EXPLORING ATTRIBUTION IN TURKISH DISCOURSE: AN ANNOTATION-BASED ANALYSIS

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

AYSU NUR YAMAN

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

SEPTEMBER 2024

EXPLORING ATTRIBUTION IN TURKISH DISCOURSE: AN ANNOTATION-BASED ANALYSIS

Submitted by Aysu Nur Yaman in partial fulfillment of the requirements for the degree of **Master of Science in Cognitive Science Department, Middle East Technical University** by,

Prof. Dr. Banu Günel Kılıç
Dean, **Graduate School of Informatics** _____

Assoc. Prof. Dr. Barbaros Yet
Head of Department, **Cognitive Science** _____

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

Examining Committee Members:

Assoc. Prof. Dr. Barbaros Yet
Cognitive Science Dept., METU _____

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

Prof. Dr. Mine Güven
Department of Translation and Interpreting, Doğuş _____
University

Date: 04.09.2024

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 : **Aysu Nur Yaman**

Signature **:** _____

ABSTRACT

EXPLORING ATTRIBUTION IN TURKISH DISCOURSE: AN ANNOTATION-BASED ANALYSIS

Yaman, Aysu Nur
MSc., Department of Cognitive Science
Supervisor: Prof. Dr. Deniz Zeyrek Bozşahin

September 2024, 101 pages

Attribution involves recognizing and crediting sources, a process integral to both written and spoken discourse. This study extends existing frameworks, particularly the Penn Discourse TreeBank (PDTB), which elucidates how sources and statements are attributed in English, to Turkish texts using the Turkish Discourse Bank version 1.2 (TDB 1.2). The aim is to understand the mechanisms of attribution in Turkish and reduce dependency on manual annotation for text analysis. Employing insights from the literature, a tailored annotation scheme was developed. Data annotation achieved strong inter-annotator agreement with Cohen's kappa coefficients: 0.83 for Arg1, 0.80 for Arg2, and 0.77 for Entire Discourse Relation (Entire Drel), indicating near-perfect to substantial agreement. Analysis of the annotated data revealed that the Other (Ot) category dominated with 296 instances in REL, followed by Arg1 (259 instances) and Arg2 (221 instances). The majority of verbs were communicative such as *de-* ('to say') 211 times, *söyle-* ('to tell') 88 times, *belirt-* ('to point out') 56 times, with communicative verbs comprising 75.9% of occurrences in relevant categories. In comparing journalistic and non-journalistic texts, the analysis found that journalistic genres had higher frequencies of attribution. News texts showed the highest number of attributions with 307 instances, followed by articles with 89 instances, and interviews with 27 instances. In non-journalistic texts, novels exhibited 296 attributions, followed by memoirs with 146, and research texts with 82 attributions. This analysis enriches the TDB and sets a foundation for future automated text analysis.

Keywords: Turkish Discourse Bank, Attribution Annotation, Annotation Schema, Computational Linguistics.

ÖZ

TÜRKÇE SÖYLEMDE ATIF İNCELEMESİ: ANOTASYON TABANLI BİR ANALİZ

Yaman, Aysu Nur
Yüksek Lisans, Bilişsel Bilimler Bölümü
Tez Yöneticisi: Prof. Dr. Deniz Zeyrek Bozşahin

Eylül 2024, 101 sayfa

Atıf, kaynakları tanıma ve onlara kredi verme sürecini içerir; bu süreç hem yazılı hem de sözlü söylemlerde bütünleşik bir rol oynar. Bu çalışma, kaynakların ve ifadelerin İngilizce'de nasıl atfedildiğini açıklayan Penn Discourse TreeBank (PDTB) gibi mevcut çerçeveleri genişleterek, Türkçe metinlere Türkçe Söylem Bankası 1.2 (TDB 1.2) kullanarak uygulamaktadır. Amaç, Türkçe'de atıf mekanizmalarını anlamak ve metin analizi için manuel anotasyon bağımlılığını azaltmaktır. Literatürdeki bulgulardan yararlanılarak, Türkçeye özgü bir anotasyon şeması geliştirilmiştir. Veri anotasyonu, Cohen kappa katsayıları ile verilerdeki atıf bulgularını işaretleyen kişiler arasıda güçlü bir anlaşma sağladı: Arg1 için 0.83, Arg2 için 0.80 ve Tüm Söylem İlişkisi (Entire Drel) için 0.77, neredeyse mükemmel ile önemli bir uzlaşmaya işaret ediyor. İşaretlenmiş verilerin analizi, REL'de Ot (Diğer) kategorisinin 296 örnekle baskın olduğunu, ardından Arg1 (259 örnek) ve Arg2 (221 örnek) geldiğini ortaya koymuştur. Çoğu fiil, *de-* 211 kez, *söyle-* 88 kez, *belirt-* 56 kez olmak üzere iletişimsel fiillerden oluşmuş ve iletişimsel fiiller, ilgili kategorilerdeki olayların %75.9'unu oluşturmuştur. Gazetecilik ve gazetecilik dışı metinler karşılaştırıldığında, gazetecilik türlerinde atıfların daha sık kullanıldığı tespit edilmiştir. Haber metinleri 307 atıfla en yüksek sayıya sahipken, makaleler 89 atıfla, röportajlar ise 27 atıfla onu takip etmektedir. Gazetecilik dışı metinlerde ise romanlar 296 atıfla öne çıkarken, anılar 146 atıfla ikinci, araştırma metinleri ise 82 atıfla üçüncü sırada yer almaktadır. Bu analiz, TDB'yi zenginleştirerek gelecekteki otomatik metin analizlerine bir temel oluşturmaktadır.

Anahtar Sözcükler: Türkçe Söylem Bankası, Atıf Anotasyonu, Anotasyon Şeması, Hesaplamalı Dilbilim.

To My Mother

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABL	Ablative
ACC	Accusative
AdvP	Adverbial Phrase
AO	Abstract Object
AR	Attribution Relation
Arb	Arbitrary
ARG1	Argument 1
ARG2	Argument 2
B	Beginning
Comm	Communicative
Comp	Complementizer
CRF	Conditional Random Field
Ctrl	Control
CVP	Converb
DC	Discourse Connective
EVID	Evidential
Ftv	Factive
GATE	General Architecture for Text Engineering
GEN	Genitive
I	Inside
INF	Infinitive
ISST	Italian Syntactic-Semantic Treebank
LOC	Locative
MSA	Modern Standard Arabic
N	Noun
NLP	Natural Language Processing
NMN	Nominative
NE	Nominated Entity
NEG	Negation
NP	Noun Phrase
O	Outside
Ot	Other
PAST	Past Tense
PAtt	Belief Proposition
PL	Plural
PP	Prepositional Phrase
PRES	Present Tense
PROG	Progressive
PSG	Passive Voice
PDTB	Penn Discourse TreeBank

REL	Relation
RST	Rhetorical Structure Theory
S	Sentence
SARC	Spoken Italian AR Corpus
SBAR	Subordinate Clause
TDB	Turkish Discourse Bank
V	Verb
VP	Verb Phrase
WSJ	Wall Street Journal
Wr	Writer

CHAPTER 1

INTRODUCTION

This chapter provides an introduction to the purpose, and research questions of the thesis on the examination of attribution in Turkish discourse. It also highlights the significance of the research and provides an outline of the thesis to acquaint the readers with the work.

1.1. Attribution

Attribution in written or spoken communication is the act of acknowledging the sources of information, ideas, or statements that are not originally generated by the author or speaker. It is an essential aspect of responsible and ethical communication, as it gives credit to the original creators or contributors and helps readers or listeners assess the reliability and validity of the information presented.

Attribution is a key idea in how we understand discussions and texts. It shows the links between abstract ideas and the people or agents connected to them. In other words, attribution involves the cognitive and linguistic process of ascribing beliefs and statements expressed to the responsible party or parties (Riloff and Wiebe, 2003; Wiebe et al., 2004, 2005).

In English, the attribution of sourced material is often achieved through the use of diverse expressions, such as a specific agent given as the source of attribution as well as specific predicates like *say* or *state*, or expressions like *according to*. These linguistic cues indicate that the presented information or ideas originate from external sources. For instance, consider the following example:

(1)   *According to* the latest research findings, climate change is intensifying at an alarming rate.

In this example, the attribution marker *according to* signals the attribution of the statement and informs readers about the source, that is, the latest research findings.

1.2. Purpose of the Work and Research Questions

The Penn Discourse TreeBank (PDTB) has explored the attribution of beliefs and statements in written text and offered a guideline for the identification of agents of attribution ('source') and the predicate type of attribution ('type'). Based on this groundwork, the current research explores attribution patterns in Turkish texts and

their distribution to different genres in the multi-genre Turkish Discourse Bank version 1.2 (TDB 1.2), an annotated discourse resource of Turkish (Zeyrek & Er 2022).

Thus, this work aims to enrich our understanding of attribution in the discourse of Turkish and enrich the discourse relation annotations that exist in Turkish Discourse Bank. This research also aims to pave the way for advanced text analysis applications that do not need to involve manual attribution annotation in Turkish.

The research questions addressed in this work include:

1. What are the patterns of placement of attribution features across journalistic and non-journalistic genres within the TDB?

2. What are the linguistic cues signifying the 'source' and 'type' features of attribution across journalistic and non-journalistic genres in TDB?

1.3. Significance of the Thesis

This research aims to contribute to the field by addressing the topic of attribution within the Turkish context. Based on the PDTB manual, it seeks to extend the TDB by annotating the unmarked attribution spans within it. The goal is to open ways for automated text applications using the enriched Turkish attribution annotation manual and the annotated dataset.

1.4. Scope of the Thesis

This thesis provides a new set of annotations over the annotations existing in the TDB 1.2 As it will be discussed later in the thesis, the TDB 1.2 is a corpus of discourse relations, annotating discourse connectives such as *ama* ('but')*, fakat* ('however')*, ayrıca* ('also') and the text spans that the connectives link. It annotates both explicit connectives (1) and implicit connectives (2), i.e. the inferred connective that relates two text spans. Argument 1 of discourse relations is indicated in italics, and Argument 2 is in bold.

(1)
Dışa karşı güçlüydü, <u>ama</u> **içe, kendi yüreğine yıkılmak üzereydi.**

Outwardly, s/he were strong, <u>but</u> **inwardly, s/he on the verge of collapsing onto their own heart**.

(2)
Koca bir duvar taşıyordun yüreğinde kimsenin aşamayacağı, aşmaya cesaret bile edemeyeceği. **Dışa karşı güçlüydü, ama içe, kendi yüreğine yıkılmak üzereydi**. (implicit: *öyle ki* 'such that')

You were carrying a great wall in your heart, one that nobody could overcome, not even dare to attempt. **Outwardly, you were strong, but inwardly, you were on the verge of collapsing onto your own heart**.

In some cases, discourse relations annotated in the TDB 1.2 involve attributed material. For example:

(3)
Nakışçıya girerken, "Hiç olmazsa ben sokaklarda gezinebiliyorum," [diye mırıldandı]. **Yaşlı adamı göremedi**.

As she/he entered the embroiderer's shop, "At least I can wander the streets," [s/he muttered], **She/he couldn't see the old man**.

The present thesis deals with examples like (3), where attribution is annotated over discourse relations. It does not deal with statements that include attribution over text segments not annotated for discourse relations. Thus, an example like (4) below is not annotated within the scope of this thesis, since it lacs discourse relations:

(4)
"Biliyorsun, eskiden beri ben de beğenirdim seni," dedi Fevzi.

"You know, I've liked you for a long time," said Fevzi.

1.5. Organization of the Thesis

To address the research objectives and questions, this thesis is organized as follows: Chapter 1, titled Introduction, provides an overview of the thesis's context, the concept of attribution, and its significance. It also outlines the structure of the thesis. Chapter 2, the Literature Review, examines the existing literature on discourse, attribution, and previous research related to computational prediction of attribution. This chapter lays the foundation for the thesis and identifies gaps in the literature. Next, Chapter 3, Methodology, describes the research methodology employed, including data collection from the TDB annotation procedures using the PTDB Manual. Chapter 4, An Evaluation and Finalization of the Dataset, presents the process of verifying annotator consistency using Cohen's kappa, refining the dataset, and converting it to Excel to ensure the data's accuracy and readiness for analysis. Chapter 5, Discussion of Quantitative Results, analyzes the annotated data to identify attribution patterns and their frequencies across various genres in the TDB. Chapter 6, Conclusion, summarizes the main findings and contributions of the study, addresses its limitations, and proposes future research direction.

CHAPTER 2

LITERATURE REVIEW

In this chapter, a review of the existing literature on attribution is provided, with a focus on studies that have examined attribution patterns and mechanisms in various languages. The review includes research conducted in the field of discourse, computational linguistics, psychology and natural language processing. The literature review contextualizes the current thesis within the broader discourse and highlights the gaps in the field that the present research aims to address.

2.1. What is Attribution?

Attribution denotes the connection between beliefs, opinions, statements within a text, and their origins, specifically the entities responsible for expressing them (Ruppenhofer et al., 2008). In the broad context of information propagation, attribution is more than merely identifying authorship within textual contexts. It establishes a link between linguistic expressions and specific individuals or groups, thereby connecting cognitive processes and linguistic mechanisms.

The importance of attribution goes beyond its complexity. According to scholars such as Riloff and Wiebe (2003), Wiebe et al. (2004), and Wiebe et al. (2005), attributing beliefs and assertions to the corresponding agents and identifying subjective language is a crucial task. Such attribution enables the differentiation of factual statements from speculative ones. It helps in recognizing the distinction between subjective information and objective facts and understanding the impact of perspective on private states, including personal thoughts and emotions (Quirk et al., 1985; Wiebe et al., 2005). These private states present the role of attribution in the accurate assessment of information.

Within discourse structure, attribution establishes an ownership between abstract objects (AOs) and agents (Prasad et al., 2007, p. 40). AOs may consist of propositions, events, or states, but the attribution relation (AR) might not align precisely with a single AO. Instead, AR can involve elements that span different AOs or are summarized within a single AO (see Section 2.3 for PDTB's annotation approach to attribution).

Different scholarly perspectives enrich the understanding of attribution. Carlson and Marcu (2001) emphasize the role of verbs and other linguistic cues as anchors for identifying attributions, although they acknowledge that verbs alone are not the sole indicators of ARs.

The PDTB approach to attribution offers a different perspective setting it apart from other frameworks in discourse structure. In the PDTB, attribution is uniquely treated, not as a discourse relation, but as a separate discourse event (Prasad et al., 2008). This means that the source of information (e.g., she said or he believes) is annotated independently from the content of the discourse.

This distinction leads to an understanding of attribution in the PDTB that enables capturing how discourse relations and their arguments can be associated with different individuals. To clarify the PDTB's approach to attribution, it is beneficial to examine the concepts of discourse semantics and sentence semantics.

Discourse Semantics and Sentence Semantics are two levels of analysis that contribute to the understanding of meaning within a text. Discourse Semantics is about the relationships that exist across sentences or clauses within a broader discourse, exploring how different textual elements interconnect to create coherence across the entire text. It aims to identify the underlying structure that links various parts of the discourse and to form a cohesive whole. On the other hand, Sentence Semantics concentrates on the internal structure of individual sentences and examines how various components within a single sentence relate to one another. While Discourse Semantics provides a macro view of textual relationships, Sentence Semantics offers a more microscopic perspective, focusing on the intricate relations within each sentence. Together, these two layers of semantic analysis provide a comprehensive understanding of how meaning is constructed at both the sentence and discourse levels.

Consider the following examples from the PDTB (Joshi, 2010):

(1) *There have been no orders for the Cray-3 so far, though the company says it is talking with several prospects.*
- Discourse semantics: There's a contrary-to-expectation relation between "there being no orders for the Cray-3" and "there being a possibility of some prospects" anchored by the connective "though". This refers to the overall relationship between these two ideas within the discourse.
- Sentence semantics: A contrary-to-expectation relation anchored by "though" exists between "there being no orders for the Cray-3" and "the company saying something". This illustrates how the two parts of the sentence are connected.
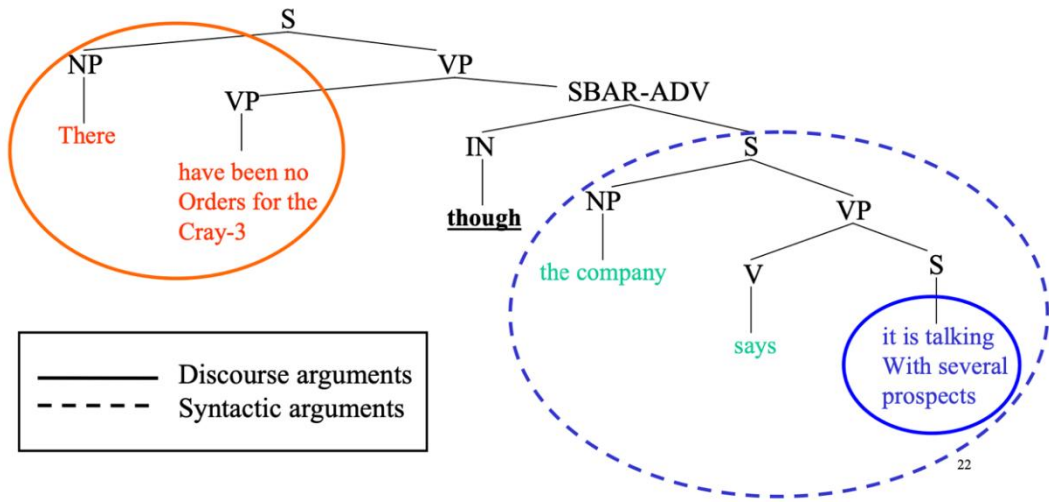
Figure 2. 1. Illustration of discourse and syntactic arguments (Joshi, 2010)

(2)   *Although takeover experts said they doubted Mr. Steinberg will make a bid by himself, the application by his Reliance Group Holdings Inc. could signal his interest in helping revive a failed labor-management bid.*

- Discourse semantics: There is a contrary-to-expectation relation between "Mr. Steinberg not making a bid by himself" and "the RGH application signaling his bidding interest". This refers to the broader relationship between these ideas in the context of the discourse.
- Sentence semantics: A contrary-to-expectation relation exists between "experts saying something" and "the RGH application signaling Mr. Steinberg's bidding interest". This refers to the relationship between these ideas within the same sentence.

-
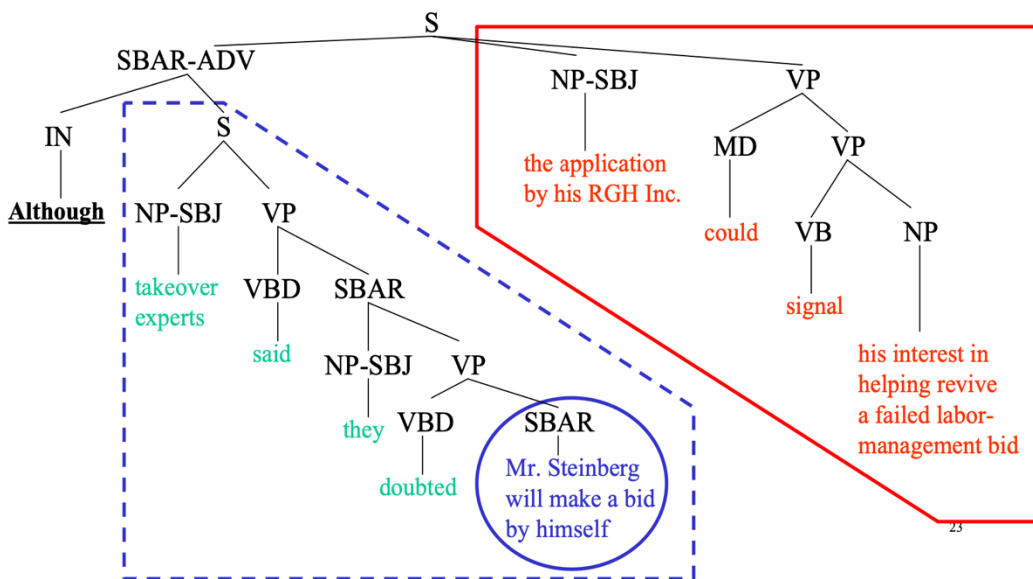
Figure 2. 2. Illustration of discourse and syntactic arguments (Joshi, 2010)

As seen in the figures above, obtained from the presentation *Dependencies at the Sentence Level and at the Discourse Level* at the Clara Workshop (Joshi, 2010), trees drawn at the syntax level differ from those at the discourse level due to attribution. Additionally, other factors such as coherence relations, topic shifts, information structure, and pragmatic context also contribute to the distinctions between sentence-level syntax and discourse-level analysis.

The PDTB's differentiation between attribution and discourse relations leads us to a more detailed analysis that captures the complexity of how statements, beliefs, or events are linked to specific sources or agents.

Mann and Thompson's Rhetorical Structure Theory (RST), introduced in 1988, explains how parts of a text are connected, with the nucleus holding the main idea and the satellite providing supporting details (Mann & Thompson, 1988). In RST, attribution occurs when a text segment attributes a statement or belief to a source, typically through the satellite, while the nucleus presents the main content. This concept will be discussed in more detail (see 2.2.2. RST Discourse Treebank).

These different perspectives on attribution highlight the multifaceted nature of the concept and its significance within discourse analysis. A comprehensive understanding of how attribution functions in discourse and its implications for communication can be gained by exploring the diverse definitions and approaches.

In summary, the main differences in attribution annotation between the PDTB and the Rhetorical Structure Theory Discourse Treebank (RST-DT) stem from their segmentation approaches and operational goals. RST-DT uses a hierarchical, text-span oriented approach emphasizing writer's intentions and the text's rhetorical structure. The PDTB focuses on adjacent sentences or within-sentence relations and prioritizes syntactic cues and explicit connectives to identify attribution. Discrepancies arise in their annotation strategies due to these foundational differences, with implicit relations showing particularly low agreement between frameworks (Rutherford et al., 2017; Scholman, 2019)

Insufficient attention has been given to studies on attribution in Turkish, including its structure, patterns, and related issues. The limited number of studies that touch upon the subject of attribution in Turkish will be discussed in detail (see Section 2.4. Attribution in Turkish).

2.2. Annotation of Attribution in Discourse Corpora

2.2.1. Penn Discourse TreeBank

The PDTB is a large-scale annotated resource focused on discourse relations across the 1-million-word Wall Street Journal (WSJ) Corpus (Prasad et al., 2008). Building on earlier work by Webber and Joshi (1998), which considered discourse connectives as predicates operating at the discourse level with inputs like events, states, and propositions (Asher, 1993). The PDTB offers a framework for identifying explicit

relations termed as discourse connectives, as well as implicit discourse relations intended by the writer for the reader's inference. This reveals how sentences or larger discourse segments interlink through relationships such as elaboration, contrast, and cause-effect.

2.2.1.1. Important notions of the PDTB annotation style

The PDTB annotation style distinguishes between several types of discourse relation classes and top-level senses, as summarized in Tables 2.1 and 2.2. These classifications serve as a cornerstone for understanding the complexity of discourse analysis within the PDTB framework.

Table 2. 1. Discourse relation types

Relation Class	Description	Example
Explicit DR	Relations signaled through lexico-syntactic elements.	The city's Campaign Finance Board has refused to pay Mr. Dinkins $95,142 in matching funds because his campaign records are incomplete.
Implicit DR	Inferred relations that can be rephrased by an overt marker.	...the government should encourage home ownership, (Implicit = by means of) including issuing bonds that guarantee holders the right to purchase an apartment.
Alternative Lexicalization (AltLex)	When the insertion of an implicit connective is perceived redundant.	After trading at an average discount of more than 20% in late 1987 and part of last year, country funds currently trade at an average premium of 6%. (The Reason:) Share prices of many of these funds this year have climbed much more sharply than the foreign stocks they hold.
Entity Relations (EntRels)	Implicit relations formed based on entity-based coherence.	Pierre Vinken, 61 years old, will join the board as a non-executive director Nov. 29. Mr. Vinken is chairman of Elsevier N.V., the Dutch publishing group.
No Relations (NoRels)	No identifiable semantic relation between adjacent text spans.	Mr. Rapanelli met in August with U.S. Assistant Treasury Secretary David Mulford. Argentine negotiator Carlos Carballo was in Washington and New York this week to meet with banks.
Hypophora	Arg1 is a question and Arg2 provides a meaningful response.	Underclass youth are a special concern. Are such expenditures worthwhile, then? Yes, if targeted.

Table 2. 2. Top-level senses

Level-1 Sense	Description	Example
Temporal	Situations described in the arguments are related temporally.	After finishing her morning routine, Sarah noticed (Implicit = then) that her cat had knocked over a vase.
Comparison	One eventuality is compared to another in terms of similarity or difference.	The new model is efficient, but investing all resources in it might be risky.
Contingency	Situations described in Arg1 and Arg2 causally influence each other.	Even though James isn't the main candidate, he's been active in politics. Therefore, it wouldn't be appropriate for him to take that position.
Expansion	Expands the discourse, moving its narrative or exposition forward.	The theory that "vaccination prevents severe illness" is gaining traction. (Implicit = for instance) A recent study found a significant reduction in hospitalizations among the vaccinated.

These tables present the foundational elements of the PDTB's approach to discourse analysis, categorizing relations and the top-level senses that define the primary nature of these relations.

