunan söylem bağlaçları incelenmiş ve sınıflandırılmıştır. Tamlayıcı sözcüklerin anlamsal ayrımları belirlenmiş ve söylem ilişkilerinin kipliği üzerinde etkileri incelenmiştir. Veri ve sınıflandırmalar ölçümlenmeye alınmış ve 10-kat- çapraz geçerlilik testi ile bir karar ağacı uygulanmıştır. Sonuçlar söylem bağlaçlarına ait tamlayıcı sözcüklerin söylem ilişkisinin kipliği açısından söylem üzerinde rolü olduğunu göstermiştir.

Anahtar Kelimeler: Söylem Bağlaçları, Söylem İlişkileri, Tamlayıcı Sözcükler, Türkçe Söylem Bankası, Karar Ağacı

ACKNOWLEDGEMENTS

I express sincere appreciation to Prof. Dr. Deniz Zeyrek Bozşahin for her excellent guidance and teaching. This work has become possible with her enthusiasm and love for linguistics and cognitive science. Thanks to Assist. Prof. Dr. Cengiz Acartürk for his guidance and helpful suggestions. Fikret Arslan and Savaş Çetin are also gratefully acknowledged for their support and help during the annotation process.

To my mother Nurcan and my father Yüksel, I offer sincere thanks for their faith in me and endless support. I am also indebted to İrem Çorum, Ant Dolay, Gözde Adar, İpek Günad, Elif Eren, Cemre Çelik and Melek Başdal for their emotional support and friendship.

All remaining errors belong to me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

TDB	: Turkish Discourse Bank
PDTB	: Penn Discourse Tree Bank
ARG1	: Argument 1
ARG 2	: Argument 2
RST	: Rhetorical Structure Theory
FP	: Focus Particle

CHAPTER 1

INTRODUCTION

This chapter introduces the terms discourse and discourse relations, explicit discourse connectives and modifiers respectively. The second part of the chapter presents the aim and purpose of the thesis and continues with the method used in the thesis.

1.1 Discourse and discourse relations

Discourse is a form of language use (Brown &Yule 1993; Van Dijk, 1997), and "It means actual instances of communication in the medium of a language" (Johnstone, 2002, p: x). Discourse can be described by the structures formed across clauses or sentences. These discourse structures constitute a collection of sentences which possess semantic relations between them. According to Kehler (2002: 3), "the semantic relationship between the arguments is *discourse relation*". While different terms for discourse relations are available in the literature such as *rhetorical predicates* (Grimes, 1975), *conjunctive relations* (Halliday and Hasan,1976), or *discourse relations* (Sanders et al.,1992); the function and meaning of all of them is the same. The term *discourse relations* is going to be used in this thesis.

Discourse relations may be expressed explicitly or implicitly (Halliday and Hasan, 1976). There are specific words or phrases which make the discourse relation explicit; they are called *discourse markers (or discourse connectives)* (Risselada and Spooren, 1998; Schiffrin, 1987). Discourse markers are composed of conjunctions, adverbs or other kinds of phrases that connect two or more discourse units (Taboada, 1998). These markers help the discourse units to form a discourse relation (Risselada and Spooren, 1998) or give hints about the discourse relations (Sanders et al., 1992).

According to Webber and Joshi (1998) and Webber (2004); discourse connectives are discourse-level predicates that express the discourse relation between the discourse units explicitly. The examples below indicate the discourse relations with explicit discourse connectives. Examples (1) and (3) are taken from Penn Discourse Tree Bank (PDTB) while the other two examples are taken from METU Turkish Discourse Bank (TDB). The discourse relation category, its type and sub-type are shown in parenthesis. The discourse relation category is capitalized, its type is shown by capitalizing the first letter, and its subtype is shown in normal fonts. This type of categorization is taken from the PDTB annotation manual (Prasad et al., 2007). The connective is underlined.

(1) But a Soviet bank here would be crippled unless Moscow found a way to settle the $188 million debt, which was lent to the country's short-lived democratic Kerensky government <u>before</u> the Communists seized power in 1917.

(TEMPORAL:Asynchronous:precedence)

(2) Okkan'ın ölmeden <u>önce</u> açıkladığı 26 kişilik tetikçi listesinde adı bulunan Sarıağaç'ın elindeki silahın, Okkan suikastında da kullanıldığı anlaşıldı.

'It was understood that the gun in the hands of Sarıağaç, who was in the 26- people hitmen list which Okkan revealed before he died, was also used in the assassination of Okkan.

(TEMPORAL:Asynchronous:precedence)

(3) In addition, its machines are typically easier to operate, so customers require less assistance from software.

(CONTINGENCY:Cause:result)

(4) Erkböke, romantizm öldü artık, ayağa düştü, direnenlerin son temsilcisi benim. Zaten şiiri bu yüzden bıraktı Kerim, diyor Metin.

'Metin says "Erkböke says that romantism had died and gone to the floor, I am the last representator of the resistants. As a matter of fact, Kerim left poetry because of this."'

(CONTINGENCY:Cause:result)

Examples (1) and (2) above have a temporal discourse relation which is clarified with the help of the explicit connectives *before* and *önce* 'before'. On the other hand, Ex. (3) and (4) express a contingency relation and the relation is presented by *so* and *bu yüzden "due to this reason"*. Note that in (4), there is an alternative lexicalization rather than a connective. An alternative lexicalization is a discourse marker which expresses the discourse relation between the sentences. However; "it also provides an implicit connective which can be alternatively lexicalized by some non-connective expressions" (Prasad et al., 2007:22) (e.g. after that, because of this)

Many connectives can occur with adverbs such as *only, at least* etc. These adverbs help the discourse connectives define the discourse relation more precisely and are called discourse connective *modifiers*. In English, modifiers are adverbs that appear together with the connective.

Below, example (5) conveys a contingency relation which makes the relation explicit by means of *since*. The adverb underlined is the modifier, specifying the meaning of the connective *since*. Example (6) has a contingency relation that is modified by *partly* . The modifier conveys that there are some other reasons for the event.

(5) *That power can sometimes be abused*, particularly **since jurists in smaller jurisdictions operate without many of the restraints that serve as corrective measures in urban areas.** (Prasad et al., 2007,p: 9)
(6) We're seeing it partly because older vintages are growing more scarce. (Prasad et al., 2007,p: 9).

The overall aim of the thesis is to identify the modifiers of discourse connectives in TDB and classify their meanings. This will allow us a means to determine the semantic effect of modifiers on discourse relations and can also help in identifying the category of the discourse relation both linguistically and in further language technology applications (e.g. automatic translation).

1.2 Types of modifiers in TDB

In this thesis, we specify the semantic distinctions between various modifiers annotated in the TDB. The TDB is a resource of written Turkish including approximately 400.000 words from various genres (Zeyrek et al., 2010; Zeyrek et al., 2013). Explicit discourse connectives and the discourse segments they connect are annotated. TDB shares the same annotation principles with PDTB in terms of annotating the discourse segments and explicit discourse connectives. PDTB itself is composed of a written corpus and it contains the annotation of argument structure, sense and attribution of discourse connectives and their arguments (Prasad et al., 2007).

In PDTB, the arguments are the clauses or sentences that are related to each other. They are annotated as ARG1 and ARG2. ARG2 is the argument that syntactically or morphologically hosts the discourse connective, and ARG1 refers to the other argument which is related by the connective (Zeyrek et al., 2013). In this thesis, ARG1 will be presented in italics and ARG2 will be presented in bold. Discourse connectives are bolded and underlined and modifiers are underlined.

In TDB, 540 modifiers are annotated, but their semantic differences have not been marked. This thesis will fill this gap by identifying the modifier's semantic distinction. This classification will contribute to Turkish discourse in a theoretical sense. It will also contribute to language automation systems by doing a decision tree study on the types of discourse connectives and their modifiers.

The examples below are some of the discourse structures annotated in TDB. In (7), the discourse connective *ardından* 'later' sets a succession temporal relation between the two arguments. The modifier of the connective *hemen* 'immediately' intensifies the temporal relation and the function of the discourse connective in the relation. The discourse in (8) is a contingency-pragmatic cause relation supported by the alternative lexicalization *bu yüzden* 'because of this'. The modifier *belki* 'maybe', which is an epistemic adverb, modifies the connective in terms of the possibility with which the eventuality may occur. The discourse relational device[1] in (9) is also a contingency-cause relation which is set by *bunun için* 'for this'. The modifier *sırf* 'just' conveys that the reason presented in the discourse relation is not the only one.

(7) *Bursa hakkında daha bir meraklı olmakta* ve hemen **ardından ansiklopedinin maddelerine bakmakta değil mi?..**

'He is becoming curious about Bursa, and immediately after that he looks at the contents of encyclopedia, doesn't he?'

(TEMPORAL: Asynchronous: succession)

(8) *Bu köyde doğup büyümüştü sanki; çınarın hışırtısı kulaklarını, damların köy alanına akan beyazlığı gözlerini, toprağın cayır cayır yanan soluğu tenini hiç etkilemiyordu.* Belki de **bu yüzden, o gün bekçiden başka hiç kimse bir yabancıyla aynı yerde bulunmanın tedirginliğini duymamıştı kendinde.**

'As if he had been born and grown up in this village; the rustle of the plane tree did not affect his ears, the whiteness of the roofs flowing over the village center did not disturb his eyes and the burning breath of soil did not affect his skin. Maybe because of this, he was not nervous about anything except being in the same place with a warder on that day.'
(CONTINGENCY: Pragmatic cause)

(9) Ekonomik bir yükmüş bu. *Bir kısmının ölmesi gerekiyormuş sessiz sedasız.* Sırf **bunun için havası yavaş yavaş boşaltılan kabinler hazırlamışlar.**

' It was an economical burden. Some of them had to die silently. Just for this, they prepared some cabinets in which the air was emptied slowly.' (CONTINGENCY: Pragmatic cause)

1.3 Content, epistemic and speech-act modality

A line of discourse research deals with whether the connective expresses an epistemic domain, or a real word event. If a sentence expresses an event in the real world, it is said that it takes place in the *content* domain (Sweetser, 1990). *Epistemic domain* presents the representation of the relation as being outside of the real world (Sweetser, 1990). It expresses the addresser's judgement or assumption on the relation, while content domain just gives information about the event without the comment of the addresser (Knott et al., 2001). This difference between content domain and epistemicity are alternatively named as *objectivity* and *subjectivity* in Knott et al. (2001), Spooren et al. (2010) and Maat and Sanders (2001); the terms content and epistemic will be used in this study. On the other

[1] In TDB, connective devices that include a deictic item such as *bunun için* 'for this' are annotated along with explicit connectives such as conjunctions (ve, sonra etc.) and discourse adverbials (ardından, ne var ki etc.). These are called phrasal expressions in TDB.

hand, there is also another relation category that holds between two sentences, where the aim is to motivate a realization in the discourse context (Sweetser, 1990). This relation category, which is called *speech-act domain*, provides the addresser to express their intention of stating the relation, and the addressee acts according to the intention of the utterance (Sweetser, 1990). The speech-act domain has differences from epistemic domain.

(10) The neighbors left for Paris last Friday. So they are not at home. (Knott et al., 2001:202) (Content domain)
(11) There may be a six-pack in the fridge, but I am not sure because Joe had friends over last night. (Sweetser, 1990, p: 70) (Epistemic domain)
(12) There may be a six-pack in the fridge, but we have work to do. (Sweetser, 1990, p: 70) (Speech-act domain)

Example (10) expresses a content domain presenting two related events in the real world. On the other hand; although examples (11) and (12) express modality, their functions are different. Example (11) expresses the possibility of the situation and the judgement of the addresser. It is an epistemic domain example. On the other hand, in example (12), the addresser implies that they should keep working even if there is a six-pack in the fridge. This example is in the speech-act domain.

1.4 Aim and Method

The aim of the current study is to identify the role of modifiers of discourse connectives in Turkish discourse. The role of modifiers and their effect on the sense of discourse connectives are a unique aspect of discourse relations since modifiers can also influence the modality of the discourse relation and help to express them as content, epistemic or speech-act modalities. Therefore, the goal of the present thesis is to find out how modifiers affect the discourse relation in Turkish. We have a two-pronged approach, a corpus-based approach and a computational part, in which the aim is to understand if there is a systematic role of modifiers on the meaning (e.g. cause-effect) and modality (i.e. content, epistemic and speech act modalities) of the discourse connective.

The research questions of this study are therefore as follows:

- Which modifiers are used with explicit discourse connectives and alternative lexicalizations?
- Do modifiers affect the meaning of explicit discourse connectives and their modality?
- Can we find a systematic effect of modifiers on the meaning of the connective/ alternative lexicalizations and their modality?

In the PDTB, although modifiers are studied in terms of discourse units, they aren't annotated in isolation, but they are taken into consideration with their discourse connectives (Prasad et al., 2007). In other words, PDTB does not annotate the type of the modifier category. In the TDB, explicit discourse connectives/ alternative lexicalizations and their arguments are annotated togehter with their two arguments and modifiers, following the principles of PDTB. However, there has not been any sense annotation of the discourse relation in TDB. The sense of modifiers and their effect on discourse relations have not been studied, either. To fill these gaps, in this study, we develop a framework for annotating the sense of discourse relations considering both the sense of explicit discourse connectives/ alternative lexicalizations and modifiers in Turkish. For example, our annotation framework has the following categories:

(13) Böylece **yurtdışına çıkışından** 19 yıl **sonra** *Karadeniz türkülerinden oluşturduğu "Lazutlar" albümünü hazırladı.*
‘ Thus; 19 years after goaing abroad, he prepared the album “Lazutlar” which he composed of Black Sea Region folk songs.’

Sense of the relation TEMPORAL: Asynchronous: succession
Sense of the modifier: quantitative
Modality of the relation together with the modifier: content
Modality of the relation without the modifier: content

In this way, we have created a set of the data which constitutes the sense and modality of the connective and the modifier at the same time. The epistemicity and semantic distribution of the connective device have been identified. We then put the data into a computational environment. A decision tree algorithm is used to classify the effect of modifiers on the sense of discourse connectives; after which we examine the effect of modifiers on a new set of data with explicit discourse connective devices. The decision tree algorithm is "a "divide-and-conquer" approach to the problem of learning from independent instances which leads naturally to a style of representation" (Witten et al., 2011, p:192). We aimed to understand the semantics of modifiers and test the predictibility of the data by means of a decision tree algorithm. Cross-validation was applied to the data to see the predictibility rate of the data.

The remainder of this thesis has the following structure. Chapter 2 introduces background knowledge about discourse relations, types of modifiers in TDB and three different domains (content, epistemic, speech-act). Chapter 3 introduces the aim and methodology of the thesis. First, the annotation process and the classifications are mentioned and then the computational aspect of the thesis is introduced. Chapter 4 introduces and discusses the results; starting with the classifications obtained from the data and the results of the decision tree algorithm. and Chapter 5 presents the main findings of the thesis.

CHAPTER 2

LITERATURE REVIEW

This chapter gives detailed information about discourse relations, connectives and modifiers, and it continues with the modality issue in discourse relations.

2.1 Discourse relations

The function and meaning of discourse relations have been dealt in detail by various authors (Grosz and Sidner, 1986; Mann and Thompson, 1987; Sanders et al., 1992; Moore and Pollack, 1992; Hobbs, 1990). However, the specific functions or relation classes have not been fully clarified, yet (Knott and Sanders, 1998). There has been several discourse relation theories and classifications, mainly by Mann and Thompson (1987), Mann and Thompson (1988), Schiffrin (1987), Grosz and Sidner (1986), Sanders et al.(1992) and Kehler (2002). Although the classifications are compatible with each other, their philosophies and approaches to the formation of discourse relations differ from each other. These will be introduced below.

2.1.1 Different approaches towards discourse relations

According to Grosz and Sidner (1986:175), "discourse structures are composed of three seperate but interrelated components" which are *linguistic structure, intentional structure* and *attentional state*. Linguistic structure deals with the sequence and constitution of the discourse segments, which are the utterances in discourse. Discourse segments are the clauses related in a semantic way. On the other hand, intentional structure determines the purpose of the discourse within the scope of the linguistic structure. While *discourse segment purpose* represents the intention of discourse segments and their function in discourse, *discourse purpose* indicates the overall purpose in discourse (Grosz and Sidner,1986). The function of attentional state is to give some information about the propositions, events or abstract objects, and it deals with the discourse itself instead of discourse segments (Grosz and Sidner,1986).

Rhetorical Structure Theory suggested by Mann and Thompson (1987;1988) is a descriptive theory which deals with the basic characteristics of the formation of a natural text. This theory is quite useful to classify the structures and relations of the text. Mann and Thompson (1987;1988) suggest that the meanings of the sentences and other units themselves are the key to understand the discourse relation between them and eventually the whole text.

