

DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

**DEVELOPMENT OF NLP BASED STORYBOARD
ENGINE**



by
Mehmet CENGİZ

August, 2020
İZMİR

DEVELOPMENT OF NLP BASED STORYBOARD ENGINE

**A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of Dokuz Eylül University
In Partial Fulfillment of the Requirements for the Degree of Doctor of
Philosophy in Computer Engineering**

**by
Mehmet CENGİZ**

August, 2020

İZMİR

Ph.D. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled “**DEVELOPMENT OF NLP BASED STORYBOARD ENGINE**” completed by **MEHMET CENGİZ** under the supervision of **ASSIST. PROF. DR. KÖKTEN ULAŞ BİRANT** and we certify that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Doctor of Philosophy.

.....
Assist. Prof. Dr. Kökten Ulaş BİRANT

Supervisor

.....
Assist. Prof. Dr. Özlem AKTAŞ

Thesis Committee Member

.....
Assist. Prof. Dr. Mete EMİNAĞAOĞLU

Thesis Committee Member

.....
Prof. Dr. Yılmaz KILIÇASLAN

Examining Committee Member

.....
Assoc. Prof. Dr. Emre ÇOMAK

Examining Committee Member

.....
Prof.Dr. Özgür ÖZÇELİK

Director
Graduate School of Natural and Applied Sciences

ACKNOWLEDGMENT

Firstly, I would like to express my sincere gratitude to my advisors Assist. Prof. Dr. Ulaş BİRANT and Assist. Prof. Dr. Samsun M. BAŞARICI for the continuous support of my Ph.D. study and related research, for their patience, motivation, and immense knowledge. Their guidance helped me in all the time of research and writing of this thesis.

Besides my advisor, I would like to thank the rest of my thesis committee: Assist. Prof. Dr. Özlem AKTAŞ, Assist. Prof. Dr. Mete EMİNAĞAOĞLU, Assoc. Prof. Dr. Emre ÇOMAK, and Prof. Dr. Yılmaz KILIÇASLAN for their insightful comments and encouragement, but especially perspectives which incited me to widen my research from various perspectives.

I must express my very profound gratitude to my parents and my brother for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them.

In addition, I thank my student Burak VELİOĞLU from Adnan Menderes University for his contribution and support to this study.

In the end, I have special thanks to my wife Ayşe Betül CENGİZ for her brilliant comments, suggestions, and making me encouraged during the development and writing phase of the thesis. Thank you.

Mehmet CENGİZ

DEVELOPMENT OF NLP BASED STORYBOARD ENGINE

ABSTRACT

Storyboards are essential to ensure that all visual elements appear in a compatible way according to the scenario. At this point, it should be discussed what might be a visual element. By choosing the easiest way, all objects can count as a visual element; but their facades may change due to the time interval of the main event. Storyboards are generally created using texted scenarios. To extract an object and their actions, Natural Language Processing Techniques can be used. Even if this information from the scenario is obtained, storyboard creation is a laborious process that requires technical knowledge. For this reason, storyboard creators must know the drawing tools well.

Knowledge extraction is a hard process. In this dissertation, the main knowledge that should be extracted is roles, which are known as thematic roles. On the other hand, word placements should be examined in accordance with their places in the sentence. This analysis is done using grammatical roles. Thus, in this thesis, Natural Language Processing is used to find thematic and grammatical roles.

In this study, it is aimed that storyboard creators will direct their talents to creativity rather than using these tools through a new tool that will be developed. Development progress of the tool contains two main stages which are Natural Language Processing and drawing of the extracted objects and its actions/motions on a storyboard.

Keywords: Natural language processing, phrase structures, thematic roles, English, NLP, storyboard

DOĞAL DİL İŞLEME TEMELLİ GÖRSEL SENARYO TASLAĞI MOTORU GELİŞTİRİLMESİ

ÖZ

Görsel senaryo taslakları, tüm görsel öğelerin senaryoya göre uyumlu bir şekilde görünmesini sağlamak için gereklidir. Bu noktada, nelerin görsel bir unsur olabileceği tartışılmalıdır. En kolay yolu seçerek, tüm nesneler görsel bir öğe olarak sayılabilir; ancak dış görünüşleri ana olayın zaman aralığı nedeniyle değişebilir. Görsel senaryo taslakları genellikle yazılı senaryolar kullanılarak oluşturulur. Bu nedenle nesneyi ve eylemlerini ayıklamak için Doğal Dil İşleme Teknikleri kullanılabilir. Senaryodan bu bilgiler elde edilse bile, bu görsel taslakları oluşturma işlemi teknik bilgi gerektiren zahmetli bir süreçtir. Bu nedenle, görsel senaryo taslağı yaratıcıları çizim araçlarını iyi bilmelidir.

Bilgi çıkarma zor bir süreçtir. Bu tezde çıkarılması gereken temel bilgi tematik roller olarak bilinen rollerdir. Öte yandan, kelime yerleşimleri de cümledeki yerlerine göre incelenmelidir. Bu analiz dil bilgisel rolleri kullanılarak yapılır. Dolayısıyla bu tezde, tematik ve dil bilgisel rolleri bulmak için Doğal Dil İşleme kullanılmaktadır.

Bu çalışmada, görsel senaryo taslağı içerik oluşturucularının geliştirilecek yeni bir araçla bu araçları kullanarak zaman kaybetmeleri yerine yeteneklerini yaratıcılığa yönlendirmeleri hedeflenmektedir. Aracın geliştirme süreci, Doğal Dil İşleme ve çıkarılan nesnelerin çizimi ve eylemleri/hareketleri bir görsel senaryo taslağı üzerinde olan iki ana aşama içerir.

Anahtar sözcükler: Doğal dil işleme, sözcük grupları, tematik roller, İngilizce, DDİ, görsel senaryo taslakları

CONTENTS

	Page
Ph.D. THESIS EXAMINATION RESULT FORM	ii
ACKNOWLEDGEMENTS	iii
ABSTRACT	iv
ÖZ	v
LIST OF FIGURES.....	ix
LIST OF TABLES	x
 CHAPTER ONE - INTRODUCTION	 1
1.1 Overview	1
1.2 Aim of Thesis	1
1.3 Brief Information about Storyboards	3
1.4 Brief Information about Natural Language Processing and Linguistics	4
 CHAPTER TWO – STORYBOARDS AND SCENARIOS.....	 11
 CHAPTER THREE – GRAMMAR THEORIES	 18
 CHAPTER FOUR – METHOD.....	 27
4.1 Parts of Speech	28
4.2 Phrase Structures	31
4.3 Grammatical and Thematic Roles	32
4.3.1 Grammatical Roles	32
4.3.2 Thematic Roles	34
4.4 Archetype Analysis	38
 CHAPTER FIVE - MODEL	 39
5.1 Controlled Natural Language	39

5.1.1 Selected Word Categories	39
5.1.1.1 Preposition Category	39
5.1.1.2 Verb Category	40
5.1.1.3 Word Category	41
5.1.1.4 Adword Category	41
5.1.2 Phrase Structures	41
5.1.2.1 Sentence (S).....	42
5.1.2.2 Determiner Phrase (DP).....	42
5.1.2.3 Noun Phrase (NP).....	43
5.1.2.4 Adjective Phrase (AdjP)	45
5.1.2.5 Adverbial Phrase (AdvP).....	45
5.1.2.6 Prepositional Phrase (PP)	46
5.1.2.7 Tense Phrase (TP).....	47
5.1.2.8 Verb Phrase (VP)	47
5.1.2.9 Auxiliary Phrase (AuxP)	47
5.1.2.10 Brief Notes about TP, VP, and AuxP	48
5.1.2.11 Algorithm of Branches of Basic Phrases	48
5.2 Grammatical and Thematic Roles	50
5.3 Class Hierarchy and Complexity of the Proposed Model	50
5.4 Implementation	51
5.4.1 Implementation of Phrase Structures.....	52
5.4.2 Thematic Role Stacks	54
5.4.3 Implementation of Storyboard	61
5.4.4 Implementation of Archetype Analysis	64
5.4.5 Complexity of Drawing Process.....	65
5.5 Results	66
 CHAPTER SIX – CONCLUSION	 70
6.1 Conclusion.....	70
6.2 Future Works.....	71

REFERENCES	73
APPENDICES	80
Appendix A: Word Types	80
Appendix A.1 Verbs	80
Appendix A.2 Nouns	81
Appendix A.3 Pronouns	82
Appendix A.4 Determiners	82
Appendix A.5 Auxiliary Verbs	84
Appendix A.6 Adjectives	84
Appendix A.7 Adverbs	85
Appendix A.8 Interjections	85
Appendix A.9 Clitics	86
Appendix A.10 Contractions	86
Appendix A.11 Adpositions	87
Appendix A.12 Conjunctions	88
Appendix A.13 Coverbs	89
Appendix A.14 Preverbs	89
Appendix A.15 Particles	89
Appendix A.16 Classifiers	90
Appendix A.17 Cardinal Numbers	90
Appendix B: Full Form of the Developed Phrase Structure Tree	91
Appendix C: Sentence Examples to Test the Developed Method	92

LIST OF FIGURES

	Page
Figure 1.1 Syllable structure tree	6
Figure 1.2 Morpheme tree of the word “unkindness”	6
Figure 2.1 Possible key object locations	17
Figure 3.1 Classification of grammar types	18
Figure 3.2 Pictorial representations of preposition “over”	19
Figure 3.3 An example of HPSG notation	23
Figure 3.4 Trees of the sentence “the cat easily escaped”	24
Figure 3.5 Full TAG tree.....	25
Figure 4.1 Lattice structure of parts of speech	29
Figure 4.2 The Hasse diagram of locatives sets	37
Figure 4.3 The Hasse diagram of figurals sets	38
Figure 5.1 Two cases of determination phrase.....	43
Figure 5.2 There cases of noun phrase.....	44
Figure 5.3 Full Adjective phrase sub-tree	46
Figure 5.4 Full tense phrase tree	48
Figure 5.5 The result of stack operations according to given sentences	57
Figure 5.6 Three static scenes example of the sentence “the girl walked from the kitchen to the room in home”	62
Figure 5.7 One static scene example of the sentence “the girl saw the dog”	63
Figure 5.8 One animated scene example of the sentence “the dog is chasing the cat”	63
Figure 5.9 One static scene example of the sentence “the cat hid under the table” ..	64

LIST OF TABLES

	Page
Table 4.1 Chomsky’s word analysis according to his 1970 work.....	29
Table 4.2 Chomsky’s early work on phrase structure rules.....	31
Table 4.3 Analysis of “Colorless green ideas sleep furiously”	31
Table 5.1 Noun cases in English and Turkish	40
Table 5.2 Effects of the concept of tense on locatives.....	56



CHAPTER ONE

INTRODUCTION

1.1 Overview

This dissertation is about creating storyboards using natural language processing techniques. In this section, a general introduction has been made about the topics of the thesis, and literature information has been given. Therefore, general information about storyboard and natural language processing is given below. Note that, in this dissertation, English is chosen as the object language. The examples and explanations given are always in English unless otherwise stated.

In this chapter of the thesis, general information about the storyboard, natural language processing, and linguistics are presented. In the second chapter, the storyboard is explained with its history. It is also explained how a storyboard is designed. In the third chapter, we discuss grammar theories. In the model we created, we developed a new approach inspired by grammar theories. In Chapter 4, we describe all the methods we use in this study. In Chapter 5, we explain how we used the methods we explained in Chapter 4 in our study. In the last chapter, Chapter 6, we discuss the results, contributions, and future works of our study.

1.2 The Aim of the Thesis

In this study, we put forward the thesis that a written text can be visualized as a storyboard using natural language processing techniques.

Nowadays, storyboard creation is a laborious process that requires technical knowledge. For this reason, storyboard creators must be knowledgeable and talented at using drawing tools well. Even if that is the case, the creation process may still take a long time. If some automated drawing tools can be created, storyboard creators can shift their capabilities to other processes.

When doing so, the texted scenario should be examined. With the help of natural language processing (NLP) techniques, scenarios can be examined in a way that a human would do. NLP has wide application areas and these areas are becoming more and more popular. Generally, NLP is mostly used for:

- Machine Translation (MT): MT aims to translate one natural language (NL) into another. A well-known example of MT is the Google Translate project (Wikipedia, 2019b).
- Natural Language Generation (NLG): NLG tries to convert data and operation results in natural languages that experts can understand (Reiter & Dale, 1997).
- Speech Studies: Speech studies include speech recognition, text to speech conversation, speech to text conversation, etc. Each of these studies examines speech signals and tries to turn them into texts.
- Summarization: This application area aims to create a shortened version of a text by a computer program (Mani, 2001).

The main goal of this study is to develop an engine that parses a given scenario and creates a storyboard. Another goal of this study is to develop a mathematical model that connects the syntax level to the semantic and pragmatic levels. Thus, this thesis can be an application for NLG, MT, and summarization areas. Also, the input language of the developed engine is chosen as English.

Although we mainly use NLP techniques, we have many other approaches in our study. Thematic role and archetype concepts come first. We created our phrase structure trees, inspired by grammar theories. These trees are tested using a controlled natural language. We examined the phrases, which are obtained with the help of our PST, with the stack data structure that is created according to our approach. Thus, we tried to achieve thematic roles in the most accurate way. Finally, we made archetypal analyzes by spreading thematic role analysis to the context. So a relatively new technique is proposed to visualize a text in the natural language as a storyboard including some semantic and pragmatic matters.

Some approaches have been exhibited in order to realize the thesis we have put forward. The biggest contribution that this study creates to the literature is the blending of NLP, computer science, and software engineering techniques with linguistic and literary concepts.

1.3 Brief Information about Storyboards

Sometimes writers illustrate their works, whether they write scripts or novels. These illustrations help authors to see their plot's progress. The concept of telling a story through a series of sequential drawings goes back to Egyptian hieroglyphics, even back to the cave man's drawings of stampeding cattle (Hart, 2008). The usage of sequential drawings to tell stories can be called a storyboard. The Bayeux Tapestries (1066), which is a picture sequence of William the Conqueror's invasion of England, is the first storyboard to be officially accepted (Wikipedia, 2019a). Therefore, it can be said that storyboards are a narrative prototype. They narrate a scenario illustrating how the use cases and functional requirements are played out in the context of a user's actual tasks (Arnowitz, Arent, & Berger, 2007). Almost all storyboards are created for scenarios in the movie industry.

'Scenario' was introduced into the common language as a term to describe a movie setting (Coates, 2000). Storyboards are essential to ensure that all visual elements appear in a compatible way according to the scenario. This study focuses on the visualization of prepared scenarios.

As aforementioned, storyboards illustrate visual elements of scenarios. By the light of this information, it should be discussed what might be a visual element. By choosing the easiest way, all objects can count as a visual element, but their facades may change during the time interval of the main event.

On the other hand, storyboards should also show actions, motions in some way. To make a very general definition, motion is the displacement of an object in space over time and it is a continuous phenomenon. Motion is an objective entity and can be

recorded with a motion capture system. Action is a goal-directed motion (or force) sequence, for example picking up a stone from the ground, playing a piano tone (Jensenius, 2018).

1.4 Brief Information about Natural Language Processing and Linguistics

This dissertation uses written scenarios. Thus, Natural Language Processing techniques are used for object extraction and action/motion detection from a text. “Natural Language Processing (NLP)”, which is one of the application areas of computer technologies, can be defined as the construction of a computing system that processes and understands human natural language (Aktaş, 2010). Before discussing NLP, a little piece of information about the areas of linguistics and computational linguistics should be given. This study focuses on the syntactic level but at some points, there is an obligation to address phenomena concerning the semantic or pragmatic level.

Computational linguistics (CL) is the study of linguistics using computational techniques. The goal is to learn new knowledge about natural languages. NLP is the creation of tools, algorithms, databases that solve problems. The goal is to build useful tools that work on natural languages. While CL is about science, NLP is about engineering. On the other hand, CL is the superset that encompasses NLP.

Although NLP has a history of about 60 years, there is still much to do in this field of study. We still cannot easily and efficiently communicate with machines in natural language. To overcome this shortcoming, NLP needs to work with the linguistics field. An unbounded study cannot be expected. For this reason, this thesis also includes researches in the field of linguistics such as work on phrase structures.