Moving from different types of relations and main categories, the PDTB organizes these into a hierarchy similar to Asher's (1993) Hierarchy of Abstract Objects, shown in Figure 2.3. This presents a structured approach to understanding discourse semantics.
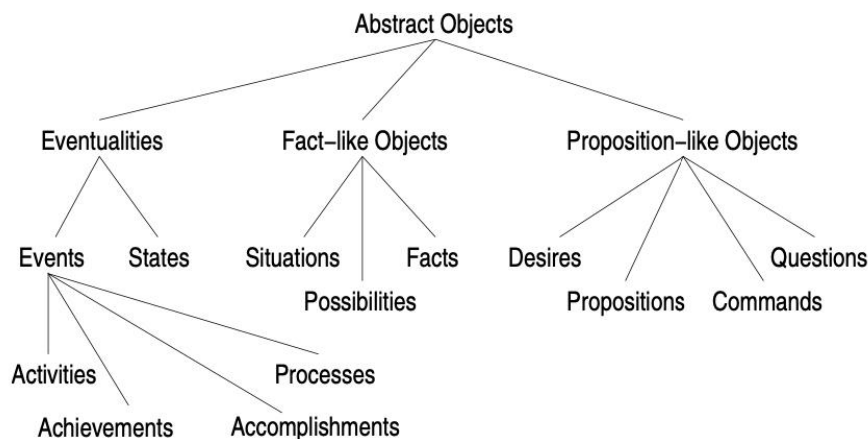
Figure 2. 3. Hierarchy of Abstract Objects (Asher, 1993)

10

This hierarchy organizes the top-level senses into a systematic structure, where each sense can include a range of more specific sub-relations, as shown in Figure 2.4 below. The way the PDTB organizes its top-level senses into a hierarchy clearly explains each category's details. Temporal relations are divided into asynchronicity and synchronicity, with further distinctions such as precedence and succession. The Comparison category includes Contrast, with subcategories like juxtaposition and opposition, and Concession, which further delineates expectations and counter-expectations. Contingency encompasses Cause, with reasons and results, and various forms of Conditions, ranging from hypothetical to factual scenarios. Finally, Expansion is categorized into Conjunction, Instantiation, Restatement, and other forms such as Alternatives and Exceptions, each with their own subdivisions that articulate the specific nature of the discourse expansion. Therefore, this snapshot (Figure 2.4) serves as a clear visualization of the intricate hierarchy and presents how the PDTB captures the complexity of discourse relations.

However, as emphasized by Biran & McKeown (2015), in the PDTB, while specific discourse relationships like cause-effect or contrast are organized into a detailed hierarchy, the overall structure of a document does not form a rigid, tree-like hierarchy, unlike frameworks such as RST. This means that in the PDTB, the connections between different parts of the text are categorized but do not always link up into a single, overarching tree structure. Instead, the text structure in the PDTB is more like a network with various connections that are not strictly parent-child in nature and that reflects the layered way discourse often functions in real texts.

Another important notion regarding the PDTB annotation style lies in its protocol for argument naming. A discourse connective, which acts as a discourse-level predicate, has two primary elements to consider: the adjunct and the matrix clause.

The adjunct is a clause or phrase that provides additional information or clarification about the main idea presented in the discourse. In the PDTB's established convention, this adjunct is labeled as Arg2. It is defined as the material being syntactically attached to the connective. On the other hand, the matrix clause, which the adjunct is modifying or supplementing, stands as the primary proposition or statement in the discourse. It is labeled as Arg1 within the PDTB framework. To clarify the notions, an example (3) is provided below.

(3) *The city's Campaign Finance Board has refused to pay Mr. Dinkins $95,142 in matching funds* <u>because</u> **his campaign records are incomplete**.
In this example, "because his campaign records are incomplete" serves as the adjunct (Arg2) and provides further details about why the Campaign Finance Board refused to pay Mr. Dinkins. In the example (3) and throughout this thesis, Arg1 is rendered in italics, Arg2 in bold fonts. The discourse connective is underlined.

Level-1	Level-2	Level-3
TEMPORAL	SYNCHRONOUS	–
	ASYNCHRONOUS	PRECEDENCE
		SUCCESSION
CONTINGENCY	CAUSE	REASON
		RESULT
		NEGRESULT
	CAUSE+BELIEF	REASON+BELIEF
		RESULT+BELIEF
	CAUSE+SPEECHACT	REASON+SPEECHACT
		RESULT+SPEECHACT
	CONDITION	ARG1-AS-COND
		ARG2-AS-COND
	CONDITION+SPEECHACT	–
	NEGATIVE-CONDITION	ARG1-AS-NEGCOND
		ARG2-AS-NEGCOND
	NEGATIVE-CONDITION+SPEECHACT	–
	PURPOSE	ARG1-AS-GOAL
		ARG2-AS-GOAL
COMPARISON	CONCESSION	ARG1-AS-DENIER
		ARG2-AS-DENIER
	CONCESSION+SPEECHACT	ARG2-AS-DENIER+SPEECHACT
	CONTRAST	–
	SIMILARITY	–
EXPANSION	CONJUNCTION	–
	DISJUNCTION	–
	EQUIVALENCE	–
	EXCEPTION	ARG1-AS-EXCPT
		ARG2-AS-EXCPT
	INSTANTIATION	ARG1-AS-INSTANCE
		ARG2-AS-INSTANCE
	LEVEL-OF-DETAIL	ARG1-AS-DETAIL
		ARG2-AS-DETAIL
	MANNER	ARG1-AS-MANNER
		ARG2-AS-MANNER
	SUBSTITUTION	ARG1-AS-SUBST
		ARG2-AS-SUBST

Figure 2. 4. PDTB-3 Sense Hierarchy (Webber et al., 2019)

2.2.1.2. Attribution annotation in the PDTB

In the PDTB, attribution is not marked as a separate discourse relation; it's annotated only when overlapping with other discourse relations. The PDTB manual provides extensive guidelines to assist annotators in capturing attribution, such as source identification, attribution type, scopal polarity, determinancy (elaborated in Table 2.6 PDTB's Attribution Annotation Scheme). It's important to note that while the source

12

attribute is annotated in the PDTB, other aspects like type, scopal polarity, and determinacy are not explicitly annotated. Still, the manual offers detailed guidance and enables annotators to capture attribution within the discourse.

In the PDTB, attribution annotation involves indicating the ownership of an expression to its source, which can be attributed to the writer, someone explicitly mentioned in the text, or unspecified individuals referred to in the text. Sources typically comprise individuals or groups such as agents, reports, letters, articles, viewpoints, and the like. Source types and examples from the PDTB are presented in the table below (Table 2.3).

Table 2. 3. Source types and examples from the PDTB

Attribution Source	Source Type	Examples from the PDTB
Writer	Individuals or Groups, e.g., authors, agents	Since **the British auto maker became a takeover target last month**, *its ADRs have jumped about 78%*. (0048)
Other	Individuals or Groups specifically mentioned in the text	"*The public is buying the market* <u>when</u> in reality **there is plenty of grain to be shipped,**" [said Bill Biedermann, Allendale Inc. director]. (0192)
Arbitrary	Nonspecific Entities alluded to in the text. These can refer to unnamed individuals or vague sources, often expressed through adverbs like 'reportedly' or passive constructions like 'it has been stated.'	*East Germans rallied* <u>as</u> **officials** [reportedly] sought **Honecker's ouster**. (2278)
Inherited	Arguments inheriting source value from the relation	*Factory orders and construction outlays were largely flat in December* <u>while</u> [purchasing agents said] **manufacturing shrank further in October**. (0178) (The source value of Arg1 is inherited from the relation)

Attribution links an abstract object to one or more source entities. These abstract objects correspond to those serving as arguments to discourse relations and are classified into four sub-types: assertion propositions, belief propositions, facts, and eventualities leading to different inferences about the degree of factuality of the AO.

Assertion Proposition AOs. AOs categorized as assertion propositions are associated with a communication-type attribution, abbreviated as Comm. This attribution is manifested through the use of conventional verbs of communication, such as *say*, *mention*, *claim*, *argue* and *explain*. Levin (1993) extensively discussed this classification and the specific verbs employed in expressing assertion propositions.

Belief Proposition AOs. Belief proposition abstract objects are connected to a type of ascription known as the belief category. This ascription is conveyed through the utilization of propositional attitude verbs, including *believe*, *think*, *expect*, *suppose*, and *imagine*. Scholars such as Hintikka (1969) have extensively studied and referred to this manner of attribution as PAtt for brevity.

*Fact AOs.*Fact AOs are distinguished by attributing an assessment or awareness of a proposition to an agent, presuming the truthfulness of the proposition without questioning it. This form of attribution is recognized as a Factive attribution, often denoted as Ftv. It is commonly conveyed through the use of factive and semi-factive verbs. Notable works by Kiparsky and Kiparsky (1971) as well as Karttunen (1971) have examined and provided insight into these verbs, which encompass terms like *regret*, *forget*, *remember*, *know*, *see*, *hear* among others.

Eventuality AOs. The attribution of an event or action to an agent signifies the intention or attitude of the agent towards it. These attributions, referred to as eventuality AOs, are typically associated with control verbs, denoted as Ctrl. Scholars such as Sag and Pollard (1991) have proposed categorizing these control verbs into three distinct categories based on their characteristics. The first category includes verbs of influence, such as *persuade*, *permit* and *order*, involving one agent influencing another to perform or refrain from an action. The second category is centered around commitment verbs such as *promise*, *agree*, *try*, *intend*, *refuse*, and *decline*, where an agent commits to carrying out (or not carrying out) an action. The third category consists of verbs of orientation, including *want*, *expect*, and *wish*, which reflect an individual's desires, expectations, or similar mental dispositions towards specific states of affairs. These distinctions aid in analyzing and comprehending how agents attribute control in diverse contexts.

Abstract object categories and the typical (English) verbs that convey them are presented below (Table 2.4).

Table 2. 4. Abstract Object Categories

Abstract Object Category	Types	Examples
Assertion Propositions	Propositions conveying assertions	"Judge O'Kicki <u>says</u> that..."
Belief Propositions	Propositions representing beliefs	"The researcher <u>believes</u> that..."
Facts	Assessment of an agent regarding or understanding of a statement assumed to be true	"Turkish people <u>know</u> that..."
Eventualities	Intention or attitude of an agent towards a contemplated event, condition, or action	"Mr. Craven had <u>persuaded</u> them to…"

Attribution is not bound strictly by conventional paradigms. Rather, it can adopt various structural forms to convey its intent. Table 2.5 illustrates additional cues of attribution, demonstrating the range of structures that can be employed to attribute information in English.

Table 2. 5. Other Attribution Components

Attribution Component	Examples
Adverbs	extensively, reportedly, allegedly, supposedly
Appositive NP	Jane, a respected expert in the field, …
Prepositional Phrases	according to X

All in all, based on the information provided in the PDTB manual regarding the attribution annotation scheme, Table 2.6 that represents the key properties of attributions and their features are presented below.

Table 2. 6. PDTB's Attribution Annotation Scheme

Feature	Description
Source	Distinguishes between different types of agents:
	- Wr (Writer)
	- Ot (Other agent)
	- Arb (Arbitrary individual(s) indicated via non-specific reference)
	- Inh (source value inherited from the relation)
Type	Signifies the nature of the relationship between agents and AOs, reflecting their factuality:

Feature	Description
	- Comm (Assertion proposition AOs, conveyed by standard verbs of communication)
	- PAtt (Belief proposition AOs, conveyed by propositional attitude verbs)
	- Ftv (Fact AOs, conveyed by factive and semi-factive verbs)
	- Ctrl (Eventuality AOs, conveyed by control verbs)
Scopal Polarity	Indicates cases where verbs of attribution are negated on the surface, but the negation (Neg) reverses the polarity of the attributed relation or argument content. For example, consider this sentence: "*Having the dividend increases is a supportive element in the market outlook*, <u>but</u> [I don't think] **it's a main consideration**," he says. (0090) In this case, the polarity for the entire relation and Arg1 is considered Null, which refers to a neutral state where no clear positive or negative polarity is assigned to the relation or argument., while Arg2 is specifically annotated as Neg.
Determinacy	Indicates whether an attribution over a relation or argument can be canceled in specific contexts. Indeterminacy is marked by "Indet," while determinate contexts use "Null." For instance, in the sentence, "It is a silly libel on our teachers [to think] *they would educate our children better* <u>if only</u> **they got a few thousand dollars a year more**," the idea that teachers would improve with higher pay is not attributed to anyone, even arbitrarily (Arb). Instead, it is a conjecture, arising from the infinitival context ("would educate"), indicating a potential rather than a definite claim. This distinction clarifies the strength and certainty of attributions.
Attribution Span	Annotates the text span associated with the attribution. Includes all non-clausal modifiers of the elements contained in the span (e.g., adverbs and appositive NPs). Connectives may be excluded.

The following examples illustrate different aspects of attribution annotation in the PDTB:

(4) [Mr. Marcus believes] *spot steel prices will continue to fall through* early *1990* **and** <u>then</u> **reverse themselves.** (Prasad et al., 2008)

	REL	Arg1	Arg2
[Source]	Ot	Inh	Inh
[Type]	PAtt	Null	Null

The example (4) illustrates how attribution is annotated in the PDTB by defining the source and type features of attribution. In the manual (Prasad et al., 2008), "Mr. Marcus believes" is squared to demonstrate the attribution, and since Mr. Marcus is the specified agent of the belief, it is marked as Ot. Both arguments are marked as

Inh (inherited). It indicates that the source is inferred from the relation (REL). The type feature characterizes the connection between an agent and an AO and it is labeled as PAtt because the verb *believe* falls into the category of belief proposition AOs. Since there's no separate attribution for Arg1 or Arg2, like in Example (4), the type value for these arguments are marked as Null.

(5) Eward and Whittington had planned to leave the bank earlier, but [Mr. Craven had persuaded them] *to remain* until **the bank was in a healthy position.** (Prasad et al., 2008)

	REL	Arg1	Arg2
[Source]	Ot	Inh	Inh
[Type]	Ctrl	Null	Null

Example (5) illustrates a control type of attribution, where an implied subject in one clause (often an infinitival clause) is determined or controlled by an explicit subject or object in a related main clause, without the need for it to be overtly stated in the controlled clause (Chomsky, 1993).

The phrase "Mr. Craven had persuaded them" is highlighted to represent the attribution. The source stands for the individuals who hold certain intentions or attitudes. Since the source is explicitly stated in the text, it is labeled as Ot in the example. The verb *persuade* falls into the eventuality AOs category, expressing the agent's intention or attitude toward an event or state. Consequently, the relation's type is marked as Ctrl. Both Arg1 and Arg2 are inherited, meaning that their source values are derived from the relation. Their types are marked as Null because there is no separate occurrence of attribution on these arguments.

(6) "*Having the dividend increases is a supportive element in the market outlook,* but [I don't think] **it's a main consideration***,"* he says. (Prasad et al., 2008)

	REL	Arg1	Arg2
[Source]	Ot	Inh	Ot
[Type]	Comm	Null	PAtt

Example (3) illustrates a communication type of attribution that refers to an agent conveying or transmitting a specific message or piece of information. The phrase "I don't think" is written in square brackets to present the span of attribution that includes the source and type. However, this attribution relation specifically applies to Arg2 of the overall relation. The entire sentence is attributed to a particular source, characterized by the communicative type of attribution verb *say*. Thus, the relation is marked as Ot for the source and Comm for the type. In contrast, Arg2 has its own separate attribution, marked as Ot and PAtt, since the verb *think* falls under the category of belief proposition AOs.

2.2.1.3. Advancements in Discourse Analysis: From PDTB 2.0 to 3.0

The upgrade from PDTB 2.0 to PDTB 3.0 represents an important enhancement in the field of computational linguistics in terms of three main areas: dataset expansion, annotation methodology, and category refinement. With approximately 13,000 additional tokens, the new version offers a larger and more varied dataset. Advanced annotation methods of the PDTB 3.0 improve the precision and detail of data interpretation. They allow for more accurate annotations and standardized comparisons across different textual elements. These enhancements collectively aim to make PDTB 3.0 a more comprehensive tool for text analysis and improve the capacity for detailed discourse analysis in academic research. The comparison of features between the PDTB 2.0 and the PDTB 3.0 is detailed in the table below (2.7).

Table 2. 7. A Comparison of Features between the PDTB 2.0 and the PDTB 3.0

Feature	PDTB 2.0	PDTB 3.0
Dataset	Smaller, less varied	Larger, more varied
Annotation Methods	Basic	Advanced
Categories	Limited	Expanded
Aim	Basic understanding	More accurate and comprehensive analysis
Tool Improvement	Limited	Enhanced
Guidelines	Basic	Detailed
Impact	Limited usability	Improved accessibility and usefulness
Reflects	Partial complexity	Ongoing effort towards precise representation
Advancement	Incremental	Substantial
Enables	Basic studies	Detailed and sophisticated studies

2.2.2. RST Discourse Treebank

Rhetorical Structure Theory is a foundational linguistic framework that enables analysis of textual composition through the segmentation and hierarchical ordering of consecutive clauses (Mann & Thompson, 1988). This mechanism, which identifies specific semantic relations such as Elaboration, Evidence, Solutionhood, and Background, among others (Den Ouden, 1998), finds practical application in various linguistic fields, including attribution.

The RST Discourse Treebank, a collection compiled by Carlson and Marcu (2001), represents a rich resource in this context. It consists of 385 news articles drawn from the Wall Street Journal section of the Penn Treebank (Marcus et al., 1993).

RST provides a structure for linguistic analysis, particularly in the delineation and structured arrangement of successive clauses (Mann & Thompson, 1988). As mentioned briefly earlier (see 2.1 What is Attribution?), fundamental to RST is the differentiation between the nucleus and the satellite. The nucleus can be seen as the primary or central component of a text segment that carries essential significance, whereas the satellite provides additional information or support to the nucleus (Mann & Thompson, 1988). For instance, consider the sentence:

(7)   *Pollilution is harmful*, <u>as</u> **it can lead to various respiratory diseases**.

Here, "Pollution is harmful" is the nucleus since it presents the main claim, while "as it can lead to various respiratory diseases" serves as the satellite, supporting the claim.

As for attribution, this analytical model delineates structures where the attribution predicate is designated as the satellite, and the content being attributed, or the reported message, serves as the nucleus (Potter, 2019). This framework highlights how elements of attribution are systematically organized within texts. To elucidate this structural arrangement, consider the following passage:

(8)
- a. Senator Chris Coons, the Delaware Democrat, told me
- b. that his longtime colleague [Senator Lindsey Graham] is "hysterically funny" and "personally engaging."

In this structure, the initial part (a) acts as the satellite, containing the attribution predicate, whereas the second part (b) takes on the role of the nucleus, holding the content or message being attributed. This setup highlights the fundamental difference between the attributing agent and the attributed content within the framework of attribution as posited by Carlson and Marcu (2001) (Potter, 2019).

The methodology of Carlson and Marcu (2001) is demonstrated in examples (1a) and (1b) drawn from Marcu's Wall Street Journal Corpus. The direction of the arrow indicates the nucleus:

(1) a. wsj_1377

19

b. wsj_1157

1-2

Attribution

The shares represented 66% of his Dun & Bradstreet holdings,

according to the company.
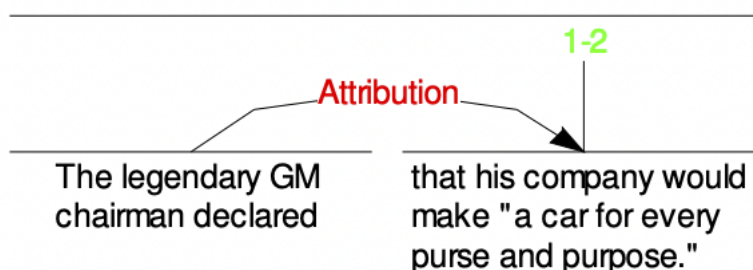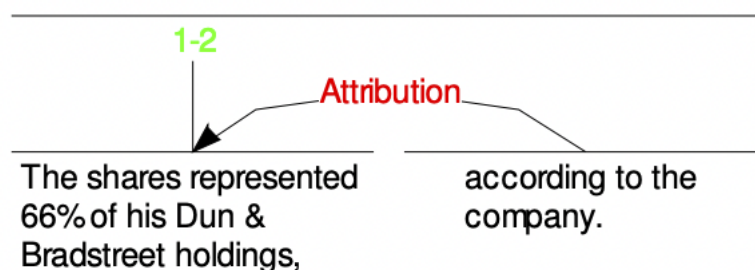
However, the classic RST does not distinguish between the complements of verbs and parenthetical speech reporting clauses from their corresponding host clause. Therefore, traditional RST presents difficulties in integrating speech report attributions. It often misses the subtle differences between verb complements and clauses that report speech parenthetically in relation to their main clause. RST marks attribution the same way it marks, for example, the cause relation.

Redeker & Egg (2006) provide a critique of the RST's attribution annotation. They argue that the attribution relation is defined not only with speech verbs but also includes cognitive predicates. The reasoning behind this approach is clear: *verba cogitandi* ('cognitive verbs') such as *to believe* or *to expect*, are frequently utilized (in the guise of grammatical metaphor) in clauses reporting speech. However, the attribution relation is equally applied to speech verb uses and mental verb uses. Additionally, all these segments that are attributed are given nuclear status. As a result, there is no dominant node available to prevent incorrect interpretations in situations where the context is not clear.

Pareti (2015) also criticizes the RST's approach to attribution and argues that attribution is marked only at the intra-sentential level. Attribution is primarily restricted to attributing verbs and the phrase *according to*, while excluding other cues. On the other hand, prepositional phrases such as *in the words of* or *as described by* can signify attribution. Also, control verbs such as *promise* or *convince* inherently require attribution, as they express intentions, expectations, and relational dynamics between interacting parties. Adverbials such as *reportedly* or *allegedly* also possess attributional importance, as they add an element of uncertainty to the information they accompany. Moreover, passive constructions and references to published works, like *in a recent study*, also enhance the range of potential cues for attribution. Therefore, this narrow focus of RST's approach might cause misunderstandings in research related to attribution, as researchers could mistakenly view these limited indicators as representing the full scope of attribution.

The complexities of attribution are more apparent in the range of scholarly discussions. Skadhauge and Hardt (2005) suggested that attribution relations could be identified purely through syntax, and argued against the necessity of specific annotations within the RST corpus. This viewpoint supports understanding

attribution through linguistic and structural methods. On the other hand, For instance, Pardo and Nunes (2003) and Pardo et al. (2004) explore attribution through RST to understand discourse organization. Similarly, Afantenos et al. (2012) highlight the interpretive role of attribution within RST.

From a structural perspective, Stede et al. (2017) argue that attribution also displays complexities. They assert that, syntactically, the so-called "reporting unit" is not an autonomous discourse entity. Instead, it typically manifests itself as a noun clause, exemplified by "that a representative's duty was…" in (9) below. This clause is not merely an adjunct; it is, in fact, a syntactic complement of the verb *explained*.

(9)   Madison explained that a representative's duty was to speak not for the narrow interests of one group but instead for the common good (Stede et al., 2017).

To make the concept clearer, Stede et al. (2017) used the straightforward phrase "Madison explained it" as an example. Here, "it" acts similarly to a more complex noun clause in an earlier example and serves as the object of the verb "explained." They emphasize the syntactic and semantic connection between the reporting verb and its complement and argued against dividing such clauses during discourse analysis. This approach highlights the importance of maintaining the unity between reporting verbs and their complements to accurately understand their interplay in discourse.

Stede et al. (2017) also highlight the complexities involved in analyzing direct speech. Unlike indirect or reported speech, which offers a more straightforward representation of a statement, direct speech can integrates multiple elements in a manner that is not always linear or contiguous. Taking the provided example (10) as an illustration, several layers of embedded clauses and relationships are observed.

(10)   Katsumoto says to Nathan on the dawn of battle, "You think a man can change his destiny?" to which Cruise replies, "I believe a man does what he can, until his destiny is revealed" (Stede et al., 2017).

Firstly, there are the primary reporting verbs – *says* and *replies*. These verbs not only link the speaker and the spoken content but also interact to indicate a sequence within the dialogue.

Within the dialogue itself, the two directly quoted statements present additional layers of meaning and structure. But identifying the nature of the relationship between the reporting verbs (*says* and *replies*) and the content of the speech is challenging. While it's clear that the speech is attributed to the speakers (Katsumoto and Cruise) through these verbs, defining the relationship beyond this attributive connection presents a challenge. The speech does not merely convey information; it embodies emotions, rhetorical questions, beliefs, and responses, adding layers of depth that go beyond basic analysis.

These examples reveal the complex nature of attribution and the difficulties involved in segmenting complement clauses related to reporting verbs as well as in classifying direct speech with intertwined structures (Stede et al., 2017).

Further exploration of attribution within the RST framework was conducted by Potter (2019), who addressed the ongoing debates and complexities concerning attribution. Through methodical analysis, Potter (2019) identified three central challenges: the Discourse Units Issue, the Nuclearity Issue, and the Relation Identification Issue. These findings highlighted the diverse nature of attributional relations and emphasized that the classification and understanding of these relations depend heavily on the writer's intentions.