To be able to characterize the discourse relation, the theory first determines the hierarchic structure in the text, which means understanding the sentence or discourse units in isolation. The second phase of the classification is to idenfity the relations between the two sentences. The last step to be done to classify the discourse relation is to clarify the transition between the discourse units with the extent of the items correlated (Mann and Thomson, 1986). This process provides the relations among the sentences in a text. Within the scope of these sets of methods step by step, Mann and Thompson (1987) obtained a very detailed classification of discourse relations. The classification of discourse relations constituted by Mann and Thompson (1987) is shown in the Table 2.1 below.

Table 2.1 The classifications of discourse relations by Mann and Thompson (1987)

Circumstance Solutionhood Elaboration Backgroung Enablement and Motivation ➢ Enablement ➢ Motivation Evidence and Justify ➢ Evidence ➢ Justify Relations of clause ➢ Volutional cause ➢ Non-volutional cause ➢ Volutional result ➢ Non-voluntional result ➢ Purpose	Antithesis and Concession ➢ Antithesis ➢ Concession Condition and Otherwise ➢ Condition ➢ Otherwise Interpretation and Evaluation ➢ Interpretation ➢ Evaluation Restatement and Summary ➢ Restatement ➢ Summary Other relations ➢ Sequence ➢ Contrast

Sanders et al.(1992) claim that the two approaches above are not convincing enough to determine the discourse relations due to the fact that they are more focused on *descriptive adequacy*, which is the appropriate structure to represent the discourse relation. However, discourse relations include more than just descriptive adequacy. Besides descriptive adequacy, *psychological plausibility* is also another inevitable part of discourse relations. "Psychological plausibility concerns the status of discourse relations as cognitive entities" (Sanders et al., 1992:3). Discourse relations and discourse structures affect the cognitive representation of discourse and this leads to discourse understanding, which is the function of discourse relations. It is possible for the hearer to interpret the relation in a different way, even misunderstand. Sanders et al.(1992) conclude that discourse relation may not be understood or interpreted by dealing with a kind of hierarchical structure, because discourse units may not mean anything in isolation and the whole relation may mean a compatible content by taking all the parts into consideration at the same time. Instead of classifications determined by Mann and Thompson (1987;1988), Sanders et al.(1992) constitute a taxonomy of discourse relations based on cognitive representations, obtain combinations of different discourse relations and name them seperately by applying the taxonomy to characterize a specific discourse relation. The criteria for the taxonomy are explained:

- Basic operation: causal or additive

While causal operations express some reason and result relations, additive operations provide a relation attaching some other information to the currennt structure.

- Source of coherence: semantic or pragmatic

  Semantic source presents a propositional content which is about the real world. On the contrary, pragmatic source includes an awareness or increase in the hearer's mind, which is an illocutional force.

- Polarity: positive or negative

  As understood from the terms, polarity helps to identify whether the relation between the discourse units is positive or negative.

- Order of segments: basic or nonbasic

Order of the discourse units can give some clues about the discourse relations. When S1 (the first sentence) expresses P and S2(the second sentence) expresses Q in the logic format (P Q), the order is basic. On the other hand, the situation in which S2 expresses P and S1 expresses Q is a nonbasic order.

As understood from the theories above; the approach to discourse relations vary, since discourse understanding depends on both the addresser and addressee. The interpretation of discourse can also vary according to the participants or intention (Edmonson, 1999). Hobbs (2010) states that the ambition to form a discourse relation is the need of addresser to be understood, which focuses on the producer, and the addressee also has responsibilities such as finding out the intention and best explanation of the relation by interpreting it, and this aspect is called the intentional perspective. This perspective constitutes a full discourse relation by getting together with the informational perspective, which gives information about the real world (Hobbs, 1996).

Moore and Pollack (1992) combine all of the opinions above and arrive at a final opinion that the notions of intention and discourse relation are different, where a discourse relation is the awareness of the intentions in the communication. That is why, intention and discourse relation should not be considered together. Besides this, it is also emphasized by Moore and Pollack (1992) that information in a discourse relation is integrated with intention; as a result of this, RST is a detailed discourse relation theory, but there should be a multi-level analysis because of the importance of intention and hearer interpretation in discourse relation.

2.1.2 Discourse connectives and modifiers

Although there are some limitations of discourse understanding regarding the meaning that the sentences form, it is quite possible in discourse that the addresser may not express themselves or the addressee may misunderstand. Some discourse markers can be used to prevent this confusion(Zeevat, 2003).

Discourse markers can analyse, classify and sometimes constitute a discourse relation, and their function makes them an essential part of discourse relations (Taboada, 1998; Risselada and Spooren, 1998). According to Taboada (2009), they are "any conjunction, adverb or other type of phrase that frequently links two or more units of discourse"(as quoted in Renkema, 2009:131).

Discourse connectives indicate that the subsequent sentence should be connected to the other sentence (Renkema, 2004). Connectives represent the relation between propositions, events or factual objects (Asher, 1993). They give obvious clues about the discourse relations and make the relation more understandable by preventing any kind of misinterpretation or misunderstanding (Schiffrin, 1987).

Halliday & Hasan (1976) are one of the first authors to emphasize the fact that discourse connectives determine the relation in general. The examples below show the discourse relations with explicit discourse connectives. Examples (14) and (16) are taken from PDTB while the other two examples are taken from TDB. The sense of the connectives is also shown. The sense of categories are taken from PDTB (Prasad et al., 2007).

(14) No matter who owns PS of New Hampshire, **after it emerges from bankruptcy proceedings** *its rates will be among the highest in the nation*, he said.

(TEMPORAL:Asynchronous:succession)

(15) Halil, **epeyce koştuktan sonra** *yüksekçe bir duvardan atlayarak mahallenin dışındaki kadınlar hamamının bahçesine girdi.*

'Halil, entered the yard of the woman's bath located out of the neighborhood by jumping over a high wall after running so much.'

(TEMPORAL:Asynchronous:succession)

(16) *Most bond prices fell on concerns about this week's new supply and disappointment that stock prices didn't stage a sharp decline.* **Junk bond prices moved higher, however.**

(COMPARISON:Contrast:opposition)

(17) *Dışa karşı güçlüydü*, **ama içe, kendi yüreğine yıkılmak üzereydi**.

'He was strong against outside, but he was about to collapse inside, on his own heart.'

(COMPARISON:Contrast:opposition)

Examples (14) and (15) have a temporal relation supported by *after* and *sonra* 'after.' The temporal relation between the two sentences is clarified explicitly by the discourse connectives. Examples (16) and (17) have a comparison relation which is made explicit by the discourse connectives *however* and *ama* 'but'.

Many discourse connectives can take adverbs such as *just, at least, very* and so on, and these adverbs can specify or clarify the discourse connective (Prasad et al., 2007). An adverb is a word type. The adverbs that modify the discourse connectives are called discourse connective modifiers (Prasad et al., 2007). In English and Turkish, modifiers occur with their connective heads. The function of these modifiers is to specify and clarify the discourse relation in terms of detailed information or determinacy (Zeyrek et al., 2012).

Example (18) taken from PDTB and (19) taken from TDB have the modifiers *even* and *bile* 'even'. Example (18) has a contingency relation which is indicated by the discourse connective *if*. However, the modifier *even* has an effect on the whole discourse relation and it makes the discourse relation a comparison relation. Example (19) has a temporal relation supported by *sonra* 'after'. In this case, the modifier *bile* 'even' specifies the temporal relation.

(18) *You can do all this* even **if you're not a reporter or a researcher or a scholar or a member of Congress.**

(COMPARISON:Concession:expectation)

(19) Kitap sevgisi, kitaplarla insanların dünyasının derinliğine inmek, onları **okuduktan sonra** bile *tüm canlılığı ile yüreğinde taşımak*, insanı öteki insanlar arasında sadece farklı yapıyordu.

'The love for books, going deep down in the people's world with books, carrying them in heart with all their liveliness even after reading them just make the person different from the others.'

(TEMPORAL: Asynchronous: succession)

2.1.3 Types of modifiers in Turkish Discourse Bank

In this thesis, the semantic distinction between different kinds of modifiers which are annotated in TDB are specified and classified. Although modifiers are annotated independently in TDB, their semantic distribution has not been identified.

Examples (20) and (21) taken from TDB have discourse relations made explicit by discourse connectives and their modifiers. The contingency relation in example (20) is clarified by the alternative lexicalization *o nedenle* ' for that reason'. The modifier *belki de* 'maybe' expresses that the relation is not certain but possible. Example (21) has a temporal relation determined by the discourse

connective *sonra* ‘later’ and the modifier *24 saat* ’24 hours later’ gives more specific information about the discourse relation.

(20) *Kendine inanç duyan bir ulusun bireyleri oldukları vurgulanan bir dönemde kişilikleri belirlenmişti.* Belki de **o nedenle, Türkiye'nin sanki bir müstemleke muamelesi görmesini bir türlü hazmedemezdi.**

‘Their characteristics were determined in a period in which their being individuals of a nation believing in themselves was emphasized. Maybe because of that, he could not get over with the behaviour of dependency towards Turkey.’

(CONTINGENCY:Pragmatic cause)

(21) Kantini basıyorlar, bir kız öğrenci (kantinde oturan) yaralanıyor*, acile kaldırılıyor*. **Boyun omuru koruyucusu ile** 24 saat **sonra taburcu edildi**.

‘ They raid a canteen, a female student (sitting at the canteen) gets injured, is taken to the emergency. 24 hours later, she was discharged from the hospital with a rigid collar.’

(TEMPORAL: Asynchronous: succession)

2.2 Content, epistemic and spech-act modality

The function of connective in terms of expressing *content*, *epistemic* and *speech-act domain* is important for the purposes of the current thesis. These notions are introduced in more detail below.

2.2.1 Content domain

Content domain, which is defined as the semantic domain by Sweetser (1990), expresses events from the real world (Knott, 1998). This type of relation takes place “between the propositional content” of the two related utterances (Knott, 1998:3). The content domain has two different subtypes, which are *nonvolitional* and *volitonal* relations. Although they have diferences, both types express a real event in the world (Maat and Degand, 2001).

Nonvolitional relations present just the event in the real world without any effect of the protoganist. However, volitional relations include the decision making of the protoganist even if they express a real event (Sweetser, 1990; Maat and Degand, 2001). Both of the events in the examples below are expressed in the content domain. While example (22) has a nonvolitional relation with an event and its result, example (23) has a volitional relation with the decision of the protoganist.

(22) There were landslides in Malibu last week. Four neighborhoods lost their electricity. (Mann and Thompson, 1986:5) (nonvolutional)
(23) Her/my plane takes off tomorrow at 6 a.m. she/I left the party early. (Maan and Degand, 2001:219) (volutional)

2.2.2 Epistemic domain

Epistemic domain is defined as a pragmatic domain by Sweetser (1990) and it presents the determinacy or possibility of an event (Bybee and Fleischman, 1995). This modality may express an event which is outside of the real world (Sweetser, 1990), or the assumption or conclusion of the addresser is indicated by epistemic domain (Bybee and Fleischman, 1995; Maat and Degand, 2001). Epistemic domain can be expressed by linguistic units such as modals and evaluative adverbs (Maat and Sanders, 2001). These units show the conclusion or assumption of the addresser explicitly. The utterance (24) below is in the epistemic domain indicated by the modal verb *must*. The addresser concludes a situation according to the event.

(24) He must be on holiday because his car is gone. (Spooren et al., 2010:5)

2.2.3 Speech-act domain

Speech-act domain is the other pragmatic domain (Sweetser, 1990). Speech-acts are "the minimal units of communication" (Searle & Vandervaken, 1985:109), and they reflect the social meaning in terms of addresser's performing an activity by uttering (Brown & Yule, 1993). There are social relationship roles besides the relation in the event, and the addresser wants the addresse to conclude or infer the message (Maat & Degand, 2001). According to Sweetser (1990), this modality also deals with the ability or willingness of the addressee's to respond (Spooren et al, 2010). In example (25), the speech-act motivates a conversation while the role of the addresser is to ask and offer; the role of the addressee is to infer that there is an offer and to respond the question.

(25) There is a good movie on. Did you already have plans for tonight? (Maat & Degand, 2001:225)

The role of discourse connective modifiers is important for the whole discourse relation because they can also help the discourse connectives to express the relation as in the domain of content, epistemic or speech-act. Besides giving specified information about the discourse relation, modifiers can also influence the determinacy or modality of the relation.

The next chapter will present the methodology we used to deal with this problem in Turkish.

CHAPTER 3

METHODOLOGY: THE ANNOTATION SCHEME

This chapter introduces the aim and methodology of the thesis. Section 3.1 introduces the aim and scope and Section 3.2 introduces the methodology by describing the classifications of relations, modifiers and modality. Section 3.3 introduces the computational aspect of the thesis.

3.1 Aim and scope

Although discourse units such as arguments of discourse connectives and their modifiers are annotated in TDB, there has not been any semantic distribution and sense annotation of connective modifiers. However, these modifiers have not been tagged separately and they are all tagged as modifiers. The overall aim of this current thesis is to determine the role of discourse connective modifiers in Turkish discourse. Another aim of this thesis is to examine statistically whether there is a systematic distribution of the effect of modifiers on the meaning (e.g. temporal) and modality (i.e. content, epistemic and speech-act modalities). The statistical part will contribute to discourse studies by providing a language automation system within the light of semantic distribution of modifiers identified by a decision tree algorithm.

3.2 The annotation process

Our annotation scheme includes the following categories: *sense of the relation, sense of the modifier*, *modality of the relation with the modifier* and *modality of the relation without the modifier*. PDTB 2.0 is used to classify the relation types (See Figure 3.1). The tag *sense of modifier* is determined in the light of the semantic distribution of adverbs by Ernst (2000; 2004) and Quirk et al., (1985). Lastly, the *modality* tags are classified according to *content*, *epistemic* and *speech-act* classifications of Sweetser (1990). After the data are classified according to these tags, the whole data and the tags with their explanations are given to two secondary annotators[2] to annotate the data independently. Secondary annotators annotated the whole data. The results from the annotators were compared and disagreements were resolved. Table 3.1 presents the inter-rater annotator agreement results (IAA) of modified discourse relations with their types and subtypes. Table 3.2 indicates the IAA of the semantic distribution of modifiers. Table 3.3 shows the IAA of the modality of the relations. As the tables indicate, the annotators are mostly in perfect agreement in temporal, comparison and expansion relations (>.80). On the other hand, the agreement on contingency relations is not as high as the other ones, but the rate is acceptable (.70).

Table 3.1 IAA of modified discourse relations with their types and subtypes.

Semantic class	**Semantic Type**	**Semantic subtype**	**IAA**
Temporal			
	Asynchronous	Succession	99.13%
		Precedence	97.36%
	Synchronous		100%
Semantic class	**Semantic Type**	**Semantic subtype**	**IAA**
Total			**97.82%**
Contingency			
	Pragmatic cause		75.4%

[2] Both of these secondary annotators are graduate Cognitive Science students at METU experienced in discourse annotation.

	Cause	Reason	70.2%
		Result	57.1%
Total			**67.5%**
Comparison			
	Contrast	Opposition	100%
	Concession	Contra-expectation	81.25%
Total			**85.71%**
Expansion			**99%**
	Conjunction		90.24%
	Manner[3]		82.60%
	Alternative	Disjunctive	100%
	List		0%
	Instantiation		100%
Total			**99%**

Table 3.2 Inter-rater annotator agreement for the semantic distribution of modifiers

Semantic distribution of modifiers	**IAA**
Focusing	99.22%
Aspectual	92.24%
Speaker-oriented	97.14%
Quantitative	100%
Exocomparative	100%

Table 3.3 Inter-rater annotator agreement for the modality of modifiers

	Content	**Epistemic**	**Speech-act**
IAA	99.11%	100%	87.87%

Example (26) indicates a prototype of the annotation.

(26) *Müşteri gibi davranan üç kadın da, bir yere uğramaları gerektiğini ve dönüşte alacaklarını söyleyerek, dışarıda kendilerini bekleyen otomobile binerek uzaklaştılar.* Bir süre **sonra kasayı kontrol ettiğinde soyulduğunu anlayan Yüksel, soluğu karakolda aldı.**
'*Three women who acted like clients said that they had to go to a place and they would take it when they came back, and they moved away by getting on the bus waiting for them outside.* **After** a while, **Yüksel understood that he was robbed when he checked the case and he immediately went to the police station.**'

Sense of the relation: TEMPORAL: Asynchronous: succession

Sense of the modifier: quantitative

Modality of the relation with the modifier: content

Modality of the relation without the modifier: content

[3] The discourse relations expressing manner of the arguments are annotated as EXPANSION in PDTB without any types. However; it is necessary in our data to include the type *manner* within the light of PDTB.

3.2.1 Sense of the relation

The sense tag of the relation specifies the meaning of the discourse relations. For example, example (27) expresses a cause relation; however, it also indicates the evaluation of the addresser with the help of the modifier.