Linguistics is the scientific study of natural languages. Linguistics has six subfields: phonetics, phonology, morphology, syntax, semantics, and pragmatics. As a result of a personal interview with Yılmaz Kılıçaslan, he proposes to add a seventh subfield between the semantics and the pragmatics to this list: the study of computational

ontologies, which addresses issues concerning how to formally define shared conceptualizations.

Phonetics is the study of speech signals. The main concern is the physical aspects of human sounds. Phonetics is concerned with the physical properties of sounds and focuses on the production and classification of speech sounds. The production stage takes place in vocal organs, for example, lips, tongue, throat, etc. There is an alphabet named International Phonetic Alphabet (IPA), which is a framework that uses a single symbol to describe each distinct sound in the language and can be found in dictionaries and textbooks worldwide. In the classification stage, the focus is on the sorting of speech sounds in accordance with IPA. IPA is an alphabet system that holds Latin-like alphabet equivalents of phonemes. For example, the word “cash” has four letters according to the Latin alphabet, but in IPA, it is presented as three sounds: k æ ʃ.

Phonology is the study of the use of sounds and sound patterns to generate meaningful expressions. It concerns phonemes which are the smallest unit of speech distinguishing one word or word element. Phonology also examines acceptable sound combinations and differences of these combinations from other languages. The equivalents of sound in a language are examined by the phonology stage. The syllables are also formed in this stage. In language, syllables come together to produce the smallest independent element, words.

A syllable consists of three parts: an onset, a nucleus, and a coda. Nucleus and coda are grouped into rimes. In some languages, the onset is optional. The nucleus is the obligatory element of a syllable. It is a vowel. The coda is a suffix that follows the nucleus. Like the onset, it is optional in some languages. Figure 1.1 shows the constituent structure of a syllable.

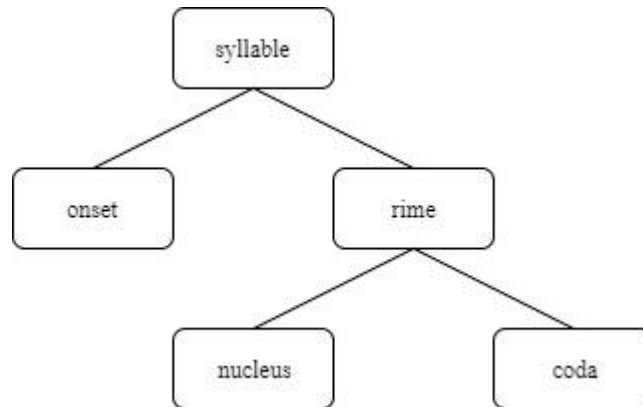


Figure 1.1 Syllable structure tree

Morphology is concerned with the internal structure of words. A word consists of morphemes, the smallest meaningful units. There are two types of morphemes: free and dependent. Free morphemes can appear alone as a simple word. Dependent morphemes must be attached to another morpheme as an affix. For example, “kind” is a free morpheme; “un-” and “-ness” in “unkind” and “unkindness” are dependent morphemes. Likewise, syllables can be shown using a tree structure such as in Figure 1.2.

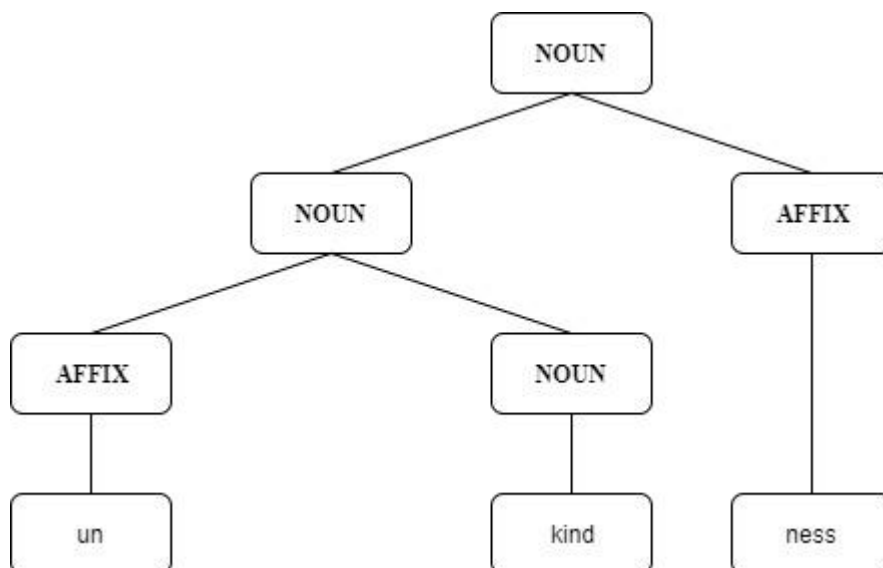


Figure 1.2 Morpheme tree of the word “unkindness”

Note that, morphemes and syllables are different concepts. A syllable is a sound unit; a morpheme is a meaning unit. Some morphemes also are syllables. For example,

the word “unfortunately” has five syllables: *un*, *for*, *tu*, *nate*, and *ly*. However, it has four morphemes, which are “un”, “fortune”, “ate”, “ly”.

The syntax is the study of word sequences, the principles and rules of sentence formation in natural languages. Also, it examines the flexibility of sentences in the language. Syntax imposes structural rules on the language and requires that sentences be formed following these rules. In the traditional grammar of English, there are four types of sentences:

- Simple Sentence: It has a subject-verb format such as “The boy jumped.”
- Compound Sentence: Compound sentences have more than one subject or verb as in “The boy jumped and the girl sang.” Compound sentences consist of a combination of two simple sentences that are connected with conjunctions.
- Complex Sentence: This type of sentence contains a subordinating clause such as “The boy jumped even though he was little.”
- Compound-Complex Sentence: These sentences contain two independent and more dependent clauses such as “Even though he was little, the boy jumped and he reached the top.”

At this point, syntax and diction should not be confused with each other. Diction is the style of speaking or writing. It includes the accent and the intonation. An accent is an identifiable style of pronunciation which is often varying regionally. Likewise the accent, the intonation is about the phonetics level. It is used for distinguishing one type of phrase or sentence from another.

Semantics is the study of linguistic meaning. The semantics concerns context-independent meanings of words; however, pragmatics concerns the contextual meanings. As aforementioned, Kılıçaslan suggests that the human brain has a lattice structure. This means all words are connected somehow. He argues that the context-independent meaning is first produced and then the contextual meaning is added. In short, the brain resolves a text according to its main meaning at first then the

ontological stage establishes a connection between words and context. As a result of this connection, the pragmatic result appears.

This dissertation mainly works on the syntactic level. Syntax examines the word sequences. In English, word sequences can be defined using some methods. The grammatical relations among words can be defined following the syntactic rules of generative grammars.

The syntax is a productive, recursive, and infinite system (Carnie, 2006). At the syntactic level, we can produce and understand sentences we have never heard before such as Chomsky's example "Colorless green ideas sleep furiously". This case shows us the productivity of the syntax. It is possible to embed a sentence into another sentence recursively. For example, the sentence "John is running" can be embedded into a sentence such as "He saw John is running" and also "I think he saw John is running". Lastly, the syntax is an infinitely productive system, theoretically an adjective, for instance, can be preceded by infinitely many adverbs such as "a very very ... fast car".

The dominant theory of syntax is due to Noam Chomsky and his colleagues, starting in the mid-1950s and continuing long years (Carnie, 2006). Although there are many other names such as Transformational Grammar, Government and Binding Theory, etc., the generative grammar is the syntactic approach of Chomsky, which tries to predict sequences that give combinations of structures that form a grammatical sentence (Müldür, 2016). In the late 1950s, Noam Chomsky introduced a generative grammar theory involving phrase structure rules (Chomsky, 1957). According to this approach, all sentences can be split into phrases such as noun phrases, verb phrases, etc. In the 1980s, Chomsky revised his earlier theories under a new theory named Government and Binding and he proposed his X-Bar Theory the first time (Chomsky, 1982). In 1993, Chomsky proposed a program named the Minimalist Program to eliminate complexities in previous works (Chomsky, 1992).

Many scientists have developed new approaches from Chomsky's studies. In the early 1980s, non-transformational theories such as Lexical-Functional Grammar (Kaplan & Bresnan, 1982), Generalized Phrase Structure Grammar (Gunji, 1986), Head-Driven Phrase Structure Grammar (Pollard & Sag, 1994), and Tree Adjoining Grammar (Joshi & Schabes, 1997) were also developed.

It can be said that NLP is a reflection of linguistics on computer science, thus NLP has some stages nearly as same as linguistics. It consists of four main analysis levels: Morphology, Syntax, Semantics, and Pragmatics.

Morphology deals with word-based analysis such as defining word types and analyzing parts of words. The parts of a word are root, suffix, and prefix. Word stems are also examined in morphological analysis.

The syntactic analysis reviews sentences and focuses on the relationship of the words within a sentence. This level ignores the meanings of words and only analyses sequences of words, thus at this level, word orders and agreements are important.

The semantic analysis discovers the meanings of the words according to the syntactic structure of the sentence. Sometimes, at this level, a syntactic structure may be updated because of misunderstandings.

The pragmatic analysis is the highest and hardest level of the NLP stages. This level finds out meanings as in the semantic level, but word meanings of the pragmatic level depend on the contextual information. For example, the sentence "Dogs howl." If this sentence is in a dog-breeding book, the dog breed is considered but if this sentence is in a novel, a herd of dogs can be considered.

In order to ensure the integrity and a better understanding of the study, two other concepts should be emphasized: toy grammar and controlled natural language (CNL). Toy grammar is one that is developed for particular research interests. It contains a relatively small set of lexical and/or phrasal rules. Developing natural language

applications from scratch may lead to maximal control and understanding, but it is also time-consuming, requires good programming skills rather than insight in natural language phenomena, thus, toy systems are vital (van Eynde, Schuurman, & Schelkens, 1998). Like toy grammar, CNL is based on natural language but applies restrictions on vocabulary, grammar, and/or semantics. Toy grammars often have features that do not occur in a natural language to test the limits of the language. In our framework, we use a controlled natural language which includes a toy lexicon and a toy grammar. A lexicon is a dictionary, which is a set of lexical entries. Lexical entries or lexemes are the minimal distinctive units in the semantic system of a language. The toy lexicon is a customized lexicon for the study to be done. Toy grammars generally ignore some complex rules of languages to prevent the complexity of the study.

CHAPTER TWO

STORYBOARDS AND SCENARIOS

Storytelling is perhaps the first conscious use of language. All cultures have their own stories and also myths. Culture is considered strong with a multitude of stories. All stories have different types of characters, but all cultures have similar types, which are called archetypes.

In 1902, the term “collective unconsciousness” is introduced by Carl Gustav Jung. This term is used to explain some kind of human behaviors that are passed on by ancestors to one's own and affect decisions made in daily life without realizing it (Jung, 2014b). The contents of the collective unconscious are known as archetypes. Archetypes, according to Jung, are model characters that symbolize personalities or behaviors. These characters not only explain personality but also give direction to personalities. Although Jung classifies the archetypes, similar studies date back to Philo Judaeus born in 25 B.C (Judaeus, 1981). Philo Judaeus mentioned an archetype named Imago Dei, in English God-image, in man.

Jung studied the symbols and myths of many different cultures and fourteen archetypes have been found. They are frequently found in the literature. However, it should be noted that there is no complete list of archetypes. We worked on the types we obtained based on Jung's works. It can be said that Odyssey was the first example that he worked on. Archetypes are the Mother, the Child, the Shadow, the Persona, Anima/Animus, the Father, the Hero, the Virgin, the Wise Old Man, the Animal, the Swindler, the Hermaphrodite and, the Self (Başarıcı & Kılıçaslan, 2017).

The hero is an archetype who overcomes obstacles and achieves certain goals. Jung defines the hero as only one who has risked the fight with the dragon and is not overcome by it wins the hoard, the "treasure hard to attain." (Jung, 2014a). Almost all stories have at least one hero. In myths, the hero's goal is to find the treasure, princess, elixir of life, etc. Luke Skywalker from the Star Wars series is a good example of it.

The mother archetype is seen in two different variations. One is a human who is related to the hero. This type of mother can be a personal mother, grandmother, or mother-in-law, etc. The second is the goddess such as Hera from Greek myths, Isis from Egyptian myths. In Christian myths, the goddess figure is known as the Virgin. Except for Christian myths, the mother can help the hero.

The Father takes the form of any god, giant, king, or father, etc. The father can be divided into two types: positive and negative. A positive father guides and protects those who are under his care and is able to put ideas into fruition. Until the moment of death, James Potter from the Harry Potter series was a positive father. On the other hand, a negative father uses his authority for abuse or bad purposes. One of the best-known examples is Don Vito Corleone from The Godfather series.

The Child is a potential future. It represents our efforts to deal with the problem of growing up, aided by the illusion of eternal fiction (Jung, 1964). In more recent years, author Caroline Myss has worked about archetypes and she suggested that the child has seven different subcategories that are Wounded, Orphan, Magical/Innocent, Nature, Eternal, Dependent, and Divine (Myss, 2019). The Wounded Child holds the childhood traumas. The Orphan Child such as Snow White reflects the feeling of not belonging to the family. The Magical or Innocent Child always sees the beauties of life. It is always positive, even if bad events happen such as in Pollyanna. The Nature Child hears nature's calling. Mowgli from The Jungle Book can be an example. The Eternal Boy/Girl guides how to stay young and enjoy life at all times. Peter Pan is the best example of this. The Dependent Child refers to the unfulfilled desires of our lives. Even if The Dependent Child and The Wounded Child are similar, The Dependent Child has no past traumas. The Divine Child is similar to The Magical Child, but it is distinguished from it by its redemptive mission.

As its name suggests, The Shadow is the dark side of the character's personality. It reflects the internal conflicts of those who have stayed in difficult situations. Jung said "One does not become enlightened by imagining figures of light, but by making the darkness conscious." about The Shadow. According to him, the shadow is essential for

the hero to find himself. The best-known duality about personality is Dr. Jekyll and Mr. Hyde. Although an elixir is used in the story, the person's personality transformations are explained in the subtext. And also, Dr. Jekyll developed the elixir to solve the problems that he could not overcome. The shadow can affect all archetypes.

The Persona is the mask worn by the hero inside the social life or public. It is a kind of social adaptation that plays an important role in dealing with peers. Generally, the hero wears different masks for different situations or groups. For example, the DC superhero universe has many masked heroes. These heroes have different personas due to protect their family or people they love.

Women have contra sexuality, which is masculine, in nature, is called The Animus, and vice versa, men have feminine contra sexuality, in nature, that is called The Anima. The anima refers to ideas such as soul, spirit, or vital force, etc. The animus describes ideas such as life, mind, courage, or rationality, etc. According to Jung, each has four different levels (Jung, 2003). The Anima has Eve, Helen, Mary, and Sophia levels. Eve indicates maternity. Helen refers to talent fullness. Mary provides virtue by the perceiving man. Sophia is the level of wisdom. The animus has physical power, romance, clergyman, and spiritual guide levels. The first step of the animas is seen as a physical power such as in Tarzan. In the second level, the animus has the capacity for planned actions. In the third stage, the animus turns a wise man such as the clergyman. The highest level of The Animus is spiritual guidelines as Sophia from the anima. According to Jung, Hermes from Greek mythology, the messenger of the gods is at this level.

The Virgin is associated with purity. Although this term is often attributed to young girls, boys can also be a virgin. In Christian myths, The Virgin Mary has ultimate purity to bring forth the perfect form of male life, a god.

The Wise Old Man is characterized by being old, father-figure with great judgment and wisdom. The Wise Old Man seldom intervenes, but always gives helpful advice.

In The Lord of The Rings series, the Gandalf figure is one of the best examples of this archetype.

The Animal archetype symbolizes the character's instinctual sides. The Animal often symbolizes the shadow sides. It is part of the character's nature. As a feature of culture, each animal has a meaning. For example, snakes are identified with insidiousness. Thus, in many religions, the devil figure has a snake form. On the other hand, the animal also can be seen everywhere in real life such as country or team symbols.

The Swindler is a sidekick character in many stories. Nevertheless, it can be the hero's itself as Robin Hood. Swindlers help the hero when its interests overlap. The good type of swindler can sometimes help unrequited, which does not suit its character. In the Doctor Who television series, Captain Jack Harkness is a good example of this archetype.

The Hermaphrodite archetype represents the union of opposites. It often appears as a male-female combination. Nevertheless, it symbolizes the war-the peace, the light-the shadow.