In summary, the analysis and understanding of attribution within the RST framework is a multifaceted and complex domain. While important advancements have been made, certain limitations and disagreements persist. The ongoing discussions and research on this subject highlight the dynamic and sophisticated nature of attribution in linguistic studies. These complexities not only contribute to our comprehension of language structures but also influence the methodology and interpretation of attribution in various rhetorical contexts.

2.2.3. Corpus Studies of Pareti

In their study titled *Annotating Attribution Relations: Towards an Italian Discourse Treebank*, Pareti and Prodanof (2010) describe the construction of an annotation schema for attribution relations in Italian discourse. Departing from previous approaches that incorporated attribution within broader discourse relations, Pareti advocates for treating attribution as an independent discourse relation. This method enables a comprehensive analysis of attribution. The proposed annotation schema is applied to a subset of the Italian ISST corpus and serves as an initial step towards the creation of an Italian Discourse Treebank. This research highlights the significance of attribution relations and their implications for the interpretation of attributed material.

Pareti's scholarly contributions extend beyond schema development to the establishment of valuable resources in the attribution domain. In her study titled *A Database of Attribution Relations*, Pareti (2012) addresses the lack of comprehensive resources available for attribution studies. To bridge this gap, she collects and annotates a database of attribution relations sourced from the PDTB. This study is designed to offer researchers a comprehensive resource that will contribute to creating strong systems for attribution extraction. Pareti highlights the importance of this resource in driving progress in the field of attribution research

Pareti's research also includes the annotation and analysis of ARs in spoken language contexts. Cervone, Pareti and their colleagues (2014) outline a methodology for annotating ARs in spoken informal dialogues. This research represents a shift from prior studies that primarily focused on attribution in formally written corpora. By exploring informal spoken language, Pareti broadens the scope of attribution

relations and addresses new challenges. Her theory suggests that combining prosodic features with linguistic indicators can offset the disjointed nature of speech and improve the analysis of attribution. The results from studying the Spoken Italian Attribution Relations Corpus (SARC) support this theory emphasize the role of prosodic cues in identifying attribution phenomena.

Pareti et al. (2013) distinguish three types of quotations. Direct quotations are fully enclosed in quotation marks and reproduce the original utterance verbatim. Indirect quotations paraphrase the original utterance and have no quotation marks. Mixed quotations contain both verbatim and paraphrase content and may thus contain quotation marks. Note that the type of a content span is assigned automatically based on its surface form using the definitions just given. Table 2.8 provides an overview of their annotation scheme.

Table 2. 8. The annotation scheme of Pareti and Prodanof (2010)

Component	Description
Source Span	Reference to the entity that is the subject of the attributed content.
Cue Span	The lexical element that serves as the anchor for the relation, expressing the source's attitude toward the attributed content.
Content Span	The attributed text itself.
Supplement Span	Additional element relevant to the interpretation of the AR, such as expressing information (e.g., recipient or circumstantial information).

Moreover, Pareti's work includes the development of automated techniques for extracting and attributing indirect and mixed quotations. In her study titled *Automatically Detecting and Attributing Indirect Quotations*, Pareti and her team (2013) conduct extensive experiments on the extraction and attribution of these types of quotations on a large scale. They propose two distinct methodologies for extracting various types of quotations from news articles and assess their effectiveness using annotated corpora. These two methods are explained below.

2.2.3.1. Token-Based Approach

The Token-Based Approach conceives of quotation extraction as analogous to Named Entity (NE) tagging, where tokens are individually labeled as either inside (I), beginning (B), or outside (O) a quotation. This approach processes the entire document as one continuous sequence and allows the inclusion of quotations that extend across sentences and paragraphs. A linear chain Conditional Random Field (CRF) was deployed as the learning algorithm, and each token in a sequence (such as a word in a sentence) is labeled based on both its own features and the labels of neighboring tokens, forming a linear sequence of dependencies. The features were incorporated to consider lexical information (including unigrams and bigrams),

sentence characteristics (such as the presence of quotation marks or pronouns), dependency relations, external knowledge, and specific syntactic properties, among others.

Evaluation involved using F-score metrics, where the model outperformed baselines and achieved higher F-scores in predicting direct, indirect, and mixed quotations. Specifically, the token approach demonstrated strict F-scores of 59% and 60% for indirect quotations and partial F-scores of 76% and 74% for mixed quotations in two different corpora.

2.2.3.2. Constituent-Based Approach

Constituent-Based Approach classifies whole phrase-structure nodes, instead of individual tokens, as either part of a quotation or not. If a quotation spans multiple constituents, all relevant subspans are labeled. A post-processing step then merges adjacent or overlapping predicted spans within a sentence. A maximum entropy classifier with $L1$ regularization was used, specifically trained on indirect and mixed quotations. Like the token-based method, this approach incorporated common features such as lexical, sentence, and dependency characteristics but also added constituent-specific features, including the length of the span, the label, number of descendants, and contextual features of the node.

The study investigated the extraction of direct, indirect, and mixed quotations from the PARC and SMHC corpora, comparing various approaches including the B(rule) method, a baseline utilizing a rule-based approach, identifies text between quotation marks that consists of at least three tokens, provided that the tokens are not all title-cased and exclude stopwords and proper nouns.

The Token-Based Approach for quotation extraction, evaluated on both the PARC and SMHC corpora, achieved notable results. Using precision, recall, and F-score metrics for both strict and partial matching, the approach demonstrated strong performance in extracting indirect quotations. In the PARC corpus, the approach achieved an F-score of 70% for indirect quotations in strict matching and 76% in partial matching, while in the SMHC corpus, it attained an F-score of 78% in strict matching and 79% in partial matching. These results highlight the effectiveness of the Token-Based Approach in accurately identifying and extracting indirect quotations from textual data, whereas the constituent-based method faced problems like class imbalance and labeling inaccuracies.

Overall, Silvia Pareti has significantly advanced the understanding, annotation, and extraction of attribution relations in discourse analysis. Her work spans multiple linguistic domains and addresses attribution challenges in different contexts.

2.2.4. Arabic Attribution Annotation

The pioneering study "Annotating Attribution Relations in Arabic" (Al-Saif et al., 2018) offers a comprehensive examination of the challenges of annotating attribution

in Modern Standard Arabic (MSA). Their motivation stemmed from the realization that while many systems, such as those for authorship identification, information retrieval, and opinion mining, rely on accurately identifying attributed arguments, this domain remains largely unexplored in Arabic texts.

Their methodology was grounded in the PDTB, which they adapted to the unique characteristics and requirements of MSA. The guidelines they proposed emphasized three core elements of attribution: the cue (lexical anchors that connect the source to the content), the source (the entity making the claim), and the content (the actual claim or statement). These guidelines were enriched with semantic features and practical conventions.

A significant outcome of this research was the development of a new annotation tool designed specifically for Arabic. This tool was important in ensuring that all instances of attribution in the corpus were annotated reliably. However, the research had some challenges. A primary concern was the absence of a comprehensive theory of attribution and the lack of a gold standard annotated corpus for Arabic.

Al-Saif et al. (2018) evaluated their corpus through a pilot annotation study involving two annotators who underwent training using the ESNAD annotation tool. They measured inter-annotator agreement using observed agreement, F-score, and Kappa coefficient, achieving high agreement for various attribution elements such as cue supplement, source, and content. Results showed 98% agreement on attribution cases, with discrepancies primarily in labeling source types due to default settings in the tool. Despite this, the study demonstrated the reliability of the annotation process and provided a valuable corpus for further linguistic and computational studies in Arabic.

The paper ended with a call to develop the first comprehensive gold standard attribution corpus for MSA news articles. Insights from their pilot annotation were important in leading to refinements in both the guidelines and the annotation tool. The goal was not only to create a corpus, but to ensure it serves as a foundation for further linguistic and computational studies in Arabic, especially in the context of news articles.

2.3 Attribution in Turkish

Attribution in the context of Turkish discourse and its computational analysis have been relatively overlooked in previous studies. However, in a recent study, Can and Ercan (2020) examined the transfer structures in Turkish op-ed (opinion editorial) articles based on Richardson's (2007, pp. 102-105) classification of transfer methods, which include direct/indirect, strategic, indirect transformational, and pseudo-direct quotation. The authors investigated various discourse functions in op-ed articles, that are accomplished through different forms of reported speech. They identified five categories that align with functions found in previous research (Lin, 1999; Kuo, 2001; Bloor and Bloor, 2007; Richardson, 2007). These functions include supporting of the writers' arguments to enhance their credibility, dramatization of events,

evasion of responsibility, establishment of an authoritative and traditional environment, description and disparagement of the reported individuals' characteristics, drawing attention to opposing opinions, providing background information, and describing and praising the characteristics of the quoted individuals. Notably, for the specific function of evading responsibility by distancing themselves from the source of information, the authors observed the linguistic utilization of the evidential marker –*mIş* (Can & Ercan, 2020).

Güven's (2023) recent study focused on the use of attribution in news texts. Güven (2023) examined the sources of opinions consulted in news texts, the methods employed to quote or indirectly convey these opinions, and the implications of such choices on the portrayal of various social groups. The main objective of the study was to identify reporting actions and attribute speech to specific individuals or institutions. The adoption of Fairclough's (1995) discourse representation forms was explored. The database includes more than 28 thousand words, and the observed frequency of many of the verbs are comparable to the present thesis. The study found that direct speech exhibits clear differentiation between the reporter's voice and the represented speaker, while indirect speech lacks explicit attribution, resulting in ambiguity. Additionally, unmarked reporting requires readers to determine the source of speech. The study suggests that the transmission of opinions is influenced by social structures and power dynamics, resulting in certain social groups, particularly those with limited access to power (e.g., women), being underrepresented and marginalized in media coverage (Caldas-Coulthard, 1994).

Furthermore, Kunduracı's (2008) study on Turkish control explains the eventuality component of attribution in the current work. According to her findings, lexical control begins with verbs that have lexically specified units and are accompanied by infinitive verb complements. However, not all verbs have the ability to exert control, and control is not universally permitted in all languages. In Turkish, lexical control is specifically restricted to the use of *-mEk* phrases as complements for control verbs. It is worth noting that the complement V-*mEk* must be in the verbal and infinite form, unlike the nominal and finite form of V-*mE*. The concept of lexical control demonstrates the interaction between syntax and semantics, where the meanings conveyed by control verbs contribute to the semantic aspect, while the required form of complements contributes to the syntactic aspect. Lexical control occurs when a control verb combines with a particular type of complement, specifically an infinitival phrase.

Kunduracı's approach to lexical control and eventuality within the Turkish linguistic framework was applied in this current thesis. We utilized her insights into control verbs and their syntactic and semantic interactions with infinitive complements in Turkish, specifically focusing on the use of -mEk phrases. This methodology allowed us to examine the structures of attribution in Turkish discourse and reveal distinct patterns and constraints specific to this language.

Building upon the insights and methodologies from the existing literature, including the specific approaches to eventuality in Turkish, this thesis developed an attribution

annotation scheme tailored to the specifics of Turkish discourse. This scheme is special in the field of attribution in Turkish because it allows a more specific analysis of how attribution operates within the discourse structure of this language. A detailed description of the annotation scheme, including guidelines on where and how to apply it, will be presented in the methodology chapter of this work (see 3.4. Annotation Style).

2.4. Attribution in Psychology

The synthesis of psychological theories of attribution with linguistic analysis provides a framework for understanding how language influences and reflects the cognitive processes involved in attributing causality, responsibility, and mental states.

The foundation of psychological attribution theory was laid by Heider (1958), who introduced the concept of internal (dispositional) versus external (situational) attributions, providing a lens through which individuals interpret actions and events. This framework was further developed by Jones and Davis (1965) and Kelley (1967), who elaborated on the processes through which people infer causes of behavior. These theories focus on the cognitive mechanisms individuals use to make send of their social world and suggest that attributions affect emotions, motivations, and behavior (Heider, 1958; Jones & Davis, 1965; Kelley, 1967).

In terms of linguistics, attribution refers to how language assigns or implies causality and responsibility, or how it expresses beliefs and intentions. Ruppenhofer et al. (2008) note the use of specific verbs, passive and active constructions, and direct or indirect speech to convey attribution. This linguistic encoding not only reflects the speaker's intentions but also influences the listener's interpretation, playing a key role in depicting the psychological states or the state of mind of individuals within a discourse.

In Turkish discourse, the linguistic mechanisms of attribution, ranging from verb choices to syntactic structures, reflect these cognitive processes. The act of attributing causality or intent in language, whether through direct statements or linguistic cues, mirrors the cognitive processes. So, examining how Turkish language encodes these attributions sheds light on the linguistic practices and also on the cognitive and cultural contexts.

Attribution theory, as conceptualized by Fritz Heider in the 1950s, explains how individuals assign explanations to behaviors and events. The theory distinguishes between internal (dispositional) and external (situational) attributions, in which individuals attribute behavior to either personal traits or environmental factors. This conceptualization has been further expanded by Bernard Weiner's Attribution Theory of Achievement Motivation, which emphasizes the impact of attributions on motivation and performance (Heider, 1958; Weiner, 1974).

Understanding the relationship between psychological theories of attribution and linguistic attribution has significant implications for various fields. In education, Weiner (1985) discusses how attributions of success and failure, influenced by linguistic framing, impact students' motivation and self-perception. In clinical psychology, the way individuals explain the causes of their experiences can impact their mental health. This concept, studied by Peterson et al. in 1982, indicates that depressive symptoms can correlate with a person's tendency to attribute negative events to unchangeable, personal, and pervasive causes. Additionally, exploring how different cultures use language in social interactions helps us understand these differences and provides valuable insights into improving intercultural empathy and understanding, according to Nisbett et al. in 2001. They suggest that linguistic structures across cultures reflect and shape these attributional patterns.

The relationship between linguistic structures and psychological attribution is explained by how language both shapes and mirrors our conceptualization of causality and agency. Fausey and Boroditsky (2010) provide evidence that subtle linguistic cues, such as the choice between active and passive voice, can significantly impact perceived blame and responsibility. This ties back psychological attribution theory by demonstrating how language can influence attribution processes.

The synthesis of psychological attribution theories with linguistic analysis deepens our comprehension of language as a mechanism for articulating and shaping cognitive and social dynamics. Investigating how choices in language express and influence perceptions of causality, responsibility, and mental states allows for a better understanding of language's effects on perception, emotion, and behavior. This mix of disciplines creates opportunities to explore how cognitive processes support language usage.

2.5. Attribution in Computational Linguistics

In recent years, there has been a growing interest in automating attribution using computational methods. Researchers have explored the potential of natural language processing (NLP) techniques and machine learning algorithms in identifying attributions in text.

Previous investigations carried out by Wiebe et al. (2005), Prasad et al. (2005), Riloff et al. (2005), and Stoyanov et al. (2005) have presented the importance of discerning and depicting the origin and factual essence of information within distinct NLP domains. These domains encompass sentiment analysis, information extraction, question answering, and text summarization. In addition, the work of Mamede and Chaleira (2004) and Elson and McKeown (2010) on narrative texts, along with Pouliquen, Steinberger, and Best (2007) and Sarmento, Nunes, and Oliveira (2009) on news texts, has contributed significantly by focusing on lexical terms and syntactic rules to infer the author of quoted text.

Furthermore, Pareti et al. (2013) and O'Keefe et al. (2012) made notable contributions to the field by developing machine learning classifiers that effectively

discerned between direct and indirect quotations through the utilization of advanced machine learning methodologies as mentioned earlier (see 2.2.3 Corpus Studies of Pareti). The studies by Elson and McKeown (2010), Fernandes, Motta, and Milidiú (2011), and O'Keefe et al. (2012) further improved the accuracy of attribution analysis by leveraging NLP approaches such as rule-based and statistical machine learning of syntactical structure features.

Moreover, Krestel et al. (2008) addressed the importance of reported speech, both direct and indirect, in traditional newspapers and modern media like blogs. It outlines the processing steps for reported speech analysis and presents a General Architecture for Text Engineering (GATE)-based implementation, consisting of a Reporting Verb Marker and Reported Speech Finder. The system achieves an 83% recall and 98% precision for reported speech extraction, with error analysis revealing challenges in handling complex circumstantial information and misguiding quotation marks. The components can be integrated into GATE pipelines for further application-specific processing, emphasizing the role of reported speech in conveying evidential scope and contributing to information attribution in textual content.

In the field of novels, He et al. (2013) tackled the challenge of speaker identification using a supervised machine learning approach that incorporates novel features like speaker alternation patterns and vocatives. Their method outperforms previous approaches by considering the sequential nature of utterances and incorporating syntactic and lexical clues. The authors evaluate their system on novels like *Pride and Prejudice*, *Emma*, and *The Steppe*, showing improved accuracy, especially when incorporating neighboring utterances. The results show improved accuracy, especially when incorporating neighboring utterances. Additionally, the system is extended to extract family relationships from the novels. They suggest future work involving advanced statistical models, automatic generation of character aliases, and joint approaches to multiple tasks for a deeper understanding of complex plots and stories.

Almeida et al. (2014) proposed a joint model for quotation attribution and coreference resolution to accurately assign quotes to speakers in text mining and media monitoring. Current systems often face inaccuracies and error propagation due to working at the mention level or assuming pre-detected coreferent mentions. The model combines entity-level quotation attribution and coreference resolution, leveraging correlations between the two tasks. Quotations are crucial in news stories for providing perspectives and objectivity, and efficiently extracting and organizing them is essential for various applications. The model formulates the problem as a logic program, considering paths in a quotation-coreference tree and combining features for mention-mention arcs, mention-quotation arcs, and paths between nodes. Experimental results demonstrate the effectiveness of the joint model in addressing challenges in entity-level quotation attribution and coreference resolution. The article concludes with an error analysis and outlines future work, including tackling indirect quotations and exploring connections to semantic role labeling.

Scheible et al. (2016) questioned the efficacy of linear-chain conditional random fields (CRFs) in quotation detection due to the Markov assumption and proposed novel models that significantly outperformed CRF models by considering global features. Two novel models, a boundary-based and a semi-Markov model, are proposed to address challenges in detecting quoted speech. The GREEDY model makes local decisions, while the SEMIMARKOV model, considering global features, significantly outperforms both CRF and GREEDY, highlighting the benefit of relaxing Markov assumptions. Implementations of all models are made public, emphasizing the importance of global features. The study, conducted on the Penn Attribution Relations Corpus, concludes that considering global features enhances quotation detection model performance.

Muzny et al. (2017) present a deterministic sieve-based system for quote attribution in literary texts, focusing on determining speakers of quotes. Their two-stage algorithm efficiently links quotes to mentions and mentions to entities, achieving an average F-score of 87.5% across three novels and surpassing previous systems. Precision tuning yields an impressive 90.4%. The paper details sieves used in the Quote→Mention and Mention→Speaker stages, employing techniques like Trigram Matching, Dependency Parses, and Coreference Disambiguation. Results demonstrate the system's efficacy, with an ablation test showing a high-precision system (95.6%). The conclusion highlights the need for expanding datasets and suggests future improvements like automatic character list extraction.

The paper of Newell et al. (2018) introduce an automated system designed to extract and analyze quotes from news articles, addressing the challenges faced by journalists in the evolving landscape of information dissemination. The system employs classifiers for verb-cues, quote content, and sources, facilitating the automated identification of key elements in news quotes. A two-stage process for quote attribution, involving content and source resolvers, enhances the accuracy of associating quotes with specific sources. The system's design incorporates metadata extraction, including a news document classifier and a semantic tagging system, providing additional context for each quote. User requirements, elicited through journalist feedback, emphasize the need for features such as quote highlighting, topic-specific searches, and comparative analysis. While the system demonstrates efficient processing and offers valuable insights into journalistic research, the authors propose future work to enhance performance and explore broader applications, such as anomaly detection and contradictions in news content.

Zhang & Liu (2022) introduce the DirectQuote corpus, a dataset containing 19,760 paragraphs and 10,279 manually annotated direct quotations from online news media. It aims to address challenges in quotation extraction and attribution in news articles, essential for fact-checking, media monitoring, and combating fake news. Direct quotations are highlighted for enhancing news credibility. The corpus, linked to Wikidata, serves various downstream tasks and proposes baseline methods like CRF, LSTM, CNN, and BERT for quotation extraction and attribution. The BERT model shows superiority. The construction involves diverse news collection, text processing, and manual annotation, validated for consistency. The article underscores

broader applications of direct quotations in natural language processing and suggests future research directions for dataset scaling and large-scale automatic systems.

Janicki et al. (2023) focus on recognizing and attributing quotes in Finnish news media for large-scale analysis. With a corpus of 1500 articles, it compares rule-based (dependency trees) and BERT machine learning methods. BERT excels, achieving 95% F-score for direct quotes and 84% for indirect ones, despite limited training data. Challenges include Finnish morphology, with BERT's flexibility highlighted. Evaluation shows a slight advantage in attribution accuracy for the rule-based model due to BERT's limitations. Future work aims to improve attribution, anticipating BERT's continued improvement. The discussion suggests enhancing entity detection, treating entity resolution separately, and recognizing nested entity complexity. The conclusion highlights presenting two methods for recognizing Finnish news quotes, marking a significant contribution for in-depth media voice analysis.

However, while advancements have been made, it's observed that a lion's share of research in computational linguistics has been principally channeled towards attribution analysis within English texts. This trend is evident even in large-scale projects like the PDTB and its adaptations in languages like Chinese, Arabic, Hindi, Czech, and Turkish. Although researchers like Pareti have ventured into predicting attribution in discourse, such endeavors are scarcely represented in languages other than English, particularly in Turkish.

In summary, this chapter has shown that limited attention has been given to attribution patterns particularly in non-English languages, including Turkish. This leads to the need for a comprehensive investigation into attribution phenomena in Turkish discourse, taking into account the specific linguistic characteristics of the Turkish language. Although this thesis does not carry out a computational analysis, the work laid out here on the Turkish attribution sets a stage for future research. Future research could analyze large corpora of Turkish texts to uncover patterns, frequencies, and variations of attribution expressions. Such analyses have the potential to reveal deeper insights into the cultural and linguistic specificities of attribution in Turkish. Therefore, this thesis enriches research on discourse and paves the way for applying computational techniques to better understand attribution in Turkish language and discourse.

CHAPTER 3
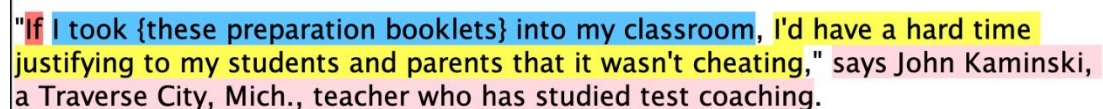
METHODOLOGY

This chapter provides an overview of the methodology employed in this research. The methodology includes attribution annotation, a comparison between the PDTB annotation approach and the proposed method, utilization of the TDB dataset, annotation style, inter-annotator agreement assessment, the statistical analysis of the data, and computational analysis using a Bayesian approach. Each of these components plays an important role in achieving the research objectives. This section aims to establish a clear framework for conducting the current investigation.

3.1. Comparison of the PDTB Annotation and the Approach of This Thesis

The PDTB annotation convention (Prasad et al., 2008) captures a variety of expression forms, including verbs that convey types of attribution like *believe*, *say*, and *know*, prepositional indicators such as *according to*, and specific noun phrases like *insights from* and *findings of*. However, it neither annotates the type and source features of sources and attribution, nor attribution span, that is, the text span associated with attribution.

To better understand, a typical relation annotation involving attribution is presented below (Figure 3.1). In the snapshot of this annotation (Figure 3.1), Arg1 is designated by the color yellow, while Arg2 is highlighted in blue. The explicit discourse marker *if* (shown in red) conveys the discourse sense of Contingency:Condition, and the attribution span involving both the source and the attribution verb *say* are marked in pink. The entire discourse relation (REL) is attributed to John Kaminski, and the arguments inherit this source-attributive material, which is shown in pink.

"If I took {these preparation booklets} into my classroom, I'd have a hard time justifying to my students and parents that it wasn't cheating," says John Kaminski, a Traverse City, Mich., teacher who has studied test coaching.

Figure 3. 1. PDTB Snapshot of a relation annotation involving attribution

Considering that the entire sentence pertains to John Kaminski, this source would be marked as other (abbreviated as Ot). Furthermore, since the type of attribution corresponds to the communicative verb *say*, it is categorized as Comm. For Arg1 and Arg2, due to the inherited source, the label inh could be used, and since they lack a specific type, Null would have been the appropriate selection (see Figure 3.1).