> (27) *Annesiz geçen çocukluk yıllarından sonra ona kavuştuğunda da şefkat eksikliğini yaşıyor.* Belki **bu nedenle, kadınlara güvensizliğinden, belki yaşamındaki ilk kadının ulaşılmazlığından hiçbir kadına bağlanmıyor.**
>
> ' *After the childhood without a mother, he also experiences a lack of affection when he meets her.* Maybe **because of this, he cannot be linked to any women because of distrust of women, or the inaccessibility of the first woman in his life.**'
>
> (CONTINGENCY: Pragmatic cause)

The PDTB sense hierarchy has 4 top-level senses, which are explained below along with our annotation scheme (Temporal, Contingency, Comparison, Expansion). Figure 3.1 presents the sense hierarchy of discourse relations in PDTB.

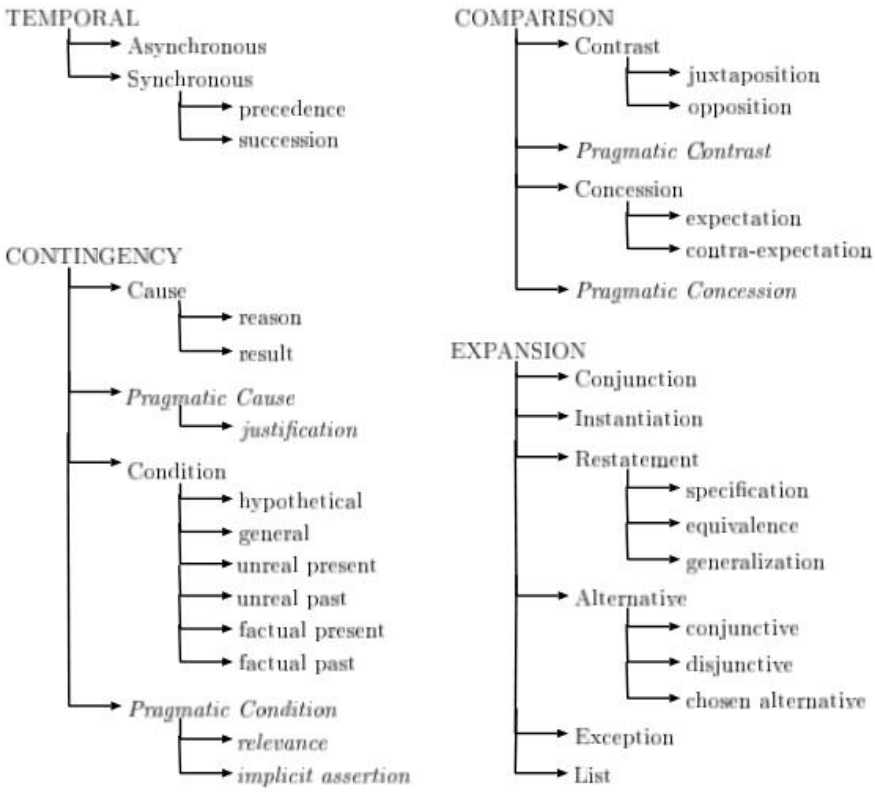

Figure 3.1 The sense hierarchy in PDTB

3.2.1.1 Temporal

The temporal class shows that the situations in arguments are temporally related (Prasad et al., 2007). The temporal class possesses two types which are *synchronous* and *asynchronous*.

- Synchronous: This relation indicates that there is an overlap between the time of the situations, it means they start and finish at the same time (Prasad et al., 2007).

  (28) *Fotoğrafa çekimsel bir sıcaklık duydum.* **Öte yandan** da **ilk defa kırmızı ışıklı evlerinin çok ünlü olduğu eski bir liman kentine geldiğimi düşündüm**.

  '*I was attracted to photography.* **And at the same time, I thought this was the first time I had come to an old seaport in which houses with red lights were very famous.**'

(TEMPORAL: Synchronous)

- Asynchronous: In this relation, the temporal order between the situations in arguments is expressed. Since this class is not sufficient to give information about the direction of the order, two subtypes *precedence* and *succession* were identified (Prasad et al., 2007).
    - *Precedence:* The situation in ARG1 happens before the situation in ARG2 in this relation (Prasad et al., 2007).

  (29) "*Sana âşık olduktan sonra karısından boşandı ya.* Bütün magazin basını aylarca bu gizli aşkı ve bu boşanmayı yazdı." Fevzi bir sigara yakmıştı. "**Demek Mithat Bey** daha **önce evliydi?**"

  ' "*He divorced his wife after falling in love with you*. All of the magazine media wrote about this secret love and divorce for months." Fevzi lit a cigarette. "**So Mr. Mithat was married before?**"'

  (TEMPORAL: Asynchronous: precedence)

    - *Succession:* The situation in ARG2 happens after the situation in ARG1 in this relation (Prasad et al., 2007).

  (30) *Bursa hakkında daha bir meraklı olmakta* ve hemen **ardından ansiklopedinin maddelerine bakmakta değil mi?..**

  '*He is becoming more curious about Bursa* and just **after that he looks at the items of the encyclopedia, doesn't he?**'

  (TEMPORAL: Asynchronous: succession)

3.2.1.2 Contingency

This relation indicates the causal influence of the situation in an argument on the other situation in the other argument (Prasad et al., 2007). The types and subtypes determined by Prasad et al.(2007) are as follows.

- Cause: This type indicates that the arguments are causally related (Prasad et al., 2007). Cause relations provide one argument to express the explanation, justification or reason of the situation in the other argument (Prasad et al., 2007). The subtypes of cause, *reason* and *result*, present the direction of the causality.

    - *Reason:* In this subtype, ARG2 indicates the causality part of the relation and ARG1 describes the result (Prasad et al., 2007).

  (31) Bak bu arkadaş bizim yeni kiracımız," dedi Nail'i göstererek. Nail'in başında kenarı yaldızlı lise şapkası vardı. Şapkası kafasına bol geliyordu. **Koca kulaklı olduğu için** de *kulakları kıvrılıyordu*.

'"Look, this is our new tenant" he said by pointing to Nail. Nail had a hat with gilted sides. His hat was loose for his head. **Because FP[4] he had big ears, his ears were curved**.'

(CONTINGENCY: Cause: reason)

- *Result:* In this type, ARG1 expresses the reason, explanation or justification of the situation in ARG2 (Prasad et al., 2007).

(32) Gelirler gayet güzeldi. Bu bahsettiğimiz bize bağlı olan grup da gayet kârlı bir gruptur. Zaten *grup, harici işlerden dolayı mali sıkıntıya girmişti.* **Bu sebepten dolayı** da **biz ortak olmuştuk**.

'Income was quite good. The group depending on us that we mentioned is a pretty profitable one. In fact, *they had already had financial problems*. **That is why FP, we became partners with them**.'

(CONTINGENCY: Cause: result)

- Pragmatic cause:In this relation with its the only subtype *justification*, one argument has the claim of the addresser about a situation while the other argument presents the evidence or justification of the claim (Prasad et al., 2007).

(33) *Bu köyde doğup büyümüştü sanki; çınarın hışırtısı kulaklarını, damların köy alanına akan beyazlığı gözlerini, toprağın cayır cayır yanan soluğu tenini hiç etkilemiyordu.* Belki de **bu yüzden, o gün bekçiden başka hiç kimse bir yabancıyla aynı yerde bulunmanın tedirginliğini duymamıştı kendinde.**

'*It seemed as if he was born and grew up in this village; he was not affected by the rustle of the plane tree, the whiteness of roofs flowing to the village zone, and the burning breath of the soil.* Maybe **because of this, nobody else except the warden was nervous about being at the same place with a stranger on that day.**'

(CONTINGENCY: Pragmatic cause: justification)

- Condition: This type expresses that the unrealized situation in an argument can cause the situation in the other argument if it comes true (Prasad et al., 2007).

(34) Both sides have agreed that *the talks will be most successful* **if negotiators start by focusing on the areas that can be most easily changed.** (Prasad et al., 2007:30).

(CONTINGENCY: Condition)

3.2.1.3 Comparison

The class comparison emphasizes the similarities and differences between the situations in the arguments (Prasad et al., 2007). There are different kinds of comparison situations and they are identified as different types, which are *contrast*, *similarity*, and *concession.*

- Contrast: In this type, the differences between the situations are described (Prasad et al., 2007; 2015). The subtypes *juxtaposition* and *opposition* specify the contrast relation.
  - *Juxtaposition:* This subtype indicates that the differences between the shared property of the arguments are the substitutes of each other (Prasad et al., 2007; 2015).

(35) *Operating revenue rose 69% to A$8.48 billion from A$5.01 billion.* **But the net interest bill jumped 85% to A$686.7 million from A$371.1 million.** (Prasad et al., 2007:33)

[4] Focus Particle

(COMPARISON: Contrast: juxtaposition)

- *Opposition:* In this relation, the shared property of the situations in the arguments are the opposites of each other (Prasad et al., 2007).

(36) *76 çeşit lisede değişik ortamlarda yetişmiş gençleri aynı sınavda aynı sorularla değerlendirmek eşitlik değildir.* Tam **tersine bu onlara yapılan adaletsizliktir, haksızlıktır.**

'*It is not fair to evaluate young people who grew up in different types of enviroment and 76 different high schools by applying the same exam with the same questions.* **Just to the contrary, it is injustice to them.**'

(COMPARISON: Contrast: opposition)

- Pragmatic contrast: This type indicates the contrast not between the situations in the arguments, but the inferences from the arguments (Prasad et al., 2007).

(37) "It's just sort of a one-upsmanship thing with some people," added Larry Shapiro. "They like to talk about having the new Red Rock Terrace one of Diamond Creek's Cabernets or the Dunn 1985 Cabernet, or the Petrus. Producers have seen this market opening up and they're now creating wines that appeal to these people." *That explains why the number of these wines is expanding so rapidly.* **But consumers who buy at this level are also more knowledgeable than they were a few years ago**. (Prasad et al., 2007:33)

(COMPARISON: Pragmatic Contrast)

- Concession: In this type, an argument expresses a situation while the other argument opposes or implies the negation of a situation (Prasad et al., 2007).

  - *Expectation:* In this subtype, ARG2 describes the situation while ARG1 refuses the situation in the argument (Prasad et al., 2007).

(38) **Although the purchasing managers' index continues to indicate a slowing economy,** *it isn't signaling an imminent recession*, said Robert Bretz, chairman of the association's survey committee and director of materials management at Pitney Bowes Inc., Stamford, Conn. (Prasad et al., 2007:34)

(COMPARISON: Concession: expectation)

  - *Contra-expectation:* In this subtype, ARG2 describes the situation while ARG1 refuses the situation in the argument (Prasad et al., 2007).

(39) *The Texas oilman has acquired a 26.2% stake valued at more than $1.2 billion in an automotivelighting company, Koito Manufacturing Co.* **But he has failed to gain any influence at the company.** (Prasad et al., 2007:34)

(COMPARISON: Concession: contra-expectation)

3.2.1.4 Expansion

This class provides an broadened information about the situations of the argument (Prasad et al., 2007). Different kinds of expansions specify the expansion relation.

- Conjunction*:* This type provides one argument to express additional information about the other situation in the other argument (Prasad et al., 2007).

(40) *Kuşkusuz, karşılığın en güzeli Tanrı katındadır!* **Ve** de **bu kata çıkmak, ölümden sonra dirilişte olacaktır.**

'*Definitely, the most beautiful response is at God's level!* **And FP to reach at this level will come true at resurrection.**'

(EXPANSION: Conjunction)

- Restatement: In this type, ARG2 expresses the situation in ARG2 in different ways and gives other information about the situation in ARG1 (Prasad et al., 2007).
    - *Specification:* In this subtype, ARG2 gives detailed description of the situation in ARG1 (Prasad et al., 2007).

(41) A Lorillard spokewoman said, "*This is an old story*. **Implicit = in fact We're talking about years ago before anyone heard of asbestos having any questionable properties.**" (Prasad et al., 2007:35)

(EXPANSION: Restatement: specification)

- *Generalization:* In this subtype, ARG1 gives detailed information about the situation in ARG2, and it means, ARG2 explains the situation of ARG1 briefly (Prasad et al., 2007).

(42) *If the contract is as successful as some expect, it may do much to restore confidence in futures trading in Hong Kong.* **Implicit = in other words. "The contract is definitely important to the exchange,"** says Robert Gilmore, executive director of the Securities and Futures Commission. (Prasad et al., 2007:35)

(EXPANSION: Restatement: generalization)

- *Equivalence:* This subtype indicates the different aspects of a shared situation in different arguments (Prasad et al., 2007).

(43) *Chairman Krebs says the California pension fund is getting a bargain price that wouldn't have been offered to others.* **In other words: The real estate has a higher value than the pending deal suggests.**

(EXPANSION: Restatement: equivalence)

- Instantiation: In this type, ARG1 describes a situation and ARG2 furthers the information about the situation by describing some of the circumstances of the situation (Prasad et al., 2007).

(44) *He says he spent $300 million on his art business this year*. **Implicit = in particular A week ago, his gallery racked up a $23 million tab at a Sotheby's auction in New York buying seven works, including a Picasso.**

(EXPANSION: Instantiation)

- Exception: When one argument describes a situation and the other argument gives some examples about circumstance where the situation does not come true, the relation is exception (Prasad et al., 2007).

(45) *Boston Co. officials declined to comment on Moody's action on the unit's financial performance this year* **except to deny a published report that outside accountants had discovered evidence of significant accounting errors in the first three quarters' results.** (Prasad et al., 2007:37)

(EXPANSION: Exception)

- Alternative: This type indicates the alternative conditions in the arguments (Prasad et al., 2007). This type has three subtypes which are *conjunctive, disjunctive* and *chosen alternative.*
    - *Conjunctive:* This subtype expresses that both of the situations in the argument are possible to come true (Prasad et al., 2007).

    (46) Today's Fidelity ad goes a step further, encouraging investors *to stay in the market* **or even to plunge in with Fidelity**. (Prasad et al., 2007:36)

    (EXPANSION:Alternative:conjunctive)

    - *Disjunctive:* This subtype indicates that one of the situations in different arguments can occur (Prasad et al., 2007).

    (47) *"İiç!" diye bağırıyor.* **Kışkırtıyor yahut** da.

    '*He shouts 'Drink!'*, **or** FP **he is being provocative.**'

    (EXPANSION: Alternative: disjunctive)

    - *Chosen alternative:* In this subtype, two alternatives are presented in arguments and one of them is excluded while the other one is chosen (Prasad et al., 2007).

    (48) Under current rules, even when a network fares well with a 100%-owned series – ABC, for example, made a killing in broadcasting its popular crime/comedy "Moonlighting" — it isn't allowed *to share in the continuing proceeds when the reruns are sold to local stations.* **Instead, ABC will have to sell off the rights for a one-time fee.** (Prasad et al., 2007:36)

    (EXPANSION:Alternative:chosen alternative)

- List: In this type, ARG1 and ARG2 are a a part of a list and they are mentioned in discourse (Prasad et al., 2007).

    (49) But other than the fact that besuboru is played with a ball and a bat, it's unrecognizable: *Fans politely return foul balls to stadium ushers;* **Implicit = and the strike zone expands depending on the size of the hitter;** (Prasad et al., 2007:37)

    (EXPANSION:List)

3.2.2 Sense of the modifier

Sense of the modifier tag is used to determine the contribution of the modifier to the discourse connective and/or the whole discourse relation. With the help of this distinction, the semantic distinction of modifiers is identified and the effect of modifiers from different semantic classifications on discourse connectives and modifiers is determined. The semantic classifications of adverbs by Ernst (2000;2004) and Quirk et al.(1985) are used in the classification. These two sources are used due to the highly comprehensive analysis of adverbs they provide.

Ernst (2000;2004) makes a semantic distribution of modifiers and divides modifiers into two basic classes, which are *predicational* and *functional* modifiers. We describe them below.

3.2.2.1 Predicational modifiers

Predicational modifiers are the ones which do not have a quantificational aspect and they take the predicates as events or propositions in arguments (Ernst, 2004). The classification of modifiers dates back to Greenbaum(1969) who examined modifiers according to their syntactic characteristics and positions with their meaning (Ernst, 2004). Jackendoff (1972) deals with this semantic distribution

where predicational modifiers are divided into four classes which are *exocomparative, manner, subject-oriented* and *speaker-oriented.* Ernst (2000; 2004) extended the classification, and predicational modifiers were classified in five classes with their types. We deal with each of them below as this classification will be the basis of our own classification of connective modifiers.

- Exocomparative modifiers

Exocomparative adverbs compare the events with other events (Ernst, 2000; 2004). Example (50) expresses that the situation in the argument is like another situation which is not mentioned in the argument, and the modifier *similarly* presents it explicitly. We have found this type of modifiers in our data.