The Prostitute archetype intimately linked with the hero's self-worth. It can appear as a human or as an undesirable job. The main point here is how much the hero puts this into account.

The Self represents the unified unconsciousness and consciousness of an individual. It often appears in dreams in stories. The character sees itself as a king, hero, or prophet. The self is the ultimate self-realization of a person.

Some characters include more than one archetype. For example, the Jeanne d'Arc figure is both the virgin and the hero. Another point of view is good-bad separation. Archetypes do not indicate good and evil. This means, there is no rule that the hero must be good as in Darth Vader from Star Wars.

Archetypes that are mentioned above are found in myths and religions. All heroes are in a plot in these myths. In 1949, Joseph Campbell examined the heroes' journeys (Campbell, 1972). He was strongly influenced by Jung's ideas. He started to analyze relationships between myths and he discovered some inferences. According to him, there are three stages:

- The hero is called to an adventure. However, he/she refuses the call. After some breaks, he/she reluctantly accepts the call with the help of the wise old man as a mentor. He/she leaves his/her old life behind.
- The hero starts to learn new skills with the help of his/her mentor. In this stage, a goddess or princess often appear. During his/her adventure, the hero is always tested by shadow. Sometimes the hero succumbs to the darkness such as Anakin Skywalker from the Star Wars series. He/she also encounters the prostitute. The prostitute might try to draw him/her to the dark side. At the end of this stage, the hero reaches the culmination of his/her powers uncontrollably, albeit with difficulties.
- In the third stage, the hero refuses to get back his/her old life back. He/she starts to question himself/herself with the help of his/her child archetypes. He/she is stuck between two worlds. At last, he/she becomes the master of his/her abilities and spreads his/her mercy all over his surroundings.

Almost all movies and novels fit into this range of events. For example, the Harry Potter series have Harry Potter as The Hero, have Dumbledore as The Wise Old Man, have Hermione Granger as The Mother or The Virgin, have The Animal as the emblems of each school, and so on.

To set an example for a novel, Hermann Hesse's *Siddhartha* can be given (Hesse, 1922). In this novel, Siddhartha's journey starts for finding his personality. He is the Hero. He has a father who is the leader and the king of a region. Different from other journeys, in this journey, The Father does not want Siddhartha to leave, but Siddhartha resists his father's pressures. He begins his quest with a companion, named Govinda, for a while, which can be count as The Swindler. On the road of his journey, he encounters The Prostitute named Kamala and she teaches him to lust. In some cases,

Kamala guides Siddhartha and it can be told Kamala is The Mother sometimes. After this point, The Shadow, which comes up as a businessman named Kamaswami, starts to affect Siddhartha and he gets some bad behaviors. As his personality deteriorates, he begins to experience internal reckoning. He decides to leave this life and finds Vasudeva, who is a wise ferryman, as The Old Wise Man. After some events, Siddhartha learns he has a boy. He sees himself in this boy and understands somethings. This boy is The Child. During his journey, Siddhartha meets different types of people and events. He has different The Personas for each of them.

A novel or a scenario can be narrated according to the point of view. There are two types of them: first-person and third-person. First or third person narration in event writings such as stories or novels is a classification made according to whether the events pass by the author or not. This study avoids first-person narration.

A scenario indicates situations that could happen. Scenarios are used to describe characters, places, events, etc. in a film. Archetypes must be adhered to during the creation of a new scenario. All existing archetypes should be explained. According to these archetypes, the path of the hero must be drawn. However, in some works, especially in art films, this path may not be obvious as in “2001: A Space Odyssey” although it is somehow visible in the Novel of Arthur C. Clarke.

The storyboard is the premiere preproduction, pre-visualization tool designed to give a frame-by-frame, shot-by-shot series of sequential drawings adapted from the shooting script. They are concept drawings that illuminate and augment the script narrative and enable the entire production team to organize all the complicated action required by the script before the actual filming is done to create the correct look for the finished film (Hart, 2008). Storyboards are generally created using scenarios.

The main purpose of drawing storyboards is to reproduce the three-dimensional reality on a two-dimensional surface. Storyboards can be likened to graphs in some way. Nodes are scenes or episodes; the edges specify transitions between them (Jantke & Knauf, 2005). In here, there are two concepts, which are scene and episode. The

main difference between them is their complexity. Scenes are primitive and episodes are groups of scenes. This study concerns only scenes.

Usually, all scenes are drawn into one frame separately. To create each frame, illustrators generally use the rule of thirds to proportion the scene. Rule of thirds says that all scenes should be divided into thirds either vertically or horizontally. After that, all the important objects should be placed along the imagery thirds lines or around their intersections (Mai, Le, Niu, & Liu, 2011). Figure 2.1 shows where the key elements should be placed.

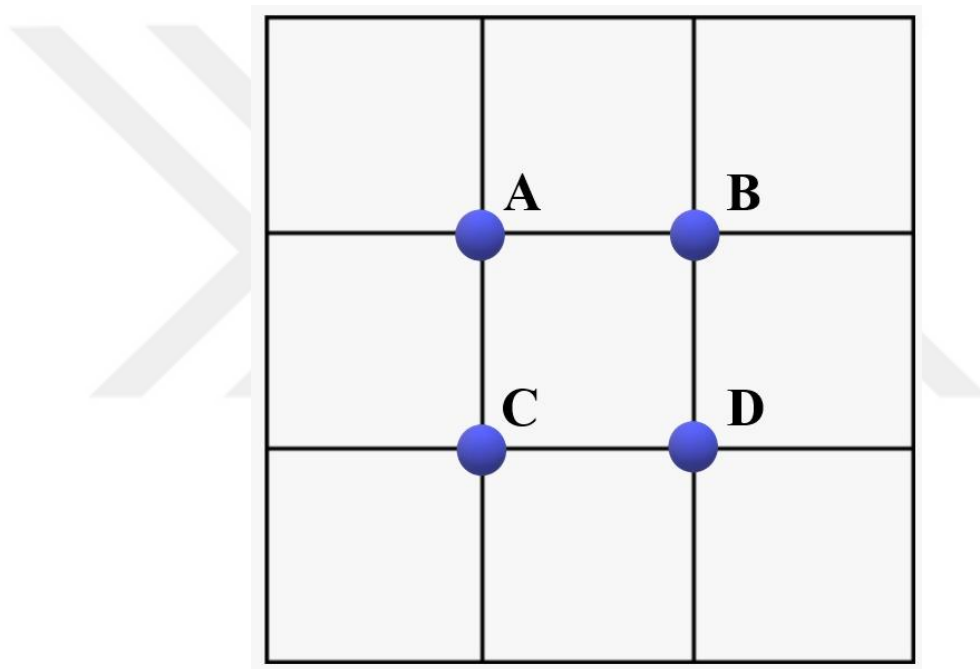


Figure 2.1 Possible key object locations

To explain Figure 2.1, imported objects should be placed along the thirds line or their intersections. If an object has a specific point such as its center, it can be placed on A, B, C, or D points. But, if it is transverse or perpendicular, it must be extended along the lines.

CHAPTER THREE

GRAMMAR THEORIES

Grammar is a set of rules by which valid sentences in a language are constructed. A grammar is a powerful tool for describing and analyzing languages. Several types of grammars can be distinguished. Their categorization may vary according to different approaches to grammar studies. This dissertation has accepted the classification shown in Figure 3.1.

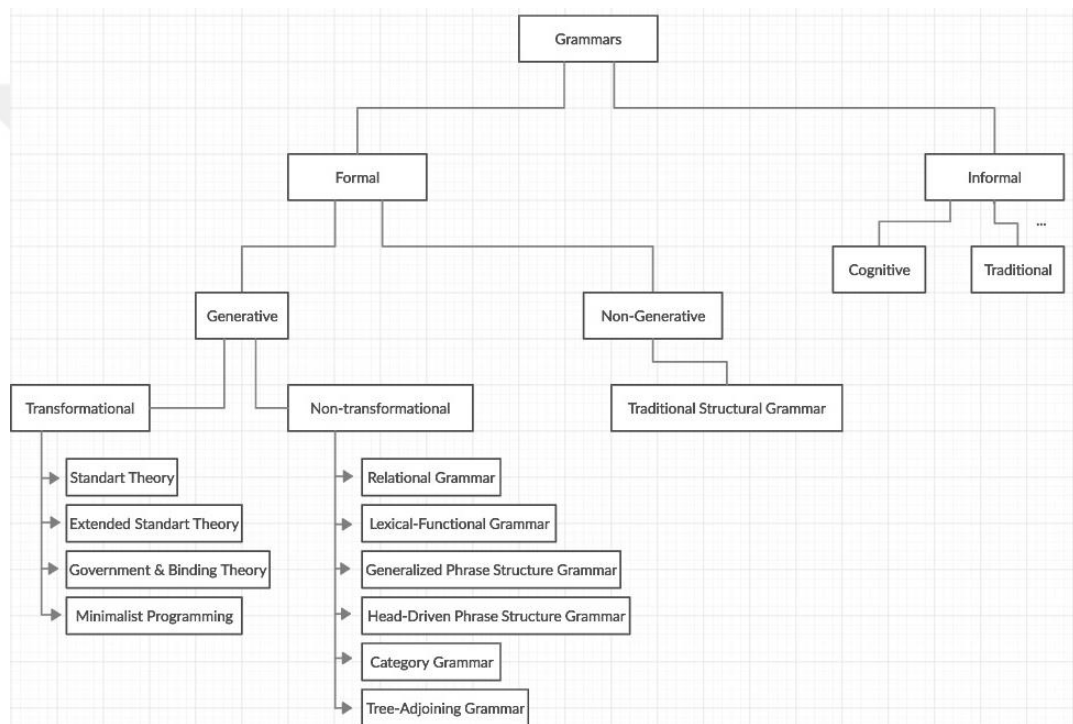
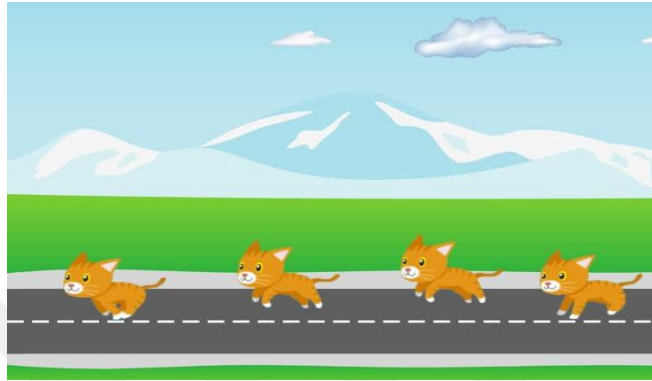


Figure 3.1 Classification of grammar types

Grammar types can be divided into two main categories that are the formal and informal grammar. Informal grammars are usage-based grammars and argue that a mathematical formula cannot be drawn in language. In informal grammars, there are two main representatives, which are traditional and cognitive (Langacker, 2008). One is the traditional grammar which is taught to every student at schools. The root of this grammar goes back to ancient Greek.

Cognitive grammar argues that language cannot and should not be analyzed into a set of formal rules and should also be given a semantic analysis. It also argues that language should be represented pictorially. For example, assuming the preposition “over”, look at Figure 3.2. According to cognitive grammar, e.g. situation “over” cannot be formalized, but picturized.



a) The cat ran over the road



b) The cat jumped over the table



c) The cat was over the table

Figure 3.2 Pictorial representations of the preposition “over”

Formal grammars can be divided into two subcategories, which are generative and non-generative. Traditional structural grammar (TSG) is the main component of non-generative grammars. The main concern of TSG is the classification of the elements of the language. It classifies words and rules using a mathematical-based approach.

In the second group are generative grammars. Generative grammars have two subcategories that are transformational and non-transformational. A generative grammar is a set of formal rules, which projects a finite set of sentences upon the potentially infinite set of sentences that constitute the language as a whole, and it explicitly does this, assigning to each a set of structural descriptions (Crystal, 2008). In short, generative grammars generate sentence whether meaningful or meaningless using their ruleset. This thesis is inspired by generative grammars.

Noam Chomsky is the principal founder and developer of the transformational grammars. He initiated the first of his studies, which he called the Principles and Parameters, with the Standard Theory, as mentioned in Chapter 1. With this study, it can be said that he introduced the concept of phrase structures into the literature. After the Standard Theory, he continued his studies and developed his next work named the Extended Standard Theory. The Government and Binding theory of him followed these studies and the phrase structure approach peaked. Subsequently, Chomsky, who reviewed his work, started the Minimalist Programming (MP) era to reduce complexity.

Generally, MP operates on the Economy Principle. As is known, every phenomenon in nature acts to provide the most benefit at the lowest cost. Taking advantage of this fact, Chomsky revised or canceled many of his costly studies in his previous theories. MP is still a program and is not a theory because it has not been tested for all languages or most languages with more accurate expression.

The main approach of the transformational grammar category is to work on a tree structure by separating sentences into phrases. Roughly, each element marked on the

tree is divided into another tree when it repeats. This process is called move. At the end of the process, all trees are combined and this process is called merge.

Noam Chomsky and his theories dominated the linguistics area for long years. Nevertheless, the second branch of the generative grammars named non-transformational grammar has also found a place in the area. This type of grammar is non-stratal: there is no transformation applied to trees. These studies suggested the method of marking and that every element could be anywhere. The first accepted study, Relational Grammar (RG), began as an alternative to transformational grammar. It was the first time that the marking process was introduced. The marking process will be referred to here as co-indexing.

RG assumes that a sentence is organized in accordance with grammatical relations. The main relations are between subject and object (Perlmutter & Postal, 1986). RG works on a tree structure, where the grammatical relations are annotated. The tree structures are formalized as follows (Kracht, 2002):

DEFINITION 1. Let G be a non-empty set and $\leftarrow$ a binary relation on G . The pair $\langle G, \leftarrow \rangle$ is called a rooted, directed acyclic graph with root r if the following holds.

1. For no x : $r \leftarrow x$.
2. For every $x \in G$ there is a sequence $x \leftarrow x_1 \leftarrow x_2 \leftarrow \dots \leftarrow x_n = r$. (If $x = r$ then $n = 0$)
3. There is no sequence $x_0 \leftarrow x_1 \leftarrow x_2 \leftarrow \dots \leftarrow x_{n-1} \leftarrow x_0$, $n > 0$.

$\langle G, \leftarrow \rangle$ is a tree if in addition for each x there exists at most one y such that $x \leftarrow y$.
 x is a leaf if there is no such y such that $y \leftarrow x$.

DEFINITION 2. Let T be a set. A relational graph over T is a quadruple $\langle G, \leftarrow, p, L \rangle$, where

1. $\langle G, \leftarrow \rangle$ is a tree,
2. p a partial function from $\leftarrow$ to T , and
3. L is a function from the leaves of $\langle G, \leftarrow \rangle$ to the lexicon.

To more information about relation grammar, see the book “Studies in Relational Grammar 1” (Perlmutter, 1986).

After RG, Lexical-Functional Grammar (LFG) was introduced by Joan Bresnan and Ronald Kaplan. LFG mainly focuses on syntax, including its relation with morphology and semantics. LFG has two main distinct formal objects: C(onstituent) and F(unctional) structure. F-structure represents grammatical functions as well as abstract features like tense and case. C-structure represents the concrete phrasal expression of these relations, governed by language-particular constraints on word order and phrase structure (Dalrymple, 2006).

After LFG, Gerald Gazdar introduced Generalized Phrase Structure Grammar (GPSG). One of the chief goals of GPSG is to show that the syntax of natural languages can be described by context-free grammars, which are mentioned later, with some suitable conventions intended to make writing such grammars easier for syntacticians (Gazdar, Klein, Pullum, & Sag, 1985). GPSG also tries to describe the syntax and semantics of natural languages.

Many theorems of GPSG were subsequently incorporated into Head-Driven Phrase Structure Grammar (HPSG). HPSG suggests bracket notation that holds all features of a word as in pronoun “she” in Figure 3.3 (Fleisher, 2007).