However, the corresponding annotation of the relation shown in Fig. 3.1 is left unannotated, as shown in Figure 3.2 below.
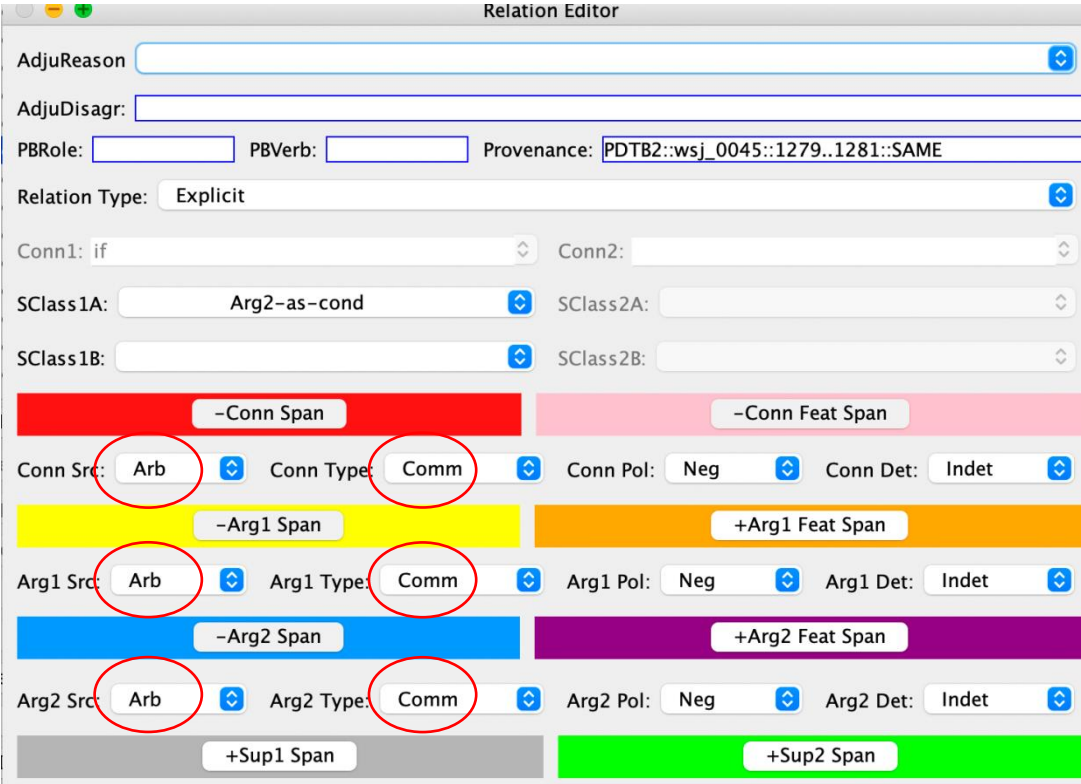
Figure 3. 2. The categories in the PDTB Relation editor that are left unmarked

In contrast, the present work adopts a more detailed annotation method enabled by the PDTB Annotation Tool (Lee et al., 2016). In this approach, both the source and the attribution verb types are explicitly annotated, and that offers a more nuanced understanding of the discourse structure. The summary of the comparison of the PDTB annotation and the approach of this thesis is presented in Table 3.1 below.

Table 3. 1. Comparative Analysis of the PDTB Annotation and Annotation Guidelines Utilized in this Research

PDTB Annotation	Our Approach
Discourse relation realization types (see Table 2.1 in Rev. of Lit) are annotated over Wall Street Journal texts together with Arg1, Arg2, and the connective, where appropriate.	Discourse relation realization types of the PDTB are annotated over TDB texts together with Arg1, Arg2, and the connective, where appropriate (see 3.2. Data: Turkish Discourse Bank)
Attribution spans are annotated without source and verb types.	Attribution spans are annotated together with the source of attribution and verb types.
Both attribution verbs and prepositional/noun phrases such as	Both attribution verbs and prepositional/noun phrases such as *-a/-e*

PDTB Annotation	Our Approach
according to are annotated as cues of attribution.	*göre* are annotated as cues of attribution.
Connective and argument features such as specific syntactic roles within arguments or detailed categorization of connective types beyond the provided semantic classes are not marked.	The focus remains on identifying and classifying discourse relations, arguments, and connectives without extending these into more detailed syntactic roles, pragmatic functions, or nuanced semantic distinctions.

3.2. Data: Turkish Discourse Bank

Linguistic mechanisms of attribution and the presence of attribution markers can differ across languages. The current investigation delves into the investigation of attribution patterns across genres within the Turkish Discourse Bank 1.2.

TDB 1.2 is a linguistically annotated corpus containing approximately 400,000 words from diverse Turkish texts, including newspapers, novels, articles, and interviews. Built on a sub-corpus of the METU Turkish Corpus, it follows the principles of the PDTB annotation framework, illustrated previously (see Table 2.1 in Rev. of Lit), providing annotations for explicit discourse connectives, implicit relations, entity relations, alternative lexicalizations as well as no relations and Hypophora (Zeyrek & Kurfalı, 2017; Zeyrek & Er, 2023). These annotations enable an in-depth analysis of discourse coherence and the structural patterns inherent in Turkish discourse.

While the TDB serves as a rich resource for research in Turkish discourse, computational linguistics, and natural language processing, it is important to note that attribution, has not been previously annotated in the corpus. This gap in the dataset has both challenges and opportunities. Turkish as a morphologically rich language with flexible word order and complex morphology demands specific adaptations in annotation style to accommodate its unique linguistic features (Zeyrek et al., 2013) and this kind of adaptation can also be extended to the annotation of attribution.

This thesis aims to extend the TDB with the annotations specifically carried out for attribution. By examining these annotations on discourse relations, we will understand how attribution is used in Turkish discourse better.
A detailed overview of the annotated material in, that is, the TDB 1.2 is presented in the table (3.2) below, including a total of 20 texts representing various discourse types. In terms of text types, the dataset comprises 7 novels, 2 articles, 1 interview, 2 memoirs, 2 research-examination and 6 news articles. Total relation count is 3,870.

Table 3. 2. Detailed Information about the Texts in TDB 1.2

No.	Turkish Discourse Bank Text	Type of Text	Word Count	Relation Count
1	00011112	Article	1335	79
2	00012112	Article	2012	164
3	00013112	Interview	1983	249
4	00001131	Novel	2005	391
5	00001231	Novel	1989	318
6	00002113	Novel	1996	218
7	00005121	Novel	2003	126
8	000065111	Novel	2019	277
9	00002213	Novel	2003	219
10	00003121	Novel	2014	242
11	00050120	Memoir	2003	222
12	00050220	Memoir	2009	247
13	00027113	Research-Exam.	2002	127
14	00027213	Research-Exam.	2016	185
15	20570000	News	2074	134
16	20580000	News	2028	118
17	20590000	News	1997	143
18	20630000	News	2062	162
19	10180000	News	1963	135
20	10330000	News	1909	114

3.3. Annotation Style

The annotation style is based on designing a set of guidelines aligned with the PDTB and applying these to annotate the TDB 1.2. This section outlines the methodology used in devising these annotation guidelines for attribution. The guidelines aim to establish criteria and rules for annotating various forms of attribution. Therefore, the guidelines not only cover verb-based attributions but also focus on noun phrases and postpositions that are specific to Turkish and indicative of attribution. Furthermore, the guidelines consider the linguistic characteristics observed in Turkish discourse to ensure effective annotation of the TDB 1.2.

3.3.1. Designing Annotation Guidelines

3.3.1.1. What are the PDTB-based annotation categories?

The present thesis adopts the PDTB as the annotation framework of the work (Prasad et al., 2008). Table 3.3 offers a detailed representation of how attribution types, as

defined by the PDTB, are annotated, complete with sub-types, corresponding descriptions, and illustrative examples in both English and Turkish.

Table 3. 3. PDTB Attribution Type Annotation

Sub-Type	Type	Description	English Examples	Turkish Examples
Assertion proposition AOs	Comm	Associated with communication type of attribution	say, mention, claim	söyle-, belirt-, iddia et-
Belief proposition AOs	PAtt	Associated with belief type of attribution	believe, think, expect	inan-, düşün-, bekle-
Fact AOs	Ftv	Associated with factive type of attribution	regret, forget, remember, know	üzül-, unut-, hatırla-, bil-
Eventuality AOs	Ctrl	Associated with control types of attribution	persuade, permit, order, etc.	ikna et-, izin ver-, emret-, etc.

In illustrating linguistic differences, it may be beneficial to present specific examples in a particular format that highlights morphological characteristics. Demonstrating the syntactic embedding differences between head-final languages (such as Turkish) and head-initial languages (like English) is important. The reason why it is important is that the position of the verb can change its importance in conveying the meaning of the sentence. Syntactically, the embedding verb becomes insignificant for discourse-semantics. Therefore, a verb that is hierarchically superior in syntax loses its prominence in semantics. For each category—namely, Assertion AOs, Belief Proposition AOs, Fact AOs, and Eventuality AOs—example sentences (1-3) are provided along with their corresponding syntax trees to illuminate the intricacies of Turkish linguistic structure.

Below, constructed examples (1) and (2) involve syntactic embedding, where the embedded part is the attributed material (shown by purple circles). These examples are provided to clarify the notion that attributed material can be syntactically embedded, but can convey the main message, the matrix verbs, *belirtti* ('indicated') and *inanıyor* ('believes'), would have a secondary role in such cases.

It should be noted that in this thesis, I follow the X-bar theory of phrase structure trees, a widely accepted model in syntactic theory. X-bar theory simplifies the structure by assuming that every phrase has a head (X), with intermediate and maximal projections (X' and XP, respectively). For practical purposes, I use basic category labels such as NP (Noun Phrase), VP (Verb Phrase), and S (Sentence) to maintain clarity and simplicity. Additionally, branching is kept binary to align with minimalist syntactic assumptions. This simplified representation allows for easier

readability and focuses on the core syntactic relations relevant to the analysis presented.

(1)	Ali	yarın akşam	yol-a	çık-acağ-ı-nı	belirt-ti
	Ali	tomorrow evening	road-DAT	set out-COMP-3Sg-ACC	indicate-PAST
	'Ali indicated that he would set off tomorrow evening.'				
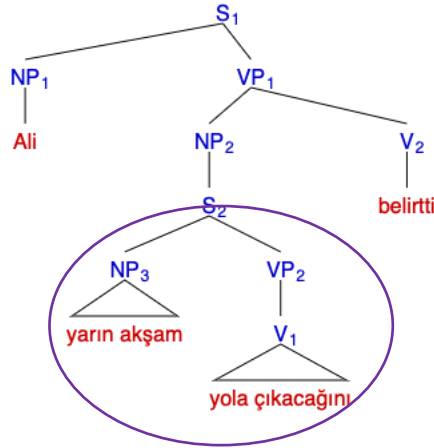
Figure 3. 3. Syntax tree of Example 1

(2)	Ayşe,	Ali'nin	doğruyu	söylediğine	inanıyor
	Ayşe,	Ali-POSS.3Sg	truth-ACC	tell-PST-NMLZ-3Sg-DAT	believe-PRES.3Sg
	'Ayşe believes that Ali is telling the truth.'				
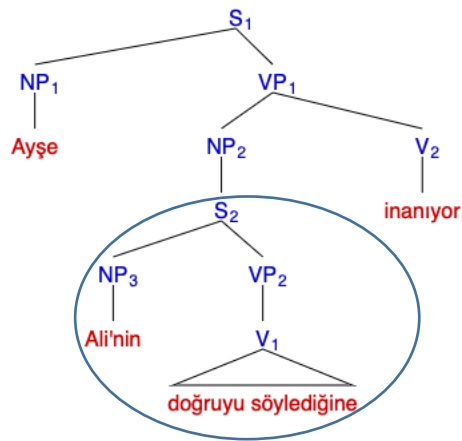
Figure 3. 4. Syntax tree of Example 2

Example (3) is a concession discourse relation anchored by *rağmen* ('despite the fact that') (underlined in the example), with its binary arguments (Arg1 and Arg2) shown

in italics and bold fonts, respectively. It is followed with a syntax tree showing the interaction between syntax and discourse (Fig. 3.5). What we are showing by our annotation scheme is that the concessive relation holds between "the winter season lasted short in South Asia" and "the deaths mainly stemmed from millions of people being homeless". The verb *belirtiliyor* ('it is indicated') is not part of the discourse arguments though it is part of syntax.

(3)	**Güney Asya'da**	**kış mevsiminin**	**kısa sürmesine**	<u>**rağ-men**</u>	*ölümlerin*	*daha çok milyonlarca kişinin*	*evsiz olmasından*	*kaynaklandığı*	[belirtiliyor]
	South Asia-LOC	winter season-POSS.3Sg-GEN	short last-VN-DAT	despite	deaths-GEN	more millions of people	homeless be-VN-3Sg-ABL	stem-VN-3Sg	indicate-PASS-PROG.3Sg
	'**Despite the fact that** the winter season lasted short in South Asia**, [it is stated] *that the deaths mainly stemmed from millions of people being homeless.*'								
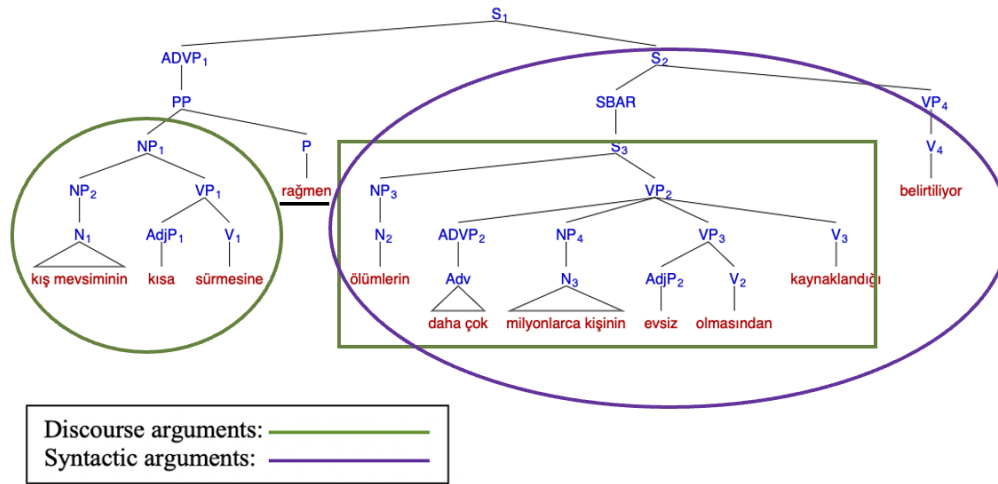
Figure 3. 5. Representation of discourse and syntactic arguments in Turkish

3.3.1.2. How are attribution and related categories annotated in Turkish?

PDTB-based categories, that is, the source of attribution and the predicate type, are annotated over TDB 1.2 by paying attention to the specific characteristics of Turkish.

3.3.1.2.1. Annotatıon of the Source and type Features

A list of potential English verbs and their Turkish translations has been made to help identify the verbs that convey various types of attribution in Turkish, namely, assertion propositions, belief propositions, facts and eventualities. This table has been utilized as the starting point for annotation the type feature (see Appendix A for the complete attribution predicate type list).

It is important to note here that, as annotators, we read each text in the corpus (20 texts in total) sentence by sentence, identifying discourse relations that include attribution, while excluding those without attribution. When we identified a

discourse relation with attribution, we annotated all relevant specifications, considering its context.

An example of annotation showing source and type features over an implicit discourse relation annotated in the TDB 1.2 is given along with its context and the representation of the relation editor on the Annotator Tool below.

For the rest of the examples, we will provide only the annotated parts, although we always considered the full context.

The context is as follows:
"*BÜYÜKELÇİLİK AYARLADI*

ABD'in Ankara Büyükelçiliği Askeri Ataşeliği'nin organize ettiği geziye AKP'lilerin yanı sıra Deniz Kuvvetleri'nden bir amiral ve beş subayla savunma sanayiiyle iş yapan üç işadamı, Japon ve Hollanda askeri ataşeleri olmak üzere toplam 15 kişi katıldı. Büyükelçilik geziye milletvekillerini özel olarak davet ederken, isimleri Gül'ün belirlediği öne sürüldü. TBMM Dışişleri Komisyonu üyesi Şirin de, büyükelçilikten kendisini arayan kişinin, 'Bu hafta sonu programınız yoksa sizi uçak gemisine davet ediyoruz' dediğini söyledi. Vekiller, Esenboğa Havaalanı'nda bindikleri uçakta, bel kemerleri dışında kollarından geçirilen kemerlerle de koltuklarına sıkıca bağlandı. Yağcı, 'Uçak o kadar hızlı fırladı ki, sadece ayaklarımız yerden kesilmedi, oturduğumuz koltuktan bile yükseldik. Sıkı bağlanmasaydık mermi gibi fırlardık' dedi."

(4)
 "[Yağcı], '*Uçak o kadar hızlı fırladı ki, sadece ayaklarımız yerden kesilmedi, oturduğumuz koltuktan bile yükseldik*. **Sıkı bağlanmasaydık mermi gibi fırlardık'** [dedi].

Figure 3. 6. The snapshot of the relation editor for Example 4

Table 3. 4. Illustration of the snapshot (Fig. 3.4)

	REL	Arg1	Arg2
[Source]	Ot	Inh	Inh
[Type]	Comm	Null	Null

The English translation of the context and the attribution annotation are as follows:

"ORGANIZED BY THE EMBASSY
Along with AKP members, a total of 15 people participated in the trip organized by the U.S. Embassy's Military Attaché in Ankara, including an admiral and five officers from the Navy, three businessmen working in the defense industry, and the military attachés from Japan and the Netherlands. While the embassy specially invited the members of parliament to the trip, it was claimed that the names were selected by Gül. Şirin, a member of the Foreign Affairs Committee of the Turkish Grand National Assembly, said that a person from the embassy called her and said, 'If you don't have any plans this weekend, we invite you to the aircraft carrier.' The MPs, who boarded the plane at Esenboğa Airport, were tightly strapped to their seats, not only with seat belts but also with belts around their arms. Yağcı said, 'The plane took off so quickly that not only did our feet lidt off the ground, we even lifted off from our seats. Had we not been tightly strapped in, we would have shot out like a bullet."

"[Yağcı said], '*The plane took off so quickly that not only did our feet lift off the ground, but we even rose from the seats we were sitting in*. **Had we not been tightly strapped in, we would have been shot out like a bullet**.'"

In example 5, an Expansion:Conjunction implicit discourse relation holds between Arg1 (shown by the yellow part) and Arg2 (shown by the blue part). This has already been annotated in the TDB 1.2. We added a further level of information to the discourse relation showing that the entire discourse relation is attributed to *Yağcı*. *Yağcı*, a proper name mentioned in the text, is the source of attributed material and annotated as other (Ot). The type of the attribution is an assertion proposition AO, since *dedi* 'said' is a communication verb (Comm).

In some discourse relations, only one of the arguments might contain attribution. In such cases, we annotate only the argument in which the attributed material is present. An example of this is provided below (Example 6):

(6)

"*ABD'nin kuzeydoğusunu etkisi altına alan soğuk hava dalgası, New York'tan Boston'a kadar uzanan şeritte ısının eksi 20'ye kadar düşmesine yol açtı.* **[Yetkililer], saatte 45 kilometre hızla esen rüzgârın etkisiyle, soğuğun eksi 60 derece gibi hissedildiğini [belirtti].**"

The English equivalent of Example (6) is provided below:

"*The cold wave affecting the northeast of the USA caused temperatures to drop as low as -20 degrees in the stretch from New York to Boston.* **[Officials stated] that due to the wind blowing at 45 kilometers per hour, the cold felt like -60 degrees**."

	REL	Arg1	Arg2
[Source]	Wr	Inh	Ot
[Type]	Comm	Null	Comm

In this illustration, an Expansion.Level-of-detail.Arg2-as-detail Implicit *hatta* ('even') discourse relation exists between Arg2 (shown by the bold part) and Arg1 (shown by the italic part). This relation has previously been noted in the TDB 1.2. In the current investigation, a further level of information is being added to the example, showing that the Arg2 is attributed to the officials, marked as other (Ot). The type of the attribution is an assertion AO, since *belirtti* ('stated') is a communicative verb (Comm). This instance is important because it shows that not just the entire discourse relation, but also its components, namely the arguments, can be traced back to a source with a specific type of attribution.

3.3.1.2.2 Characteristics Specific to Turkish

Specific characteristics of the Turkish language have been addressed, such as (i) the hearsay suffix -mIş, (ii) the formation of comment clauses in the first person singular and plural, and (iii) changes in the annotation scheme based on studies concerning Turkish control.

i.   The Turkish hearsay -mIş suffix

The Turkish past tense marker *-mIş* has an important role in conveying both temporal and evidential information.To fully understand the grammatical and functional aspects of *-mIş*, both comprehensive frameworks and specialized studies are necessary. Göksel and Kerslake (2005) provide a broad overview of Turkish grammar, while Slobin and Aksu (1982) focus on the evidential and inferential uses of *-mIş*, making these two sources key references for the analysis.

Göksel and Kerslake's *Turkish: A Comprehensive Grammar* (2005) is a well-regarded source for understanding Turkish grammar, especially the functions of the past tense marker *-mIş*. In their work, They explain that *-mIş* is mainly used to show past actions or events that the speaker did not witness directly, unlike *-dI*, which marks actions the speaker did see. Additionally, *-mIş* often shows that the speaker is passing on secondhand information or making an assumption based on evidence. Göksel & Kerslake also highlight that *-mIş* can express surprise or realization, where the speaker becomes aware of something after the fact. This use of *-mIş* is common in narrative structures, where the speaker where the speaker is distanced from the event. Their treatment of *-mIş* provides a broad framework and their work is essential for understanding the general grammatical role of this morpheme in Turkish.

In contrast, Slobin and Aksu's work *Tense, Aspect, and Modality in the Use of the Turkish Evidential* (1982) offers a more specific analysis of *-mIş* by focusing on its evidential and inferential functions. While Göksel & Kerslake provide a general overview, Slobin and Aksu delve into how *-mIş* shows indirect knowledge. They

argue that *-mIş* not only marks past events but also indicates how the speaker acquired the information—whether through inference, hearsay, or indirect evidence. This evidential function is significant because it helps differentiate between firsthand and secondhand information and reflects the speaker's stance toward the event. In addition, *-mIş* is often used to convey inference, where the speaker guesses an event has happened based on clues. Slobin and Aksu also highlight how *-mIş* i shows the speaker's understanding of an event after it has happened, reflecting on its completion.

The example sentence (7) conveys that the speaker did not witness the big festival but has heard about it. Thus, it is best to translate -mIş sentences to English with adverbs like *allegedly, apparently, supposedly* or with phrases like *I heard that*.

(7)	Köyde	büyük bir şenlik	yapılmış.
	Village-LOC	big a festival	was done-PASS.
	'I heard there was a big festival in the village.'		

Göksel & Kerslake (2005) give a comprehensive description of *-mIş* as part of the overall Turkish verbal system, while Slobin and Aksu focus more on the cognitive and evidential aspects. Slobin and Aksu's detailed insights into the use of *-mIş* to show indirect information, surprise, or realization are especially valuable for understanding how this morpheme works.

In addition, Nilüfer Şener's study, *Semantics and Pragmatics of Evidentials in Turkish* (2011), focuses on how the morpheme *-mIş* functions within the language. Şener explores how -mIş integrates aspects of tense, aspect, and epistemic modality. She argues that -mIş not only marks past events but also engages with the epistemic stance of the speaker regarding the truthfulness and source of the information. Therefore, in terms of pragmatic functionality, she examines how this marker influences the interpretation of statements within various discourse contexts. This includes how speakers use -mIş to express doubt, inference, or indirect knowledge, impacting how listeners perceive and evaluate the given information.

The inherent ambiguities of the -mIş suffix led to the decision to annotate it as an arbitrary source of attribution (Arb). The suffix's non-specificity regarding the source (Slobin and Aksu, 1982) and its interpretative flexibility support this decision.

The following example (8) demonstrates how *-mIş* is annotated as a cue for an arbitrary agent (Arb) in the PDTB Annotator Tool, along with the corresponding English translation:

(8)
"Son yıllarda petrol karşılığı mal alımı üzerindeki kurallar esnek tutularak Irak'a bir parça nefes alma şansı tanın[*mış*], ancak **para karşılığı petrol satışı sağlanamadığı için yoksulluğun önü kesileme**[**miş**]."

43

The English equivalent of Example (8) is provided below:
"[Allegedly], in recent years, by keeping the rules for goods purchases in exchange for oil flexible, a chance for some breathing space has been provided to Iraq. <u>However</u>, **[apparently], since the sale of oil for money could not be achieved, poverty has not been alleviated**."

	REL	Arg1	Arg2
[Source]	Wr	Arb	Arb
[Type]	Null	Null	Null

The use of 'allegedly' in translations of sentences with the *-mIş* suffix reflects its function of conveying information based on indirect sources or inference, not directly observed by the speaker.

In conclusion, the *-mIş* suffix plays a crucial role in expressing indirect information and serves as a key feature in attributing non-firsthand knowledge. Its ambiguity in specifying the source of information, combined with its evidential and inferential uses, makes it a central element in Turkish syntax and semantics, as explored by Göksel and Kerslake (2005), Slobin and Aksu (1982), Sener (2011) and others.

ii.   *Comment Clauses*

Comment clauses, such as *I think*, *I believe*, also known as parenthetical clauses, are linguistic constructs that allow the speaker or writer to add information, commentary, or clarification to a statement. These clauses often take the form of an observation, explanation, or opinion. They can be inserted within a sentence without changing the core meaning of the sentence. They are set off from the main clause by commas, dashes, or parentheses (Quirk et al., 1985).

For example:

(9)	The weather, I think, will be nice tomorrow

Turkish examples include:

(10)	Yarın havanın güzel olacağını düşünüyorum.
	'I think that the weather will be nice tomorrow.'

(11)	Biz bu maçın berabere bitmesi gerektiğini düşünüyoruz.
	'We think that this match should end in a draw.'

Comment clauses are of interest in literary and linguistic studies for various reasons. They offer insights into the narrative voice, modality, and discourse strategies. By

studying comment clauses, researchers can understand how writers and speakers negotiate meanings, express attitudes, and engage with readers or listeners (Brinton, 2008).