(50)Similarly, this machine makes widgets (Ernst, 2004:79)

- Manner modifiers

Manner modifiers express the judgment of the speaker about the event in the argument in terms of how the event occurs (Ernst, 2004). Example (51) indicates how the action is done in the event. Manner adverbs have not been identified in our data.

(51) They saw the sigh clearly (Ernst, 2004:43)

- Subject oriented modifiers

Subject-oriented modifiers express the judgment or information about the agent with respect to the event (Ernst, 2000; 2004). The modifier *foolishly* in example (52) indicates the behavior of the agent. This type of modifiers have not been identified in the data.

(52) The senator has been talking foolishly to reporters. (Ernst, 2004:54)

- Speaker-oriented modifiers

Speaker-oriented modifiers express the evaluation or judgment of the addresser (Ernst, 2000; 2004). The modifier *perhaps* in example (53) expresses the evaluation of the addresser.

(53) The markets will perhaps respond to lower interest rates. (Ernst, 2004:69)

3.2.2.2 Functional modifiers

Functional modifiers can express time, aspect or frequency of an event without mentioning speaker judgment or evaluation (Ernst, 2004). The functional modifier types are described below.

- Aspectual modifiers

Aspectual modifiers express the time relation between two events. While one of the events happens first, the other one refers to the first event temporally (Ernst, 2004; Quirk, 1985). The aspectual modifier *still* in example (54) indicates an aspectual relation and the argument refers to another event in the past. Aspectual modifiers are one of the most common modifier type in the data.

(54) They still were doing it yesterday. (Ernst, 2004:344)

- Frequency modifiers

Frequency modifiers "quantify the events over subsets" (Ernst, 2004:347) and they express how often the event takes place. Example (55) indicates the frequency of the event with the modifier *often*. A few frequency modifiers have been found in the data.

(55) Sarah often listens to Clayfoot Strutters tapes.

- Focusing modifiers

Focusing modifiers intensify the unit that they modify (Ernst, 2004; Quirk et al., 1985). The modifier *only* in example (56) emphasizes the unit that it modifies. Focusing modifiers are predominantly in the data.

(56) John could only see his wife from the doorway (Quirk et al., 1985:605)

3.2.2.3 Singular vs. Multiple modifiers

In Turkish, we have come to realize that discourse connectives can have multiple modifiers as well as having single modifiers. Example (57) expresses a temporal relation with the help of the discourse connective *sonra* 'after', and it takes two modifiers *daha* 'more', and *da* 'focus particle –dA'. On the other hand; although example (58) also expresses a temporal relation, it has a singular modifier *da* 'focus particle –dA'.

(57) *Kofi Annan' ın Kıbrıs için hazırladığı planın hemen başında "müzakere edilemez" bulduğunu söylemişti.* Daha **sonra** da **bu tutumunu tekrar etti**.

'*At the very beginning of the plan that Kofi Annan had prepared for Cyprus, he said that the plan was not "nonissuable".* **After that FP, he continued to his attitute.**'

(TEMPORAL: Asynchronous: succession)

(58) 'Yüzük taşı' derdik bunlara. *Mendillerimize sarıp eve götürür*, **sonra** da **ne yapacağımızı bilemediğimizden döküp saçar, sonunda kaybederdik...**

'We used to call them the 'ring stone'. *We used to pack them in our napkins*, **then FP we used to slop them because we did not know what to do. We used to lose them in the end...**'

(TEMPORAL: Asynchronous: succession)

3.2.3 Modality of the relation with the modifier and modality of the relation without the modifier

These tags together identify the modality in the relation expressed by the connective and the modifier. In these tags, connectives and modifiers are taken into consideration separately, and the modality is identified according to the modality types *content*, *epistemic* and *speech-act* by Sweetser (1990). Example (59) has a temporal relation with a modified discourse connective *zaman* 'when'. The modifier is the focus particle *da* modifying the connective, it does not contribute any addresser judgment or a motivation of conversation. That's why we identify the example as a content domain relation.

(59) Sevgilisi Mustafa Sirmen'le böyle bir karar aldıklarını doğrulayan Esra Eron, "*Daha çok telefonla görüşüyoruz ama arada sırada geliyor*. **Geldiği zaman** da **basının olmadığı yerlere gidiyoruz.**

'Esra Eron approved that they had had such a decision with her boyfriend Mustafa Sirmen, and said, "*We generally talk on the phone, but he sometimes visits me*. **When FP he comes, we go to the places where there is no media reporters.**'

(TEMPORAL: Synchronous)

On the other hand, example (60) has a temporal relation with the discourse connective *o zaman* 'at that time' where the modifier *belki* 'maybe' expresses the speaker's judgment or evaluation. In this case, the relation is in the epistemic domain due to the connective modifier.

(60) *Bu tür haberleri yapan gazeteciler, bu tartışmayı bu haberleri bir de kendilerini Songül'ün yerine koyarak gözden geçirseler iyi edecekler.* Belki **o zaman ne yaptıklarını daha iyi anlayacaklar.**

‘ *These kinds of journalists had better revise the discussion and the news by putting themselves in Songül's shoes.* Maybe **then, they would understand what they have done.** ‘

(TEMPORAL: Synchronous)

Lastly, the contingency relation in example (61) is expressed by the discourse connective *için* ‘for’ and its modifier *biraz da* ‘partly’. The connective describes the contingency relation and the modifier makes a further contribution. In this example, the modifier conveys that there are some other reasons for the situation described besides the reason given. We argue that with this contribution, the relation is in the speech-act domain.

(61) *Göz alanımın tamamen dışına çıktığımız için sizin kapı ağzında dikilmekte olduğunuzu ancak hissedebiliyorum.* Biraz da **sizi benden tümüyle kopardığı için** *lap-top'luyu öfkeyle ve pervasızca süzmeye girişiyorum.*

‘*I can just feel that you are standing up at the door because you are out of my sight.* Partly **because it definitely seperated me from you,** *I tried to look at the one with the laptop.*’

(CONTINGENCY: Cause+speech-act: reason)

So far, we have described the annotation scheme in detail. We used this scheme to further classify and annotate the 513 modifiers in TDB.. (TDB actually annotated 540 modifiers but we have decided to eliminate some of them from further analysis because the modifiers *sanki* (1 token) and *ancak* (1 token) were not used with discourse connectives. On the other hand, the negation particle *değil* (5 tokens) was also eliminated because it is not a modifier, but a syntactic particle. We also eliminated tthe modifiers which caused ambiguity in the discourse). The corpus statistics will be presented in Chapter 4.

3.3 The statistical analysis: decision tree

The aim of the statistical study (decision tree) in this thesis is to try to understand if there is a systematic semantic distinction between the discourse relations and the effect of modifiers on the modality of relations. WEKA, which is a data processing tool, is used (Witten et al., 2011). The data and the tags were written in XML format and the data was converted into the appropriate file and structure for the data processing tool. The *relation, attributes,* and *data* were entered according to the appropriate structure. The tag *relation* indicates the relation (i.e. modified discourse relations) that we are looking for. A*ttributes* represent the tags and their classifications. The attribute types are presented below in Table 3.4.

Table 3.4 the attribute types in the decision-tree algorithm

The type of attribute	**The tag it presents**	**Example**
Connective	The explicit connectives as well as alternative lexicalizations in the relation	*sonra* 'after', *bu yüzden* 'because of this' etc.
Modifier	The modifiers of the connectives	*hemen* 'immediately', *belki* 'maybe', etc.
sense_of_the_relation	Sense of the relation	*TEMPORAL:Synchonous, CONTINGENCY:Cause: reason, etc.*
sense_of_the_modifier	Sense of the modifier	*focusing, aspectual, speaker-oriented, interrogative, exocomparative, frequency*
modality_of_the_relation_without_the_modifier	modality of the relation with the modifier	*content, epistemic, speech-act*
modality_of_the_relation_with_the_modifier	modality of the relation without the modifier	*content, epistemic, speech-act*
modifier2	The second modifier in the relation (if any)	*da* 'focus particle –dA', *ise* 'as for', etc.
sense_of_the_modifier2	Sense of the second modifier	*focusing, aspectual, speaker-oriented, interrogative, exocomparative, frequency*
modality_of_the_relation_without_the_modifier2	Modality of the relation without the second modifier	*content, epistemic, speech-act*
modality_of_the_relation_with_the_modifier2	Modality of the relation with the second modifier	*content, epistemic, speech-act*

The tags and the data are written as shown in Figure 3.2. The last 4 attributes present the tags for the connectives with multiple modifiers in the relations. The tags of the second modifier were entered in the appropriate attributes when a relation has multiple modifiers. However; if a relation has a singular modifier, then we had missing values for those which do not have multiple modifiers. In this case, question marks '?' were placed for the last 4 attributes. A supervised filter in the data processing tool was applied to the data in order to normalize the data and compute the modes and means.

```
@relation sense

@attribute connective                                       { nominal values }
@attribute modifier                                         { nominal values }
@attribute sense_of_the_connective                          { nominal values }
@attribute sense_of_the_modifier                            { nominal values }
@attribute modality_of_the_relation_without_the_modifier    { nominal values }
@attribute modality_of_the_relation_with_the_modifier       { nominal values }
@attribute modifier2                                        { nominal values }
@attribute sense_of_the_modifier2                           { nominal values }
@attribute modality_of_the_relation_without_the_modifier2   { nominal values }
@attribute modality_of_the_relation_with_the_modifier2      { nominal values }

@data
aksine,tam,COMPARISON:Concession:opposition,focusing,content,content,?,?,?,?
bu_amaçla,belki_de,CONTINGENCY:Pragmatic-cause,speaker-oriented,speech-act,epistemic,?,?,?,?
bu_amaçla,de,CONTINGENCY:Pragmatic-cause,focusing,speech-act,speech-act,?,?,?,?
amacıyla,de,CONTINGENCY:Cause:reason,focusing,content,content,?,?,?,?
amacıyla,de,CONTINGENCY:Cause:reason,focusing,content,content,?,?,?,?
ardından,de,TEMPORAL:Asynchronous:succession,focusing,content,content,?,?,?,?
ardından,de,TEMPORAL:Asynchronous:succession,focusing,content,content,?,?,?,?
ardından,de,TEMPORAL:Asynchronous:succession,focusing,content,content,?,?,?,?
```

Figure 3.2 The tags and the data in WEKA

After the whole data was written in the appropriate format, was uploaded to WEKA, and the J48 decision tree algorithm,which is a slightly modified version of C4.5, was applied. The cross-validation test was used to examine and evaluate the predictibility of the data. K-fold cross validation is a test which estimates the predictibility of the data by defining a training and a test set in the data. K-fold cross validation takes every *(k-1)* number of data as the training test and the k^{th} data as the test data. In this data set, 10 was taken as the number *k*. The results of this algorithm will be presented in Chapter 4.

CHAPTER 4

RESULTS & DISCUSSION

This chapter introduces the results of the annotations we have carried out, the analysis of the modifiers, as well as corpus statistics. The results of the decision tree algorithm are also presented. Section 4.1 introduces the theoretical part of the study which consists of the semantic classifications of discourse relations and the effect of modifiers on discourse relations, and section 4.2 is about the statistical analysis of the decision tree algorithm.

4.1 Results of the annotation

This section presents the results and corpus statistics of the annotation of types of modifiers and the effects of modifiers on discourse relations.

4.1.1 Four top-level relations in TDB

Our first annotation category was the *sense of the relation (or the discourse connective*), where we annotated the modified connectives with respect to four levels in the PDTB hierarchy. The frequency of 513 modified connectives (including the *singular* and *multiple types*) with respect to 4 top level sense of relations is presented in Table 4.1.

Table 4.1 The sense of discourse relations at the top level classes

Modifiers	**Temporal**	**Contingency**	**Comparison**	**Expansion**	**Total**
Singular	**291** (56.73%)	72 (14.04%)	18(3.5%)	104(20.28%)	485 (94.55%)
Multiple	28 (5.45%)				28 (5.45%)
Total	**319** (62.18%)	72 (14.03%)	**18** (3.5%)	104 (20.28%)	513 (100%)

Modified temporal connectives are primarily singular (e.g *24 saat sonra* '24 hours later') though some multiple modifiers also exist (e.g *daha sonra da* 'and later'). According to Table 4.1, the discourse connectives that are modified are predominantly temporal, followed by expansion and contingeny connectives. Comparison connectives are modified at the lowest level.

Table 4.2 shows the modifiers in temporal relation with its types and subtypes.

Table 4.2 The number of modifiers in connectives with the Temporal sense

Type	**Asynchronous**		**Synchronous**	**Total**
Subtype	Succession	Precedence		
Modifiers	**229** (71.78%)	36 (11.26%)	54 (16.93%)	319(100%)

In terms of lower level senses, modified temporal connectives are predominantly of the succession subtype followed by the synchronous subtype. The asynchronous precedence connectives are used at the lowest frequency.

Table 4.3 below shows the modifiers with contingency connectives with the types and the subtypes.

Table 4.3 The number of modifiers in Contingency sense

Type	**Subtype**	**Modifiers**	**Total**
Cause	Reason	**26 (36.11%)**	30 (41.66%)
	Result	4 (5.55%)	
Pragmatic cause		**42 (58.34%)**	**42 (58.34%)**
Total			72 (100%)

According to Table 4.3, contingency type is predominantly pragmatic cause followed by cause type with its reason type. On the other hand, the result subtype of cause is at the lowest frequency.

Table 4.4 indicates the modifiers in comparison relation with the types and the subtypes.

Table 4.4 Frequency of modifiers in the Comparison sense

Type	**Contrast**	**Concession**	**Total**
Subtype	Opposition	Contra-expectation	
Modifiers	5 (27.8%)	**13 (**72.2%)	18 (100%)

Table 4.4 indicates that there are two types of modified comparison connectives and each of these types possesses one subtype. The type which mostly occurs is concession.

Lastly, Table 4.5 presents the modifiers in expansion relation with the types and subtypes.

Table 4.5 Frequency of modifiers in the Expansion sense

Type	**Conjunction**	**Manner**	**Alternative**	**List**	**Instantiation**	**Total**
			Disjunctive			
Total	**74 (**71.16%)	23 (22.12%)	**1 (0.96%)**	5 (4.8%)	**1 (0.96%)**	104 (100%)

According to Table 4.5, the conjunction type of expansion occurs at a high frequency, followed by the manner type. The disjunctive subtype of the alternative type is modified at the lowest level. Table 4.6 indicates the relations found in modified discourse relations in TDB.

Table 4.6 PDTB relations found in modified discourse relations

Top-level class	**Type**	**Subtype**	**Number of instances**
Temporal	Asynchronous		
		Precedence	36 (7.02%)
		succession	**229 (44.63%)**
Top-level class	**Type**	**Subtype**	**Number of instances**
	Synchronous		54 (10.53%)
Total			319 (62.18%)
Contingency	Cause		
		Reason	26 (5.06%)
		Result	**4 (0.78%)**
	Pragmatic cause		0

		Justification	**42 (8.19%)**
	Condition		0
	Pragmatic condition		0
Total			72 (14.03%)
Comparison	Contrast		
		Juxtaposition	0
		Opposition	5 (0.97%)
	Pragmatic contrast		0
	Concession		0
		Expectation	
		Contra-expectation	13 (2.53%)
Total			18 (3.5%)
Expansion	Conjunction		74 (14.25%)
	Instantiation		1 (0.19)
	Restatement		
		Specification	0
		Equivalence	0
		Generalization	0
	Alternative		
		Conjunctive	0
		Disjunctive	1 (0.19%)
		Chosen alternative	0
	Exception		0
	List		5 (0.97%)
	Manner		23 (4.49%)
	Instantiation		1 (0.19%)
Total			104 (20.28%)

As seen in Table 4.6, all kinds of modified Temporal relations have been identified while the other 3 top-level relations have missing types in the data. The second predominanty modified relation is Expansion with the type conjunction. One of the subtypes of Expansion is manner –which is not classified as an independent type in PDTB (See Footnote 3).

4.1.2 The sense class distribution of modifiers

Our second annotation category was the *sense, or sense class, of the modifier*. Table 4.7 presents the semantic distribution (sense class) of modifiers.