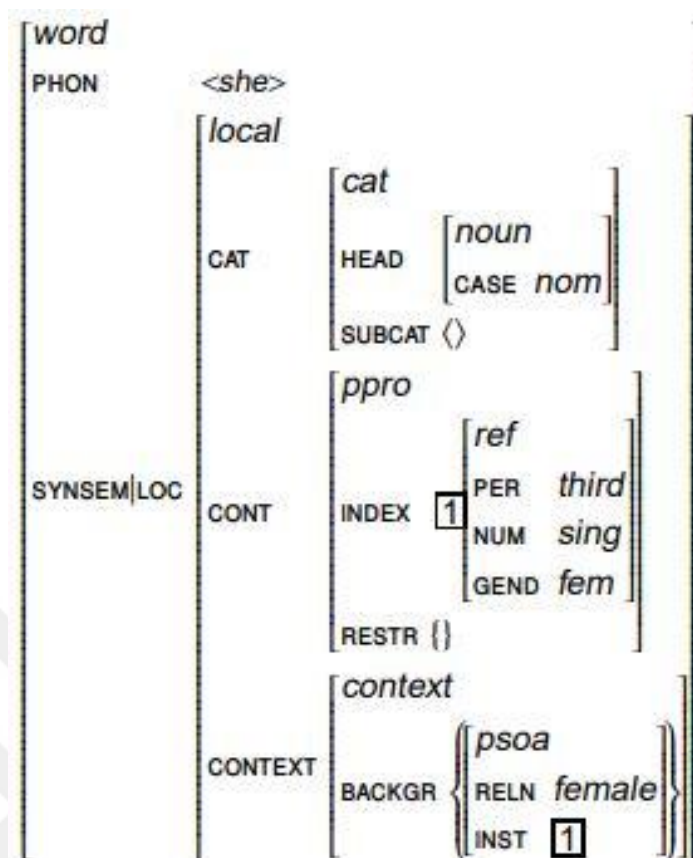
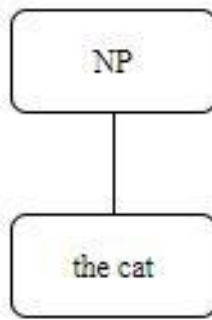


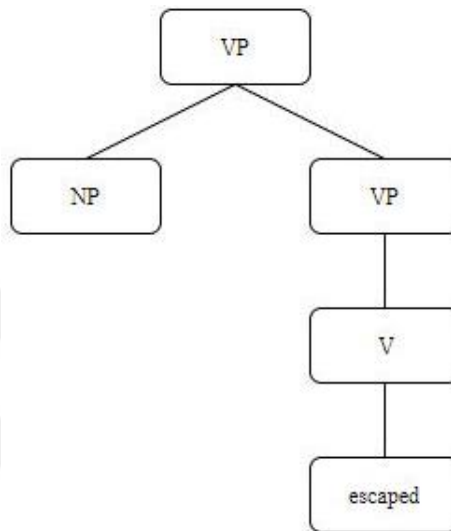
Figure 3.3 An example of HPSG notation (Fleisher, 2007)

HPSG does not have phrase structure rules or transformations. It is declarative, non-derivational, and constraint-based. HPSG suggests that agreement problem can be solved using co-indexing. For more information, refer to “Head-Driven Phrase Structure Grammar” (Pollard & Sag, 1994).

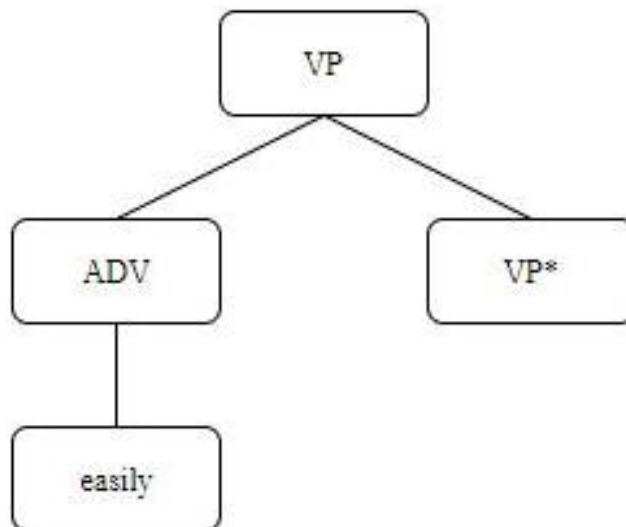
The last member, which is in the non-transformational grammar branch, is Tree-Adjoining Grammar (TAG). A TAG consists of a number of elementary trees, which can be combined with a substitution, and an adjunction operation. There are two types of trees: initial and auxiliary. Initial trees are replaced with a terminal node. This process is called substitution. Auxiliary trees are replaced with a non-terminal node using the adjunction process. For example, assuming the sentence “the cat easily escaped”, three trees are created as in Figure 3.4. For a reminder, the terminal node is a leaf, the non-terminal node is an inner node.



a) Noun phrase initial tree



b) Verb phrase initial tree



c) Adverb auxiliary tree

Figure 3.4 Trees of the sentence "the cat easily escaped"

According to rules, an auxiliary tree as in Figure 3.4.c should be replaced with second level VP of the tree that is in Figure 3.4.b. The tree in Figure 3.4.a should be

replaced with the leaf of the tree in Figure 3.4.b. As a result of this, we obtain a new tree as in Figure 3.5.

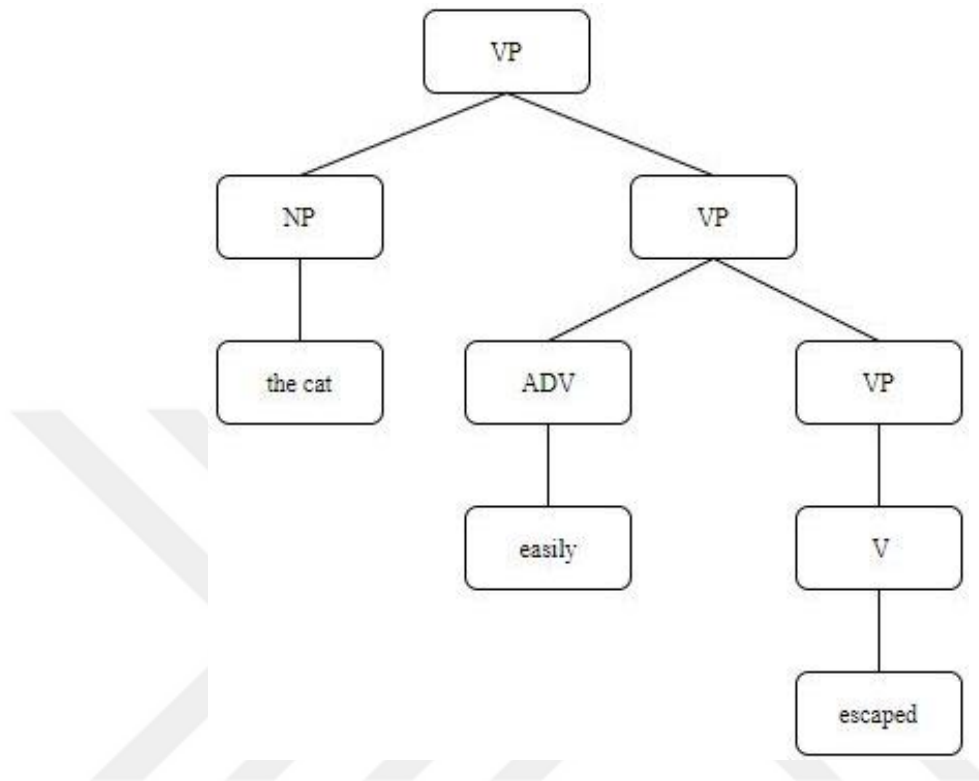


Figure 3.5 Full TAG tree

The other classification method interrupts the aforementioned classification method orthogonally. This classification method has two sub-branches that are Descriptive Grammar (DG) and Prescriptive Grammar (PG). PG argues that people should use the language following the relevant set of rules, while DG advocates that language rules should be the way people use them. In this respect, it can be said that PG is more formal and DG is more scientific. Therefore, there is no formalist grammarian in the field of PG.

The last classification method to be mentioned includes the automata theory approach. This approach divides grammar into two subcategories, which are Context-Free Grammar (CFG) and Context-Sensitive Grammar (CSG). CFGs apply all the rules regardless of context. Rules are shown in (3.1).

$$V \rightarrow w \quad (3.1)$$

This representation tells us that “V” is a single nonterminal symbol, and “w” is a string of terminals and/or non-terminals (possibly empty). “w” can be one of the infinite numbers of symbols. A context-sensitive grammar has productions that are of the form in (3.2).

$$wAx \rightarrow wyx \quad (3.2)$$

In this representation, “w” and “x” are a string of terminals and non-terminals and y is a string. Replacing “A” with “y” will only happen if given the context of “w” and “x” in the input.

CHAPTER FOUR

METHOD

Sentences like (4.1) are to be examined in three stages: tokenization, phrase structure determination, and role extraction. On the other hand, archetype investigation should be done to make semantic and even pragmatic inferences. For this reason, archetypes should be determined from the entire text.

The girl walked from the room to the kitchen in the house. (4.1)

Given a sentence, tokenization is the task of chopping it up into pieces, called *tokens*, perhaps at the same time throwing away certain characters, such as punctuation (Manning, 2008). In other words, tokenization is the process that separates words from a given sentence. Here "What is the word?" and "What are the types of words?" questions should be asked. The concept word is also known as "parts of speech". In the following section of this study, we are discussing in detail.

Phrase structure rules are discussed by Noam Chomsky in 1957 (Chomsky, 1957). Phrase structure grammars are some kind of generative devices and contain phrase structure rules.

There are two types of roles that are examined in this study, grammatical and thematic roles. Grammatical roles (GRs), also called grammatical relations, are relations between words in sentences. Thematic roles were introduced in generative grammar during the mid-1960s and early 1970s (Fillmore, 1968; Gruber, 2008; Jackendoff, 1972) as a way of classifying the arguments of natural language predicates into a closed set of participant types which were thought to have a special status in grammar ("Thematic Roles," 2018).

4.1 Parts of Speech

A word is defined as the unit of expression which has universal intuitive recognition by native-speakers, in both spoken and written language by Crystal (Crystal, 2008). Words can be in two different forms: root form and affixed form. The root is the base form of a word. Roots cannot be divided into smaller parts during the analysis. Affixed words are those that are attached affixes. Affixes come in two types: derivational and inflectional.

All languages have affixes. Affix is a general term, which includes prefix, suffix, and circumfix. Prefixes and suffixes are respectively left- and right-attached to a word while circumfixes involve both operations. For instance, considering the root *happy*, the prefix *un-*, and the suffix *-ness*. If the prefix *un-* is added to *happy*, *unhappy* is derived. *Unhappy* is also a stem. Same way, if the suffix *-ness* is added to *happy*, *happiness* is derived. If both of them are added to *happy*, we have *unhappiness*, which is an example of a circumfix. Besides, the German past participle is an example of circumfixes. For instance, the verb *spielen* has the participle *gespielt* which occurs in the circumfix *ge-* and *-t*.

The term “part of speech” is also known as the “word category”. Words are classified according to similar syntactic features or morphologies. In traditional grammars, words are categorized according to their grammatical relations. Word categorization studies date back to the 5th or 6th centuries BC. During these times, a 4-category study named Nirukta from Yaska emerged in India. After approximately two centuries from Nirukta, western, especially Greek, scholars worked on this study and they decided to divide the words into eight categories. However, these studies were language-specific, thus they could not meet the needs of all languages. Because some language families have categories of types different from those of others.

In the study by Chomsky in 1970 (Chomsky, 1970), words can be classified according to nominality and verballity features. This analysis uses $\pm$ noun and $\pm$ verb notation. In table 4.1, word classes can be seen.

Table 4.1 Chomsky's word analysis according to his 1970 work

	+N	-N
+V	Adjectives	Nouns
-V	Verbs	Prepositions

On the other hand, some words are functional. At this point, the issue of the functionality needs to be explained. Even though the functional words do not have concrete meanings, sentences would make a lot less sense without them. Andrew Radford proposes that new feature to hold functionality should be added (Radford, 1997). Thus, almost all words can be classified according to whether they contain three basic features: Nominality, verbality, and functionality. Nominals refer to words that have some of the attributes of nouns. Likewise, verbality refers to words that have some of the attributes of verbs. Functionality indicates whether words have functional features. Each word category differs according to these features. These attributes are indicated by the notation $\langle \pm n, \pm v, \pm f \rangle$. Words can be examined into two layers using the functionality feature. The first layer contains $\langle \pm n, \pm v, -f \rangle$ and the second layer contains $\langle \pm n, \pm v, +f \rangle$ features. This structure is demonstrated using a lattice structure as in Figure 4.1.

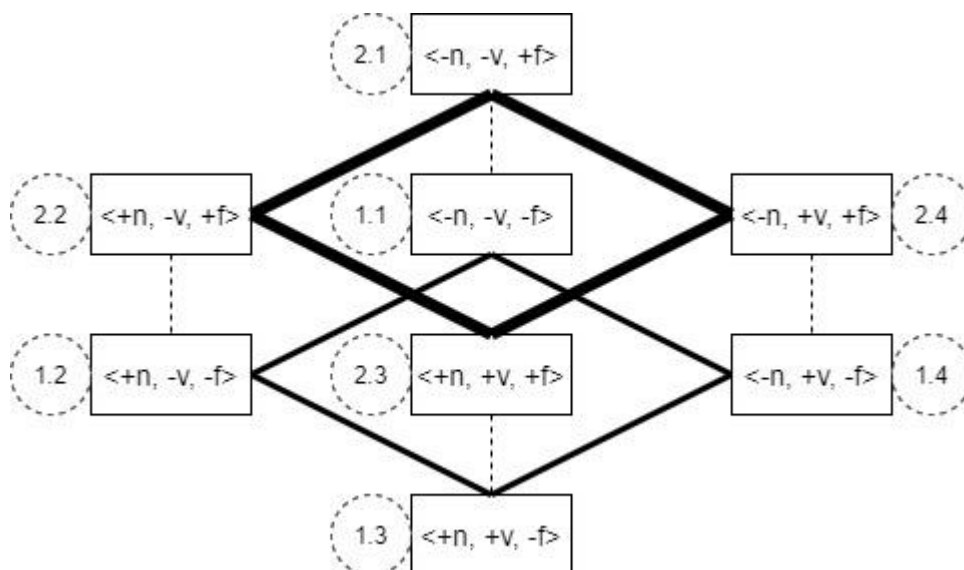


Figure 4.1 Lattice structure of parts of speech

In Figure 4.1, thicker lines connect the first layer. The thickest lines connect the second layer, the functional layer. Finally, dashed lines provide bonds between the two layers. Dashed numbers are used for indexing to help explain in the figure.

The case $\langle +n, -v, -f \rangle$ marked 1.2 and the case $\langle -n, +v, -f \rangle$ marked 1.4 indicate pure nouns and verbs respectively. If a word has both characteristics of noun and verb as in the case $\langle +n, +v, -f \rangle$ marked 1.3, it becomes an adjective. When a word lacks either feature as in 1.1, it becomes a preposition.

In the second layer which is marked 2.X in dashed circles, functional words are seen. If a word has a pure noun and pure verb with the functional feature, they are pronouns marked 2.2 and auxiliary verbs marked 2.4 respectively. If a word has both nominal and verbal features with the functional feature, it can be a quantifier or determiner. And a word has no verbal or nominal features but a functional feature, it is a case-marker. Case-markers mark roles of nouns concerning verbs and other function-like expressions.

As a result of these and subsequent studies, it can be mentioned that there are seventeen categories from different languages: adjectives, adpositions, adverbs, auxiliary verbs, cardinal numbers, classifiers, clitics, conjunctions, contractions, coverbs, determiners, interjections, nouns, particles, preverbs, pronouns, verbs. The details about the word types can be found in Appendix A.

The tokenization process helps to separate words of the sentence. And, each word is labeled according to its word category. After the tokenization process, phrases should be determined.

On the other hand, there are some cases about the labelization of words. In some cases, we can use a word as a different type. For example, we know that the word *red* is an adjective. However, if we make a sentence such as “Give me the red”, it becomes a noun. In this study, we do not concern these types of situations.

4.2 Phrase Structures

Phrase structures are built up in accordance with a set of context-free rewrite rules, as shown in (4.2). In this notation, X is a single term designating a nonterminal element.

$$X \rightarrow \dots Y \dots \quad (4.2)$$

Chomsky proposed some structure rules such as those in Table 4.2 (Chomsky, 1965).