Initially, in this thesis, we decided to exclude comment clauses from the attribution annotation process. This was because they often involve personal thoughts or beliefs rather than contributing essential information about the attribution of a statement. Including them might have introduced noise or irrelevant data. Instead, key characteristics such as subject type were documented in a supplementary notes section. For example, as shown in Example (12), in a discourse relation involving the implicit discourse marker *çünkü* ('because'), the use of *dedi* ('said') links the statement to a specific source, categorized as Comm for communication verbs. In contrast, *bil-mek* ('to know'), a factive verb, is indicated as Ftv. Therefore, in initial manual annotations, comment clauses were recorded separately and not included in the primary dataset.

(12)
"'Tanıyamadın değil mi?' [dedi]. 'Haklısın. *Az önce konuştuğun Nazlı ile en ufak bir benzer yanımızın olmadığını ben de* [biliyorum]. **İçimi, dışımı yenileyip baştan yarattı beni. Çok da başarılı oldu üstelik**. Eline sağlık.'"

Figure 3. 7. Attribution annotation of Example 12

	REL	Arg1	Arg2
[Source]	Ot	Inh	Inh
[Type]	Comm	Ftv	Null

The English equivalent of (12) is provided below:
"'You didn't recognize me, did you?' [she said]. '*You are right.* [*I also know*] *that there is not the slightest resemblance between me and the Nazlı you were just talking*

45

to. **She completely renewed me inside and out, recreated me**. And she was very successful at it, too. Kudos to her.'"

However, following further discussions and reevaluation, we decided to include the annotation of comment clauses to align with the PDTB annotation standards. This change ensures that our annotations are comprehensive and consistent with established frameworks. Recognizing that comment clauses can blur the distinction between personal and attributed statements, we found that their inclusion provides a more detailed understanding of discourse. Therefore, in this work, we annotated comment clauses as attribution spans to fit the established guidelines of the PDTB, and to enhance the overall comprehensiveness of our analysis.

iii.   *Control Verbs in Turkish*

In linguistics, control refers to a grammatical structure where the subject of a verb is determined by context. Typically, a higher-ranking verb influences the arguments of a lower-ranking one. This concept was extensively examined in the 1980s under the government and binding framework, with terms still in use today (van Riemsdijk and Williams, 1986: 128ff.; Cowper, 1992:161ff.; Borsley, 1996:126-144). Control verbs in English such as *want*, *try*, and *plan*, illustrate this phenomenon.

Kunduracı (2008) discusses the use of the *-mEk* and *-mE* suffixes in Turkish and argues that they are not used interchangeably in control structures. The -mEk suffix is used to form verb infinitives, functioning similarly to the English *to* preceding verbs, as seen in *to read* or *to write*. Conversely, the *-mE* suffix serves as a nominalizer, converting verbs into nouns or noun-like entities. This is comparable to the English gerund suffix *-ing*, which transforms verbs into nouns, such as changing *run* into *running*. Through his analysis, Kunduracı (2008) substantiates the assertion that while -mEk is integral to controlled complement phrases, *-mE*, as a nominalizer, cannot operate in the same capacity.

The examples (13 & 14) below present how verb+*mEk* (the infinitive) and verb+*mE* (the verbal noun) conflict with each other in the control structure.

(13)	Kaan	okumağı	çok	istiyor.
	Kaan	read-INF-ACC	much	want-PROG
	'Kaan wants to read very much'			

(14)	?Kaan	okumayı	çok	istiyor.
	Kaan	read-NMN-ACC	much	want-PROG
	'(Lit.) Kaan wants reading very much'			

As seen in these examples, *-mEk* and -mE are not interchangeable and cannot both function as the controlled complement of control verbs in Turkish.The sentences

clearly demonstrate that the use of *-mEk* in Example (13) results in grammatical sentences, while the use of -mE in Example (14) leads to ungrammatical constructions in control contexts.

Therefore, Kunduracı (2008) asserts that control verbs exclusively govern a specific form of complement, the infinitival phrase marked by V-*mEk*, differentiating them from the nominal and finite V-*mE*. This assertion emphasizes the importance of recognizing both syntactic and semantic aspects of control without prioritizing one over the other. In essence, the study posits that control in Turkish is not solely a syntactic or semantic phenomenon; rather, it serves as a reflection of the interplay between these two linguistic dimensions. In the light of this, attribution annotation of the eventuality AOs (Ctrl) are arranged in accordance with the control study of Kunduracı (2008).

In Kunduracı (2008), subject control and object control verbs are listed and discussed. The verbs listed in this compilation have been included in the attribution verbs list created for the purpose of this thesis (please refer to Appendix A for the detailed list).

The verbs included in this list are marked as control only when they are used with the -mEk suffix. An example (15) from the TDB is provided below:

(15)
"Adamın anlatacak bir şeyi kalmadığında sözü [Ante]'nin alması gerekmişti. *Paşa'nın sorusu ve isteği üzerine belki bininci kere yaptığı kahramanlığı anlatmak* [*zorunda kaldı*]. **Nereye davet edilse, aynı şeyleri anlatıyordu**."

English equivalent of the Example (15) is provided below:

"When the man had nothing left to tell, it was supposed to be [Ante]'s turn to speak. *Upon the Pasha's question and request, he* [*was forced*] *to recount the heroism he had performed for perhaps the thousandth time*. **Wherever he was invited, he would tell the same stories**."

	REL	Arg1	Arg2
[Source]	Wr	Ot	Inh
[Type]	Comm	Ctrl	Null

In this context, the control type of attribution appears in the first argument, and therefore only the first argument is annotated. The specific source of the control is Ante, who is the teller of the heroism he had performed, and the type of attribution is classified as control, *to be forced*. This annotation aligns with the findings of Kunduracı (2008) and demonstrates how control attribution has been annotated in the current research, reinforcing the identified pattern of control in Turkish.

3.3.1.2.3. Marking of Attributional Expressions

This section examines the encoding of attribution sources in discourse, focusing on the use of markers that signal the origin of statements across languages. It compares the PDTB framework's techniques with Turkish practices, highlighting key examples. The goal is to illuminate the diverse linguistic strategies for indicating attribution, essential for nuanced discourse analysis.

  i.   Attribution through According to

In the PDTB approach, the phrase *according to* is marked as a marker of attribution. The equivalent Turkish phrase, *-a/-e göre* is annotated in the current work. An example is provided below.

(16)	Abdi Efendi'den	başkasıyla	görüşemiyordu.
	Abdi Efendi-ABL	apart-3Sg.POSS-COM	meet-REC-ABIL-NEG-PROG-PAST
'*He/she couldn't meet with anyone else apart from Abdi Efendi.*'			

Ondan	aldığı	haberlere	göre	gemiler	batmıştı.
He/she-ABL	take-NMLZ-3Sg.POSS	news-PL-DAT	according to	ships	sink-EVID-PAST
'[**According to the news he/she got from him/her**], **the ships had sunk**.'					

  ii.   The role of converbial suffixes in attribution

We came across various occurrences of Turkish verbs in the data, such as *söyleyerek* ('saying') and *ekleyerek* ('adding') incorporating the converbial suffix -ArAk, distinguishing them as attributional. With these communicative verbs, this suffix does not merely serve as a converbial suffix but also as mechanisms for attributing statements or actions within discourse. An example (17) is presented below.

(17)	Orhan	Kemal'i	de	getirmesini	söyleyerek	evlerine	çağırmıştı	onları.
	Orhan	Kemal-ACC	also	bring-VN-3Sg-POSS-ACC	say-CONV	house-PL-3Sg.POSS-DAT	call-EVID PAST-3PSg	they-ACC
	'[*Orhan*] *had invited them to his house*, [**saying**] **also to bring Kemal**.'							

Here, the use of *söyleyerek* ('saying') provides a clear attributional role, linking the action of inviting to the act of saying to bring Kemal, thereby ensuring the coherence of the narrative through the use of -ArAk.

To provide a comprehensive approach, the guidelines have expanded to include other adverb-verb suffixes, marked as attributive phrases when used alongside attributional

verbs. The following table (Table 3.5) presents Turkish morphemes that commonly appear with attribution predicates. These morphemes, when attached to verbs, modify the verb's meaning and can have various interpretations depending on the context.

Table 3. 5. Turkish morphemes (converbial suffixes) and their English translations that appear with attribution predicates

No.	Morpheme	English equivalent
1	-Ip	and, by doing so
2	-ArAk	by, through
3	-mAdAn	without doing
4	-mAksIzIn	without
5	-dIkÇE	as, while
6	-InCA	when, as soon as
7	-AlI	since (time-related)
8	-ken	while, when
9	... -r ... -maz	no sooner... than...
10	-AsIyA	until
11	-CAsInA	as if, like

iii.   *Beyond Explicit Attributional Phrases*

Furthermore, an examination of the PDTB revealed that the source of a quotation is still indicated even without an attributional phrase or verb, only by a quotation marker. This principle has been applied in the current thesis, recognizing individual statements as attributed even in the absence of specific verbs. An example (18) of the attribution with a quotation marker is shown below:

 (18)
"   -'Sayın [Esin], sizi Aşıklı Höyük kazılarıyla tanıyoruz. Aşıklı Höyük kazılarından önceki çalışmalarınızdan kısaca söz eder misiniz?'
   -'*İÜ Prehistorya Anabilim Dalı'ndayken bir çeşit arkeometri çalışmalarının başlangıcını yaptım*. **Spektral analiz çalışmalarıyla Anadolu Maden Enstitüsü hakkında kapsamlı bir araştırma yaptım**.'"

English equivalent of Example (18) is provided below:

"   -'Ms. [Esin], we know you from the Aşıklı Höyük excavations. Could you briefly talk about your work before the Aşıklı Höyük excavations?
   -'*While I was at the Prehistory Department of Istanbul University (İÜ), I initiated some kind of archeometric studies*. **I conducted comprehensive research on the Anatolian Mining Institute through spectral analysis studies**.'"

In this example, an Expansion.Level-of-detail.Arg2-as-detail implicit discourse relation holds between Arg1 and Arg2. This relation has been previously annotated

in the TDB 1.2, and in the current investigation, we add an extra layer of information to the corpus showing that the entire discourse relation is attributed to *Esin*. *Esin*, a proper name mentioned in the text, is the source of attributed material and annotated as Ot. The type of the attribution is an assertion proposition AO, even though the verb is not present, with the help of the quotation mark, it is clear that Esin conveys the message.

3.3.2. Annotation Method

All the texts from Turkish Discourse Bank were annotated by two independent annotators for attribution. The annotators were native in Turkish. Except for the first set of files (as described below), the primary annotator (the author of the present thesis) annotated the files and an expert annotator checked and revised them.

Initially, a set of 6 files from different genres (30% of the corpus) was annotated individually by the advisor of this thesis and the primary annotator thoroughly examined and updated them using the annotation categories we developed within this research. Afterward, using the same categories and the manual, the primary annotator annotated the remaining files. The expert annotator reviewed and verified the annotations (see Figure 3.8 below for the annotation method).
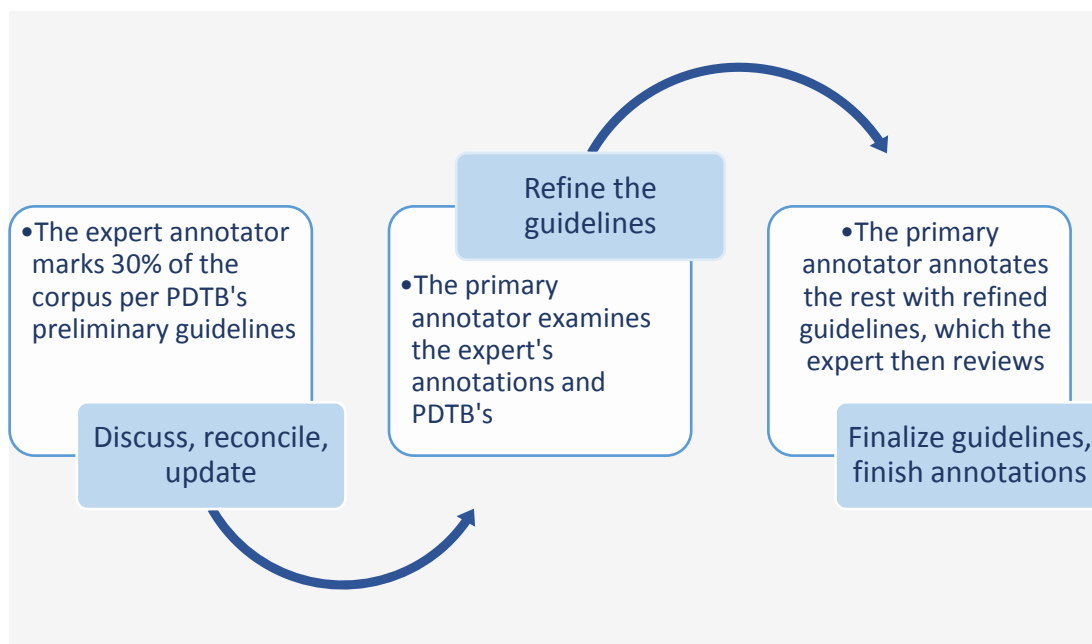
Figure 3. 8. Annotation Method

The annotation process was carried out using the PDTB Annotator tool (Fig. 3.9).
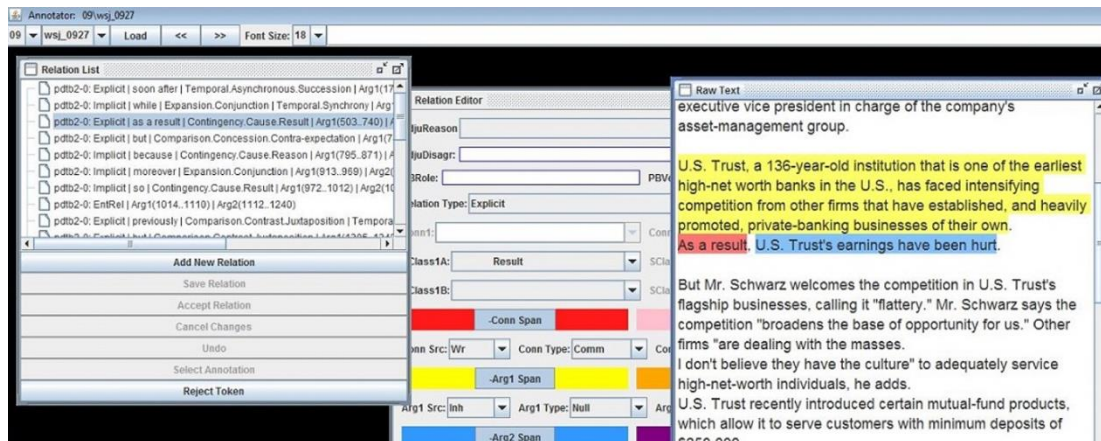
Figure 3. 9. A view of the PDTB Annotator Tool. Users may rearrange the subpanels as needed (Lee et al., 2016)

In conclusion, this chapter presented a detailed methodology framework for this work byfocusing on the evaluation and adaptation of annotation practices for discourse attribution in Turkish. It bridged the gap between established PDTB approaches and the specific linguistic features of Turkish. This chapter prepared for more analysis and improved our understanding of discourse attribution.

CHAPTER 4

AN EVALUATION AND THE FINALIZATION OF THE DATASET

This chapter examines the essential stages of evaluating and finalizing the dataset, ensuring its quality and readiness for analysis. This chapter outlines the evaluation methods used to verify data accuracy and relevance, followed by a detailed account of the finalization process, where the dataset is refined and confirmed to meet the objectives of the thesis. Through this meticulous preparation, we establish a solid foundation for the dataset, essential for the integrity and success of the findings of the present thesis.

4.1. Inter-Annotator Agreement Method

Inter-Annotator Agreement is a statistical measure utilized to determine the level of agreement between two or more annotators who classify items into specific categories independently of each other. High IAA indicates that the annotation task has been executed with clarity and reflects both the quality of the guidelines provided and the understanding of the annotators. Conversely, a lower IAA might suggest ambiguities in instructions, the inherent complexity of the data, or disparities in annotator interpretations. A detailed interpretation table is presented in the table below (Table 4.1).

Table 4. 1. Cohen's kappa interpretation table

Cohen's Kappa Value	Level of Agreement
0	No agreement
0.10 - 0.20	Slight agreement
0.21 - 0.40	Fair agreement
0.41 - 0.60	Moderate agreement
0.61 - 0.80	Substantial agreement
0.81 - 0.99	Near perfect agreement
1	Perfect agreement

For the current thesis, we have chosen the Cohen's Kappa coefficient as our measure of IAA (Cohen, 1960). Cohen's Kappa quantifies the degree of agreement between two raters while accounting for the possibility of chance agreement. Its advantage

over percentage agreement lies in its ability to provide a corrected measure. It ensures a more accurate representation of true agreement.

The equation representing the Cohen's Kappa coefficient is as follows:

$$k = \frac{p_o - p_e}{1 - p_e}$$

Where:
- *Po* represents the observed agreement.
- *Pe* denotes the expected agreement by chance.

4.1.1. Statistical Analysis of Inter-Annotator Agreement

The kappa analysis shows how consistently the annotators identify and mark attributional phrases and verbs within this work. The results cover discourse relations aspects like Arg1, Arg2, and Entire Drel and show a range of Cohen's Kappa (κ) values that indicate agreement levels from moderate to almost perfect.
The calculation of the statistical findings are shown as below.
- Arg1 Analysis: The average κ value for Arg1 across all documents stands at approximately 0.83 and indicates an "almost perfect agreement" between annotators. This level of concordance suggests a high reliability in the identification of the first argument in discourse relations. It highlights the clarity and effectiveness of the extended PDTB guidelines in guiding annotators' judgments in this aspect.
- Arg2 Analysis: The average κ value for Arg2 is around 0.80, which also falls into the range of "almost perfect agreement." Similar to Arg1, this demonstrates a strong consensus on the identification of the second argument in discourse relations and validates the annotation approach and the comprehensibility of the guidelines used.
- Entire DRel Analysis: For the Entire DRel category, the average κ value is approximately 0.77 and indicates a "substantial agreement" among the annotators. Though slightly lower than the agreement levels for Arg1 and Arg2, this still reflects a considerable degree of reliability in annotating entire discourse relations. The relatively lower kappa score might stem from the inherent complexity of annotating entire discourse relations as opposed to specific arguments.

Following the kappa calculations, we addressed and corrected the identified inconsistencies through discussion and data adjustments. A flowchart describing the processes of the IAA calculation is presented below (Fig. 4.1).
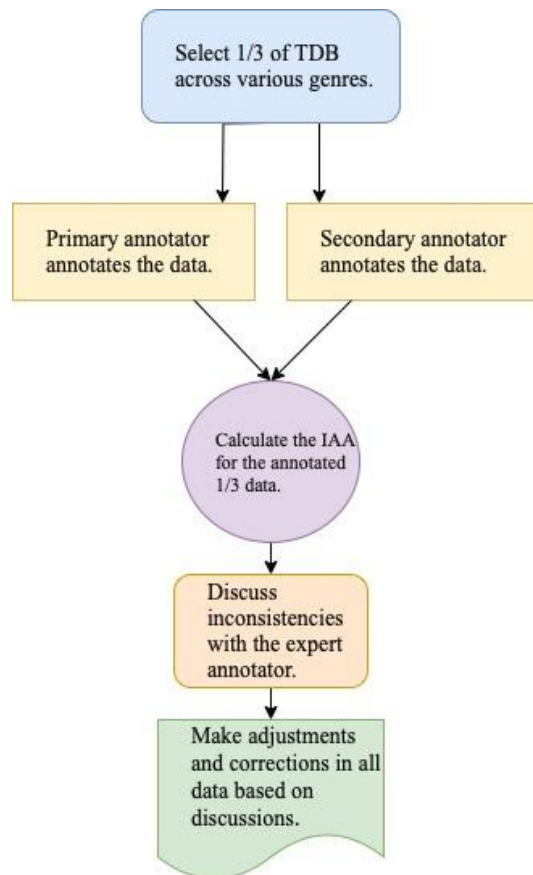
Figure 4. 1. Flowchart of the IAA process

All in all, the kappa analysis highlights the annotation framework's strength and the PDTB guidelines' success in securing high inter-annotator agreement.

4.2. Evaluation Procedure

For the evaluation procedure of this thesis, the author, identified herein as A.Y., was the primary annotator, while M. E., a graduate with a background in linguistics from Cognitive Science, helped as the secondary annotator. Over a period of approximately six months, the primary annotator engaged in a thorough examination of the PDTB annotations, holding discussions with the thesis advisor on attribution markings and noting specifics within the Turkish context. This led to the development of a detailed guideline for Turkish attribution annotation, which was subsequently shared with the secondary annotator. The secondary annotator was then provided with approximately one hour of training over two sessions, conducted jointly by the advisor and the primary annotator.

Following the training, texts from various genres were selected for inclusion in the Turkish Discourse Bank, ensuring that the total number of relations analyzed (1,270) represented one-third of the overall total (3,870). These texts span six genres: articles, interviews, novels, memoirs, research studies, and news reports. Each text was independently annotated by both the primary and secondary annotators. Subsequently, a kappa analysis was performed on these six files to assess inter-

annotator reliability. Throughout this process, any disagreements were discussed with the thesis advisor, identified as D.Z. As a result of these discussions, the data were cleaned and revised, ultimately resulting in the finalized Turkish attribution dataset. A flowchart that summarizes the evaluation process is provided below (Fig. 4.2)
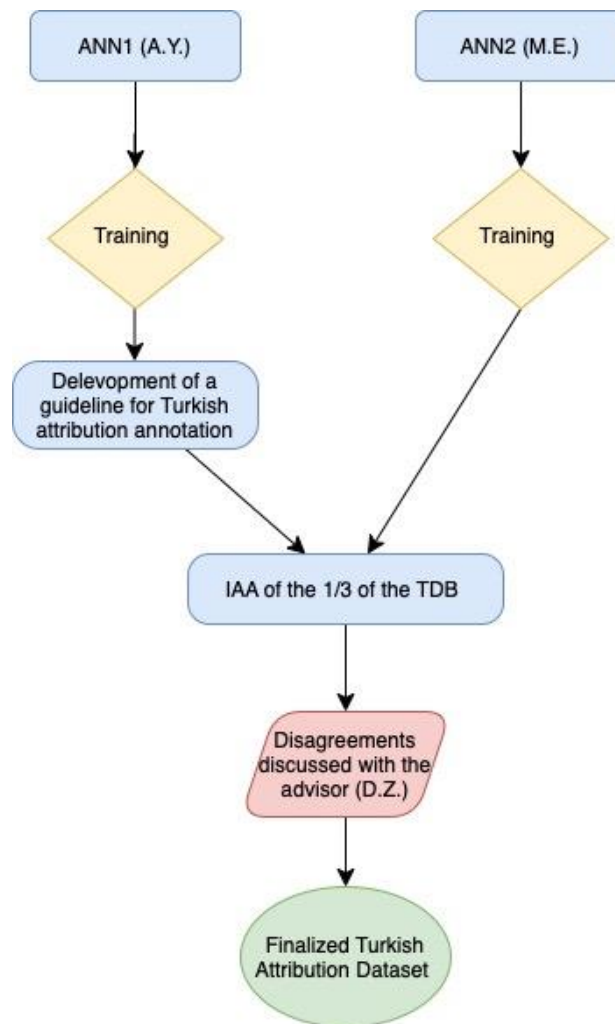
Figure 4. 2. Flowchart of the Evaluation Process

4.3. Dataset Finalization and Publication

The TDB dataset has been finalized with a focus on accuracy, consistency, and web compatibility. The review process ensured that the annotations met quality standards. The dataset is designed for easy integration into digital platforms, supported by metadata for effortless navigation and search. The TDB dataset is ready for online sharing, and hopefully will serve as a foundational resource for future linguistic research and analysis.

In conclusion, this chapter has detailed the evaluation and finalization process of the TDB dataset and validated its high quality and consistency. By using the Cohen's

Kappa coefficient, we established the dataset's reliability through inter-annotator agreement. The dataset's refinement that was characterized by detailed discussions and revisions, has enhanced its quality. The TDB dataset is an asset for linguistic research in Turkish discourse analysis.

CHAPTER 5

DISCUSSION OF QUANTITATIVE RESULTS

This chapter analyzes the TDB data that is annotated in terms of attribution. It describes how the data was annotated using a specific manual designed for Turkish, checked for accuracy and prepared for statistical analysis. This chapter presents findings on how often different types of sources and verbs are used in Turkish texts, showing trends and patterns through tables and figures.

5.1. Annotation and Data Preparation

In this study, a detailed attribution annotation manual has been prepared based on the PDTB attribution annotation by noting various nuances and language-specific morphological structures in Turkish. With the help of this manual, texts from different genres in the Turkish discourse bank (20 files) were annotated, and inter-annotator agreement was confirmed using Cohen's kappa. After the inter-annotator agreement was completed and the dataset was cleaned, the pipe-delimited format obtained from the Annotator Tool was converted to *Excel* for statistical extraction. Frequency tables for source and verb types were then generated, as presented in Table 5.1 and Table 5.2, respectively.