Table 4.7 Sense class distribution of modifiers

Sense class	**Modifiers**	**Example**
Focusing	***dA* 'focus particle –dA' (234)** *özellikle* 'especially' (2) *ise* 'as to'(3)	*sonra da* 'then' *özellikle...sonra* 'especialy after'

	tam'exact'(4) ***bile*'even if'(1)** *sırf,yalnızca,sadece* 'just'(3) *ancak*'only'(5)	*tam tersine* 'on the exact contrary' *sadece bu yüzden* 'just because of this'
Total	**258 (50.3%)**	
Aspectual	*daha*'more' (84) *hemen*'immediately' (4) *ilk*'first'(2) *en*'most'(4) *taa ki*'until'(1) *ki* 'that'(1) *işte* 'only'(12)	*daha sonra* 'later' *işte o zaman* 'at that time' *hemen sonra* 'immediately after'
Total	108 (21.05%)	
Quantitative [5]	*biraz* 'partly, a bit' (47) *birkaç* 'a few',etc. (51) *neden*'lit. why, a long time '(3) *kaç yıl*'how many years'(1) 102 (19.88%)	*2 hafta sonra* 'two weeks later' *biraz sonra* 'after a while'
Total		
Exocomparative	*sanki* 'as if'(24) *tıpkı* 'just like'(9) *aynı* 'just like' (1)	*sanki...gibi* 'as if' *tıpkı... gibi* 'just like'
Total	34 (6.63%)	
Sense class	**Modifiers**	**Example**
Speaker-oriented	*belki(de)*'maybe'(9) *çarpıcı* 'striking' (1) *tabii* 'of course' (1)	*belki de bu yüzden* 'maybe because of this' *çarpıcı örnek olarak* 'as a striking example'
Total	11 (2.14%)	

[5] Discourse connectors may occur with measure phrases such as *2 yıl* 'two years' and *biraz* 'a while'. These phrases were not a part of the adverb classification of Ernst (2000;2004) and Quirk et al.(1985). We named them as quantitative modifiers.

According to Table 4.7, modifiers are divided into 5 semantic categories. Focusing modifiers are at the highest level among all the modifiers in the data and they are followed by aspectual modifiers. The modifier type which has the least number in discourse relations is speaker-oriented modifiers with 11 examples.

4.1.3 Distribution of the modality of the relation with respect to the connective modifier

Our third annotation category was the *modality of the relation with the connective modifier*. In this section, we present the results with respect to each top-level calss. Table 4.8 presents the modality of modified connectives in temporal relations.

Table 4.8 The modality of modifiers in Temporal sense

Temporal relations	**Subtypes**	**Modality**		
		Content	Epistemic	Speech-act
Asynchronous				
	Succession	*biraz* 'a bit', *iki hafta* 'two weeks' etc. (89) *da* 'focus particle –dA'(64) *daha* 'more'(59) *en* 'the most'(4) *ise* 'as to'(2) *hemen* 'immediately'(2) *bile* 'even if'(1)		*neden* 'lit., long'(3) *ancak* 'just'(2) *özellikle* 'especially'(2) *kaç yıl* 'how many years'(1)
	Total	221 (69.27%)		8 (2.50%)
Temporal relations	**Subtypes**	**Modality**		
	Precedence	*daha* 'more' (25) *da*'focus particle – dA' (3) (*biraz* 'a bit', *iki hafta* 'two weeks' etc. (3) *ilk* 'first'(2)		

		hemen 'immediately'(2) *taa ki* 'even'(1)		
	Total	36 (11.29%)		
Synchronous		*da*'focus particle –dA' (33) *ki* (particle) (1) *işte* 'that'(1) *ise* 'as to'(1)	*belki* 'maybe'(1) *tabii* 'of course'(1)	*işte* 'that' (11) *ancak* 'only'(3) *da*'focus particle –dA' (2) 16 (5.02%)
	Total	36 (11.29%)	2 (0.63%)	
Total		**293 (91.85%)**	**2 (0.63%)**	24 (7.52%)

According to Table 4.8, the most common modified temporal connectives are in the content domain while modified temporal connectives in the epistemic domain are found at the lowest level. Aspectual modifiers and focusing modifiers are the ones that predominantly occur in temporal connectives while pragmatic and speaker-oriented modifiers occur at the lowest level. The focusing modifiers *–dA* 'focus particle –dA', *bile* 'even', and *ise* 'as to' commonly occur with temporal connectives in the content domain. On the other hand, the focusing modifiers such as *özellikle* 'especially', and *ancak* 'just/only' occur with temporal connectives mostly in the speech-act domain where they motivate the addressee to make a deduction from the argument, thus pulling the relation to the speech-act (pragmatic) domain. Aspectual modifiers *daha* 'more', *hemen* 'immediately', *ilk* 'the first', *en* 'most', *işte* 'that', *ki* 'that', *taa ki* 'until' commonly express temporal relations in the content domain. However; modifiers such as *belki* 'maybe', and *tabii* 'of course' always occur in the epistemic domain. These modifiers express speaker's judgment in the discourse relation (or in one of the arguments of the discourse relation).

Temporal relations are the only class that possesses multiple modifiers such as *daha sonra* 'later' and the focus particle *da* 'after this FP'.

Table 4.9 presents the modality of modified connectives in contingency relations.

Table 4.9 The modality of modifiers in Contingency sense

Contingency relations	**Subtypes**	**Modality**		
		Content	Epistemic	Speech-act
Pragmatic cause			*belki (de)* 'maybe' (8)	*yalnızca* 'just'(1) *bütün* 'all'(1) *çoğu kez* 'mostly'(2) *biraz da* 'partly'(2) *sadece* 'just'(1) *sırf* 'just' (1) *da* 'focus particle –da' (26)
Cause	Reason	*da* 'focus particle –da' (26)		
	Result	*da* 'focus particle –da' (4)		
Total		30 (41.66%)	**8 (11.12%)**	**34 (47.22%)**

According to Table 4.9, the modifiers predominantly occur with contingency connectives in the speech-act domain, followed by those in the content domain. The least frequent of modifiers is found in epistemic relations. Focusing modifiers occur at the highest level and quantitative modifiers are at the lowest level. The focus modifier *–dA* usually expresses a contingency relation in the content domain. However, the modifiers *yalnızca* 'only', and *sadece* 'only' change the modality of contingency relation to the speech-act domain. They provide a motivation for the addressee to infer a conclusion from the situation in one of the arguments of the discourse relation or in the whole discourse relation. Speaker-oriented modifiers such as *belki de* 'maybe' often occur in the epistemic domain. These modifiers generally express speaker judgment and present the evaluation or opinion of the addresser. On the other hand, interrogative, aspectual and frequency modifiers usually give responsibilities to the addressee and they motivate them to conclude from the situation in the arguments. As a result, they co-occur with discourse connectives conveying the speech-act domain.

Table 4.10 presents the modality of modified connectives in comparison relations.

Table 4.10 The modality of modifiers in Comparison sense

Comparison relations	**Subtypes**	**Modality**		
Contrast		**Content**	**Epistemic**	**Speech-act**
	Opposition	*tam* 'exact'(4) *da* 'focus particle – dA'(1)		
Concession				
	Contra-expectation		*sanki*(12)	*bütün*(1)
Total		5 (27.8%)	**12 (66.6%)**	**1 (5.6%)**

According to Table 4.10, the most frequent modality type in modified comparison connectives is the speaker-oriented modifiers followed by the content domain. Speech-act domain is at the lowest level with one example in the corpus. Comparison connectives are mostly modified by the exocomparative modifier *sanki* 'as if'. Exocomparative modifiers include speaker judgment and the situation is expressed from their point of view. These modifiers are used with the discourse connective *gibi* 'like'. Exocomparative modifiers may also express speaker judgment and hence they pull the modality of the connective to the epistemic domain. On the other hand, contrastive connectives, which are a type of comparison connectives, may be used with focusing modifiers and with these modifiers the connective often expresses the content domain.

Table 4.11 presents the modality of modified connectives in expansion relations.

Table 4.11 The modality of modifiers in Expansion sense

Expansion relations	**Modality**			**Total**
	Content	Epistemic	Speech-act	
Conjunction	*da* 'focus particle – dA'(61)		*da* 'focus particle – dA'(13)	74 (71.82%)
Manner		*sanki* 'as if'*(12)* *tıpkı* 'just like'(9) *aynı* 'just like'(1) *da* 'focus particle – dA'(1)		23 (22.36%)
Expansion relations	**Modality**			**Total**
List	*da* 'focus particle –			5 (4.85%)

	dA' (5)			
Alternative: disjunctive	*da* 'focus particle – dA' (1)			1 (0.97%)
Instantiation		*çarpıcı* 'striking' (1)		
Total	**67 (64.43%)**	24 (23.07%)	**13 (12.5%)**	104 (100%)

Table 4.11 indicates that the most common modality type is the content modality and the least common one is the speech-act domain. Focusing modifiers are the most common modifiers and the least occurring modifier is the exocomparative modifier *aynı* 'like'. The focusing modifier *–dA* 'focus particle –dA' generally expresses the content domain in 67 expansion relations while the focus particle is occasionally in the speech-act domain and rarely in the epistemic domain. The manner type of expansion connectives and their modifiers *sanki* 'as if', *tıpkı, aynı* 'like' express the evaluation or opinion of the addresser, and they pull the relation to the epistemic domain. Besides, the only example of instantiation relation takes a speaker-oriented modifier, *çarpıcı* 'striking', and it also pulls the relation to the epistemic domain.

4.1.4 An interesting case of *sanki* and *hem de*

Sanki...gibi (as if): In the data, there are two different kinds of relations expressed by *sanki...gibi* 'as if'. One of the relations indicated by the connective *gibi* 'like' and the modifier *sanki* 'as if' is COMPARISON:Concession: contra-expectation. This relation implies the negation of ARG2. On the other hand, in the other relation type indicated by the same connective and modifier, ARG2 expresses how the action in ARG1 is done. This relation is EXPANSION:Manner (See Appendix A).

Hem de: Two different types of the connective *hem* 'both' were identified in the data. The discourse relation that they express is the same, which is EXPANSION: Conjunction. The continuous connective *hem .... hem* 'both...and' has a content use and the singular connective *hem* 'both' has a speech act use (in 17.56% of the cases); in these cases, it expresses the judgment of the speaker and their motivation of giving the addressee a responsibility to reach a conclusion. (See Appendix A)

To wrap up this section, we provide the distribution of connectives and their modifiers with respect to the content, epistemic and speech act domains. (Table 4.12).

Table 4.12 The use of connective modifiers in the content, epistemic and speech act domains

Modifier sense class	**Content**	**Epistemic**	**Speech-act**
Focusing	*bile* 'even'(1) *ise* 'as to' (3) *tam* 'exact'(4) *da* 'focus particle-dA'(240)		*özellikle* 'especially'(2) *yalnızca,sırf,sadece* 'only'(3) *ancak* 'just'(5)
Modifier sense class	**Content**	**Epistemic**	**Speech-act**
Aspectual	*daha* 'more'(84) *hemen* 'immediately'(4) *işte* 'that'(12) *ki* 'that'(1) *ilk* 'first'(2)		

	en 'the most'(4) *taa ki* 'even'(1)		
Speaker-oriented		*belki(de)* 'maybe'(9) *çarpıcı* 'striking'(1) *tabii* 'of course'(1)	
Quantitative	*Birkaç gün* 'a couple of days', *biraz* 'a little', *kısa süre* 'a little time', *iki hafta* 'two weeks' *etc.*(90) *kaç zaman* 'so many years'(1) *kim bilir kaç kedi gün* 'who knows how many cat days'(1)		*neden* 'lit. long'(3) *kaç yıl* 'how many years'(1) *çoğu kez* 'mostly' (in contingency relations) (2) *biraz da* 'partly'(2) (in contingency relations) *bütün* 'all'(2) (in contingency relations)
Exocomparative		*sanki* as if'(24) *tıpkı* 'just like' (9) *aynı* 'just like'(1)	
Total	**448 (87.33%)**	45 (8.78%)	**20 (3.89%)**

According to Table 4.12, when used with discourse connectives, modifiers commonly express the content domain, and the modifiers which express the speech-act domain are at the lowest level. For example, the focusing modifiers *–dA* 'focus particle –dA', *bile* 'even', *ise* 'as to'and *tam* 'exact' usually express the content domain while the focusing modifiers *özellikle* 'especially', *sırf* 'just', *sadece* 'just' and *yalnızca* 'just' commonly express the speech-act domain.

Aspectual modifiers (e.g. *daha* 'more') are another modifier type which affect the content use of discourse connectives. They occur mostly with temporal and contingency connectives. Nevertheless, the aspectual modifier *bütün* 'all' also appears with comparison connectives. When aspectual modifiers appear in temporal relations, they always convey the content domain. On the other hand, they have a speech-act use when they are used with contingency or comparison connectives, i.e., aspectual modifiers appear to affect the use of connective.

Quantitative modifiers are yet another most common modifiers in Turkish. They occur mostly with temporal and contingency relations. Nevertheless, the quantitative modifier *bütün* 'all' also appears with comparison connectives. When quantitative modifiers appear in temporal relations, they always convey the content domain. On the other hand, they have a speech-act use when they are used with contingency or comparison connectives, i.e., quantitative modifiers appear to affect the use of connective.

Exocomparative modifiers sometimes occur with the connectives presenting the negation of the situations or expressing how the events or situations in the arguments of the relation happen. Exocomparative modifiers (e.g. *sanki* 'as if') affect the discourse connective in terms of epistemicity.

Speaker-oriented modifiers (belki 'maybe') directly indicate the evaluation of the addresser in any connective they co-occur with. Speaker-oriented modifiers are identified mostly in temporal and mostly contingency relations.

Above, we have mentioned multiple modifiers, which are the ones that co-occur with a connective, usully one preceeding the connective, the other following it. They only appear with temporal connectives in the data. While the first modifier used just before the discourse connective is usually aspectual, the modifier just after the discourse connective is always a focusing modifier (e.g. *daha sonra da* 'and later').

4.1.5 Distribution of the use of the connective (without the modifier) in three domains

Our fourth annotation category was *the modality of the relation without (the contribution of) the modifier*. Table 4.13 below presents the distribution of the modality of the relation in 4 top-level classes with respect to content, epistemic and speech-act domains.

Table 4.13 Distribution of the use of the connective (with the modifier) in three domains

Modality	**Temporal**	**Contingency**	**Comparison**	**Expansion**	**Total**
Content	**300**	35	5	68	395 (77%)
Epistemic	0	0	12	23	46 (8.97%)
Speech-act	19	37	1	13	72(14.03%)

Table 4.14 Distribution of the use of the connective (without the modifier) in three domains

Modality	**Temporal**	**Contingency**	**Comparison**	**Expansion**	**Total**
Content	**300**	35	5	68	408 (79.5%)
Epistemic	0	0	12	23	35 (6.84%)
Speech-act	19	37	1	13	70(13.66%)

According Table 4.13 above, the connectives with the modifiers had the following distribution with respect to the three domains:

Content (77%) > Speech act (14.03%) > Epistemic (8.97%)

According to the Table 4.14, the majority of the connectives are identified as temporal and in the content domain when considered without the modifier. The content domain is followed with the speech-act domain and finally with the epistemic domain:

Content (79.5%) > Speech act (13.66%) > Epistemic (6.84%)

The comparison of these two results suggest that while discourse connectives we have examined within the scope of this thesis have predominantly the content use, this may change slightly with the contribution of the modifier. The comparison also suggests that epistemic and speech act uses are far less frequent in the data but particularly modifiers with an epistemic sense have a role of pulling the connective towards the epistemic domain, as expected.

4.1.6 Distribution of the modified alternative lexicalizations and explicit discourse connectives

In the data, we have seen two different discourse connectors, which are *alternative lexicalizations* and *explicit discourse connectives*. Table 4.15 presents the number of alternative lexicalizations and explicit discourse connectives in the data.

Table 4.15 Distribution of the modified alternative lexicalizations and explicit discourse connectives

Alternative lexicalization	**Explicit discourse connective**
62 (12.1%)	450 (87.9%)

Table 4.15 above indicates that explicit discourse connectives are the predominant discourse connectors and there is a small number of alternative lexicalizations. While 451 modified explicit discourse connectives were identified, there were 62 modified alternative lexicalizations. The discourse relations constituted by alternative lexicalizations and explicit discourse connectives were also classified and Table 4.16 presents the results.

Table 4.16 The discourse relations of modified alternative lexicalizations (AltLex) and explicit discourse connectives

Discourse relations			**AltLex**	**Explicit discourse connective**
Temporal	Asynchronous			
		Succession	2 (0.39%)	**227 (44.24%)**
		Precedence	0	36 (7.02%)
	Synchronous		17 (3.31%)	37 (7.21%)
Total			19 (3.70%)	300 (58.47%)
Contingency				
	Cause			
		Reason	0	26 (5.06%)
		Result	2 (0.39%)	2 (0.39%)
	Pragmatic cause		**38 (7.40%)**	4 (0.78%)
Total			40 (7.79%)	32 (6.23%)
Comparison				
	Contrast			
		Opposition	2 (0.39%)	4 (0.78%)
	Concession			
		Contra-expectation	0	12 (2.33%)
Total			2 (0.39%)	16 (3.11%)
Expansion				
	Conjunction			74 (14.44%)
	Manner			23 (4.49%)
	List			5 (0.97%)
	Instantiation		**1 (0.19%)**	
Discourse relations			**AltLex**	**Explicit discourse connective**
	Alternative			
		Disjunctive		**1 (0.19%)**
Total			1 (0.19%)	103 (20.09%)

As seen in Table 4.16 above, explicit discourse connectives mostly occur with succession subtype of temporal relations and rarely occur with disjunctive subtype of expansion relations. On the other hand, the most common relation in alternative lexicalizations is pragmatic cause type of contingency relations and the least common one is instantiation type of expansion relations. Table 4.17 presents the modality of discourse relations with the modifier with respect to the discourse connectors.