Table 4.2 Chomsky's early work on phrase structure rules

Sentence $\rightarrow$ Noun Phrase + Verb Phrase
Noun Phrase $\rightarrow$ (Determiner) + Noun
Noun $\rightarrow$ (Adjective Phrase) + Noun + (Prepositional Phrase)
Verb Phrase $\rightarrow$ Verb + (Adverb)

Assuming the rules mentioned above Chomsky's example sentence which is first encountered in his book Syntactic Structure (Chomsky, 1957) can be solved. You can see the analysis of the mentioned sentence in Table 4.3.

Table 4.3 Analysis of "Colorless green ideas sleep furiously"

Sentence $\rightarrow$ Colorless green ideas (Noun Phrase) + sleep furiously (Verb Phrase)
Verb Phrase $\rightarrow$ sleep (Verb) + furiously (Adverb)
Noun Phrase ₍₁₎ $\rightarrow$ green (Adjective) + ideas (Noun)
Noun Phrase ₍₂₎ $\rightarrow$ colorless (adjective) + Noun Phrase ₍₁₎

Phrase structure rules can be denoted as shown in (4.3). In this notation, NP means noun phrase, VP means verb phrase and AP mean adjective phrase. ADJ, N, and ADV mean adjective, noun, and adverb respectively.

$$[\text{NP}[\text{ADJ N}] \text{ VP}[\text{V}] \text{ AP}[\text{ADV}]] \quad (4.3)$$

Usually, words are in a structure, although they have a role alone. Therefore, it is more correct to assign roles after the phrase structures are detected.

4.3 Grammatical and Thematic Roles

In this study, two types of roles are used as mentioned above. Sometimes the concept of thematic role is confused with the concept of grammatical role.

4.3.1 Grammatical Roles

A grammatical role is used to label to identify the syntactic position of NPs within each clause. Although common terms are used to refer to particular, grammatical relations are subject, object, ergative, absolutive, genitive, and oblique (Payne, 2006), the most studied relation is between subject and object. It can be said that grammatical relations are functional relationships between elements in a sentence (Sanfilippo, 1990).

One of the major constituents of a sentence is the subject. In a sentence, it is the person, place, thing, or idea that is doing or being something. The subject has eight semantic roles: agent, instrument, patient, recipient, locative, temporal, eventive, and empty “it”.

The most typical role of the subject is the agent which is the direct doer of the event as in the “**Ayşe** reads a book” example. An instrumental subject is the inanimate causer of the happening as in “**The ball** broke the glass” example. A patient is the one being affected by the event as in “**The door** opened” example. A recipient indicates the receiver in a situation as in “**Ayşe** got a postcard from the child” example. A locative subject denotes a location. Considering the sentence “**The bus** can hold thirty people”, it means thirty people can sit *in* the bus. A temporal subject denotes a time as in “**Today** is the first-day” example. An eventive subject denotes an event as in “**The**

concert is tomorrow” example. An empty “it” subject uses when we lack semantic content as in “**It** is getting cold” example.

Another major constituent of a sentence is the object. It is the person or thing that receives the action of the verb. The object can be two different types, which are direct and indirect. Direct objects link directly to the verb, whereas indirect objects link through a preposition. From here, it can be said that indirect objects are placed after intransitive verbs, and vice versa, direct objects are placed after transitive verbs.

In passive voice, the object is promoted to the subject position while the subject is demoted to the object position of a by-phrase. Considering the “reading book” example, the place of the object in active voice such as in (4.4) changes in passive voice such as in (4.5).

Ayşe	reads	a book.	[Active Voice]	(4.4)
Subject	Verb	Object		

A book	was read	by	Ayşe.	[Passive Voice]	(4.5)
Subject	Passive	Preposition	Object		

Like the subject, the object has semantic roles: patient, locative, effected, and recipient. A patient is a direct object which affected the event directly as in “He kicked **the door**” example. A locative object is a direct object which shows a location as in “I go to **the school**” example. An effected object is a direct object that exists only by virtue of the activity indicated by the verb as in “I am writing **a letter**” example. A recipient object is an indirect object which is implicated by the happening or state. “I gave **my brother** some money” as an example can be given.

The oblique role is used when a noun phrase is the object of either a verb or a preposition. The patient structure of an indirect object is also an oblique role as in “He gave **the door** a kick” example.

Genitive role refers to an entity the speaker is portraying as someone who is closely associated with another entity such as in the sentence “Ayşe’s book”. Here the relation is that of possession. However, the genitive role not only defines ownership. It defines different relations such as maybe Ayşe wrote that book or Ayşe used to have this book.

In some languages such as Basque, Georgian, Mayan, Tibetan, a few Indo-European languages, there is a formal parallel between the object of a transitive verb and the subject of an intransitive one. In such languages, the more patient-like argument of a transitive verb appears in the same absolutive case as the single argument of an intransitive verb, while the more agent-like argument of a transitive verb is marked differently, in what is known as the ergative case (Manning, 1994). In other words, case suffixes of a subject are determined according to whether the verb is transitive or intransitive for example in Basque. As we can see in the example, the word *Maria* has a suffix when an intransitive verb is used (4.6). Otherwise, *Maria* is nominative as in (4.7).

Mariak	liburua	irakurri	zuen.	
Maria	book	read	3 rd singular pronoun	(4.6)

‘Maria reads the book.’

Maria	korrika	
-------	---------	--

‘Maria runs.’ (4.7)

4.3.2 Thematic Roles

Although different types are created according to the needs of some studies, the most used thematic roles are given below (Başarıcı & Kılıçaslan, 2017; Dowty, 1989):

- Location: the place where something is
- Source: the place from which an entity moves
- Goal: the place toward which an entity moves
- Path: the trajectory along which an entity moves

- Theme: an entity in motion or being located
- Actor: the initiator of an action
- Patient: an entity affected by an action
- Instrument: an object with which an action is performed
- Experiencer: an individual that feels, perceives or cognizes a situation

A good example to show all these roles may be the following sentences. “*In the gloomy house, the man walked from the living room to the kitchen through a narrow hallway. He saw the maid chopping an onion with a knife.*” In this sentence, the theme is represented using “*gloomy*”. The actor is “*the man*” and this actor turns to the experiencer into the second sentence. The location, the source, the goal, and the path are “*the home*”, “*the living room*”, “*the kitchen*”, and “*hallway*” respectively. In the second sentence, “*an onion*” is the patient. Finally, the instrument is “*the knife*”.

Kılıçaslan has a theory that there is a relationship between the lattice structure in the brain and thematic roles (Kılıçaslan, Güner, & Yildirim, 2009; Kılıçaslan, Uçar, Uçar, & Güner, 2008; Kılıçaslan & Tuna, 2014; Kılıçaslan & Tuna, 2015). Below are statements that explain this relationship. He gathered thematic roles in three sets as in (4.8). He called it “*tripartition of the set of thematic roles*”.

$$\{\{\text{Location, Source, Goal, Path}\}, \{\text{Theme, Actor, Patient, Instrument}\}, \{\text{Experiencer}\}\} \quad (4.8)$$

The location, the source, and the goal are grouped because they indicate the locatives. The theme, the actor, the patient, and the instrument are grouped because these entities might be in these places. The experiencer is the individual entity, which does not fit into any set of a thematic role.

According to the situations mentioned below, thematic roles have some primitive features. Kılıçaslan, inspired by the Gestalt school, has put forward some theories on these primitive structures. He also examined thematic roles as a lattice structure.

There are two different situations about thematic roles, especially about location. Assuming the following situations, the first situation can be seen:

Jack received a book from Jill. (4.9)

I have a job from Tuesday to Friday. (4.10)

The weather turned from sunshine to rain. (4.11)

In (4.9), the preposition "from" gives possession information, even if it is a location preposition. A book is taken from the source, Jill. In (4.10), scheduling is seen. The time interval is given from Tuesday (source) to Friday (goal). Lastly, in (4.11), property changes are observed. The weather's first location is sunshine and then it reached the goal, rain. These situations show that the location role appears in other forms. Note that here, primitives are shown in the italic font; thematic roles are shown in normal font.

Locative roles as in (4.8) are the location, the source, the goal, and the path. It is argued that they have a main primitive, which is named grounding. Other primitives are location, source, and goal. As a result, the following primitive sets have emerged:

- 1) Location = {*grounding*, location}
- 2) Source = {*grounding*, location, source}
- 3) Goal = {*grounding*, location, goal}
- 4) Path = {*grounding*, location, source, goal}

In 4, the path should especially be examined. Because, in some sentence examples such as "I walked through the road", there are no goals and sources. In fact, during the walk unreached places of the road is the goal, and the crossed road is the source. In Figure 4.2, the Hasse diagram to illustrate the lattice structure of locatives is shown.

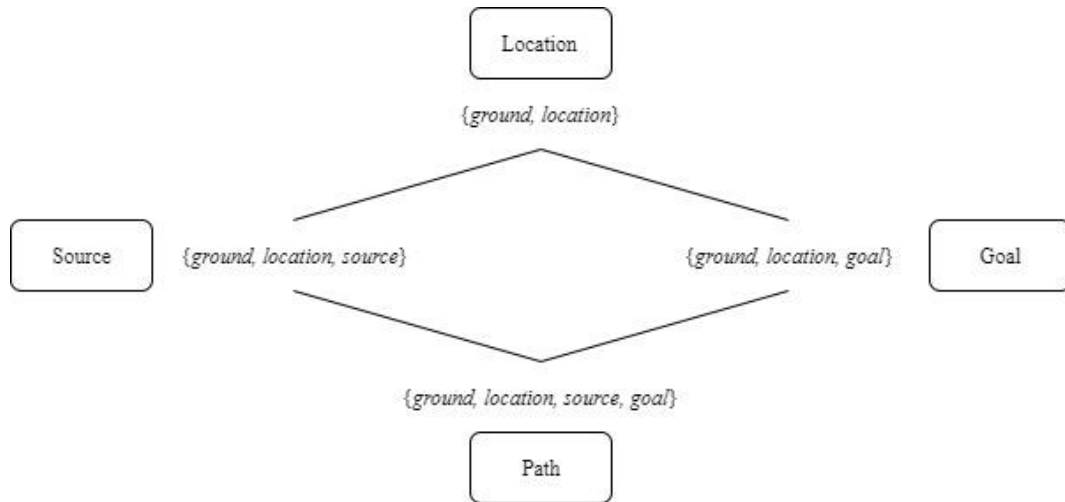


Figure 4.2 The Hasse diagram of locatives sets

The second case is about figural roles. In (4.8), there is a set that contains the actor, the patient, the theme, and the actor. These roles have a main primitive named figural. On the other hand, as in (4.9), Jill is used as a source. Thus, locatives are needed. Location, goal, and source primitives can be used here.

- 1) Theme = $\{figural, location\}$
- 2) Actor = $\{figural, location, source\}$
- 3) Patient = $\{figural, location, goal\}$
- 4) Instrument = $\{figural, location, source, goal\}$

Here, the instrument is the same as the path in locative roles. The source of the instrument is the actor, and the goal of the instrument is the patient. Figure 4.3 shows the Hasse diagram to indicate the lattice structure of figurals.

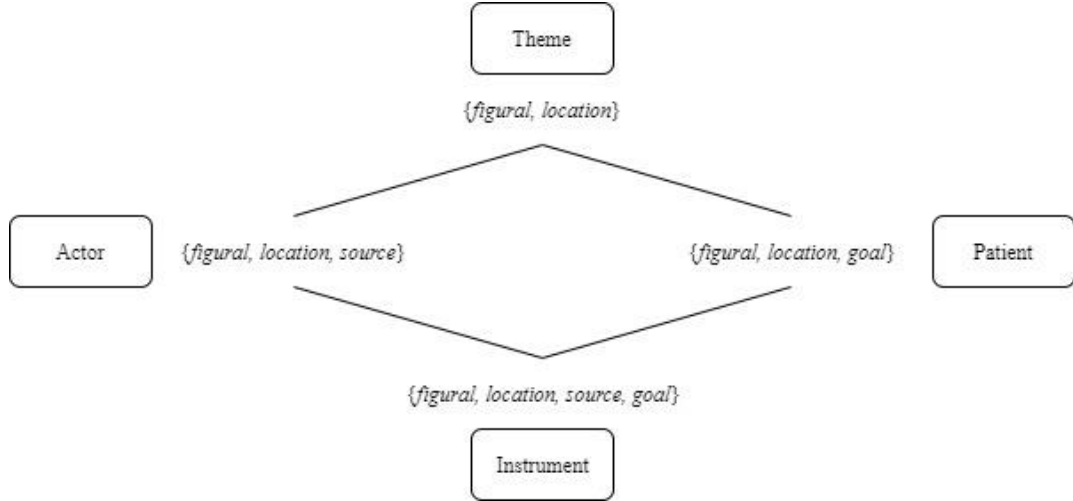


Figure 4.3 The Hasse diagram of figurals sets

There is one thematic role distinct from locatives and figurals. It is the experiencer. The experiencer has own cases. Experiencer can be counted as intangible concepts. To understand the experiencer, a new primitive named perceptual is needed. Because of its own nature, the experiencer emerges as a locative. Thus, primitives, which are location, source, and goal, are needed to create the sets below:

- 1) Perceptual_location = $\{\text{perceptual}, \text{location}\}$
- 2) Perceptual_source = $\{\text{perceptual}, \text{location}, \text{source}\}$
- 3) Perceptual_goal = $\{\text{perceptual}, \text{location}, \text{goal}\}$
- 4) Perceptual_path = $\{\text{perceptual}, \text{location}, \text{source}, \text{goal}\}$

Up to this point, a single sentence has been analyzed. After these stages, the entire text should be examined. Archetype detection is the basis of this review.

4.4 Archetype Analysis

After the main hero is determined, archetypes are determined according to this hero. Due to the plot, in some cases, the side characters appear as the main hero, but the main hero is fixed. From here, we can make assumptions about other archetypes. We will talk about this process in Chapter 5.

CHAPTER FIVE

MODEL

5.1 Controlled Natural Language

To reduce the complexity of this study, a controlled natural language is created. First of all, the types of words to be used in this study were selected. After that, our own phrase structure rules were produced inspired by the work of Chomsky's 1965 work (Chomsky, 1965).

5.1.1 Selected Word Categories

All words have some basic attributes. Each attribute holds a value out of a set of values. Values are selected according to their contemporary meanings in the dictionary. To provide this, the Longman Dictionary of Contemporary English (Pearson Education, 2019) is used. Another one is its category. According to traditional grammar, seven categories are accepted, including verbs, nouns, adjectives, adverbs, prepositions, determiners, and pronouns. As in determining the value of the word, the category selection was made according to the first category of the words in the dictionary. In terms of ease of computational progress, these categories have been rearranged under four categories. While verb and preposition categories are kept as it is, adjective and adverb categories are merged into a new category named "Adword". Noun, pronoun, and determiner categories are also gathered into a new category named "Word" category.

5.1.1.1 Preposition Category

The preposition category has four distinct attributes. These attributes indicate whether the preposition provides location, time, direction, and/or device information. A preposition may contain some or all of this information. For example, the preposition "on" consists of location and time.

As it is known, nouns have cases: nominative, accusative, dative, locative, ablative, and genitive. While, in Turkish, these cases are provided using affixes, in English, prepositions can be used, as we can see in Table 5.1. Considering these cases, a new attribute is created to hold cases.

Table 5.1 Noun cases in English and Turkish

Cases	English	Turkish
Nominative	The book is good.	Kitap güzel.
Accusative	She reads the book .	Kitap-ı okudu.
Dative	She gave a book to me .	Ban-a bir kitap Verdi.
Locative	She is at home .	O ev-de .
Ablative	She came from home .	O ev-den geldi.
Genitive	Mehmet's book is old.	Mehmet'in kitabı eski.

On the other hand, prepositions can be in two forms: simple or complex. The ones such as *on*, *up*, *behind*, etc. are in the simple form. In contrast, the ones such as *next to*, *according to*, etc. are in complex form.

5.1.1.2 Verb Category

Verbs, auxiliary verbs, and modals are grouped in this category. An important property keeps the information that the verb is an auxiliary/modal or not. This information can be provided with a simple Boolean flag. In addition, a verb can be described as transitive or intransitive based on whether it requires an object to express a complete thought or not (Traffis, 2020). This feature is also provided with a simple Boolean flag named *transitivity*.

The verb category is most significantly divided in terms of whether a verb denotes an action, a motion, or a state. For a verb to be a motion, it is acknowledged that it must be displaced when the duration is over; otherwise, it is accepted as an action. If a verb showed reflexive or uncontrollable behavior, it was considered a state. For

instance, the verb “run” is a motion and the verb “look” is an action, on the other hand, the verb “saw” is a state. This attribute is named as *MAS*.