5.2. Detailed Analysis of the Distribution of Attribution Source and Predicate Type

This table (5.1) presents the frequency of different agents ('source') of attribution used within the components of discourse relations. The category labeled Other (Ot) shows the highest counts in entire discourse relations (Rel), Argument 1 (Arg1), and Argument 2 (Arg2), with frequencies of 296, 259, and 221 respectively. This indicates a predominant use of sources that are specific and identifiable, yet distinct from the writer or generalized sources. The Writer (Wr) category, which includes both the first-person singular and plural, is the second most frequent source type. This shows that the author's voice, from both individual and collective viewpoints, is frequently present but less dominant compared to other external sources within the discourse. The arbitrary (Arb) category, which includes nonspecific sources, is prevalent in Arg2. This observation could indicate a tendency to use less specific references more frequently in supporting arguments than in main claims or discussions of the relationships between ideas.

Table 5. 1. Total source type frequency

Source	Rel	Arg1	Arg2
Ot	296	259	221
Wr	54	53	49
Arb	16	31	41

Table 5.2 outlines the frequency of verb types used for attributions across different discourse components. Communicative (Comm) verbs are the most frequent across all components -Rel, Arg1, and Arg2- with the highest occurrence in Rel (311 instances). This highlights the frequent use of verbs that denote direct communication. Belief Proposition (PAtt) verbs are also common but to a lesser extent so they show their significant but secondary role in attributions. Control (Ctrl) verbs appear more in Arg1 and Arg2 than in Rel. Factive (Ftv) verbs, have a consistent presence across all components. The category Null, indicating the absence of explicit verb types, occurs in Arg1 and Arg2. It points to situations where attributions are carried out using cues other than verbs.

Table 5. 2. Total verb type frequency

Verb Type	Rel	Arg1	Arg2
Comm	317	229	227
PAtt	40	37	27
Ctrl	2	41	28
Ftv	7	22	14
Null	0	13	15

The pie chart (Figure 5.1) visualizes the attribution types' distribution. As seen from the chart, Comm verbs are the most prevalent verb type. This is succeeded by PAtt, Ctrl, and Ftv verbs. The Null category, representing attributions that are carried out with attributional phrases instead of attribution verbs, has the smallest share. Based on these findings, it can be concluded that attribution is primarily used to convey statements, secondarily to express beliefs and thoughts, then to indicate intention or attitude towards an event or situation, and lastly in cases involving assurance of information accuracy or evaluation of that information.
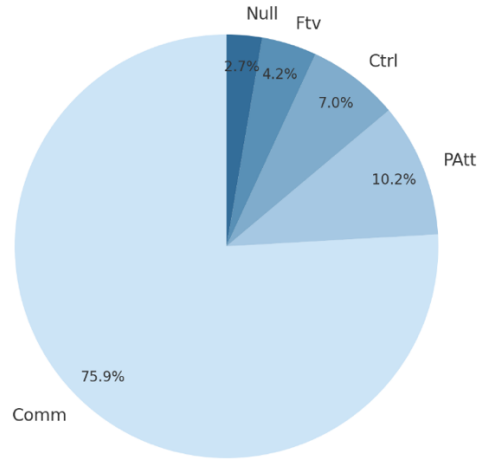
Figure 5. 1. Distribution of Attribution Types

Figure 5.2 presents a bar chart detailing the frequency of key attribution features (Source, Type) annotated in TDB 1.2, complete with English equivalents. As for the verbs conveying types of attribution, *de-* ('to say') appears most frequently, with 211 occurrences, indicating a dominant usage. *Söyle-* ('to tell') follows with 88 instances and *belirt-* ('point out') with 56 and these results highlight their significant roles in attributive contexts. Lesser-used verbs include iste- ('to want'), anla- ('to understand'), and others like devam et- ('to continue'), açıkla- ('to explain'), sor- ('to ask'), vurgula- ('to emphasize'), and anlat- ('to describe'), showing a diverse range of predicate types such as "*Bakkaldan kendisine bir sigara almamı istedi.*" ('He asked me to buy a cigarette for him from the grocery store.'). Figure 5.2 provides a quantitative insight into the verb usage patterns within attributions annotated over the TDB 1.2, all of which are thoroughly presented in Appendix B of the thesis.
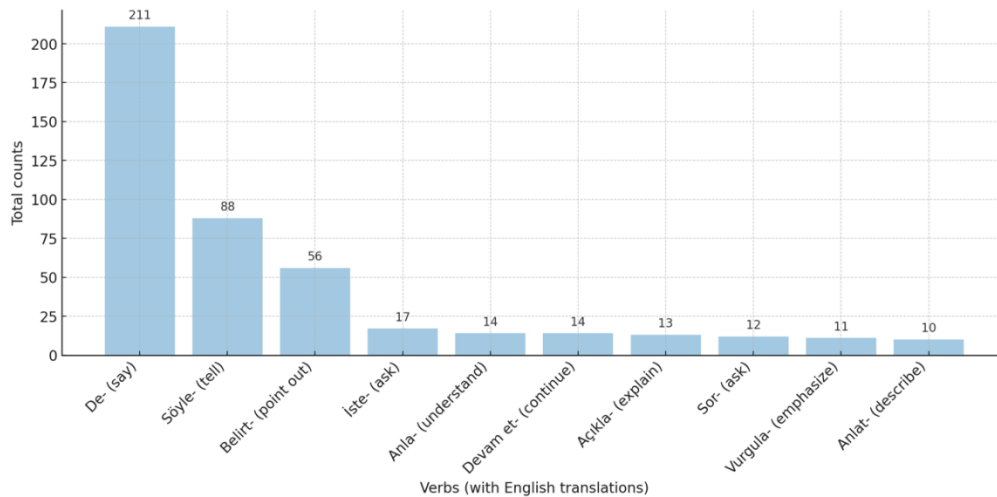
Figure 5. 2. Top verbs by frequency

5.3. Comparative Analysis of the Distribution of Attribution Features Across Genres

This thesis aimed to reveal the distribution of attribution features across different genres, as the TDB contains texts from six different genres, namely novels, news, memoirs, research-examination, articles, and interviews. These texts can be analyzed in two separate groups—journalistic and non-journalistic—in addition to the overall results for a better comparison. This distinction is important because the nature of attribution may differ based on the formality and purpose of the text. Journalistic texts include news, articles, and interviews, whereas non-journalistic texts include novels, memoirs, and research-examination texts. The analysis began by examining the overall distribution and then proceeded to focus on the journalistic and non-journalistic categories.

Figure 5.3 displays the attribution frequency across a variety of genres, both journalistic and non-journalistic, including novel, news, memoir, research, article, and interview. The genres are ordered by the total frequency of their relations. The data reveals that novels and news reports have the most frequent use of attribution over discourse relations. This frequency suggests a detailed reporting in these genres, and it is characterized by extensive use of direct and indirect speech. Memoirs are also high on the list and that indicates the use of personal narrative and retrospective explanations in terms of attribution. Research-examination, interviews, and articles tend to have fewer attributions, which likely has to do with the structural demands of these genres.
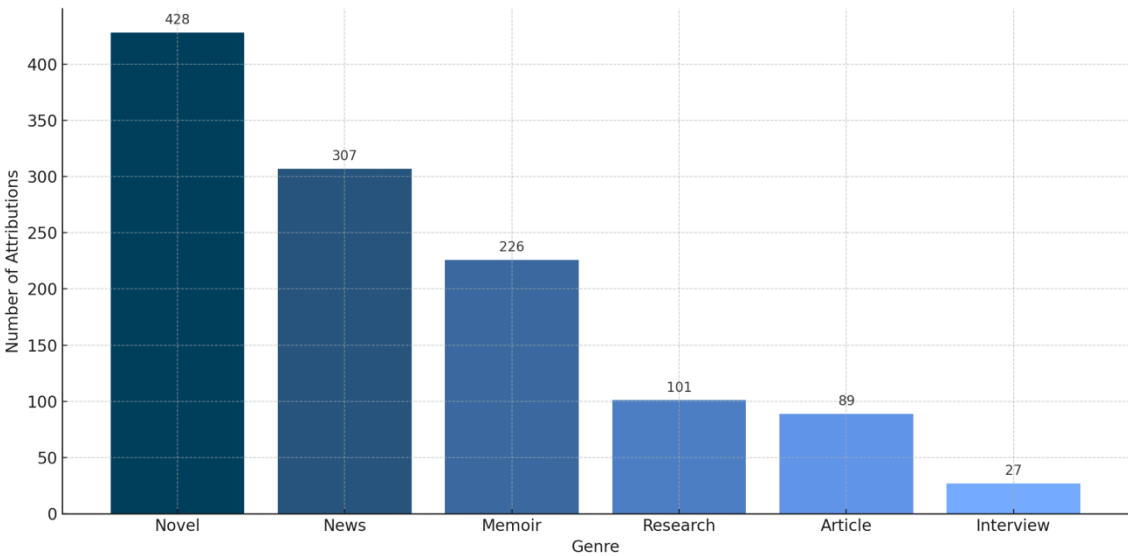
Figure 5. 3. Distribution of attribution annotations by genre

In Figure 5.4, the distribution of attribution annotations over the journalistic texts in the TDB is presented. News has the highest number of attributions (which means the total number of samples annotated for attribution source and predicate type) with a total of 307 samples. This reflects its need for frequent citations and source

verification to maintain credibility. Articles follow with 89 attributions, as they often reference external data and expert opinions to support analysis. Interviews have the fewest attributions (27). This may be due to their focus on the perspective of a single individual, resulting in a reduced need for external references. This distribution highlights the varying dependence on sources across these genres.
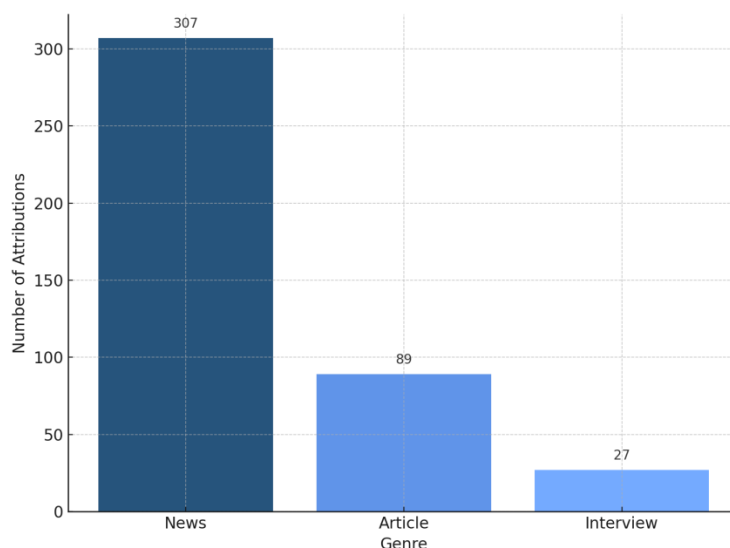
Figure 5. 4. Frequency of attribution features across journalistic genres

The distribution of attribution annotations across the non-journalistic texts in the TDB is analyzed individually for novels, memoirs, and research texts, as shown in Figure 5.5. Novels show the highest number of attribution features, likely influenced by the specific types of novels in the corpus. It can be hypothesized that the genre of novels in the corpus contributes to this higher count of attribution features. Memoirs also exhibit a notable number of attributions, which may be due to their inclusion of personal reflections and insights. Research-examination texts, on the other hand, contain the fewest attributions. This suggests that the type of content and purpose of each genre heavily influence the frequency of attributions.
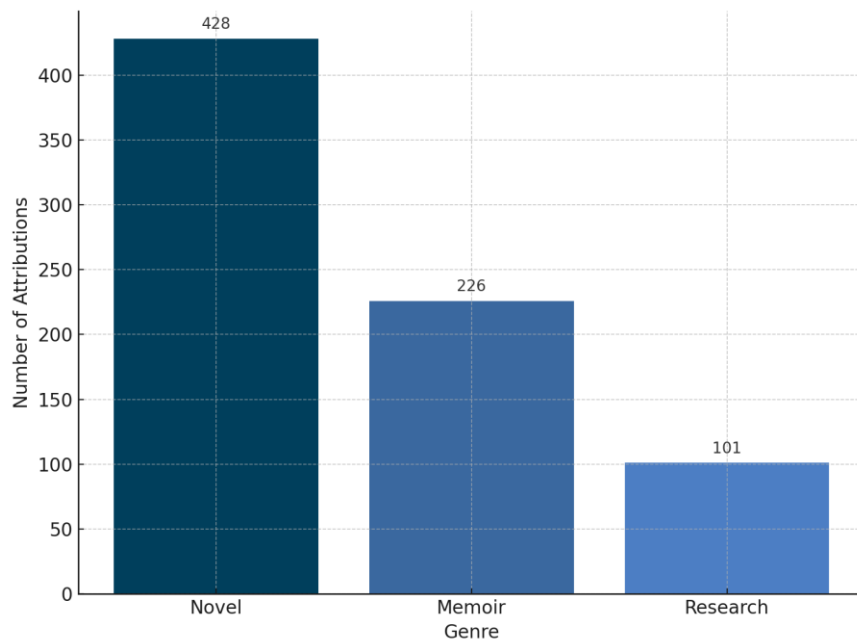
Figure 5. 5. Frequency of attribution features across non-journalistic genres

5.3. Patterns of Attribution in Turkish

One of the central research questions explored in this thesis revolves around identifying the patterns of attribution in Turkish discourse. By patterns, we refer to the structural and functional placement of attribution, including both the source of attribution and type of attribution, within discourse relations: does it appear with Arg1, Arg2, or does it span the entire discourse relation, encompassing both arguments?

The figure below (Figure 5.6) visually explains the possible placements of attribution. The attribution may appear solely with Arg1, solely with Arg2, or, as shown in the section marked REL at the bottom, the entire discourse relation can be attributed to a single source. Understanding these patterns is crucial as it provides insights into how Turkish discourse marks the source of information, which, in turn, affects interpretation and the flow of discourse.
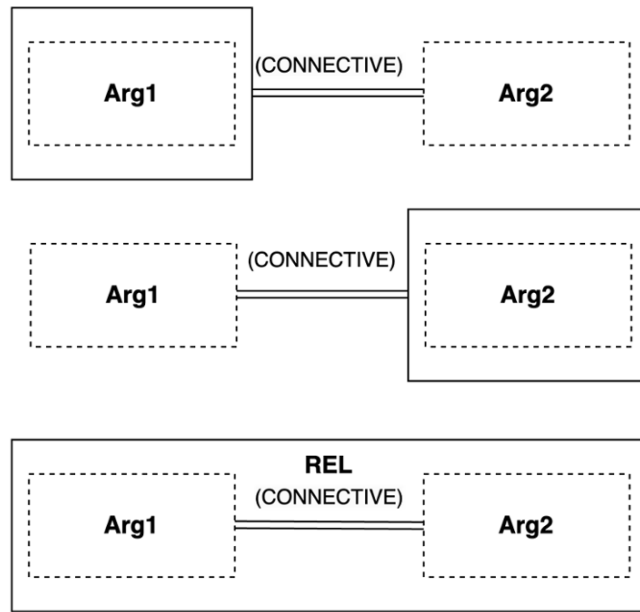
Figure 5. 6. Attribution placement patterns (boxes with undotted lines show the placement of attribution)

The analysis started with the overall distribution and continued with the journalistic and non-journalistic categories.

Based on the total numbers of all the texts, the overall distribution of attribution to Arg1, Arg2 and the entire relation in the corpus is as follows (Figure 5.7). Rel has the highest number of attributions (360) in both journalistic and non-journalistic texts, and that might suggest that a significant portion of attribution is spread across the entire discourse relation. Arg1 and Arg2 also play significant roles, with Arg1 being more common for presenting both the source and content upfront, a pattern that aligns with more narrative styles in non-journalistic texts.
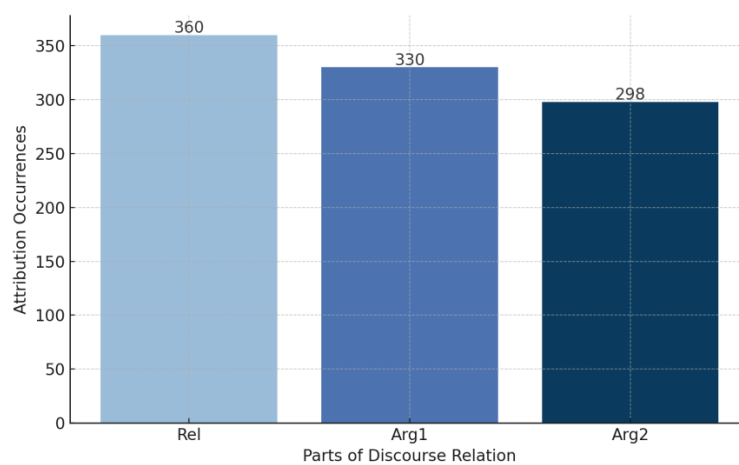
Figure 5. 7. The pattern of attribution throughout all the texts

In journalistic texts, the pattern of attribution in a discourse relation is shown as below (Figure 5.8). The highest number of attributions occurs in the entire discourse relation (Rel), with 199 instances. This suggests that journalistic texts frequently distribute attribution across both Arg1 and Arg2 as a cohesive unit. This pattern likely reflects how journalists structure information to ensure clarity and responsibility in reporting, where both the source and the type of attributed statement form a complete discourse relation.
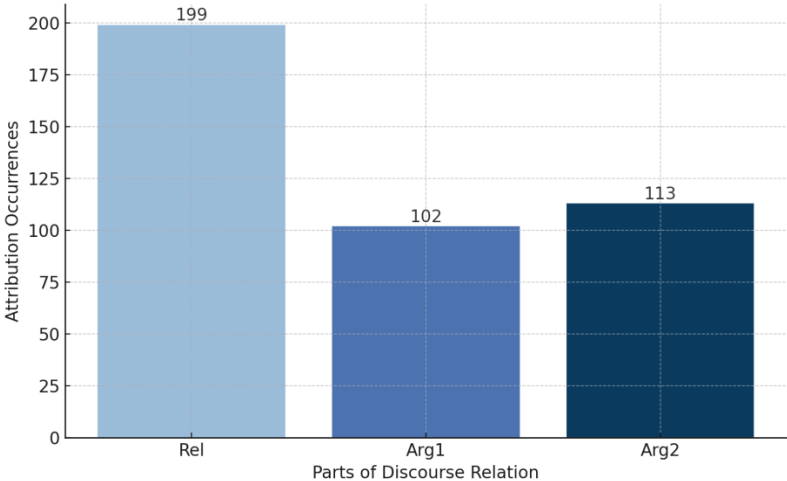
Figure 5. 8. The pattern of attribution in journalistic texts

In non-journalistic texts, the pattern of attribution in a discourse relation is shown as below (Figure 5.9). The highest number of attribution occurrences appears in Arg1, showing a strong preference for introducing the source of information and the statement in this component of the relation.
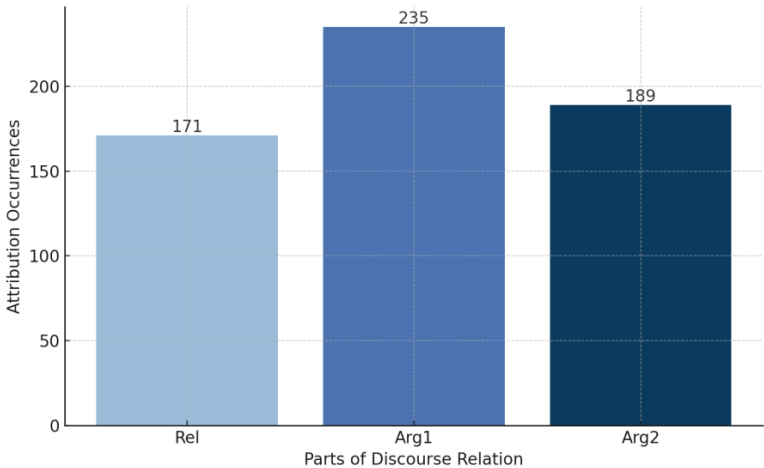
Figure 5. 9. The pattern of attribution in non-journalistic texts

The attributions annotated in this study, as detailed in Appendix B, and their distribution across the corpus are a significant part of the findings of this work. They improve the understanding of the various ways of attribution in Turkish discourse.

This appendix includes a comprehensive list of Turkish that are commonly used to attribute ideas, statements, or thoughts to various sources, each accompanied by their English translation. Beyond predicate types of attribution such as *açıkla (*'explain'*) anlat- (*'tell'*)*, we also annotated attributive phrases, by which we mean cues other than verbs such as 'according to'. In Turkish, they range from those marked by a converb such as -ArAk ('by'), as in *düşünerek* ('by thinking'), - IncE ('when') as in *yanıt verince* ('when responding') to those used to attribute information to a source without a specific predicate such as *-a/e göre* ('according to'). Additionally, phrases such as *X'in de düşündüğü gibi* ('As X also thought') and X *bildiren bir mektup* ('a letter stating X'), used to indicate agreement with someone's thoughts or to convey information stated in a different form of communication, are annotated. By cataloging these phrases, the thesis enhances the toolkit available to researchers for analyzing Turkish texts.

Lastly, the distribution of attribution features across different discourse relations (Explicit, Implicit and Altlex) is examined. While there are 35 different variations of attribution that occur with Explicit discourse connectives, 52 attributions were found with Implicit connectives and only 12 attributions in Altlex. By attribution variation, we refer to the unique sets of feature values created by different combinations of source and type arrangements within Arg1, Arg2, and Rel. Significantly, Implicit connectives not only show the most varied ways of expressing attribution but also have more distinct attribution fatures and a higher total count of attributions compared to Explicit and Altlex connectives. Therefore, it can be concluded that Implicit connectives play an important role in conveying attributions within discourse, surpassing both Explicit and Altlex connectives in the diversity and frequency of attribution expressions. The details of this analysis can be examined in appendix C, D and E.

Overall, this work has enhanced the understanding of the complexity involved in annotating and analyzing discourse. The frequent use of communicative verbs points to a dynamic and direct style of communication in Turkish discourse. Moreover, the data showed that attribution frequently occurs in Rel, with the highest overall occurrences, particularly in journalistic texts, where clarity and accountability are essential. All in all, the findings add useful knowledge about Turkish discourse that can help create better text analysis tools, which might work for other similar languages.The study highlights important tendencies in terms of forms and patterns of attribution in Turkish. This could inform future research and practical applications in language technology.

CHAPTER 6

CONCLUSION

6.1. Summary

In this thesis, I explored attribution in Turkish by identifying source of attribution and type of attribution in accordance with the PDTB guidelines. I examined attribution within the TDB dataset.

In summary, this thesis presented the analysis in five chapters. In the introduction, I outlined the objectives and scope of the thesis. I detailed how this thesis would fill the gap in the literature of attribution in Turkish discourse.

Chapter 2 provides a comprehensive literature review and establishes the work within the context of attribution analysis. Since attribution is relevant in many fields, I presented studies in psychology, computation, and the Turkish language, starting with key discourse models such as the PDTB and RST.

In Chapter 3, I detailed the methodology employed in this research by adapting and extending the PDTB annotation guidelines to align with Turkish. When annotating the data, we read each text sentence by sentence, and whenever attribution was found in a discourse relation, we annotated all relevant features by considering the context. Source of attribution and type of attribution were categorized by prioritizing their semantic meanings within the context. In this chapter, I also elaborated on the use of the TDB dataset and its expansion to include annotations for attribution and readdressed the gap in the study of Turkish discourse.

Chapter 4 focuses on the evaluation and finalization of the dataset, where I ensured the data's accuracy and reliability. Employing the Cohen's Kappa coefficient, I measured the inter-annotator agreement, and reflected the precision and consistency of the annotation process. In this chapter, I explained the steps taken to refine the dataset and its potential as a resource for future linguistic research.

In Chapter 5, I analyzed the TDB data annotated for attribution and presented the findings. I found trends in the source and type of attribution in Turkish texts, and presented them quantitatively and qualitatively. Also, I highlighted the differences in the distribution of attribution across genres. Moreover, I revealed distinct differences in how attribution is distributed to components of discourse relations, often encompassing both Arg1 and Arg2 or appearing in entire relations.

Overall, this thesis shows how the PDTB guidelines can be applied to Turkish discourse effectively, as demonstrated by the reliably annotated TDB dataset. The high consistency scores confirm its accuracy. Therefore, this thesis helps future studies in computational linguistics and cross-linguistic research by showing the complexities of Turkish discourse.

6.2. Discussion of the Results

This thesis aimed to uncover patterns of attribution in Turkish texts, focusing on both journalistic and non-journalistic genres in the TDB 1.2. The analysis helped answer the research questions presented at the beginning of the study.

The first research question aimed to find out the distribution pattern of attribution to different genres in the corpus Turkish texts. The results showed clear differences between journalistic and non-journalistic genres. In journalistic texts, such as news articles and interviews, attribution often covers the entire discourse relation (Rel), with a heavy use of communicative verbs. This shows that journalistic writing relies on clearly identifying sources and making sure the information comes from credible, external voices.

On the other hand, in non-journalistic texts like novels and memoirs, attribution tends to appear in Arg1, where the source is introduced before further details are given. This reflects the narrative style of these genres, where the focus is often on establishing who is speaking or providing the information at the start, before elaborating on their statement.

However, despite these differences, the pattern of attribution across the entire relation (Rel) was observed consistently across both genre types. This suggests that, regardless of genre, Turkish discourse often integrates attribution in a manner that encompasses the full discourse relation.