Table 4.17 Modality of discourse relations with the modifier with respect to the discourse connectors

Modality	**Alternative lexicalizations**	**Explicit discourse connectives**
Content	**3 (4.84%)**	**392 (86.9%)**
Epistemic	10 (16.13%)	**36 (8%)**
Speech-act	**49 (79.03%)**	23 (5.1%)
Total	62 (100%)	451 (100%)

Table 4.17 indicates that while the modality of the discourse connective with the modifier is mostly speech-act in alternative lexicalizations, the modality is mostly the content domain with explicit discourse connectives. On the other hand, the least frequently seen domain is the content domain in

alternative lexicalizations. Table 4.18 shows the modality distribution of discourse relations without the modifiers regarding the discourse connectors.

Table 4.18 Modality of the discourse relations without the modifier with respect to the discourse connectors

Modality	**Alternative lexicalizations**	**Explicit discourse connectives**
Content	**5 (8.1%)**	**403 (89.36%)**
Epistemic	0	35 (7.76%)
Speech-act	**57 (91.9%)**	**13 (2.88%)**
Total	62 (100%)	451 (100%)

Table 4.18 indicates that discourse relations constituted by alternative lexicalizations are mostly in the speech-act domain without modifiers. But discourse relations constituted by explicit discourse connectives are mostly in content domain without modifiers.

4.1.7 Genre distribution of the three domains

Table 4.19 presents the modality of discourse relations with respect to the genre of the data that has modified discourse connectors.

Table 4.19 Modality of discourse relations with respect to genre

Genre	**Sense of the modifier**	**Modality of the relation with the modifier**		
		Content	**Epistemic**	**Speech-act**
Memoir	Aspectual	0	0	1 (0.19%)
Genre	**Sense of the modifier**	**Modality of the relation with the modifier**	**Genre**	**Sense of the modifier**
	Focusing	5 (0.97%)	0	2 (0.39%)
	Quantitative	2 (0.39%)	0	0
Research	Quantitative	2 (0.39%)	0	3 (0.58%)
	Focusing	15 (2.92%)	0	5 (0.97%)
	Speaker-oriented	0	1 (0.19%)	0
	Aspectual	11 (2.14%)	0	1 (0.19%)
Travel	Focusing	5 (0.97%)	0	2 (0.39%)
	Quantitative	3 (0.58%)	0	0
Diary	Exocomparative	0	1 (0.19%)	0
	Focusing	2 (0.39%)	0	1 (0.19%)
News	Aspectual	**68 (13.28%)**	0	**1 (0.19%)**
	Speaker-oriented	**0**	**5 (0.97%)**	0
	Exocomparative	**0**	**17 (3.32%)**	0
	Focusing	**88 (17.18%)**	0	**18 (3.51%)**
	Quantitative	**53 (10.35%)**	0	0
Article	Aspectual	3 (0.58%)	0	0
	Focusing	5 (0.97%)	0	2 (0.39%)
	Quantitative	1 (0.19%)	0	1 (0.19%)

Story	Aspectual	5 (0.97%)	0	8 (1.56%)
	Speaker-oriented	0	2 (0.19%)	0
	Exocomparative	0	9 (1.75%)	0
	Focusing	30 (5.85%)	0	11 (2.14%)
	Quantitative	9 (1.75%)	0	5 (0.97%)
Novel	Aspectual	8 (1.56%)	0	0
	Speaker-oriented	0	3 (0.58%)	0
	Focusing	55 (10.93%)	0	9 (1.75%)
	Exocomparative	0	8 (1.56%)	0
	Quantitative	23 (4.49%)	0	1 (0.19%)
Interview	Aspectual	**1 (0.19%)**		
	Focusing	**1 (0.19%)**		1 (0.19%)
Total		**395 (76.9%)**	46 (8.98%)	72 (14.1%)

Table 4.19 shows that the most common genre with respect to modified relations is news followed by novel. While modified connectives in research and article genres are rarely in the speech-act domain, the amount of modified connectives in the epistemic domain is higher in the novel and story genres. Regarding modified connectives, the predominantly occuring genre is news and the relations are mostly in the content domain followed by the same genre with modified relations in speech-act domain. The interview genre is the least common genre in the data and the modified connectives in this genre are in the content domain.

4.2 The decision tree results

This section introduces the decision tree results of the data in accordance with the application of cross-validation test. Figure 4.1 provides the decision-tree.

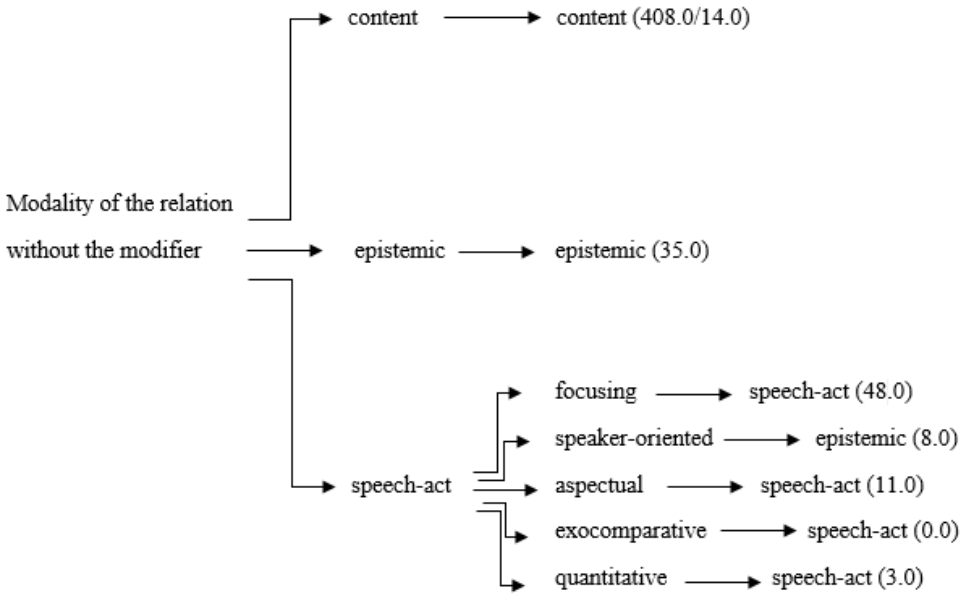

Figure 4.1 The decision-tree of the data

The numbers in the decision tree indicate the number of instances classified correctly and incorrectly. While the numbers on the left in the paranthesis show the number of correctly classified instances, the numbers on the left indicate incorrectly classified instances. The decision tree indicates that the modality of the relation with the modifier is in the content domain if the modality of the relation without the modifier is in the content domain. However; 14 instances were classified incorrectly among the 408 instances. On the other hand; according to the decision tree, speaker-oriented modifiers always pull the relation to epistemic domain (all of the 35 instances were classified correctly). On the other hand; when modality of the relation without the modifier is speech-act, modality of the relation with the modifier becomes speech-act except with speaker-oriented modifiers, which pull the relation to epistemic domain. Table 4.19 indicates the summary of the cross-validation test.

Table 4.20 The summary of the cross-validation test

Correctly classified instances	499 (97.2625%)
Incorrectly classified instances	14 (2.1744%)
Kappa statistics	0.9251
Correlation coefficient	0.9259
Mean absolute error	0.0355
Root mean squared error	0.1335
Relative absolute error	13.8438 %
Root relative squared error	37.3858 %
Total number of instances	513

The test results of the statistical study indicate that 493 out of 513 instances were classified correctly (97.2625%) by the application of the 10-fold cross-validation test. According to the results, the classification and effect of modifiers on discourse relations are not a coincidence, but there is a systematic distribution among the modifier types and their functions in the discourse relations (*Kappa statistics*: 0.9251). Correlation coefficient result indicates that the classifications are positively correlated (0.9259). Mean absolute error indicates how far are the predicted valus and observed. In this data, predicted and observed values are close to each other (*Mean absolute error:* 0.0355). Root mean squared measure indicates the accuracy of the model and as seen in Table 4.19, the error of the accuracy is quite low (*Root mean squared error:* 0.1335). Relative squared error indicates the error rate of the average of actualy values and the value is low (*Relative squared error:* 13.8438 %).

Table 4.21 presents accuracy levels by the classes.

Table 4.21 Accuracy by the classes

	TP Rate	**FP Rate**	**Precision**	**Recall**	**F-Measure**	**ROC Area**	**Class**
	1	0.118	0.966	1	0.983	0.919	content
	0.935	0	1	0.935	0.966	0.952	epistemic
	0.849	0	1	0.849	0.919	0.898	Speech-act
Weighted Avg.	0.973	0.09	0.974	0.973	0.972	0.919	

According to the detailed results indicating the accuracy of the data by classes, the overall data is quite predictable. When the three modality classes are evaluated independently, the accuracy of the data is again high. In the content domain, the TP Rate (which is also the same as Recall) indicates the true content domain instances which are classified as content in the algorithm and means that all of the content instances classified as content by the algorithm are in the correct class *(TP Rate:* 1). The TP

Rate of epistemic and speech-act modalities indicate that the rate of true positives that are correctly identified is high (*TP Rate for epistemic:* 93.5%, *TP Rate for speech-act:* 84.9%). The FP Rate shows that the negative instances which are classified as positive are at a very low level (*FP Rate for content:* 11.8%, *FP Rate for epistemic:* 0%, *FP Rate for speech-act:* 0%). As seen in the results, while content and speech-act domains have a a small number of negative instances classifed as positive, epistemic domain does not have any negative instances classified as positive. Precision presents the rate of positive predictive values and the results indicate that the values which are predicted are at high levels with 0.966 for content, 1 for epistemic and 1 for speech-act domain. F-measure is the combined result of precision and recall which presents the accuracy of the test. As seen in Table 4.13, the accuracy of the model is quite high (0.983 for content, 0.966 for epistemic and 0.952 for speech-act). Lastly, ROC area indicates the ability of the test to distribute the instances into the classes, and the ability rate is also high in the test (0.919 for content, 0.952 for epistemic and 0.898 for speech-act).

Table 4.22 provides the confusion matrix of the classes.

Table 4.22 Confusion matrix of the classes

a	b	c	<=classified as
394	0	0	a = Content
3	43	0	b = Epistemic
11	0	61	c = Speech act

The confusion matrix above indicates the detailed performance of the decision tree algorithm. The top row indicates the predicted values while the column on the right shows the actual values. The results indicate that all of the 394 content relations were classified as content. On the other hand, out of 46 epistemic relations, 43 of them were classified correctly while 3 of the epistemic relations were classified in the content domain in the cross-validation test. Out of 72 speech-act relations, 11 relations were identified as content and the rest 61 relations were identified as speech-act relations in the data. It can be concluded from the confusion matrix above that there is a systematic distinction between content, epistemic and speech-act modalities in modified discourse relations.

CHAPTER 5

CONCLUSION

Discourse relations are any kind of "semantic relationship between the arguments" (Kehler, 2002). Discourse relations can be understood by understanding the discourse segments in isolation and constitute the relation between the segments (Mann and Thompson, 1986). Discourse connectives are important discourse markers which connect two propositions, events or ideas (Asher, 1993). They make the discourse relation explicit (Halliday and Hasan, 1976). Connectives can appear with modifiers such as *very, quite, just, etc.* which specify, clarify or indicate the determinacy of the discourse relation (Prasad et al., 2007; Zeyrek et al., 2012).

The aim of this current thesis was two-fold: to understand the role of modifiers in TDB and to do a decision tree analysis as proof of evidence. Connectives can have three uses, referred to as the content, epistemic and speech-act domains by Sweetser (1990). The content domain refers to the situation in the actual word, while the epistemic domain expresses the speaker judgment or evaluation. The speech-act domain encourages a two-way communication between the addresser and the addressee. If an utterance is in the speech-act domain, the addressee makes an inference from the utterance.We have analyzed the role of modifiers with respect to these three domains. We assumed that the default use of a discourse connective is the content domain, and argued that the modifiers play a role in giving the connective an epistemic use or a speech act use. We do not claim, however, that these uses can be contributed by means of modifiers only. An analysis of the epistemic and speech act uses of connectives without modifiers was out of scope of the current thesis.

This thesis includes a two-pronged approach, one of which is the corpus-based part based on new annotations on TDB, dealing with the sense of discourse connectives, and the contribution of modifiers to the sense of connectives. The second part of the thesis is the statistical part in which have carried out a decision tree analysis to determine whether there is a systematic relation between the role of modifiers and the modality of the discourse connectives.

Methodologically, we developed an annotation framework for annotating the 513 discourse relations (already annotated on TDB). Our annotation scheme consisted of:

1. *sense of the relation*
2. *sense of the modifier*,
3. *modality of the relation with the modifier*
4. *modality of the relation without the modifier*

5.1 Results of corpus annotation

Our annotation effort has yielded the following results.

5.1.1 Sense of the relation

To annotate the senses, the PDTB sense hierarchy was used (Prasad et al., 2007). We identified all 4 top-level discourse relations in the data, which are temporal, contingency, comparison and expansion. The distribution of these senses are as follows, indicating that the temporal relations are highly modified in Turkish.

Temporal (60.34%) > Expansion (19.36%) > Contingency (16.92 %) > Comparison (3.38%)

Among the modified temporal relations, we have identified 298 singular modifiers, 28 multiple modifiers. We have not found any multiple modifiers co-occuring with any other discourse connectives in the data.

5.1.2 Sense (or semantic class) of the modifier

The data contains 5 semantic types of modifiers, which are focusing, aspectual, exocomparative, speaker-oriented and quantitative modifiers. The distribution of these classes in our data shows that focusing modifiers are predominant in the data, followed by aspectual and quantitative modifiers. The remaining class of modifiers appear very rarely.

Focusing (48.6%) > Aspectual (38.6%) > Quantitative (19.92%)

5.1.3 Modality of the relation (as indicated by the discourse connective and the modifier)

The occurrence of the three uses in discourse relations with modified discourse connectives is as follows:

Content (84.22) > Epistemic (9.58%) > Speech act (20%)

Our investigations have also shown that

- modified connectives with the temporal sense are predominantly in the content domain;
- modified connectives with the contingency sense tend to have more speech act and epistemic uses than the content use;
- modified connectives with the comparison sense tend to have the epistemic use more than the content and speech act uses;
- modified connectives with the expansion sense mostly have a content use, with less frequent uses of the speech act and epistemic uses.

5.1.4 Modality of the relation considering the discourse connective (without the modifier)

The results concerning the distribution of the modality of the relation when only the discourse connective was concerned showed that the content use was predominant, followed by the epistemic and speech act uses:

Content (77%) > Speech act (14.03%) > Epistemic (8.97%)

The comparison of these results with those in Section 5.1.3 suggest that with or without a modifier, discourse connectives are in the content domain, with the epistemic and speech act uses occurring far less frequently. However, our results suggest that modifiers have a role in pulling the domain of the discourse connective towards the epistemic or speech act domain.

5.1.5 Distribution of the modified alternative lexicalizations and explicit discourse connectives

The results indicate that while most of the modified discourse relations contain explicit discourse connectives (87.9%), modified alternative lexicalizations also exist in the data (12.1%). Our analysis have shown that modified temporal relations appear mostly with explicit discourse connectives (58.58%) and modified contingency relations are expressed predominantly by alternative lexicalizations.

The discourse relations expressed by modified alternative lexicalizations are predominantly in the speech-act domain when their sense is considered with the modifiers (79.03%) and rarely in the content domain (4.84%). Explicit discourse connectives are mostly in the content domain (86.9%) and rarely in the epistemic domain (8%) when their sense is considered with the modifiers. Modified alternative lexicalizations are mostly in the speech-act domain (91.9%) even without the modifiers and

explicit discourse connectives are generally express content domain when considered without the modifiers (89.36%).

5.1.6 Genre distribution

According to the results, modified connectives were identified in 9 different genres, which are *memoir, research, travel, diary, news, article, story, novel,* and *interview.* News is the most common genre (48.62%) and it also represents all the semantic domains in the highest frequencies (40.81% content domain, 4.29% epistemic domain and 3.51% speech-act domain). On the other hand, the least frequent use of content domain is in interview genre (0.38%). Research and diary genres have the least number of epistemic domain with one example in each genre (0.19%). Lastly, the diary genre has the least number of speech-act domain (0.19%) with modified connectives.

5.2 The decision tree analysis

The results of the decison-tree analysis and 10-fold cross-validation test show that the modality distinction and the role of modifiers on discourse relations are more than coincidence. The data and the classifications are highly predictable.