Other information is the tense and the aspect of the verb. Tenses are divided into three categories: present, past, and future. The concept of tense shows the location of a verb in time. Aspect refers to the duration of an event within a particular tense. This holds four different categories: simple, progressive, perfect, and perfect progressive.

5.1.1.3 Word Category

Nouns, pronouns, and determiners are stored in this category. Using a “type” attribute, they are separated. Determiners are added into this category because they are dependent on nouns and pronouns.

In this category, there are two dimension attributes for nouns and also pronouns. Because of agreement, dimensions of a pronoun may vary. Another attribute keeps the singularity of the words. As known, all words of these categories have a plural form such as “sheep” for a noun, “they” for a pronoun, “some” for a determiner. Another attribute holds proper name information. This information is helpful in determining whether to use determiner before a noun.

5.1.1.4 Adword Category

Adjectives and adverbs are put into this category. They are separated using a Boolean flag. Because adverbs modify verbs and adjectives and adjectives modify nouns, they were decided to be put into this category. They are modifiers.

5.1.2 Phrase Structures

Phrases are rearranged differently from the literature in this study because of computational complexity. Nine types of phrases are used: sentence (S), determiner phrase (DP), tense phrase (TP), noun phrase (NP), adjective phrase (AdjP), adverbial

phrase (AdvP), prepositional phrase (PP), verb phrase (VP), and auxiliary phrase (AuxP). Note that, we use the asterisk symbol “*” to indicate that the chosen phrase should use zero or more times. If a phrase is in brackets such as “()”, this phrase is optional.

5.1.2.1 Sentence (S)

The sentence is the biggest syntactic unit. Texts are a combination of sentences, but they are not a phrasal unit. We have accepted that the sentence is the combination of a determiner phrase and tense phrases. The full tree of a sentence can be seen in Appendix B. The sentence structure can be seen in (5.1).

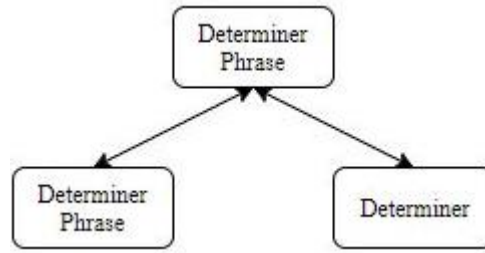
$$S \rightarrow \text{Determiner phrase} + \text{Tense Phrase} \quad (5.1)$$

5.1.2.2 Determiner Phrase (DP)

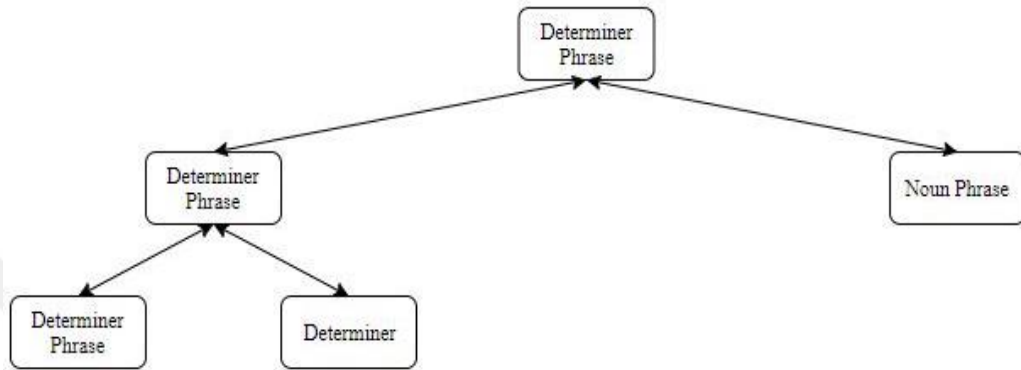
DP is one of the two main phrases with the tense phrase. It has two special cases. According to these cases, it can be in different branches of the tree. One case holds the determiner groups and it is marked as DP' (5.2.a). Another case holds DP' and noun phrase (5.2.b).

$$\begin{aligned} \text{a) } DP &\rightarrow (DP')^* + \text{Determiner} \\ \text{b) } DP &\rightarrow (DP')^* + \text{Noun Phrase} \end{aligned} \quad (5.2)$$

This structure tells us if there is more than one determiner, they are grouped in one tree branch as in (5.2.a). This case is shown in Figure 5.1.b. (5.2.b) indicates that the left side of the sub-tree is DP' and the right side of the sub-tree is a noun phrase as in Figure 5.1.a.



a) DP' sub-tree



b) Full DP sub-tree

Figure 5.1 Two cases of determination phrase

Some languages such as Portuguese, use articles before proper names. In English, there is no situation like this. At this point, the null determiner shows up with certain kinds of nouns, most notably the bare plurals like *books*, mass nouns such as *lunch*, or proper names such as *Mehmet*. Based on Chomsky's ideas on this subject, it can be assumed that in the most primitive times, each language used determiners before a noun. According to him, this structure is broken in terms of ease of speech in use over time. In our method, we accept this theory, thus, there should be a DP before each noun phrase. Because of that case, differently from many studies about this area (Chomsky, 1957, 1965), it is accepted that DP is a primary phrase structure.

5.1.2.3 Noun Phrase (NP)

Noun phrases hold the adjective, adverbs, prepositions, and, of course, nouns. As we can see in (5.3.a), the main NP's left side is another NP case that is marked as NP' and the right side is the prepositional phrase. If there is no prepositional phrase, the right sub-tree is pruned. As in (5.3.b), NP' keeps adjective phrase at the left sub-tree

of it and keeps another noun phrase case, which groups nouns, at the right sub-tree of it. This third case is marked as DP" as in (5.3.c).

- a) $NP \rightarrow NP' + \text{Prepositional Phrase}$
- b) $NP' \rightarrow \text{Adjective Phrase} + NP''$ (5.3)
- c) $NP'' \rightarrow (NP''')^* + \text{Noun}$

As mentioned earlier, most of syntax theorem accepts NP is the main phrase, which comes after S. In this study, it is decided that NP is chosen as a secondary phrase. Its tree structure is seen in Figure 5.2. In Figure 5.2.c, the full NP sub-tree is seen, in 5.2.b, NP' is seen and in 5.2.a, NP'' is seen. Other sub-trees will be mentioned below.

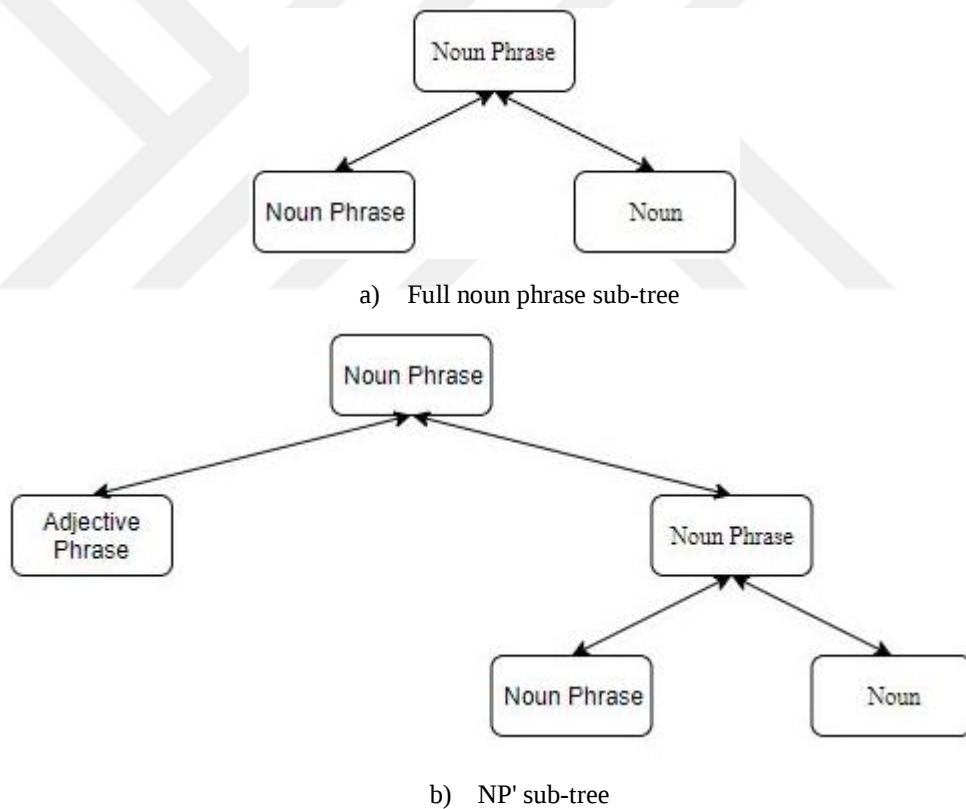
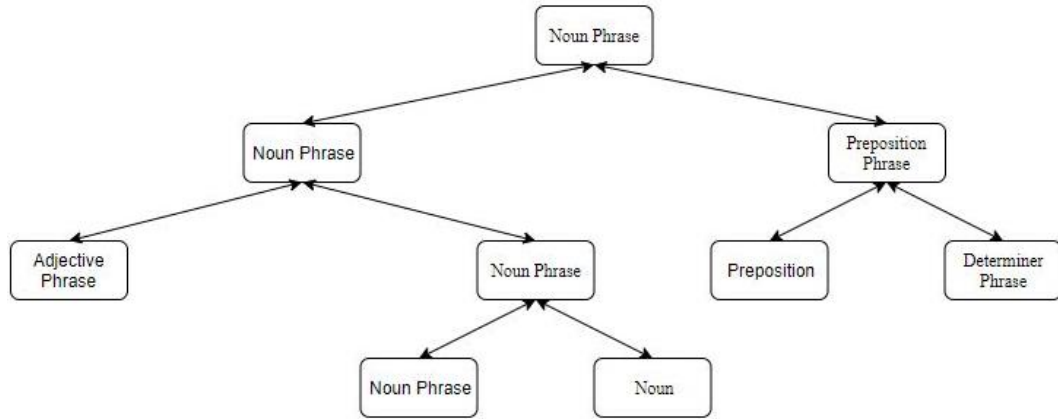


Figure 5.2 There cases of the noun phrase



c) NP" sub-tree

Figure 5.2 continues

5.1.2.4 Adjective Phrase (AdjP)

The adjective phrase has two forms. One of them holds adverbs, which form the adverbial phrase. The second one holds a group of adjectives. The rules that include adverbs can be seen in (5.4.a) and (5.4.b).

- a) $\text{AdjP} \rightarrow (\text{AdvP})^* + \text{Adjective}$
b) $\text{AdjP} \rightarrow (\text{AdjP})^* + \text{Adjective}$ (5.4)

All cases of the AdjP including adverbial phrases in a sub-tree is seen in Figure 5.3.

5.1.2.5 Adverbial Phrase (AdvP)

Adverbial phrases have only one case which holds the adverb groups. Multiple adverbs can be used before modifying their complements. The AdvP rule can be seen in (5.5).

$$\text{AdvP} \rightarrow (\text{AdvP})^* + \text{Adverb} \quad (5.5)$$

In English, AdvP can come after or before a sentence. Nevertheless, within the framework of our study, this situation is out of our concerns because of our

tokenization approach. Adjective phrase and adverbial phrase sub-trees can be seen in Figure 5.3. As we can see, in the deepest level of the sub-tree, adverbial phrases are located.

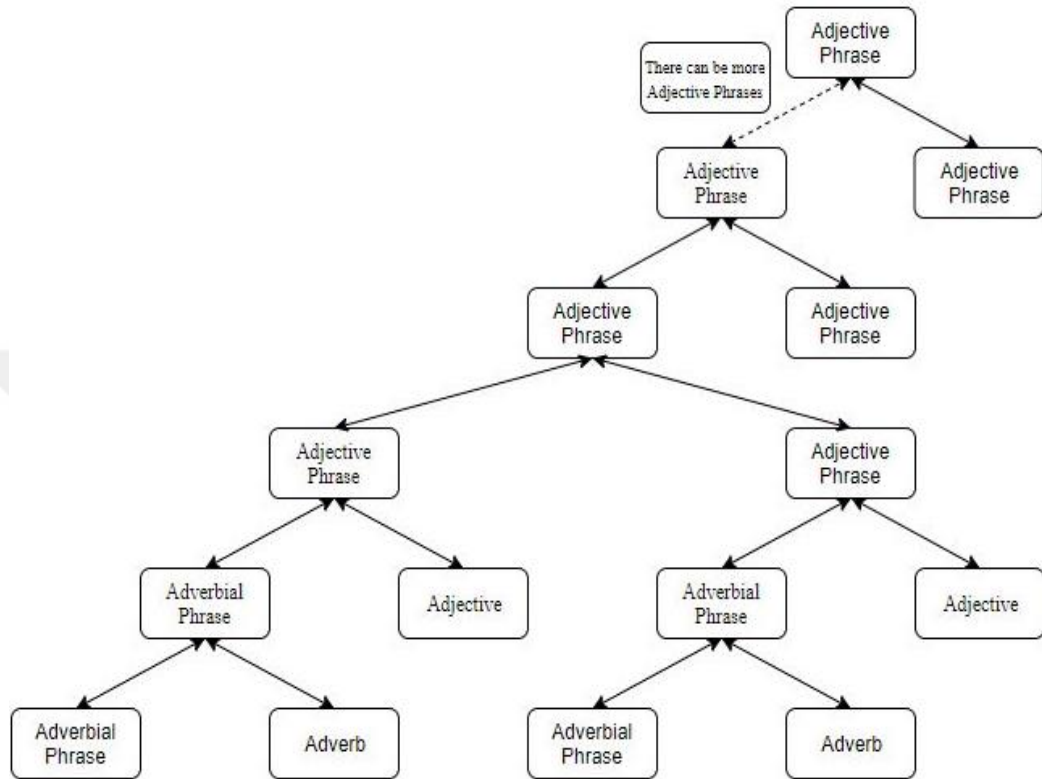


Figure 5.3 Full Adjective phrase sub-tree

5.1.2.6 Prepositional Phrase (PP)

PPs are originally a modified form of an NP. They have one case and hold a preposition at the left sub-tree and an NP at the right sub-tree such as in (5.6).

$$PP \rightarrow \text{Preposition} + NP \quad (5.6)$$

As it can be seen that in Figure 5.2.a the right sub-tree of NP sub-tree is the place of the prepositional phrase.

At this point, one situation about the ambiguity of prepositions should be discussed. Considering the sentence “*the dog is running from the garden*”, there are two cases.

According to our approach, the model examines this sentence in two ways. One is the dog is running and it is coming from the garden. Another is there is a dog in the garden and it is running. In other words, the initial position of the dog in the second case and the direction of the movement of the dog in the first case are understood. We solved this ambiguity according to the contextual information we have. If we know that the dog has gone to the garden before according to the context, we accepted the second case otherwise the first case is accepted.

5.1.2.7 Tense Phrase (TP)

TP is another main phrase of the sentence. Tense phrases have two sub-trees, which are a verb phrase and a DP. In this study, it is accepted that TP has only one case such as in (5.7).

$$TP \rightarrow \text{Verb Phrase} + DP \quad (5.7)$$

5.1.2.8 Verb Phrase (VP)

VP has two sub-trees that are auxiliary phrase and a verb as a leaf. Because it is accepted that the clauses and conjunctions are out of concern of this work, there is only one verb in a sentence. VP's structure can be seen in (5.8).

$$VP \rightarrow \text{Auxiliary Phrase} + \text{Verb} \quad (5.8)$$

5.1.2.9 Auxiliary Phrase (AuxP)

AuxPs hold both auxiliary verbs and modals. In English, multiple modal and auxiliary can be used such as “had been” or “should be”. To this situation, the left sub-tree of AuxP should be again itself as in (5.9).

$$\text{AuxP} \rightarrow (\text{AuxP})^* + \text{Auxiliary} \quad (5.9)$$

The full TP including VP, AuxP, and DP can be seen in Figure 5.4.