The second research question focused on identifying the linguistic markers that signify attribution in Turkish texts. The results showed that communicative verbs (*de-*, 'to say'; *söyle-*, 'to tell') were the most frequent markers, especially in journalistic texts. This prevalence underscores the direct and explicit nature of attribution in Turkish journalistic discourse, where the source and the action of communication are made clear to ensure transparency and accuracy in reporting.

Non-journalistic texts, however, exhibited greater variation in the use of attribution verbs. In addition to communicative verbs, verbs expressing intention, thought, or perception (*iste-*, 'to want'; *anla-*, 'to understand'; *belirt-*, 'to point out') were also common. This suggests a broader range of attribution strategies in genres like fiction and memoirs, where attribution often serves to convey subjective thoughts, beliefs, or attitudes.

6.3. Contributions to the Field

Firstly, this work work adapted and applied the PDTB annotation framework to Turkish texts. The PDTB guidelines were expanded to involve the morphological and semantic nuances of Turkish, so it resulted in a more comprehensive set of guidelines for analyzing attribution in Turkish discourse. These methodological extensions provide a thorough approach for examining Turkish attribution.

Secondly, the absence of extensive research on attribution in Turkish presented a clear gap in the field. By filling this gap, this thesis contributes to Turkish linguistic studies and offers new insights and frameworks for analysis.

Thirdly, I created a rich dataset with fully annotated instances of attribution. This sets a strong foundation for future research. The dataset is crucial for developing computational tools that can automatically detect and analyze attribution in Turkish texts. It paves the way for significant advancements in both linguistic research and natural language processing applications

6.4. Limitations

The thesis has several contraints, including limitations tied to the Annotator tool's default settings for attributions. The tool automatically sets the source as writer and the type as assertion. This automatic setting could lead to inaccuracies in the data, particularly in computational analysis in whcih finer distinctions are necessary. In the tool, when annotating attribution for entire relations that consist of a phrase without a verb such as *according to* there is no option to mark it appropriately. This is because a Null category exists only within the argument (Arg1 and Arg2) verb type categories, not for the entire relation section. Consequently, even though phrases corresponding to 'according to' do not contain a verb yet signify an attribution, they are inaccurately tagged as Comm.

Moreover, the IAA only involves two annotators and that may affect the reliability of the data. A more comprehensive IAA process with multiple annotators would enhance the validity of the analysis. These findings emphasize the need for more adaptable and detailed tools to properly capture the complexities of annotating language data.

6.5. Future Directions

This thesis highlights several areas for future research that could extend the current understanding of attribution in Turkish discourse. First, expanding the corpus to include a broader array of genres and text types could deepen insights into how attribution is employed across different contexts in Turkish. Additionally, conducting comparative studies to analyze attribution practices across various languages would provide different perspectives on cross-linguistic patterns. Lastly, the application of these findings in computational linguistics could lead to the development of

automated models specifically designed for detecting and analyzing attribution in Turkish texts.

6.6. Concluding Remarks

In conclusion, this thesis not only extends our understanding of Turkish discourse through analytical methods but also contributes to the broader field of computational linguistics by providing a methodological framework that is unique to the specific features of Turkish language. The insights gained from this work promise to enhance both academic research and practical applications. Therefore, it leads to future innovations in natural language processing technologies.

REFERENCES

Afantenos, S. D., Asher, N., Benamara, F., & Taboada, M. (2012). Discourse parsing for multi-party chat dialogues. *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, 156-164.

Almeida, M. S., Almeida, M. B., & Martins, A. F. (2014, April). A joint model for quotation attribution and coreference resolution. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 39-48).

Al-Saif, A., Alyahya, T., Alotaibi, M., Almuzaini, H., & Algahtani, A. (2018, May). Annotating attribution relations in Arabic. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Asher, N., & Asher, N. (1993). From events to propositions: a tour of abstract entities, eventualities and the nominals that denote them. *Reference to abstract objects in discourse.*

Biran, O., & McKeown, K. (2015, September). Pdtb discourse parsing as a tagging task: The two taggers approach. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*.

Bloor, M., & Bloor, T. (2013). *The practice of critical discourse analysis: An introduction*. Routledge.

Borsley, R. D. (1996). *Modern phrase structure grammar*. Blackwell.

Caldas-Coulthard, C. R. (1994). On reporting reporting: The representation of speech in factual and fictive narratives. In M. Coulthard (Ed.), *Advances in written text analysis* (pp. 295–308). Routledge.

Carlson, L., & Marcu, D. (2001). Discourse tagging manual. *ISI Technical Report ISI-TR-545*. University of Southern California, Information Sciences Institute.

Can, Ö., & Ercan, G. S. (2020). Reported Speech Types and their Discursive Functions in Turkish Op-Ed Articles. *Dilbilim Arastirmalari Dergisi*, *31*(2).

Cervone, A., Pareti, S., Bell, P., Prodanof, I., & Caselli, T. (2014). Detecting attribution relations in speech: a corpus study. *Detecting attribution relations in speech: a corpus study*, 103-107.

Chomsky, N. (1993). *Lectures on government and binding: The Pisa lectures* (No. 9). Walter de Gruyter.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, *20*(1), 37-46.

Cowper, E. A. (1992). *A concise introduction to syntactic theory: The government-binding approach*. University of Chicago Press.

Demberg, V., Asr, F. T., & Scholman, M. (2017). How compatible are our discourse annotations? insights from mapping rst-dt and pdtb annotations. *arXiv preprint arXiv:1704.08893*.

Den Ouden, J. N., Van Wijk, C. H., Terken, J. M. B., & Noordman, L. G. M. (1998). Reliability of discourse structure annotation. *IPO Annual Progress Report*, *33*, 129-138

Elson, D., & McKeown, K. (2010, July). Automatic attribution of quoted speech in literary narrative. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 24, No. 1, pp. 1013-1019).

Fausey, C. M., & Boroditsky, L. (2010). Subtle linguistic cues influence perceived blame and financial liability. *Psychonomic Bulletin & Review, 17*(5), 644–650.

Fernandes, W. P. D., Motta, E., & Milidiú, R. L. (2011). Quotation extraction for portuguese. In *Proceedings of the 8th Brazilian Symposium in Information and Human Language Technology*.

Göksel, A., & Kerslake, C. (2004). *Turkish: A comprehensive grammar*. Routledge.

Güven, M. (2023). Çevrimiçi Haberlerde Aktarım: Kimin Sözü? Kimin Sesi? [Transfer in Online News: Whose Words? Whose Voice?]. 36. Ulusal Dilbilim Kurultayı'nda Sunulmuş Bildiri, Erciyes Üniversitesi, Kayseri, Türkiye.

He, H., Barbosa, D., & Kondrak, G. (2013). Identification of speakers in novels. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1312–1320). Association for Computational Linguistics.

Heider, F. (1958). *The Psychology of Interpersonal Relations*. Wiley.

Hintikka, J., & Hintikka, J. (1969). *Semantics for propositional attitudes* (pp. 87-111). Springer Netherlands.

Janicki, M., Kanner, A., & Mäkelä, E. (2023, May). Detection and attribution of quotes in Finnish news media: BERT vs. rule-based approach. In *Proceedings of the 24th Nordic Conference on Computational Linguistics (NoDaLiDa)* (pp. 52-59).

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. *Advances in Experimental Social Psychology, 2*, 219-266.

Joshi, A.K. (2010, December 15-17). Dependencies at the Sentence Level and at the Discourse Level [Conference presentation]. CLARA WORKSHOP, Prague.

Karttunen, L. (1971). Some observations on factivity. *Research on Language & Social Interaction*, *4*(1), 55-69.

Kelley, H. H. (1967). Attribution theory in social psychology. *Nebraska Symposium on Motivation, 15*, 192-238.

Kiparsky, P., & Kiparsky, C. (1971). Fact. In D. D. Steinberg & L. A. Jakobovits (Eds.), *Semantics: An interdisciplinary reader in philosophy, linguistics, and psychology* (pp. 345–369). Cambridge University Press.

Krestel, R., Bergler, S., & Witte, R. (2008). Minding the source: Automatic tagging of reported speech in newspaper articles. *Reporter*, *1*(5), 4.

Kunduracı, Aysun. 2008. Syntax and Semantics on Control in Turkish. The 14th International Conference on Turkish Linguistics (Tam Metin Bildiri/Sözlü Sunum)

Kuo, S. H. (2001). Reported speech in Chinese political discourse. *Discourse Studies*, *3*(2), 181-202.

Lee, A., Prasad, R., Webber, B., & Joshi, A. (2016, December). Annotating discourse relations with the PDTB annotator. In *Proceedings of COLING 2016, The 26th International Conference on Computational Linguistics: System Demonstrations* (pp. 121-125).

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

Lin, N. (2017). Building a network theory of social capital. *Social capital*, 3-28.

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

Mamede, N., & Chaleira, P. (2004). Character identification in children stories. In *Advances in Natural Language Processing: 4th International Conference, EsTAL 2004, Alicante, Spain, October 20-22, 2004. Proceedings 4* (pp. 82-90). Springer Berlin Heidelberg.

Marcus, M. P., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2), 313–330.

Muzny, G., Fang, M., Chang, A., & Jurafsky, D. (2017, April). A two-stage sieve approach for quote attribution. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers* (pp. 460-470).

Newell, C., Cowlishaw, T., & Man, D. (2018). Quote extraction and analysis for news. In *Proceedings of the Workshop on Data Science, Journalism and Media, KDD* (pp. 1-6).

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review, 108*(2), 291. DOI: 10.1037/0033-295X.108.2.291

O'Keefe, T., Pareti, S., Curran, J. R., Koprinska, I., & Honnibal, M. (2012, July). A sequence labelling approach to quote attribution. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning* (pp. 790-799).

Quirk, R., Greenbaum, S., Leech, G., Svartvik, J. (1985) A Comprehensive Grammar of the English Language. London: Longman.

Pardo, T. A. S., & Nunes, M. G. V. (2003). A framework to support discourse analysis. *Proceedings of the 4th SIGdial Workshop on Discourse and Dialogue*, 80-89.

Pardo, T. A. S., Nunes, M. G. V., Rino, L. H. M., & Souza, M. L. (2004). Rhetorical structure theory and the analysis of discourse. *Proceedings of the 5th Discourse Anaphora and Anaphor Resolution Colloquium (DAARC)*, 93-98.

Pareti, S. (2012, May). A Database of Attribution Relations. In *LREC* (pp. 3213-3217).

Pareti, S., O'keefe, T., Konstas, I., Curran, J. R., & Koprinska, I. (2013, October). Automatically detecting and attributing indirect quotations. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (pp. 989-999).

Pareti, S. (2015). Attribution: a computational approach.

Pareti, S., & Prodanof, I. (2010, May). Annotating Attribution Relations: Towards an Italian Discourse Treebank. In *LREC*.

Peterson, C., Semmel, A., von Baeyer, C., Abramson, L. Y., Metalsky, G. I., & Seligman, M. E. P. (1982). The Attributional Style Questionnaire. *Cognitive Therapy and Research, 6*(3), 287-299.

Potter, A. (2019). *The rhetorical structure of attribution*. In A. Zeldes, D. Das, E. M. Galani, J. D. Antonio, & M. Iruskieta (Eds.), *Proceedings of the Workshop on Discourse Relation Parsing and Treebanking 2019* (pp. 38–49). Association for Computational Linguistics.

Prasad, R., Miltsakaki, E., Dinesh, N., Lee, A., Joshi, A., & Robaldo, L. (2007). The Penn Discourse Treebank 2.0 Annotation Manual.

Redeker, G., & Egg, M. (2006, July). Says who? On the treatment of speech attributions in discourse structure. In *Proceedings of the Workshop Constraints in Discourse* (pp. 140-146).

Richardson, J. E. (2007). *Analysing newspapers: An approach from critical discourse analysis*. Palgrave Macmillan.

Riemsdijk, H. C. V., & Williams, E. (1986). Introduction to the Theory of Grammar.

Ruppenhofer, J., Somasundaran, S., & Wiebe, J. (2008, May). Finding the sources and targets of subjective expressions. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)* (pp. 2781-2788). ELRA.

Rutherford, A., Demberg, V., & Xue, N. (2017, April). A systematic study of neural discourse models for implicit discourse relation. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*.

Pouliquen, B., Steinberger, R., & Best, C. (2007, September). Automatic detection of quotations in multilingual news. In *Proceedings of Recent Advances in Natural Language Processing*.

Sag, I. A., & Pollard, C. (1991). An integrated theory of complement control. *Language*, 63-113.

Sarmento, L., Nunes, S., Teixeira, J., & Oliveira, E. (2009, September). Propagating fine-grained topic labels in news snippets. In *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology* (Vol. 3, pp. 515-518). IEEE.

Scheible, C., Klinger, R., & Padó, S. (2016, August). Model architectures for quotation detection. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1736-1745).

Scholman, M. C. J. (2019). Coherence relations in discourse and cognition: comparing approaches, annotations and interpretations.

Skadhauge, P. R., & Hardt, D. (2005). Syntactic identification of attribution in the RST Treebank. In *Proceedings of the Sixth International Workshop on Linguistically Interpreted Corpora (LINC-2005)*.

Slobin, D., & Aksu-Koç, A. (1982). Tense, aspect and modality in the use of the Turkish evidential.–Paul Hopper (ed.), Tense-Aspect: Between Semantics and Pragmatics.

Sener, N. (2011). Semantics and pragmatics of evidentials in Turkish.

Stede, M., Taboada, M., & Das, D. (2017). Annotation guidelines for rhetorical structure. *Manuscript. University of Potsdam and Simon Fraser University*.

van Riemsdijk, H., & Williams, E. (1986). *Introduction to the theory of grammar*. Basil Blackwell.

Webber, B. L., & Joshi, A. K. (1998). Anchoring a lexicalized tree-adjoining grammar for discourse. *arXiv preprint cmp-lg/9806017*.

Webber, B., Prasad, R., Lee, A., & Joshi, A. (2019). The penn discourse treebank 3.0 annotation manual. *Philadelphia, University of Pennsylvania*, *35*, 108.

Weiner, B. (1974). *Achievement motivation and attribution theory*. Morristown, NJ: General Learning Press.

Weiner, B. (1985). *An attributional theory of achievement motivation and emotion*. New York, NY: Springer-Verlag.

Wiebe, J., Wilson, T., & Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language resources and evaluation*, *39*, 165-210.

Zeyrek, D., & Kurfalı, M. (2017, April). TDB 1.1: Extensions on Turkish discourse bank. In *Proceedings of the 11th Linguistic Annotation Workshop* (pp. 76-81).

Zeyrek, D., & Er, M. E. (2022). A description of Turkish Discourse Bank 1.2 and an examination of common dependencies in Turkish discourse.

Zhang, Y., & Liu, Y. (2021). DirectQuote: A Dataset for Direct Quotation Extraction and Attribution in News Articles. *arXiv preprint arXiv:2110.07827*.

APPENDICES

APPENDIX A: POTENTIAL ATTRIBUTION PREDICATE TYPES SEARCHED IN THE DATA

A.1 Communication Verbs (Comm Verbs)

Table A. 1. Communication verbs

English Verb	Turkish Equivalent	Turkish Synonyms
Say	söylemek	demek, ifade etmek
Mention	bahsetmek	söz etmek
Claim	iddia etmek	ileri sürmek
Argue	tartışmak	savunmak, ispat etmek
Explain	açıklamak	izah etmek
Inform	bilgilendirmek	haberdar etmek
Suggest	önermek	teklif etmek
Describe	tanımlamak	tarif etmek
Propose	teklif etmek	önermek
Concede	kabul etmek	teslim etmek
Complain	şikayet etmek	yakınmak
Deny	inkar etmek	reddetmek
Promise	söz vermek	taahhüt etmek
Warn	uyarmak	ikaz etmek
Apologize	özür dilemek	af dilemek
Conclude	sonuçlandırmak	sonuca varmak
Acknowledge	kabul etmek	itiraf etmek
Reveal	ortaya çıkarmak	açığa vurmak
Confirm	doğrulamak	onaylamak
Demand	talep etmek	istemek
Inquire	sormak	araştırmak
Boast	övünmek	böbürlenmek
Criticize	eleştirmek	tenkit etmek
Comment	yorumlamak	yorum yapmak

English Verb	Turkish Equivalent	Turkish Synonyms
Refuse	reddetmek	geri çevirmek
Agree	aynı fikirde olmak	kabul etmek
Disagree	fikir ayrılığına düşmek	katılmamak
Insist	ısrar etmek	diretmek
Object	itiraz etmek	karşı çıkmak
Request	rica etmek	talep etmek
Remind	hatırlatmak	anımsatmak
Report	rapor vermek	bildirmek
Reprimand	azarlamak	kınamak
Threaten	tehdit etmek	korkutmak
Urge	ısrar etmek	tavsiye etmek
Advocate	savunmak	desteklemek
Congratulate	tebrik etmek	kutlamak
Disapprove	onaylamamak	tasvip etmemek
Discuss	tartışmak	görüşmek
Estimate	tahmin etmek	değer biçmek
Imply	ima etmek	anlam çıkarmak
Reassure	güvence vermek	rahatlatmak
Recommend	tavsiye etmek	önermek
Rejoice	sevinmek	mutlu olmak
Emphasize	vurgulamak	altını çizmek
Exaggerate	abartmak	büyütmek
Foresee	öngörmek	tahmin etmek
Insinuate	ima etmek	üstü kapalı söylemek
Qualify	şart koşmak	sınırlamak

A.2. Belief Proposition AOs (PAtt Verbs)

Table A. 2. Belief proposition verbs

English Phrase	Turkish Meaning	Turkish Synonyms
believe	inanmak	güvenmek, kabul etmek
think	düşünmek	sanmak, tasarlamak
expect	beklemek	ummak, tahmin etmek

English Phrase	Turkish Meaning	Turkish Synonyms
suppose	varsaymak	farzetmek, zannetmek
imagine	hayal etmek	düşlemek, tasavvur etmek
hope	ummak	temenni etmek, arzulamak
decide	karar vermek	karara varmak
figure	şekil, rakam; anlamak	biçim, sayı; kavramak
predict	tahmin etmek	önceden söylemek, kestirmek
project	tasarlamak	plan; düzenlemek
speculate	spekülasyon yapmak	tahminde bulunmak
feel	hissetmek	duymak
double	iki katına çıkmak; çift	katlamak; ikili
anticipate	önceden görmek	beklmek, ummak
suspect	şüphelenmek	kuşkulanmak
theorize	teori oluşturmak	kuram geliştirmek
conclude	sonuçlandırmak	bitirmek, tamamlamak
estimate	tahmin etmek	değer biçmek
fear	korkmak	endişelenmek
reckon	hesaplamak; düşünmek	saymak; sanmak
seemed to believe	inanmış gibi görünmek	-
X's argument/belief/assumption/belief is	X'in argümanı/inancı/varsayımı/inancı şu ki	-
X saw (perceived) other problems	X başka sorunlar gördü (algıladı)	-
it proves X [that...]	bu, X'in [...] olduğunu kanıtlar	-
X is convinced/concerned/confident/sure	X ikna olmuş/endişeli/emin/emniyetli [ki...]	-
X interpreted something to be...	X bir şeyi ... olarak yorumladı	-
X is optimistic that	X, ... konusunda iyimser	-
X is advancing a chilling casuistry that	X, ... olduğuna dair ürpertici bir safsata ileri sürüyor	-
X has taken the view that	X, ... görüşünü benimsemiş	-
under the theory	teorinin altında	-
the theory (idea) is that	teori (fikir) şu ki	-

English Phrase	Turkish Meaning	Turkish Synonyms
opinion is mixed [that]	görüşler karışık [ki]	-
rumors have been [that]	söylentiler var [ki]	-

A.3. Eventuality AOs (Ctrl Verbs)

A.3.1. ORDER/PERMITtype [object control]:

Table A. 3. Order/permit type object control verbs

English Verb	Turkish Definition	Turkish Synonyms
order	emir vermek	sipariş etmek, düzenlemek, komuta etmek
persuade	ikna etmek	kandırmak, inandırmak, razı etmek
bid	teklif vermek	teklif etmek, fiyat vermek, davet etmek
charge	suçlamak	görevlendirmek, ücret talep etmek, doldurmak
command	komuta etmek	emretmek, yönetmek, talimat vermek
direct	yönlendirmek	doğrudan etkilemek, emretmek, yönetmek
enjoin	emretmek	yasaklamak, buyurmak, talimat vermek
instruct	talimat vermek	öğretmek, yönergeler vermek, talimat vermek
advise	tavsiye vermek	öğüt vermek, bilgilendirmek, önermek
authorize	yetkilendirmek	onaylamak, izin vermek, belgelemek
mandate	görevlendirmek	zorlamak, emretmek, talimat vermek
convince	ikna etmek	inandırmak, kandırmak, güven sağlamak
impel	zorlamak	itmek, sebep olmak, harekete geçirmek
induce	sebep olmak	teşvik etmek, ikna etmek, harekete geçirmek
influence	etkilemek	etki etmek, tesir etmek, etkisinde bırakmak
inspire	ilham vermek	aşılamak, canlandırmak, coşturmak
motivate	motive etmek	harekete geçirmek, teşvik etmek, ilham vermek
move	harekete geçirmek	sarsmak, taşımak, duygulandırmak
pressure	baskı yapmak	zorlamak, baskı uygulamak, etkilemek
prompt	hızlandırmak	teşvik etmek, hatırlatmak, hızlandırmak
sway	sarsmak	etkilemek, yönlendirmek, hareket ettirmek
stir	harekete geçirmek	kışkırtmak, heyecanlandırmak, çalkalamak
talk (into)	ikna etmek	konuşarak ikna etmek, ikna etmek için konuşmak
compel	zorlamak	mecbur etmek, baskı yapmak, zorunda bırakmak
press	baskı yapmak	sıkıştırmak, zorlamak, ısrar etmek

English Verb	Turkish Definition	Turkish Synonyms
propel	itmek	harekete geçirmek, ilerletmek, iteklemek
push	itmek	zorlamak, bastırmak, sevk etmek
spur	teşvik etmek	tetiklemek, kışkırtmak, uyarıcı olmak
encourage	cesaretlendirmek	teşvik etmek, özendirmek, canlandırmak
exhort	uyarmak	nasihat vermek, öğütlemek, teşvik etmek
goad	teşvik etmek	dürtmek, kışkırtmak, harekete geçirmek
incite	tahrik etmek	kışkırtmak, hızlandırmak, tetiklemek
prod	dürtmek	teşvik etmek, zorlamak, ileri itmek
urge	ısrar etmek	teşvik etmek, tavsiye etmek, çağrıda bulunmak
bring	getirmek	getirmek, sebep olmak, yol açmak
lead	öncülük etmek	yönetmek, yol göstermek, takip etmek
signal	sinyal vermek	işaret vermek, belirtmek, haber vermek
ask	Talep etmek	talep etmek, rica etmek, istemek
empower	yetki vermek	güçlendirmek, izin vermek, yetkilendirmek
appeal (to)	başvurmak	çağrıda bulunmak, hitap etmek, başvuruda bulunmak
dare	cesaret etmek	cüret etmek, cesaretlendirmek, meydan okumak
defy	meydan okumak	karşı gelmek, itiraz etmek, meydan okumak
beg	yalvarmak	rica etmek, istemek, yalvarışta bulunmak
prevent (from)	engellemek	önlemek, engel olmak, engelleme yapmak
forbid	yasaklamak	men etmek, izin vermemek, engellemek
allow	izin vermek	kabul etmek, müsaade etmek, onaylamak
permit	izin vermek	müsaade etmek, onaylamak, izinli kılmak
enable	olanak sağlamak	mümkün kılmak, yetkili kılmak, etkinleştirmek
cause	neden olmak	sebep olmak, yol açmak, tetiklemek
force	zorlamak	mecbur etmek, baskı yapmak, kuvvet kullanmak

A.3.2. PROMISE type [subject control]:

Table A. 4. Promise type subject control verbs

English Verb	Turkish Definition	Turkish Synonyms
promise	söz vermek	vaat etmek, taahhüt etmek, garanti vermek
swear	yemin etmek	ant içmek, and içmek, yemin etmek
agree	kabul etmek	aynı fikirde olmak, anlaşmak, mutabık kalmak

English Verb	Turkish Definition	Turkish Synonyms
contract	sözleşme imzalamak	anlaşma yapmak, taahhüt etmek, anlaşma sağlamak
pledge	söz vermek	taahhüt etmek, vaat etmek, kefalet vermek
vow	yemin etmek	and içmek, ant içmek, yemin etmek
try	denemek	çaba göstermek, uğraşmak, gayret etmek
intend	niyet etmek	amaçlamak, planlamak, hedeflemek
refuse	reddetmek	geri çevirmek, reddetmek, kabul etmemek
choose	seçmek	tercih etmek, seçilmek, karar vermek
decline	reddetmek	geri çevirmek, kabul etmemek, ret etmek
decide	karar vermek	karar vermek, hüküm vermek, belirlemek
demand	talep etmek	istemek, gerektirmek, emretmek
endeavor	çabalamak	gayret etmek, uğraşmak, çalışmak
attempt	denemek	girişimde bulunmak, çaba göstermek, denemek
threaten	tehdit etmek	korkutmak, ürkütmek, gözdağı vermek
undertake	üstlenmek	görevlendirmek, işe girişmek, yüklenmek
propose	teklif etmek	önermek, teklif sunmak, teşvik etmek
offer	teklif etmek	sunmak, önermek, vermek
aim	hedeflemek	amaçlamak, hedefe yönelmek, niyetlenmek