5.3 Limitations and further research

This thesis is not without any limitations. First of all, it is limited with the 513 instances of modified connectives in TDB. Further research with more instances of annotated modifiers may reveal more results. Secondly, it is limited with written language. The issue of modifiers can also be investigated in a spoken corpus of Turkish. Thirdly, a psycholinguistic analysis can be run to understand the online responses of language users with respect to modifiers in discourse relations.

Despite these limitations, this thesis provides an initial investigation of corpus-based modified discourse connectives and the role of modifiers in Turkish and contributes to a deeper understanding of Turkish discourse to the extent explicit connectives and alternative lexicalizations are concerned.

REFERENCES

Asher, N. (1993). *Reference to Abstract Objects in Discourse*. Kluwer Academic Publishers, Dordrecht.

Brown, G. and Yule, G. (1993.) *Discourse Analysis*. Cambridge: Cambridge University Press.

Bybee, J., and Fleischmann, S. (1995). *Modality in Grammar and Discourse.* John Benjamins Publishing.

Demirşahin, Işın, Çallı A. B. S., Balaban H. Ö., Çakıcı R., & Zeyrek D. (2012). Turkish Discourse Bank: Ongoing Developments. *Proceedings, Workshop on Turkic Languages, Istanbul: LREC 2012*. http://www.lrec-conf.org/proceedings/lre.

Edmonson, W. (1999). If Coherence Is Achieved, Then Where Doth Meaning Lie?. In Bublitz, W., Lenk, U. and Ventola, E. *Coherence in Spoken and Written Discourse: How to create it and how to describe it.* John Benjamins Publishing.

Ernst, T. (2000). *Semantic Features and the Distribution of Adverbs. In Approaching the Grammar of Adjuncts (ZAS Papers in Linguistics 17)*, ed. C. Fabricius-Hansen, E. Lang, and C. Maienborn, pp. 79–97. Berlin: ZAS.

Ernst, T. (2004). *The Syntax of Adjuncts.* Cambridge University Press.

Greenbaum, S. (1969). *Studies in English Adverbial Usage.* Coral Gables, Fla.: University of Miami Press.

Grimes, J. E. (1975). *The Thread of Discourse* The Hague & Paris: Mouton.

Grosz, B.J., & Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3): 175-204.

Halliday, M.A.K. and Hasan, R. (1976). *Cohesion in English*. London: Longman.

Hobbs, J., R. (1996). On the Relation Between the Informational and Intentional Perspectives on Discourse. *Computational and Conversational Discourse Volume 151 of the series NATO ASI Series* pp 139-157.

Hobbs, J., R. (2010). Clause-internal coherence. In Benz, A. and Sidner, C., L. *Constraints in Discourse 2.* John Benjamins Publishing.

Jackendoff, R. (1972). *Semantic Interpretation in Generative Grammar.* Cambridge, Mass.: MIT Press.

Jonhstone, B. (2002). *Discourse Analysis.* Malden, Mass. : Blackwell

Kehler, A. (2002). *Coherence, Reference, and the Theory of Grammar*. Stanford, CA: CSLI.

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

Knott, A., Sanders, T. and Oberlander, J. (2001). Levels of Representation in Discourse Relations. *Special Issue of Cognitive Linguistics.* Berlin Mouton de Gruyter.

Maat, H. P. and Sanders, T. (2001). Subjectivity in causal connectives: An empirical study of language in use. *Cognitive Linguistics* 12-3, 247-273.

Maat, H., P. and Degand, L. (2001). Scaling causal relations and connectives in terms of speaker involvement. *Cognitive Linguistics 12-3*, 211-245.

Mann, W.C. & Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text* 8(3): 243-281.

Mann, W.C. and Thompson, S.A. (1986). *Rhetorical Structure Theory: Description and Construction of Text Structures* (Technical Report No. ISI/RS-86–174). Marina del Rey, CA: Information Sciences Institute.

Mann, W.C. and Thompson, S.A. (1987). *Rhetorical Structure Theory: A Theory of Text Organization* (No. ISI/RS-87–190). Marina del Rey, CA: Information Sciences Institute.

Moore, J.D. and Pollack, M.E. (1992). A Problem for RST: The Need for Multi-level Discourse Analysis, *Computational Linguistics* 18(4): 537–44.

Prasad, R., Miltsakaki, E., Dinesh, N., Lee, A., Joshi, A., Robaldo, L., et al. (2007). The Penn Discourse Treebank 2.0 Annotation Manual. Technical Report 203, Institute for Research in Cognitive Science, University of Pennsylvania, Philadelphia, Pennsylvania.

Quirk, Randolph, S. Greenbaum, G. Leech, and J. Svartvik. (1985). *A comprehensive grammar of the English language.* London: Longman.

Renkema, J. (2004). *Introduction to Discourse Studies.* John Benjamins Publishing..

Renkema, J. (2009). *Discourse, of course.* Amsterdam: John Benjamins.

Risselada, R. and Spooren, W. (1998) Introduction: Discourse markers and coherence relations. *Journal of Pragmatics* 30 (2), 131-133.

Sanders, T., Spooren, W. and Noordman, L. (1992). Toward a Taxonomy of Coherence Relations, *Discourse Processes* 15: 1–35.

Schiffrin, D. (1987). *Discourse markers* Cambridge: Cambridge University Press.

Searle, J. and Vanderveken, D. (1985). *Foundations of illocutionary logic.* Cambridge, England: Cambridge University.

Spooren, W., Sanders, T., Huiskes, M. and Degand., L. (2010) Subjectivity and causality: A corpus study of spoken language *Empirical and experimental methods in cognitive/functional research*, 241-255.

Sweetser, E. (1990). *From etymology to pragmatics : metaphorical and cultural aspects of semantic structure*. Cambridge [England] ; New York : Cambridge University Press.

Taboada, M. (1998). Rhetorical Relations in Dialogue. In Moder, C., L. And Martinovic-Zic, A.(ed.) (2004). *Discourse Across Languages and Cultures*. John Benjamins B. V.

Taboada, M. (2009). Implicit and Explicit Coherence Relations. In J. Renkema (ed.) *Discourse, of Course*. Amsterdam/Philadelphia: John Benjamins. 127-140.

Van Dijk, T.A. (1997). *Discourse studies : a multidisciplinary introduction.* London ; Thousand Oaks, Calif. : Sage Publications.

Webber, B. (2004). D-LTAG: Extending lexicalized TAG to discourse. *Cognitive Science*, *28*(5), 751-779.

Webber, B., and Joshi, A. (1998). Anchoring a lexicalized tree-adjoing grammar for discourse. In Stede, M., Wanner, L., Hovy, E. (Eds) *Discourse Relations and Discourse Markers: Proceedings of the Conference*, pp. 86–92. Association for Computational Linguistics.

Witten, I. H., Eibe, F. and Hall, M., A. (2011). *Data mining : practical machine learning tools and techniques.* Morgan Kaufmann Publishers.

Zeevat, H. (2003). Particles: Presupposition triggers, context markers or speech act markers. Optimality theory and pragmatics, 91-111.

Zeyrek, D., Demirşahin, I., Sevdik-Çallı, A. B., Balaban, H. Ö., Yalçınkaya İ.,& Ümit D. Turan (2010). The annotation scheme of the Turkish discourse bank and an evaluation of inconsistent annotations. In *Proceedings of the Fourth Linguistic Annotation Workshop*, pages 282-289, Uppsala, Sweden.

Zeyrek, D., Demirşahin, I., Sevdik-Çallı, Ayışığı, B., Çakıcı, R. (2013). Turkish Discourse Bank: Porting a discourse annotation style to a morphologically rich language. *Dialogue & Discourse. Vol. 4, No. 2:* 174-184.

APPENDIX A

This appendix presents the examples of the discourse relations in the data.

Temporal relations

1) *Politikası ne olursa olsun, uzun yıllar sonra iktidara tek partinin geçmesi kimileri için yeni bir "umudun" dayanağı oldu...* Hemen **ardından Kopenhag 'da alınan "umut kırıcı" sonuçlar ise aynı siyasi söylemde "AB sorgulamasını" öne çıkardı...**

'*Whatever policy they have, the single party regime after so many years became the basis of a new "hope" for some people.* Right **after this, the "disappointing" results taken in Kopenhagen drove the "EU questioning" in the same political statement.'**

(TEMPORAL: Asynchronous: succession)

2) *Kurgu konusuna aşağıda yine döneceğim ama* daha **önce Sevgili Arsız Ölüm'ün başka bir yönüne değinmek istiyorum.**

'*I will come back to the fiction issue* but **before that, I want to mention another aspect of Sevgili Arsız Ölüm**'

(TEMPORAL: Asynchronous: precedence)

3) *Bense tutup onunla yarışmak isterdim.* ve bu isteğimin nereden geldiğini hiçbir zaman anlayamadım. **Büyüklük taslardım bir yandan** da, **uzaklaş, sen önde başla, derdim.**

'*And I used to compete with him*, and I could never understand where this desire came from. **At the same time, I used to patronize and say him to move away and start early.**'

(TEMPORAL: Synchronous)

Contingency relations

4) Burada evler sık sık iklim değişmesinden dolayı tamir görmüş. **Tamir gördüğü için** de, *biz bunların yapı evrelerini takip edebiliyoruz*.

'The houses here hav been repaired so many times because of the climate change. **Since they have been repaired,** *we can follow the construction processes of these houses.*'

(CONTINGENCY: Cause: reason)

5) *Adeta sorulabilecek bütün büyük soruların sorulup tüketildiği, bilimin mevcut düzleminin ulaşılabilecek en yüksek düzlem olduğu,* **dolayısıyla** da **teorinin son bulduğu bir dönemde yaşıyoruz.**

'In fact, **we live in an era in which** *all the possible questions have been asked and consumed, science is the highest platform that its platform can reach,* **so theory ends up.**'

(CONTINGENCY: Cause: result)

6) Bu köyde doğup büyümüştü sanki; çınarın hışırtısı kulaklarını, damların köy alanına akan beyazlığı gözlerini, toprağın cayır cayır yanan soluğu tenini hiç etkilemiyordu.

Belki de **bu yüzden, o gün bekçiden başka hiç kimse bir yabancıyla aynı yerde bulunmanın tedirginliğini duymamıştı kendinde.**

'As if he had been born and grown up in this village; the rustle of the plane tree did not affect his ears, the whiteness of the roofs flowing over the village center did not disturb his eyes and the burning breath of soil did not affect his skin. Maybe because of this, he was not nervous about anything except being in the same place with a warder on that day.'

(CONTINGENCY: Pragmatic cause)

Comparison relations

12) Buna rağmen, Falih Rıfkı Bey "bütün davayı çıkaran sensin" deyince, Feyzi Bey *böyle bir durumda mebusluktan istifa edeceğini,* **aksi halde** de **Falih Rıfkı Bey'in istifa etmesi gerektiğini** belirtmiş ve Fırka Grubu'nun büyük desteğini almıştır.

'Although this; when Mr. Falih Rıfkı said "you are the one who has caused the case", Mr. Feyzi stated that *he would resign from being a member of the parliament in this situation* **otherwise Mr. Falih Rıfkı had to resign** and he took a huge support of the Party Group.'

(COMPARISON: Contrast: opposition)

13) Sanki **bu ülkede bir inanç sorunu varmış gibi,** *"Benim vatandaşım göğsümü gere gere Müslümanım diyemeyecek mi?"* diyen siyasal parti başkanları, baş sorumlulardır.

' **As if there were a belief problem in this country**, the political party leaders who say "*Can't my citizen proudly say that he/she is Muslim?"*are the main responsibles.

(COMPARISON: Concession: contra-expectation)

Expansion relations

14) Devrimci tekne insanı **hem** *düşünce açlığından kurtarır*, **hem** de **doyurur.**

'A revolutionist boat **both** *saves the person from thought hunger* **and satisfies him/her.**'

(EXPANSION: Conjunction)

15) Ardından sanki, **öpücük yollar gibi** *elini beyaz dudaklarına koyuvermişti*, ama tam emin değildi öpücük yolladığından.

' And then *he put his hand on his white lips* **as if he was sending a kiss**, but he was not certainly sure that he was sending a kiss.'

(EXPANSION)

16) *ABD'ye niçin üs vermemeliyiz* **ve** de **savaşa katılmamalıyız!**

'*We should not give any exponent to the USA* **and we should not crusade war!**'

(EXPANSION: List)

18) *"İiç!" diye bağırıyor.* **Kışkırtıyor yahut** da.

'*He shouts 'Drink!*', **or provokes.**'

(EXPANSION: Alternative: disjunctive)

The Special Case of *sanki* and *hem de*

Sanki

19) İşte hep söylediği gibi, sanki **insanoğlu değilmiş, su içmez, uyumazmış gibi** *renkli taş parçalarını yan yana getirerek yeni yaşamlar oluşturuyor; hep bir şeyleri kendince ölümsüzleştiriyor.*
‘ As he always says, *he forms new lives by ordering the colorful stone parts side by side and immortalizes something all the time in his opinion;* **as if he were not a human, did not drink water or sleep**.’

(COMPARISON: Concession: contra-expectation)

20) “Benimle ölür müsün?" demişti. *Bunu* sanki **"Benimle birlikte yaşar mısın? Benimle evlenir misin? Benimle sevişir misin? Benimle gelir misin? Benimle yürür müsün?" der gibi** *söylemişti.*
‘He said “Will you die with me?”. He said this as if he said “Will you live with me? Will you marry me? Will you make love with me? Will you come with me? Will you walk with me?”’

(EXPANSION)

Hem de

21) Her insan **hem** *toplumda onay görmek ve herkes gibi olmak* **hem** de **farklı, özgün ve üstün olmak** ister.
‘Every person wants to be **both** *approved, like everybody*, **and different, unique, outstanding.**’

(EXPANSION: Conjunction)

22) *Söyledi bana bunları, biliyor musun?* **Bir gece zamanı, yataktayken söyledi hem** de...
‘ *Do you know that he had told me these?* **Besides, while we were in the bed at a night...**’

(EXPANSION: Conjunction)

APPENDIX B

This appendix presents the gold data which is totally agreed by the independent annotators.

Discourse relation sense	**Discourse connector**	**Modifier**	**Modality with the modifier**	**Modality without the modifier**	**Number**
Temporal					
Asynchronous					
Succession	sonra	biraz, birkaç vs.	content	content	89
	sonra	focus(da, ise, bile)	content	content	59
	sonra	aspectual	content	content	63
	ardından	focus(da)	content	content	8
	ardından	aspectual(hemen)	content	content	2
	bundan sonra	focus (ise)	speech-act	speech-act	1
	bundan sonra	focus(ancak)	speech-act	speech-act	1
Total					223
Precedence	önce	biraz,birkaç vs	content	content	3
	önce	aspectual	content	content	28
	önce	focus(da)	content	content	1
	kadar	focus(da)	content	content	1
	kadar	aspectual(taa ki)	content	content	1
	kadar	aspectual(daha)	content	content	1
Total					35
Synchronous					
	bir yandan	focus(da)	content	content	27
	öte yandan	focus(da)	content	content	2
	diğer yandan	focus(da)	content	content	1
	o zaman	focus(da,ancak)	speech-act	speech-act	4
	o zaman	aspectual(işte)	speech-act	speech-act	11
	o zaman	speaker-oriented(belki)	epistemic	speech-act	1
	zaman	focus(da)	content	content	2
	zaman	aspectual(işte)	content	content	1
	ne zaman	aspectual(ki)	content	content	1
	şimdi	focus(ise)	content	content	1
	şimdi	focus(da)	content	content	1
Total					52
Discourse relation sense	**Discourse connector**	**Modifier**	**Modality with the modifier**	**Modality without the modifier**	**Number**
Contingency					
Cause					
Reason	amacıyla	focus(da)	content	content	2

	dolayısıyla	focus(da)	content	content	1
	için	focus(da)	content	content	23
Total					26
Result	dolayısıyla	focus(da)	content	content	2
	sonuç olarak	focus(da)	content	content	1
Total					3
Pragmatic cause					
	bu sebepten dolayı	focus(da)	speech-act	speech-act	1
	bundan dolayı	focus(da)	speech-act	speech-act	1
	bunun için	focus(da,sırf)	speech-act	speech-act	7
	bu nedenle	focus(da)	speech-act	speech-act	10
	bu nedenle	speaker-oriented(belki)	epistemic	speech-act	2
	o nedenle	speaker-oriented(belki)	epistemic	speech-act	2
	o nedenle	focus(da)	epistemic	speech-act	1
	bu nedenlerle	speaker-oriented(olsa gerek)	epistemic	speech-act	1
	bu nedenlerle	bütün	speech-act	speech-act	1
	bundan ötürü	focus(da)	speech-act	speech-act	2
	bu yüzden	focus(da)	speech-act	speech-act	5
	bu yüzden	speaker-oriented(belki de)	epistemic	speech-act	3
	bu yüzden	biraz da	speech-act	speech-act	1
	bu yüzden	focus(sadece)	speech-act	speech-act	1
Total					38
Comparison					
Contrast					
Opposition	aksine,tersine	focus(tam)	content	content	4
	aksi halde	focus(da)	content	content	1
Total					5
Concession					
Discourse relation sense	**Discourse connector**	**Modifier**	**Modality with the modifier**	**Modality without the modifier**	**Number**
Contra-expectation	bunlara rağmen	bütün	speech-act	speech-act	1
	gibi	Exocomparative (sanki)	epistemic	epistemic	12
Total					13
Expansion					
Conjunction					
	hem...hem	focus(da)	content	content	60
	hem	focus(da)	content	content	13
	ayrıca	focus(da)	content	content	1
	ne...ne	focus(da)	content	content	27
Total					101
Manner					
	gibi	Exocomparative	epistemic	epistemic	11

		(sanki)			
	gibi	Exocomparative (tıpkı)	epistemic	epistemic	9
	gibi	Exocomparative (aynı)	epistemic	epistemic	1
	gibi	focus(da)	epistemic	epistemic	1
Total					22
Alternative					
Disjunctive	yahut	focus(da)	content	content	1
Total					1
Instantiation					
	örnek olarak	speaker-oriented(çarpıcı)	epistemic	content	1
Total					1

APPENDIX C

This appendix indicates the discourse relations expressed by modified discourse connectors.