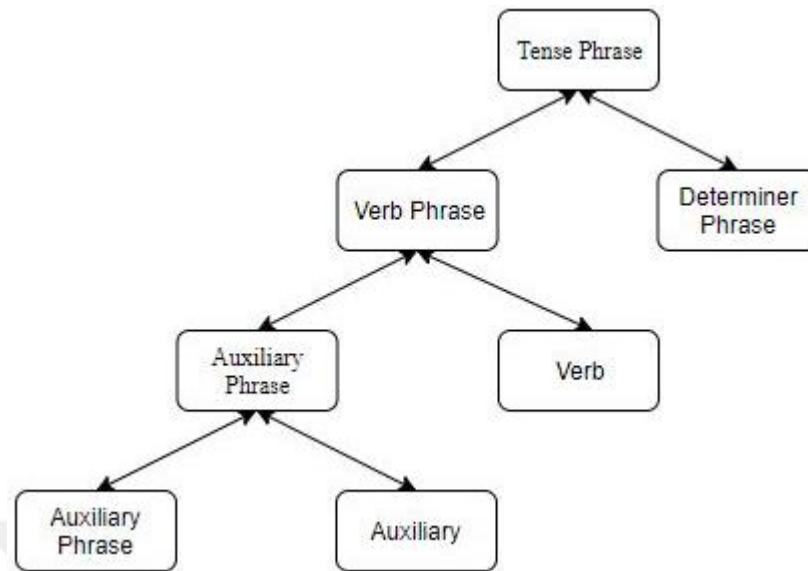


Figure 5.4 Full tense phrase tree

5.1.2.10 Brief Notes about TP, VP, and AuxP

At this point in the study, it is worth noting some issues about TPs, VPs, and AuxPs. Almost all studies about structure phrase rules accept that one of the main phrase structures is VP (Chomsky, 1957, 1965). Some works show the auxiliaries under root. According to this approach, the sentence node has three children. Thus, the designed tree is ternary at the root level and binary in the remaining stages. This approach was avoided as it was not suitable for our approach in terms of computation. Care has been taken to ensure that the designed tree is binary at all levels. Therefore, the auxiliaries had to be shown somewhere in the tree. For that reason, a new phrase named *auxiliary phrase* was designed and was merged with the verb phrase. This combined structure was processed as a tense phrase.

5.1.2.11 Algorithm of Branches of Basic Phrases

Our approach accepts that the most basic phrase structures are DP', NP'', AuxP, and AdvP. Their structures contain only their type of words. Therefore, these four phrase structures are examined in one method such as in (5.10).

```

        Basic-Phrases(wordList, node)
1  IF(wordList.count != 0)
2      IF(node == null)
3          node = wordList[0]
4          node.parent = Remaining(wordList)
5          node.parent.right = node
6          node = node.parent
          node = Basic-Phrases
7              (Remaining(wordList),
node)
8      ELSE
9          node.parent = Remaining(wordList)
10         node = node.parent.left
11         node = node.parent
12         node.right = wordList[0]
          node = Basic-Phrases
13             (Remaining(wordList),
node)
14     ENDIF
15 ENDIF
16 RETURN node

```

(5.10)

In (5.10), a bottom-up approach can be seen. If it is assumed that there is a clause such as “extremely very surprisingly big dog”, “big” is an adjective and “dog” is a noun. The remaining words are adverbs. According to this pseudocode, the first adverb is located at the bottom of the tree and the remaining adverbs are placed at higher levels as in Figure 5.3.

Other phrases are a combination of these basic phrases and their types. Thus, while creating these phrases, one branch of the trees contains one basic phrase and the other branch contains another basic phrase. This means the sub-branches of this branch are created from the top of the branch. Therefore, this technique is a top-down technique.

5.2 Grammatical and Thematic Roles

Grammatical roles are vital for the proposed model. With the help of grammatical roles the subject, object, and predicate of the sentence are found.

Thematic roles link the predicate and other arguments in the description of an event. To develop the proposed model, thematic roles are highly used. Drawable features are extracted using thematic roles. First, the predicate of the clause is detected. Then thematic roles are extracted with the help of grammatical roles. Phrases that contain experiencer usually have clauses. It should be mentioned that clauses are out of focus of this work, therefore the experiencer role is not used with all of its features.

5.3 Class Hierarchy and Complexity of the Proposed Model

The proposed method has two main classes and four auxiliary classes. All words are kept in a data structure. This structure is provided using one auxiliary class named *Node*. One main class creates the phrase structure trees named *PhraseStructureTree* (PST) for each sentence with the help of *Node* class. These trees are combined using another auxiliary class. Another main class named *ThematicRoleExtractor* (TRE) examines the created trees by thematic roles. The complexity of the auxiliary classes with respect to the main classes can be neglected.

In PST class, in the beginning, a lexicon is created and phrases are divided. These processes are linear; thus, their time consumption depends on the number of words so complexity is $O(n)$. Phrase creation processes vary. Most of them are created as a bottom-up tree, but in some cases, the top-down approach is used.

DP', NP'', AuxP, and AdvP are created using the bottom-up approach with recursion. The time complexity of them is in (5.11).

$$T(n) = c + T(n - 1) \quad (5.11)$$

The c value in (5.11) indicates the number of some unnecessary processes. The n value holds the number of incoming words. So, we can say that complexity in the worst case will be $O(n)$.

The time complexity of PP, AdjP, and VP depends on their other branches which have a sub-tree. Their one branch includes only one word. Thus, the time complexity of the one-word side is negligible. The time complexity of the sub-tree side is $O(n)$.

DP, NP, NP', and TP have sub-trees both branches, so their time complexities are such as in (5.12). According to (5.12), the worst cases of these sub-trees are accepted $O(n)$.

$$T(n) = c + T_{left}(n) + T_{right}(n) \quad (5.12)$$

TRE class finds the thematic roles using the depth-first search (DFS) method. As known, the time complexity of DFS in the worst case is $O(n)$ because DFS is a kind of linear search. Therefore, the time complexities of each finding process are $O(n)$.

5.4 Implementation

To implement the proposed model, three different technologies are used. Lexicon is stored in a database. SQLite is preferred for database operations. Phrase structure trees and thematic roles are created using the C# programming language in .NET framework. To draw the storyboard Unity Game Engine is used.

.NET is an open-source and cross-platform development platform, which has a collection of languages and libraries. In short, .NET is a framework. .NET and Microsoft .NET are the same things. With .NET, multiple languages, editors, and libraries can be used to build web, mobile, desktop, gaming, and IoT.

.NET has three programming languages that are C#, F#, and Visual Basic. C# is a simple, modern, object-oriented, and type-safe programming language. F# is a cross-

platform, open-source, functional programming language for .NET. It also includes object-oriented and imperative programming. Visual Basic is an approachable language with a simple syntax for building type-safe, object-oriented applications (Microsoft Corporation, 2019b).

C# is an elegant and type-safe object-oriented language that enables developers to build a variety of secure and robust applications that run on the .NET Framework. To create Windows client applications, XML Web services, distributed components, client-server applications, database applications, etc., C# can be used. Sometimes C# is also called C# .NET (Microsoft Corporation, 2019a).

SQLite is an in-process library that implements a self-contained, serverless, zero-configuration, transactional, and most importantly open-sourced SQL database engine (SQLite, 2019). SQLite is very light and can be integrated with any platform.

Unity is a cross-platform game engine for creating games in both 2D and 3D. Unity supports building games for many platforms such as iOS, Android, Windows, PlayStation, etc. Since its release in 2005, Unity is one of the most popular game engines. Unity gives the user the ability to create games and experiences in both 2D and 3D. For the main scripting in Unity, C# is used. Unity allows the specification of texture compression and resolution settings for each platform that the game engine supports. Unity editor is supported on Windows and MacOS. An editor is available for Linux as well but it is in an experimental stage.

5.4.1 Implementation of Phrase Structures

Proposed algorithms are implemented using the C# programming language. The main reason to choose C# is its strong object-oriented structure. The 3-tier architecture is preferred as software architecture. In this way, the written program can be moved to any environment as DLLs (dynamic link libraries). In some points, we have used design patterns to strengthen to OOP paradigm.

Developed 3-tier architecture has four levels. The 0th tier is the database level. The physical database is here. The 1st tier provides the connection between the database and the higher levels. All database queries are stored anonymously in this layer. In this way, each query can be executed independently from database tables.

The 2nd tier of the architecture is a middle layer which creates the data models. This layer provides the integrity of the data types from the 3rd layer and the data types from the database. The 3rd layer contains unique variables. These variables cannot adapt to the database without some conversion. Therefore, the 2nd layer is vital. This layer also generates the model-based database queries and passes them to the 1st layer.

The highest level, also known as the 3rd layer, is the main process layer. In this stage, first of all, a lexicon is created. This lexicon allows us to check whether words in a sentence are registered in the database. It also holds the properties of words.

At the same time, since this work is an engineering work, design patterns were used in the software development phase. An abstract factory pattern is used to instantiate each word because our words are divided into four categories. On the other hand, new features should be added to the words from time to time. A decorator pattern is designed to add them. In addition, sometimes the usage areas of words change. Therefore, a strategy pattern is used. A singleton pattern has been used to keep the database connection stable.

Secondly, encountered sentences are parsed and the model of the word is filled. Until now, word structures have been obtained. Phrase structures are created thereafter. These phrase structures are merged as a tree. Now we have a tree.

Lastly, this tree is traversed to find thematic roles. The depth-first search method is used to find roles. However, to reduce traverse time, some adjustments have been made. In English and also in our controlled natural language, actors always come before the predicate and the patient always comes after the predicate. Accordingly, only the left child of the root was traversed to find the actor, and only the right child

of the root was traversed to find the patient. Locative roles can be anywhere in the sentence such as in (5.13). So, to find them, some static approaches are used mentioned below.

The girl **at home** walked.
The girl walked **into the home**. (5.13)

To determine these cases, preposition and noun relation is analyzed. Considering (5.13), the word “home” is a place, and prepositions “at” and “into” have the place feature. This information is held in the database. Another method is to define dimensions for objects. Bigger objects are assumed as a place, so they can be used as a locative.

Another case about the location is multiplicity, e.g. considering two sentences such as “the girl walked to the kitchen into home” and “the girl saw the dog”; the location in the first sentence is home and the goal is the kitchen. In the second sentence, there is no information about the location, but it is known that the girl is in the kitchen and also at home. This was solved in the drawing step as will be explained later.

5.4.2 Thematic Role Stacks

Thematic roles can be determined differently by the reviewer. Therefore, we will act on our assumptions from this stage. As mentioned in the previous section, thematic roles have a dynamic form. The role, formerly a patient, could then be an actor or the location of an object can be changed. These changes should be traced. On the other hand, due to the ambiguity in the sentence, the determination of the roles can also be made wrong. For example, considering the sentence “the girl from the room is walking”, it can be extracted two meanings. One is about the location of the girl where we remember. The second one is about the starting point of the walking event. To solve these problems, we need a rule-based approach. Therefore, ten stack structures were developed for predicate and each thematic role. All determined roles are put into the corresponding stack.

An initial assignment is made to the stack with the push operation. When Unity starts to run, the program detects the tree structure and determines thematic roles according to a given sentence. *ThematicRoleExtractor* class has *FindPredicate*, *FindActor*, *FindPatient*, *FindSource*, *FindGoal*, *FindLocation*, and *FindInstrument* functions that find the predicate, the actor, the patient, the source, the goal, the location, and the instrument respectively. In Unity, these functions should be called in the right order. These obtained thematic roles are placed hypothetically in the corresponding stack. We call them *initial roles*. The actual location of a role or the type of the subject becomes definite after the second step, the interpretation phase. The tense information modifies locatives. The aspect information modifies the drawing process. The MAS and transitivity information modify the actor, the experiencer, the patient, and the theme. With the help of this information, we replaced all roles.

Considering three sentences such as 1- “the girl walked from the room to the kitchen through the hallway in the house”, 2- “the girl walks from the room to the kitchen through the hallway in the house”, and 3- “the girl will walk from the room to the kitchen through the hallway in the house”, the location in time of the sentence 1 is in past, the sentence 2 is in present, and the sentence 3 is in future. We accept that the location role of 1 is the same as the goal role because the event has finished. Thus, the last location of the girl is *the kitchen*. Likewise, the location role of 3 is the same as the source role (the room), because the girl has not gone to the kitchen yet. In sentence 2, the location role remains as it is because this sentence indicates a general behavior and we know that the girl is always in the house in this case. These changes can be seen in Table 5.2.

Table 5.2 Effects of the concept of tense on locatives

Sentence	Thematic Roles	The Initial Placement	The Interpreted Placement
the girl walked from the room to the kitchen through the hallway in the house	Source	The room	The room
	Goal	The kitchen	The kitchen
	Location	The house	The kitchen
the girl walks from the room to the kitchen through the hallway in the house	Source	The room	The room
	Goal	The kitchen	The kitchen
	Location	The house	The house
the girl will walk from the room to the kitchen through the hallway in the house	Source	The room	The room
	Goal	The kitchen	The kitchen
	Location	The house	The room

There is another case about locatives, as we can see in Table 5.2, there is a locative hierarchy. It can be seen that the house covers the room, the kitchen, and the hallway. This hierarchy helps us when there is a situation that no locative information such as an ongoing sentence “the girl is out of the kitchen”. We accept that the last location of the girl is *the house*.

With the same approach, figural roles are relocated in stacks using the MAS and transitivity attributes. The predicate must be checked first, and the attribute to be checked is the MAS attribute. If the predicate's MAS is a state, the actor replaced in the experiencer stack and the patient replaced in the theme stack. The second priority is on the transitivity attribute. If its value is false, we accept that the actor is a theme. Otherwise, all figural roles stay as they are.

Each role is interdependent at its level. If a role has not changed, an upper position in the stack is marked with a null value. For example, if an actor has not changed and no new information has been received about the location, that actor's location is still its old position, because one upper index in the location stack is marked null. We used the symbol ‘/0’ to mark the null value. On the other hand, if the actor and location have changed, information is entered into the corresponding stacks. Thus, the position of

the old actor remained constant, but the new actor has its location. We can see all stack operations in Figure 5.5 according to example sentences:

1. The girl walked from the room to the kitchen through the hallway in the house.
2. The girl saw the dog.
3. The dog was chasing the cat.
4. The cat ran to the garden.
5. The girl ate a potato with a fork.
6. The fork fell.
7. The cat hid.