A.3.3. WANT/EXPECT type [subject control]:

Table A. 5. Want/expect type subject control verbs

English Verb	Turkish Definition	Turkish Synonyms
want	istemek	arzulamak, gereksinim duymak, ihtiyaç duymak
desire	arzu etmek	istemek, özlemek, tutkuyla istemek
fancy	hoşlanmak	sevmek, beğenmek, hayran olmak
wish	dilemek	temenni etmek, arzu etmek, istemek
ache	acı çekmek	sızlamak, ağrımak, hatta özlem duymak
hanker	özlemek	istemek, arzulamak, çok istemek
itch	kaşınmak	kaşınmak, tahriş olmak, istek duymak
long	özlemek	hasret çekmek, özlemek, istemek
need	ihtiyaç duymak	gerekmek, gereksinim duymak, zorunlu olmak
hope	umut etmek	umutlanmak, ummak, umuda kapılmak
thirst	susamak	susamak, özlem duymak, hasret çekmek

English Verb	Turkish Definition	Turkish Synonyms
yearn	özlem duymak	istemek, arzulamak, hasret çekmek
hate	nefret etmek	düşman olmak, hoşlanmamak, sevmemek
aspire	arzulamak	istemek, özlemek, yüksek hedeflere sahip olmak
expect	beklemek	ummak, beklemek, tahmin etmek

A.3.4. Turkish Control Verbs, Their Complement Type & Case by Aysun Kunduracı (2008)

Table A. 6. Turkish control verbs by Aysun Kunduracı (2008)

Verb	English Approximation
başar-	succeed
dene-	try
becer-	manage, accomplish
çalış-	work, strive
iste-	desire, request
uğraş-	struggle, deal with
alış-	get used to
utan-	be ashamed
unut-	forget
düşün-	think
bık-	get tired
diren-	resist
heveslen-	be enthusiastic
bırak-	leave, abandon

A.3.4.1. Subject-Control Verbs:

Table A. 7. Subject control verbs by Aysun Kunduracı (2008)

Verb	English Approximation
um-	hope
başla-	begin
boşla-	empty
çabala-	strive

çekin-	hesitate
kalkış-	attempt
arzula-	desire
hedefle-	target
hoşlan-	like
öğren-	learn
sev-	love
bil-	know
arzu et-	wish for
ümit et-	hope for
tercih et-	prefer
inat et-	persist
ihmal et-	neglect
nefret et-	hate
hakim ol-	dominate
ikna ol-	be convinced
mecbur ol-/kal-	be obliged
zorunda ol-/kal-	be forced
üzere ol-	be about to
karar ver-	decide
vazgeç-	give up
göze al-	take into consideration
söz ver-	promise

A.3.4.2. Object-Control Verbs:

Table A. 8. Object control verbs by Aysun Kunduracı (2008)

Verb	English Approximation
zorla-	force
yasakla-	forbid
yönlendir-	direct
ikna et-	convince
mecbur et-/bırak	compel/leave
zorunda bırak-	make obliged
emret-	command
menet-	prohibit

kurtar-	save, rescue
söz ver-	promise

A.4. Factive AOs (Ftv Verbs)

Table A. 9. Factive Verbs

Verb	Turkish Meaning	Turkish Synonyms
Regret	Pişman olmak	Hatasını anlamak, üzgün hissetmek, arzu ettiği bir şeyi yapamadığı için üzülmek
Forget	Unutmak	Hatırlamamak, aklından çıkarmak, hatırlamamak
Remember	Hatırlamak	Unutmayın, anımsamak, bellekte tutmak
Hear	İşitmek	Duymak, dinlemek, kulak vermek
Know	Bilmek	Anlamak, farkında olmak, tanımak
See	Görmek	Bakmak, izlemek, seyretmek
Report	Rapor etmek	Bildirmek, haber vermek, rapor sunmak
Significant	Önemli	Ciddi, etkileyici, mühim
Odd	Garip	Tuhaf, anormal, alışılmadık
Tragic	Trajik	Acıklı, üzücü, dramatik
Exciting	Heyecanlı	Heyecan verici, coşkulu, eğlenceli
Relevant	İlgili	Alakalı, konuyla ilgili, önemli
Matter	Önemli olmak	Mühim olmak, etkili olmak, hesaba katılmak
Count	Saymak	Önemsemek, hesaba katmak, saygı göstermek
Makes sense	Anlam ifade etmek	Mantıklı olmak, anlaşılır olmak, tutarlı olmak
Suffice	Yeterli olmak	Kâfi gelmek, yetmek, yeterli olmak
Amuse	Eğlendirmek	Keyif vermek, güldürmek, neşelendirmek
Bother	Rahatsız etmek	Canını sıkmak, rahatsız etmek, dert vermek
Be aware (of)	Farkında olmak	Farkında olmak, haberdar olmak, bilinçli olmak
Comprehend	Anlamak	Kavramak, idrak etmek, anlaşılmak
Take into	Dikkate almak	Düşünmek, hesaba katmak, göz

Verb	Turkish Meaning	Turkish Synonyms
consideration		önünde bulundurmak
Take into account	Hesaba katmak	Dikkate almak, göz önünde bulundurmak, hesaplamak
Bear in mind	Akılda tutmak	Hatırlamak, unutmamak, aklında bulundurmak
Ignore	İgnor etmek	Görmezden gelmek, umursamamak
Forget (about)	Unutmak	Unutmak, aklından çıkarmak, hatırlamamak
Deplore	Kınamak	Üzülmek, acımak, lanetlemek
Resent	Kızmak	Tepki göstermek, öfkelenmek, içerlemek
Care (about)	Önemsemek	Dikkate almak, umursamak, ilgilenmek
Make clear	Açıklamak	Anlatmak, netleştirmek, aydınlatmak
Mind	Dikkat etmek	Farkında olmak, aldırmak
Be glad	Memnun olmak	Mutlu olmak
Be sorry	Üzülmek	Üzgün olmak, Özür dilemek
Be surprised	Şaşırmak	Şaşkına dönmek

APPENDIX B: PREDICATE TYPES AND ATTRIBUTIONAL PHRASES EXTRACTED FROM THE CORPUS

This appendix presents an organized overview of the various types of predicates and attributional phrases extracted from the analyzed corpus. The predicates are divided into different categories based on their usage in conveying communication, control over events, expression of beliefs, and factual knowledge. Additionally, specific attributional phrases that play a role in reporting or referencing statements and opinions are listed.

The tables below contain Turkish verbs and phrases commonly used for attribution, along with their English translations, and their respective frequencies as observed in the corpus.

B.1. Communication Verbs

Table B. 1. Communication verbs extracted from the corpus

Turkish Verb	English Equivalent
Açıkla-	To explain
Anlat-	To tell
Atıl-	To leap; to throw oneself
Bağır-	To shout
Belirt-	To point out
Bildir-	To inform
De-	To say
Demekten kendisini alama-	To not help saying
Dikkat çek-	To attract attention
Dile getir-	To express; to articulate
Duyur-	To announce
Ekle-	To add
Eleştir-	To criticize
Formüle et-	To formulate

Turkish Verb	English Equivalent
Göster-	To show
Haber ver-	To notify
Haykır-	To scream
Hatırlat-	To remind
İddia et-	To claim
İfade et-	To express
İleri sür-	To put forward
İşaret et-	To point out
Karşılık ver-	To respond
Kaydet-	To record
Konuş-	To speak
Mesaj ilet-	To convey a message
Mesaj ver-	To give a message
Mırıldan-	To murmur
Ortaya koy-	To put forth
Rica et-	To request
Sevinç çığlıkları at-	To shriek with joy
Sonuca var-	To conclude
Soru getir-	To pose a question
Sor-	To ask
Söz al-	To take the floor
Söz et-	To mention

Turkish Verb	English Equivalent
Sözü kap-	To interrupt
Söylenmeye başla-	To start to complain
Sustur-	To silence
Tanımla-	To define
Vurgu yap-	To emphasize
Vurgula-	To accentuate
Yaz-	To write
Öne sür-	To suggest
Özetle-	To summarize
İfade et-	To express

B. 2. Eventuality AOs (Ctrl Verbs)

Table B. 2. Eventuality verbs extracted from the corpus

Turkish Verb	English Equivalent
Amaçla-	To aim
Ant iç-	To swear an oath
Başar-	To succeed
Çalış-	To try
Hedefle-	To aim
İste-	To want; to wish
Karar ver-	To decide
Mecbur et-	To compel

Turkish Verb	English Equivalent
Mecbur ol-	To be obliged
Planla-	To plan
Sev-	To love
Zorunda ol-	To have to; to be forced

B. 3. Belief Proposition AOs (PAtt Verbs)

Table B. 3. Belief proposition verbs extracted from the corpus

Turkish Verb	English Equivalent
Anla-	To understand
Bil-	To know
Düşün-	To think
İnan-	To believe
İnancında ol-	To hold a belief
Kanıtla-	To prove
Kork-	To fear
San-	To assume
Zannet-	To suppose
Görüşünde ol-	To be of the opinion

B. 4. Factive AOs (Ftv Verbs)

Table B. 4. Factive verbs extracted from the corpus

Turkish Verb	English Equivalent
Anımsa-	To recall; to remember

Anlaşıl-	To be understood
Bil-	To know
Duy-	To hear
Farkında ol-	To be aware of
Gör-	To see
Hatırla-	To remember
Öğren-	To learn
Şaşır-	To be surprised
Görül-	To be seen

B. 5. Other Cues for Attribution: Attributional Phrases

Table B. 5. Attributional phrases extracted from the corpus

Turkish Phrase	English Translation
Düşünerek	by thinking
X-a/-e göre	according to X
söyleyince	when said
yanıt verince	when responding
anlattığı sırada	while telling
bildiren X	X who states
soru	question
X şöyledir	X is as follows
X'in de düşündüğü gibi	As X also thought
-mış	-ed (past participle marker)
demekti	it meant
dedikten sonra	after saying
deyip	saying
şöyle bir konuşma geçmişti	such a conversation had taken place
savunan X	X that defends
öğrendiğimde	when I learned
demesi üzerine	upon his/her saying
iddia eden X	X who claims
belirterek	by indicating

Turkish Phrase	English Translation
anlatan X	X who tells
dikkat çeken X	X who attracts attention
X (sonuçlarından birkaçı) şöyledir	X (a few of the results) are as follows
diyen X	X who says
kaydederek	by recording
planlarken	while planning
X'in sözleri	words of X
Anımsatan X	X who reminds
Belirten X	X who indicates
Söylerken	While saying
Maksadı	Its purpose
X (görüşler/ifadeler) şöyle	X (opinions/statements) are as follows
Aynen şöyle	is exactly as follows
X: "…"	X: "…"
ifade eden X	X who expresses
ifade edilen X'te (raporda/mektupta)	in the X (report/letter) that is stated

B. 6. Frequency of the Attribution Features Annotated in this Thesis

Table B. 6. Frequency of the attribution features annotated in this thesis

Turkish Verb	English Translation	Frequency
De-	To say	211
Söyle-	To tell	88
Belirt-	To point out	56
X (Görüşler/ifadeler) şöyle	Statements/expressions are like this	38
İste-	To want	17
Anla-	To understand	14
X: "…"	X: "…"	14
Devam et-	To continue	14
Açıkla-	To explain	13
Sor-	To ask	12
Vurgula-	To emphasize	11
Özetle-	To summarize	11
Öğren-	To learn	10
Anlat-	To narrate	10

Turkish Verb	English Translation	Frequency
İddia et-	To claim	9
Düşün-	To think	8
İleri sür-	To propose	8
Konuş-	To speak	8
Bildir-	To report	8
Karşılık ver-	To respond	8
Kaydet-	To record	8
Belirten X	X who indicates	8
Savun-	To defend	7
İfade et-	To express	7
Anımsa-	To remember	7
Çalış-	Try	7
Şaşır-	To be surprised	7
Karar ver-	To decide	6
Aynen şöyle	is exactly as follows	6
X'te (gazetede/ görüşünde/ X'in kuramında)	In (newspaper/view/in X's theory)	6
X -e/-a göre	According to X	5
Gör-	To see	5
Bil-	To know	5
Öne sür-	To suggest	5
İnan-	To believe	5
Mesajını ver-	To give the message	5
-mış	-ed (past participle marker)	5
Sonuçlara var-	Torrive at conclusions	4
İşaret et-	To indicate	4
Şöyle yaz-	To write like this	4
Oluş-	To form	4
Deyip	Saying	4
Görül-	To be seen	4
Şöyle/şeklinde konuş-	To speak like this	4
Başar-	To succeed	4
Ortaya koy-	To put forward	3
Duy-	To hear	3
Zannet-	To assume	3
Mırıldan-	To murmur	3
Formüle et-	To formulate	3
Demekten kendisini alama-	To not help saying	3

Turkish Verb	English Translation	Frequency
Hatırlat-	To remind	3
Önü kesil-	To be interrupted	3
Kaydederek	By recording	3
Düşünerek	By thinking	2
Dikkat çeken X	X who draws attention	2
Söz et-	To mention	2
Kendi deyimiyle	In his/her own words	2
Kanıtla-	To prove	2
Vurgu yap-	To emphasize	2
Kullan-	To use	2
Savunan X	X that defends	2
Şöyle bir konuşma geç-	To have a conversation like this	2
Eleştir-	To criticize	2
Dile getir-	To state	2
Sözlerine ekle-	To add to his/her words	2
Hedefle-	To target	2
Mesaj ilet-	To convey the message	2
Ol-	To be	2
Kork-	To fear	2
Ant iç-	To swear	2
X (sonuçlarından birkaçı) şöyledir	X (a few of the results) are as follows	2
Soru	To ask a question	2
Ayrıntılan-	To detail	2
Göster-	To show	2
Haykır-	To scream	2
Sevinç çığlıkları at-	To shriek with joy	2
İfade eden X	X who expresses	2
Ekle-	To add	2
Mecbur et-	To compel	2
Dikkat çeken X	X who attracts attention	2
Hatırla-	To remember	2
Söz al-	To take the floor	1
Mecbur ol-	To be obliged	1
Atıl-	To leap	1
Öğrendiğimde	When I learnt	1
Sustur-	To silence	1
Demesi üzerine	Upon his/her saying	1

Turkish Verb	English Translation	Frequency
Zorunda ol-	To be have to	1
İnancında ol-	To hold a belief	1
X'in de düşündüğü gibi	As X also thought	1
Bildiren X	X who states	1
İddia eden X	X who claims	1
Anımsatan X	X who reminds	1
İfade edilen X'te (raporda/mektupta)	In the X (report/letter) that is stated	1
Biçiminde özetlenebilecek	Can be summarized as	1
Maksadı	Its purpose	1
Mecbur ol-	Be obliged	1
Sev-	To love	1
Haber ver-	Inform	1
Belirterek	By indicating	1
Tanımla-	To define	1
Anlatan X	X who tells	1
Başla-	To start	1
X'in de düşündüğü gibi	As X also thought	1
Bekle-	Wait	1
Planlarken	While planning	1
Amaçla-	To aim	1
X'in sözleri	X's words	1
Anlaşıl-	To be understood	1
Farkında ol-	To be aware	1
Duyur-	To announce	1
Söylenmeye başla-	To start to complain	1
Yanıt ver-	To reply	1
Diyen X	X who says	1
Söyleyince	When (someone) says	1
Söylerken	When saying	1
Demekti	It meant	1
Görüşünde ol-	To be of the opinion	1
Sözü kap-	To interrupt	1
Rica et-	To request	1
San-	To assume	1
Şansı tanın-	Be given a chance	1

The appendix presents how different combinations of attribution features are applied to Explicit connectives and their arguments, with the complete breakdown shown in Table C.1. A sum of 35 unique combinations of these feature values is documented in the fourth column's total. Null means that no specific attribution feature is assigned for that field, indicating an absence of a defined value in the annotation.

Table C. 1. Combinations of attribution features for explicit connectives and arguments

Rel	Arg1	Arg2	Total
Arb.Comm.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	8
Arb.Ftv.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	1
Ot.Comm.Null.Null	Arb.Comm.Null.Null	Inh.Null.Null.Null	1
Ot.Comm.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	93
Ot.Comm.Null.Null	Inh.Null.Null.Null	Ot.Comm.Null.Null	1
Ot.Comm.Null.Null	Ot.Comm.Null.Null	Inh.Null.Null.Null	3
Ot.Ctrl.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	1
Ot.Ctrl.Null.Null	Ot.Ctrl.Null.Null	Inh.Null.Null.Null	1
Ot.Ctrl.Null.Null	Ot.PAtt.Null.Null	Inh.Null.Null.Null	2
Ot.Ftv.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	1
Ot.PAtt.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	1
Ot.Patt.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	1
Wr.Comm.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	1057
Wr.Comm.Null.Null	Inh.Null.Null.Null	Inh.PAtt.Null.Null	1
Wr.Comm.Null.Null	Inh.Null.Null.Null	Ot.Comm.Null.Null	21
Wr.Comm.Null.Null	Inh.Null.Null.Null	Ot.Ctrl.Null.Null	2
Wr.Comm.Null.Null	Inh.Null.Null.Null	Ot.Ftv.Null.Null	1
Wr.Comm.Null.Null	Inh.Null.Null.Null	Ot.PAtt.Null.Null	2
Wr.Comm.Null.Null	Inh.Null.Null.Null	Wr.Comm.Null.Null	1
Wr.Comm.Null.Null	Inh.Null.Null.Null	Wr.Ctrl.Null.Null	2
Wr.Comm.Null.Null	Inh.Null.Null.Null	Wr.PAtt.Null.Null	1
Wr.Comm.Null.Null	Inh.PAtt.Null.Null	Ot.Comm.Null.Null	1
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Inh.Null.Null.Null	46
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Ot.Comm.Null.Null	5
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Ot.Ctrl.Null.Null	1
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Ot.Null.Null.Null	1
Wr.Comm.Null.Null	Ot.Ctrl.Null.Null	Inh.Null.Null.Null	15
Wr.Comm.Null.Null	Ot.Ftv.Null.Null	Inh.Null.Null.Null	5
Wr.Comm.Null.Null	Ot.Null.Null.Null	Inh.Null.Null.Null	4
Wr.Comm.Null.Null	Ot.PAtt.Null.Null	Inh.Null.Null.Null	10
Wr.Comm.Null.Null	Ot.Patt.Null.Null	Inh.Null.Null.Null	1
Wr.Comm.Null.Null	Wr.Comm.Null.Null	Inh.Null.Null.Null	2
Wr.Comm.Null.Null	Wr.Ctrl.Null.Null	Inh.Null.Null.Null	2
Wr.Comm.Null.Null	Wr.Ftv.Null.Null	Inh.Null.Null.Null	2
Wr.Comm.Null.Null	Wr.PAtt.Null.Null	Inh.Null.Null.Null	3

APPENDIX D

This appendix presents the variety of attribution feature value sets marked for Implicit connectives and their arguments. The entire distribution is presented in Table C.2, with a sum of 52 unique feature value sets noted.

Table C. 2. Combinations of attribution features for implicit connectives and arguments

Rel	Arg1	Arg2	Total
Arb.Ftv.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	1
Ot.Comm.Null.Null	Arb.Comm.Null.Null	Arb.Comm.Null.Null	1
Ot.Comm.Null.Null	Inh.Comm.Null.Null	Inh.Null.Null.Null	2
Ot.Comm.Null.Null	Inh.Null.Null.Null	Arb.Comm.Null.Null	2
Ot.Comm.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	90
Ot.Comm.Null.Null	Ot.Comm.Null.Null	Inh.Null.Null.Null	1
Ot.Comm.Null.Null	Ot.Ctrl.Null.Null	Inh.Null.Null.Null	1
Ot.Comm.Null.Null	Wr.Comm.Null.Null	Inh.Null.Null.Null	1
Ot.PAtt.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	15
Wr.Comm.Null.Null	Arb.Comm.Null.Null	Inh.Null.Null.Null	4
Wr.Comm.Null.Null	Arb.Ftv.Null.Null	Inh.Null.Null.Null	2
Wr.Comm.Null.Null	Arb.Null.Null.Null	Arb.Null.Null.Null	1
Wr.Comm.Null.Null	Arb.Null.Null.Null	Inh.Null.Null.Null	3
Wr.Comm.Null.Null	Arb.PAtt.Null.Null	Arb.Comm.Null.Null	1
Wr.Comm.Null.Null	Inh.Comm.Null.Null	Ot.Comm.Null.Null	1
Wr.Comm.Null.Null	Inh.Comm.Null.Null	Wr.PAtt.Null.Null	1
Wr.Comm.Null.Null	Inh.Ctrl.Null.Null	Inh.Null.Null.Null	1
Wr.Comm.Null.Null	Inh.Null.Null.Null	Arb.Comm.Null.Null	10
Wr.Comm.Null.Null	Inh.Null.Null.Null	Arb.Ctrl.Null.Null	1
Wr.Comm.Null.Null	Inh.Null.Null.Null	Arb.Ftv.Null.Null	2
Wr.Comm.Null.Null	Inh.Null.Null.Null	Arb.Null.Null.Null	4
Wr.Comm.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	1328
Wr.Comm.Null.Null	Inh.Null.Null.Null	Ot.Ctrl.Null.Null	2
Wr.Comm.Null.Null	Inh.Null.Null.Null	Ot.Comm.Null.Null	59
Wr.Comm.Null.Null	Inh.Null.Null.Null	Ot.Ctrl.Null.Null	5
Wr.Comm.Null.Null	Inh.Null.Null.Null	Ot.Ftv.Null.Null	5
Wr.Comm.Null.Null	Inh.Null.Null.Null	Ot.Null.Null.Null	3
Wr.Comm.Null.Null	Inh.Null.Null.Null	Ot.PAtt.Null.Null	6
Wr.Comm.Null.Null	Inh.Null.Null.Null	Wr.Comm.Null.Null	3
Wr.Comm.Null.Null	Inh.Null.Null.Null	Wr.Ctrl.Null.Null	2
Wr.Comm.Null.Null	Inh.Null.Null.Null	Wr.Ftv.Null.Null	3
Wr.Comm.Null.Null	Inh.Null.Null.Null	Wr.PAtt.Null.Null	3
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Arb.PAtt.Null.Null	1
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Inh.Comm.Null.Null	1
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Inh.Null.Null.Null	41
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Ot.Comm.Null.Null	28

Wr.Comm.Null.Null	Ot.Comm.Null.Null	Ot.Ftv.Null.Null	1
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Ot.PAtt.Null.Null	1
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Wr.PAtt.Null.Null	1
Wr.Comm.Null.Null	Ot.Ctrl.Null.Null	Inh.Null.Null.Null	12
Wr.Comm.Null.Null	Ot.Ctrl.Null.Null	Ot.Comm.Null.Null	1
Wr.Comm.Null.Null	Ot.Ctrl.Null.Null	Ot.Ctrl.Null.Null	1
Wr.Comm.Null.Null	Ot.Ftv.Null.Null	Inh.Null.Null.Null	3
Wr.Comm.Null.Null	Ot.Null.Null.Null	Inh.Null.Null.Null	5
Wr.Comm.Null.Null	Ot.PAtt.Null.Null	Inh.Null.Null.Null	12
Wr.Comm.Null.Null	Wr.Comm.Null.Null	Inh.Null.Null.Null	4
Wr.Comm.Null.Null	Wr.Comm.Null.Null	Ot.Comm.Null.Null	2
Wr.Comm.Null.Null	Wr.Ctrl.Null.Null	Inh.Null.Null.Null	4
Wr.Comm.Null.Null	Wr.Ctrl.Null.Null	Ot.Null.Null.Null	1
Wr.Comm.Null.Null	Wr.Ftv.Null.Null	Arb.Comm.Null.Null	1
Wr.Comm.Null.Null	Wr.Ftv.Null.Null	Inh.Null.Null.Null	2
Wr.Comm.Null.Null	Wr.PAtt.Null.Null	Inh.Null.Null.Null	5

The appendix details the range of unique attribution feature sets assigned to Altlex connectives and their corresponding arguments. This complete range is presented below (Table C.3) , and there are 12 distinct feature sets in total.

Table C. 3. Combinations of attribution features for altlex connectives and arguments

Rel	Arg1	Arg2	Total
Ot.Comm.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	4
Wr.Comm.Null.Null	Inh.Comm.Null.Null	Inh.Null.Null.Null	2
Wr.Comm.Null.Null	Inh.Comm.Null.Null	Ot.Comm.Null.Null	1
Wr.Comm.Null.Null	Inh.Null.Null.Null	Inh.Null.Null.Null	116
Wr.Comm.Null.Null	Inh.Null.Null.Null	Ot.Comm.Null.Null	1
Wr.Comm.Null.Null	Inh.Null.Null.Null	Ot.Null.Null.Null	1
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Arb.Comm.Null.Null	1
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Inh.Null.Null.Null	2
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Ot.Comm.Null.Null	1
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Ot.Ctrl.Null.Null	1
Wr.Comm.Null.Null	Ot.Comm.Null.Null	Ot.Null.Null.Null	1
Wr.Comm.Null.Null	Wr.Ctrl.Null.Null	Inh.Null.Null.Null	1