Discourse connector	**Discourse relation**		**Modified discourse connectives**	**Number**
Explicit discourse connective	TEMPORAL / Asynchronous	Succession	*yıllar sonra* 'many years later'(4), *yıllar sonra da* 'and many years later'(1), *15 yıl sonra* '15 years later'(1), *15-20 gün sonra* '15-20 days later'(1), *16 yıl sonra* '16 years later'(1), *19 yıl sonra* '19 years later'(1), *2 saat sonra* '2 hours later'(1), *2 yıl sonra* '2 years later'(1), *20 gün sonra* '20 days later'(1), *25 sene sonra* '25 years later'(1), *25 yıl sonra* '25 years later'(1), *5 yıl sonra* '5 years later'(1), *8 saat sonra* '8 hours later'(1), *aylar sonra* 'many months later'(1), *az sonra* 'after a while' (4), *az sonra da* 'and after a while' (1), *sonra bile* 'even after' (1), *bir gün sonra* 'a day later'(2), *bir gün sonra da* 'and a day later'(1), *bir hafta sonra* 'a week later' (2), *bir hafta sonra da* 'and a week later' (1), *bir süre sonra* 'after a while' (18), *bir süre sonra da* 'and after a while' (2), *bir yıl sonra* 'a year later'(1), *bir yıl sonra da* 'and a year later'(1), *bir zaman sonra* 'after some time'(2), *biraz sonra* 'after a short time'(3), *birkaç gün sonra* 'a couple of days later'(2), *birkaç yıl sonra da* 'and a couple of years later'(1), *çok sonra* 'after so many times'(1), *daha sonra da* 'and later'(10), *daha sonra* 'later' (49), *sonra da* 'after' (57), *ardından da* 'after that' (8), *hemen ardından* 'immediately after' (2), *iki gece sonra* 'two nights later'(1), *iki*	227

			gün sonra ‘two days later’(2), *iki hafta sonra* ‘two weeks later’(1), *iki hafta sonra da* ‘and two weeks later’(1), *iki hafta ya da bir ay sonra* ‘two weeks or a month later’(1), *iki yıl sonra* ‘two years later’(1), *sonra ise* ‘and after’(1), *kısa bir süre sonra* ‘after a short time’(6), *kısa bir süre sonra da* ‘and after a short time’(2), *kısa süre sonra* ‘after a short time’(1), *ön dört yıl sonra da* ‘and fourteen years later’(1), *sadece üç ay sonra* ‘just three years later’(1), *uzunca bir zaman sonra* ‘after a long time’(1), *üç ay sonra* ‘three months later’(1), *üç gün sonra* ‘three days later’(1), *üç hafta sonra* ‘three weeks later’(1), *üç saat sonra* ‘three hours later’(1), *üç yıl sonra* ‘three years later’(1), *yarım saat sonra* ‘half an hour later’(2), *yedi sene sonra* ‘seven years later’(1), *ancak+ARG2+sonra* ‘only after’(1), *kaç yıl sonra* ‘how many years later’(1), *kaç zaman sonra* ‘many years later’(1), *kim bilir kaç kedi günü sonra* ‘who knows many cat days later’(1), *neden sonra* ‘somehow later’(3), *özellikle+ARG2+sonra* ‘especially after’(2) *en sonra da* ‘lastly’(4)	
Discourse connector	**Discourse relation**		**Modified discourse connectives**	**Number**
		Precedence	*yıllar önce* ‘many years ago’(1), *daha+ARG2+kadar* ‘until when’(1), *kadar da* ‘until’(1), *taa ki+ARG2+kadar* ‘until when’(1), *az önce* ‘a while ago’(1), *bir gün önce* ‘a day ago’(1), *daha önce* ‘previously’(24), *önce de* ‘before’(1), *hemen önce* ‘just before’(2), *ilk önce* ‘firstly’(2),	36

			uzun yıllar önce ‘many years ago’(1),	
	TEMPORAL / Synchronous		*öte yandan da* ‘ on the other hand’(2), *bir yandan da* ‘at the same time’(27), *tabii bir yandan da* ‘of course at the same time’(1), *diğer yandan da* ‘on the other hand’(1), *ne zaman ki* ‘when’ (1), *zaman da* ‘and at that time’(2), *işte+ARG2+zaman* ‘and at that time’(1), *şimdi de* ‘and now’(1), *şimdi ise* ‘and now’(1)	37
	CONTINGENCY/ Cause			
		Reason	*amacıyla da*‘in order to’(2), *dolayısıyla da* ‘so that’(1), *için de* ‘ and because of ‘(23),	26
		Result	*dolayısıyla da*‘as a result(2)	2
	CONTINGENCY/ Pragmatic cause		*biraz da +ARG+için* ‘partly because’(1), *çoğu kez+ARG2+için* ‘mainly because’(1), *yalnızca+ARG2+için* ‘just because of’ (1), *çoğu kez+ARG2+sonucunda* ‘mostly as a result of’(1)	4
	COMPARISON/ Contrast	Opposition	*tam aksine* ‘on the exact contrary’(1), *tam tersine* ‘on the exacy contrary’(3)	4
	COMPARISON/ Concession	Contra-expectation	*sanki+ARG2+gibi* ‘as if’(12)	12
Discourse connector	**Discourse relation**		**Modified discourse connectives**	**Number**
	EXPANSION/ Conjunction		*ayrıca da* ‘besides’(1), *gerek+ARG1+gerekse de* ‘both+ARG1+and’(1), *hem+ARG1+hem de* ‘both+ARG1+and’(33), *ne+ARG1+ne de* ‘neither+ARG1+nor’(26), *hem de* ‘besides’(13)	74
	EXPANSION/ Manner		*sanki+ARG2+gibi* ‘just like’(12) *tıpkı+ARG2+gibi* ‘just like’(9), *aynı+ARG2+gibi* ‘just like’(1), *gibi de* ‘like’(1)	23
	EXPANSION/	Disjunctive	*yahut da* ‘or’(1)	1

	Alternative			
	EXPANSION/ List		*ve de* ‘and’(5)	5
	EXPANSION/ Instantiation			0
Total				451
Alternative lexicalization	TEMPORAL / Asynchronous	Succession	*ancak bundan sonra* ‘only after this’(1), *bundan sonra ise* ‘and after this’(1)	2
		Precedence		0
	TEMPORAL / Synchronous		*ancak o zaman* ‘only at that time’(3), *işte o zaman* ‘at that exact time’(11), *o zaman da* ‘and at that time’(2), *belki o zaman* ‘maybe at that time’(1)	17
	CONTINGENCY/ Cause			
		Reason		0
		Result	*sonuç olarak da* ‘and as a result of’(1), *bu amaçla da* ‘and for this’(1)	2
Discourse connector	**Discourse relation**		**Modified discourse connectives**	**Number**
	CONTINGENCY/ Pragmatic cause		*belki de bu amaçla* ‘maybe for this’(1), *belki de bu yüzden* ‘maybe because of this’(3), *belki de o nedenle* ‘maybe because of that’(2), *o nedenle de* ‘and because of that’(1),*belki bu nedenle* ‘maybe because of this’(1), *belki de bu nedenle* ‘maybe because of this’(1), *bütün bu nedenlerle* ‘because of all of these’(1), *bu nedenle de* ‘and because of this’(10), *bu sebepten dolayı da* ‘and because of this’(1), *bundan dolayı da* ‘and because of this’(1), *bundan ötürü de* ‘and because of this’(2), *bunun için de* ‘and for this’(6), *sırf bunun için* ‘just	38

			because of this'(1), *sadece bu yüzden* 'just because of this'(1), *bu yüzden de* ' because of this'(5), *biraz da bu yüzden* 'partly because of this'(1)	
	COMPARISON/ Contrast	Opposition	*aksi halde de* 'in other case'(1)	1
	COMPARISON/ Concession	Contra-expectation	*bütün bunlara rağmen* 'in spite of all of these' (1)	1
	EXPANSION/ Conjunction			0
	EXPANSION/ Manner			0
	EXPANSION/ Alternative	Disjunctive		0
	EXPANSION/ List			0
	EXPANSION/ Instantiation		*çarpıcı örnek olarak* 'as a striking example'(1)	1

APPENDIX D

This appendix indicates the modified discourse connectors and the modifiers that they take.

Type of discourse connector	**Discourse connector**	**The modifiers it takes**	**Number**
Explicit discourse connective	*Sonra* 'after, later'	*yıllar* 'many years'(4), *yıllar...da* 'and many years'(1), *15 yıl* '15 years'(1), *15-20 gün* '15-20 days'(1), *16 yıl* '16 years'(1), *19 yıl* '19 years'(1), *2 saat* '2 hours'(1), *2 yıl* '2 years'(1), *20 gün* '20 days'(1), *25 sene* '25 years'(1), *25 yıl* '25 years'(1), *5 yıl* '5 years'(1), *8 saat* '8 hours'(1), *aylar* 'many months'(1), *az* 'a while' (4), *az ...da* 'and ...a while' (1), *bile* 'even' (1), *bir gün* 'a day'(2), *bir gün ...da* 'and a day'(1), *bir hafta* 'a week later' (2), *bir hafta ... da* 'and a week' (1), *bir süre* 'a while' (18), *bir süre ... da* 'and ... a while' (2), *bir yıl* 'a year'(1), *bir yıl... da* 'and a year'(1), *bir zaman* 'some time'(2), *biraz* 'a short time'(3), *birkaç gün* 'a couple of days'(2), *birkaç yıl ... da* 'and a couple of years'(1), *çok* 'so many times'(1), *daha ... da* 'and more'(10), *daha* 'more' (49), *da* 'focus particle-dA' (57), *iki gece sonra* 'two nights'(1), *iki gün* 'two days'(2), *iki hafta* 'two weeks'(1), *iki hafta ... da* 'and two weeks'(1), *iki hafta ya da bir ay* 'two weeks or a month'(1), *iki*	217

		yıl ‘two years’(1), *ise* ‘and’(1), *kısa bir süre* ‘a short time’(6), *kısa bir süre ... da* ‘and... a short time’(2), *kısa süre* ‘a short time’(1), *ön dört yıl ... da* ‘and fourteen years’(1), *sadece üç ay* ‘just three years’(1), *uzunca bir zaman* ‘a long time’(1), *üç ay* ‘three months l’(1), *üç gün* ‘three days’(1), *üç hafta* ‘three weeks’(1), *üç saat* ‘three hours’(1), *üç yıl* ‘three years’(1), *yarım saat* ‘half an hour’(2), *yedi sene* ‘seven years’(1), *ancak+ARG2+* ‘only’(1), *kaç yıl* ‘how many years’(1), *kaç zaman* ‘many years’(1), *kim bilir kaç kedi günü* ‘who knows many cat days’(1), *neden* ‘somehow’(3), *özellikle+ARG2+* ‘especially’(2) *en ... da* ‘lastly’(4)	
	Ardından ‘after, later’	*da* ‘focus particle – dA’(8), *hemen* ‘immediately’(2)	10
	Önce ‘before,ago’	*yıllar* ‘many years’(1), *az* ‘a while’(1), *bir gün* ‘a day’(1), *daha* ‘more’(24), *de* ‘focus particle - dA’(1), *hemen* ‘just’(2), *ilk* ‘firstly’(2), *uzun yıllar* ‘many years’(1),	33
	Kadar ‘until’	*daha+ARG2+* ‘when’(1), *da* ‘focus particle - dA’(1), *taa ki+ARG2+* ‘when’(1)	3
Type of discourse connector	**Discourse connector**	**The modifiers it takes**	**Number**
	Öte yandan da ‘on the other hand’	*da* ‘ focus particle – dA’(2)	2
	Bir yandan da ‘on the other hand, at the same time’	*da* ‘focus particle - dA’(27), *tabii ... da* ‘of course ...’(1)	28
	Diğer yandan ‘on the other hand’	*da* ‘focus particle -dA’ (1)	1
	Ne zaman ‘when’	*ki* ‘just’(1)	1
	Zaman ‘when’	*da* ‘focus particle -dA’	3

		(2), *işte+ARG'*+ 'and'(1)	
	Şimdi 'now'	*de* 'focus particle -dA'(1), *ise* 'and'(1)	2
	Amacıyla 'in order to'	*da* 'focus particle -dA' (2)	2
	Dolayısıyla 'so that'	*da* 'focus particle -dA' (3)	3
	İçin 'because'	*de* 'focus particle -dA' (23), *biraz da +ARG+* 'partly'(1), *çoğu kez+ARG2+* 'mainly'(1), *yalnızca+ARG2+* 'just'(1),	26
	Sonucunda 'as a result of'	*çoğu kez+ARG2+* 'mainly'(1)	1
	Aksine 'on the contrary'	*tam* 'exact'(1)	1
	Tersine 'on the contrary'	*tam* 'exact'(3)	3
	Gibi 'like'	*sanki...* 'as if'(24), *tıpkı* 'just'(9), *aynı* 'the same'(1), *de* 'focus particle -dA' (1)	35
	Ayrıca 'besides'	*da* 'focus particle -dA' (1)	1
	Gerek+ARG1+gerekse 'both+ARG1+and'	*de* 'focus particle -dA' (1)	1
	Hem+ARG1+hem 'both+ARG1+and'	*de* 'focus particle -dA' (33)	33
	Hem 'besides'	*de* 'focus particle -dA' (13)	13
	Ne+ARG1+ne 'neither+ARG1+nor'	*de* 'focus particle -dA' (26)	26
	Yahut 'or'	*da* 'focus particle -dA' (1)	1
	Ve 'and'	*da* 'focus particle -dA' (5)	5
Total			451
Type of discourse connector	**Discourse connector**	**The modifiers it takes**	**Number**
Alternative lexicalization	*Bundan sonra* 'after this'	*ancak* 'only'(1), *ise* 'and'(1)	2
	O zaman 'at that time'	*ancak* 'only'(3), *işte* 'and'(11), *da* 'focus particle -dA' (2), *belki* 'maybe'(1)	17
	Sonuç olarak 'as a result'	*da* 'focus particle -dA' (1)	1
	Bu amaçla 'for this'	*da* 'focus particle -dA' (1), *belki de* 'maybe'(1)	2
	Bu yüzden 'because of this'	*belki de* 'maybe'(3), *sadece* 'only'(1), *de*	10

		‘focus particle -dA’ (5), *biraz da* ‘partly’(1)	
	O nedenle ‘because of that’	*belki de* ‘maybe (2), *de* ‘focus particle -dA’ (1),	3
	Bu nedenle ‘because of this’	*belki de* ‘maybe’(1), *belki* ‘maybe’(1), *de* ‘focus particle -dA’ (10),	12
	Bu nedenlerle ‘because of these’	*bütün* ‘all’(1)	1
	Bu sebepten dolayı ‘because of this’	*de* ‘focus particle -dA’ (1)	1
	Bundan dolayı ‘because of this’	*de* ‘focus particle -dA’ (1)	1
	Bundan ötürü ‘because of this’	*de* ‘focus particle -dA’ (2)	2
	Bunun için ‘for this’	*de* ‘focus particle -dA’ (6), *sırf* ‘just’(1)	7
	Aksi halde ‘in other case’	*de* ‘focus particle -dA’ (1)	1
	Bunlara rağmen ‘in spite of these’	*bütün* ‘all’ (1)	1
	Örnek olarak ‘as an example’	*çarpıcı* ‘striking’(1)	1
Total			62