6	hid	the cat	/0	/0	/0
5	fell	the fork	/0	/0	/0
4	ate	the girl	/0	/0	a potato
3	ran	the cat	/0	/0	/0
2	was chasing	the dog	/0	/0	the cat
1	saw	the girl /0	/0 the girl	/0 the dog	the dog /0
0	walked	the girl	/0	/0	/0
	PREDICATE	ACTOR	EXPERIENCER	THEME	PATIENT

6	/0	/0	/0 the garden	/0	/0
5	/0	/0	/0 the kitchen	/0	/0
4	/0	/0	/0 the kitchen	/0	a fork
3	/0	the garden	/0 the garden	/0	/0
2	/0	/0	/0 the kitchen	/0	/0
1	/0	/0	/0 the kitchen	/0	/0
0	the room	the kitchen	the house the kitchen	the hallway	/0
	SOURCE	GOAL	LOCATION	PATH	INSTRUMENT

Figure 5.5 The result of stack operations according to given sentences

The detection of locatives depends on the context. In the stories, the location information is given in the first place and is not mentioned again and again until a new

location receives. For this reason, in the stacks containing the locatives, the previous location information is frequently carried. If there is no location information, we use the source or the goal information. Our main approach to finding the locatives is based on the tense of the verb. If the tense is past, we accept that the event is done and the goal is reached. So, we can say that the last location of the actor is the goal. Likewise, we can say that the last location is the source if the tense is future. Because the actor has not done an event right now but it will do and it is in its first location. If the tense is present, we cannot use the source or goal due to the fact that the event points a habit or regular action. Therefore we used the location information itself if there is. On the other hand, if we have no aforementioned information, we carry the previous locative information such as the (5.14) algorithms.

```

Rearrange-Locatives(initial_roles, stacks)          (5.14)
1  i = index of PREDICATE(peak())
2  nil = '/'
3  loc = nil
4  IF(predicate.Tense == PAST)
5      IF(goal != nil)
6          LOCATION.pop()
7          LOCATION.push(goal)
8  ELSEIF(location == nil)
9      IF(loc = (Find the place of actor in ACTOR)
        != nil)
10     ELSEIF(loc = (Find the place of actor in
        PATIENT) != nil)
11     ELSEIF(loc = (Find the place of actor in
        EXPERIENCER) != nil)
12     ELSEIF(loc = (Find the place of actor in
        PATIENT) != nil)
13     ELSEIF(loc = (Find the place of actor in
        INSTRUMENT) != nil)
14     ENDIF

```

```

15     LOCATION.pop()
16     LOCATION.push(loc)
17     ENDIF
18 ELSEIF(predicate.Tense == FUTURE)
19     IF(source != nil)
20         LOCATION.pop()
21         LOCATION.push(source)
22     ELSEIF(location == nil)
23         IF(loc = (Find the place of actor in ACTOR)
           != nil)
22         ELSEIF(loc = (Find the place of actor in
           PATIENT) != nil)
23         ELSEIF(loc = (Find the place of actor in
           EXPERIENCER) != nil)
24         ELSEIF(loc = (Find the place of actor in
           PATIENT) != null)
25         ELSEIF(loc = (Find the place of actor in
           INSTRUMENT) != nil)
26     ENDIF
27     LOCATION.pop()
28     LOCATION.push(loc)
29     ENDIF
30 ELSEIF(predicate.Tense == PRESENT)
31     IF(location != nil)
32         LOCATION.pop()
33         LOCATION.push(location)
34     ELSEIF(location == nil)
35         IF(loc = (Find the place of actor in ACTOR)
           != nil)
36         ELSEIF(loc = (Find the place of actor in
           PATIENT) != nil)

```

```

37     ELSEIF(loc = (Find the place of actor in
                     EXPERIENCER) != nil)
38     ELSEIF(loc = (Find the place of actor in
                     PATIENT) != nil)
39     ELSEIF(loc = (Find the place of actor in
                     INSTRUMENT) != nil)
40     ENDIF
41     LOCATION.pop()
42     LOCATION.push(loc)
43 ENDIF
44 ENDIF

```

A similar approach is used to push the figurals in stacks. In this case, we use the MAS and the transitivity attribute of the verb. As mentioned earlier, we do not fill the experiencer and theme stacks at the beginning. We place the subject and object in the actor and patient stacks. After that, as can be seen in (5.15), we replaced them in the corresponding stacks.

```

Rearrange-Locatives(initial_roles, stacks)
1  nil = '/'0'
2  IF(predicate.MAS == STATE)
3      EXPERIENCER.pop()
4      EXPERIENCER.push(ACTOR.pop())
5      ACTOR.push(nil)
6      IF(predicate.Transitivity == TRUE)
7          THEME.pop()
8          THEME.push(PATIENT.pop())
9          PATIENT.push(nil)
10     ENDIF
11 ELSEIF(predicate.Transitivity == FALSE)
12     THEME.pop()
13     THEME.push(ACTOR.pop())

```

(5.15)

```
14     ACTOR.push(nil)
15 ENDIF
```

After all of these operations, the places of the thematic role are definite. We can start to draw our storyboard.

5.4.3 Implementation of Storyboard

The storyboard is created using the Unity Game Engine (Unity) as mentioned earlier. Unity has its code editor and the C# programming language can be used in it. The codes, which create phrase structure trees, are converted to DLLs. These DLLs and database files are imported into Unity. The thematic role extraction process is written into Unity because Unity has its own OOP approach.

There is a concept named `GameObject` in unity. `GameObjects` are the fundamental objects that represent characters, props, and scenery (GeeksforGeek, 2019). They act as containers for components, which implement the real functionality. There is another concept named `Prefab` in Unity. It is, actually, an asset type. It allows storing a `GameObject` object completely with components and properties. Prefabs are similar to an object in object-oriented programming. Prefabs provide the reusability of assets.

It is needed to create a `Prefab` database, which keeps the object in the database. To make this, several prefabs are created. Unity can only use Prefabs in its file system. Thus, they are stored in a folder in Unity.

FindPredicate function returns the tense phrase of the sentence. At this point, some assumptions are needed on the tense phrase. In this work some rules are determined. The pseudocode of these is seen in (5.16).

```

    Draw-Scenes(all stacks)
1  predicate = PREDICATE.peak()
2  IF(predicate.MAS != MOTION)
3      DRAW(1 static scene)
4  ELSE
5      IF(predicate.Aspect == SIMPLE or PERFECT)
6          DRAW(3 static scenes)
7          ELSEIF(predicate.Aspect == PROGRESSIVE or
8                  PERFECT_PROGRESSIVE)
9              DRAW(1 animated scene)
10     ENDIF
11 ENDIF

```

(5.16)

In (5.13), some points need to be explained. In our work, there are two types of scenes: static and animated. Static scenes are just an image; however, animated scenes contain an animation. An example is explained below.

Considering these four sentences: (1) “the girl walked from the room to the kitchen in the house”, (2) “the girl saw the dog”, (3) “the dog is chasing the cat”, and (4) “the cat hid under the table”. (1)’s predicate is “walked”, the actor is “the girl”, location is “the house”, the source is “the room”, and the goal is “the kitchen”. Type of the predicate is motion because there is a relocation situation. The type of tense is past and the aspect is simple. Therefore, three static scenes are needed such as in Figure 5.6.



Figure 5.6 Three static scenes example of the sentence “the girl walked from the kitchen to the room in the house”

(2)'s predicate is “saw”, the patient is “the dog”, and the actor is “the girl”. This sentence has no source, goal, or location information, but it is known from (1) that the girl's last location is “the kitchen”. In consequence, location information is carried into (2). The type of predicate is action, the tense is past and the aspect is simple. Conforming this, Figure 5.7-like scene should be drawing.



Figure 5.7 One static scene example of the sentence “the girl saw the dog”

The predicate of (3) is “chasing”. Its actor is “the dog” and its patient is “the cat”. The verb *chase* is a motion and the aspect is progressive. So, one animated scene such as Figure 5.8 is needed. Note that, because animations cannot be shown here, it is illustrated.



Figure 5.8 One animated scene example of the sentence “the dog is chasing the cat”

As we can see in Figure 5.7, location information was again carried. The event still is in the kitchen. Furthermore, since no information has been given about the previous actor, it is assumed that she is still in the kitchen. She is somehow an experiencer.

(4)'s predicate is "hid", the actor is "the cat". The type of predicate is action and its tense is simple, thus, one static scene such as in Figure 5.8 is needed.



Figure 5.9 One static scene example of the sentence "the cat hid under the table"

It should be noted that some thematic roles were carried. They are extracted using another function named *ExtractSemanticComponents*. As mentioned previously, some semantic implications are made.

5.4.4 Implementation of Archetype Analysis

Archetype analysis is a context-dependent process. A given context is analyzed to extract archetypes. At this point, we used tagging and gathering information techniques. Although tagging is used to find part of speeches, we used it here to find archetypes.

The most important archetype is the hero. The character that appears most often is tagged as the hero. This means this character is generally the actor thematic role. The other archetypes are tagged according to their behaviors and thematic roles. Therefore, we tagged verbs as *archetype encounterer* (archen). Archen verbs are generally verbs

that do not do the hero. The character which has an archen verb tagged null-archetype at the beginning. After the gathering information process, its archetype is assigned.

For example, consider J. K. Rowling's Harry Potter series. The hero is Harry Potter. Another character is Albus Dumbledore. When Albus Dumbledore first appears as the patient, it is marked null-archetype with the help of its archens. As we examine the archens, it is seen that the null-archetype tagged character gives missions to our hero and mentor in some places. After this process, its archetype is updated as the old wise man.

5.4.5 Complexity of Drawing Process

The time complexity of the drawing process depends on Unity. On the other hand, in the implementation stage, stacks are used. They are the main elements that create time complexity in this application. So, it is needed to analyze stacks.

Stacks are linear, array-based structures, thus their time complexities are lower than space complexities. To extract thematic roles, the push operation is used highly and each process' time complexity is $O(1)$. Also, there are nine stacks in our application, so $9 * O(1)$ times push operation is executed.

In the drawing process, the pop operation is used. Because of our design, the first sentence to be drawn is at the bottom of the stack. Therefore, the pop operation needs to be performed n times. Originally, the pop operation's time complexity is $O(1)$, however, in our work, we need n time pop and the time complexity is calculated as $n * O(1)$. Moreover, when comparing the location information, we perform n times pop.

The space complexity of the developed stack is dependent on the number of sentences. Thus, the application takes up nine stacks of space.

5.5 Results

In experiments with approximately fifty sentences, the proposed model has often successfully created phrase structure trees. All sentences are in Appendix C. In some situations, phrase structure trees have been ill-formed. The errors are mostly created by homonyms. Homonyms have the same form but differ in meaning such as bat (= animal, club).

As a result of the investigations, it was found that these errors are due to adjectives and adverbs which are homonyms. For example, the word *very* can be an adjective or an adverb. In normally, “very” is an adverb because it is used to emphasize an adjective, adverb, or phrase, however, in an example such as “*this very room*”, it can be used as an adjective.

The errors mentioned above are only problematic in meaning dimension thanks to the generative grammar approach. Out of this, the tree-building process is very successful. In (6.1), an analysis of an example sentence can be seen.

```
+ - the girl walked from the kitchen to the room      (6.1)
in the house
| +- the girl
| | +- the
| | | +- <null>
| | | +- the
| | | | +- <null>
| | | | +- <null>
| | +- girl
| | | +- girl
| | | | +- <null>
| | | | +- girl
| | | | | +- <null>
| | | | | +- girl
```

| | | | | +- <null>
 | | | | | +- <null>
 | | | +- <null>
 | +- walked from the kitchen to the room in the
 house
 | | +- walked
 | | | +- <null>
 | | | +- walked
 | | | | +- <null>
 | | | | +- <null>
 | | +- from the kitchen to the room in the house
 | | | +- <null>
 | | | +- from the kitchen to the room in the
 house
 | | | | +- <null>
 | | | | +- from the kitchen to the room in
 the house
 | | | | | +- from
 | | | | | +- <null>
 | | | | | +- <null>
 | | | | | +- the kitchen to the room in the
 house
 | | | | | +- the
 | | | | | | +- <null>
 | | | | | | +- the
 | | | | | | | +- <null>
 | | | | | | | +- <null>
 | | | | | | +- kitchen to the room in the
 house
 | | | | | | +- kitchen
 | | | | | | +- <null>
 | | | | | | +- kitchen


```

| | | | | | | | | | | | | | | +-
<null>
| | | | | | | | | | | | | | | +-
<null>
| | | | | | | | | | | | | | +- house
| | | | | | | | | | | | | | +- house
| | | | | | | | | | | | | | +-
<null>
| | | | | | | | | | | | | | +-
house
| | | | | | | | | | | | | |
+- <null>
| | | | | | | | | | | | | |
+- house
| | | | | | | | | | | | | |
+- <null>
| | | | | | | | | | | | | |
+- <null>
| | | | | | | | | | | | | | +-
<null>

```

CHAPTER SIX

CONCLUSION

6.1 Conclusion

In this dissertation, it is suggested that a syntactic structure can be created by using scenarios and some semantic elements can be identified and a storyboard can be formed according to the result.

The main argument of the thesis has been justified within a delimited framework resting on some experiments. Here, the concept of the restricted frame refers to the controlled natural language created. Again, to remind, created controlled natural language includes the rules mentioned below:

1. This study works in English.
2. Accepted word types are verbs, determiners, nouns, adjectives, adverbs, prepositions, and auxiliary verbs including modals.
3. Accepted sentence structures are only simple sentences. Clauses are out of the concern. Therefore, the experiencer thematic role is not examined.
4. All nouns that can be drawn are accepted.
5. Abstract concepts are out of concern. This point is important because abstract concepts usually cannot be drawn in storyboards.
6. Abstract concepts are used to determine archetypes.
7. We only examined a few of archetypes such as the hero and the wise old man.
8. Phrase structure rules are redefined to ease the computation.
9. Selected contexts should be third-person narration.

This study shows we can reach handle and pragmatic levels in a limited framework using syntactic structures, thematic roles, and archetypes.

We extracted each thematic role of each sentence from the context with the help of phrase structures. These roles allow us to find archetypes. The archetype determination

process is very pragmatic, thus we accepted some assumptions. Therefore, we obtained the wrong results sometimes. The archetype determination should be considered and more mathematical approaches should be taken.

In this study, the most appropriate method in terms of computation has been developed by inspiring from many studies in the field of linguistics and natural language processing. Our approach tries to connect the syntax level to the pragmatic level. We can say that we have fulfilled this claim within our limited frame.

6.2 Future Works

Research in NLP, which is based on and goes hand in hand with linguistic studies, is a lifelong interest. Therefore, this thesis should be treated as a starting point for further studies. As new findings are obtained in the field of linguistics, the field of NLP also develops. Radical decisions can be taken and models can be created from the beginning. Consequently, this work will be updated many times in the future.

On the other hand, the drawing process is more static than the desired stage. This stage can be fed by artificial intelligence studies and attain a more flexible structure. Instead of using ready-made objects for the next stages, new images can be drawn using artificial intelligence methods.

Another field of study that should be exploited when analyzing thematic roles is that of computational ontology. Computational ontologies, including semantic webs, can be highly useful to come up with a holistic model of thematic relations and, thereby, with a sound basis for semantic analyses.

Determiners can have semantically crucial meanings. For example, in the case of articles, semantic expressions can be added by means of some small additions. For example, the article "the" points to a specific object. Starting from this, in the expression "the dog", a collar can be added to the dog in the drawing. In this way, it

can be tried to prove ownership of the dog. Although there are a lot of other possibilities, this suggestion may at least be the beginning.

Besides, verbs and auxiliary verbs can help in some matters. When creating the storyboard, shadows can be made according to the times in the context. Past tenses can be drawn in a gray background. Auxiliary verbs in case of possibility may also be performed in less significant drawings.

Although syntax is taken to be autonomous, some conclusions can be made with the help of syntactic analysis and the hero's journey.

As is known, in the hero's journey, The Hero can be observed almost everywhere. Based on this, the actor role can be expected to be the hero himself at many points. We can assume that we are opposed to another archetype at the points where the actor information changes. By looking at the actions of this new character, it can be made an assumption about which archetype it is.

REFERENCES

- Aikhenvald, A. Y. (2003). *Classifiers: A typology of noun categorization devices*. *Oxford Studies in Typology and Linguistic Theory*. Oxford: Oxford University Press.
- Aktaş, Ö. (2010). *Rule-Based natural language processing methods for Turkish*. PhD Thesis, Dokuz Eylül University, İzmir.
- Arnowitz, J., Arent, M., & Berger, N. (2007). Storyboard prototyping. In *Effective prototyping for software makers* (294–315). Massachusetts: Morgan Kaufmann.
- Başarıcı, S., & Kılıçaslan, Y. (2017). Hero's journey as a lattice structure: a case study of star wars. *Journal of Balkan Libraries Union*, 5 (2), 17–23.
- Bybee, J., Perkins, R., & Pagliuca, W. (1994). *The evolution of grammar*. Chicago: The University of Chicago Press.
- Campbell, J. (1972). *The hero with a thousand faces* (2nd ed.). New York: Pantheon Books.
- Carnie, A. (2006). *Syntax: A generative introduction* (2nd ed.). New Jersey: Blackwell Publishing.
- Chomsky, N. (1957). *Syntactic structures* (2nd ed.). Berlin: Walter de Gruyter.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Massachusetts: MIT Press.
- Chomsky, N. (1980). Remarks on nominalization. In *Studies on semantics in generative grammar* (11–62). Berlin: Walter de Gruyter.

- Chomsky, N. (1982). *Some concepts and consequences of the theory of government and binding*. (S. J. Keyser ed.). Massachusetts: MIT Press.
- Chomsky, N. (1992). *A minimalist program for linguistic theory*. MIT Working Papers in Linguistics.
- Coates, J. F. (2000). From my perspective: Scenario planning. *Technological Forecasting and Social Change*, 65 (1), 115–123.
- Crystal, D. (2008). *Dictionary of linguistics and phonetics* (6th ed.). New Jersey: Blackwell Publishing.
- Dalrymple, M. (2006). Lexical functional grammar. In *Encyclopedia of language & linguistics* (82–94). Amsterdam: Elsevier.
- Dowty, D. R. (1989). On the semantic content of the notion of ‘thematic role.’ In *properties, types and meaning* (69–129). Dordrecht: Springer.
- Eagles. (2018). *Thematic roles*. Retrieved December 23, 2019, from <http://www.ilc.cnr.it/EAGLES96/synlex/node62.html>
- Fillmore, C. J. (1968). The case for case. In *Universals in Linguistic Theory* (1–88). Newyork: Holt McDougal.
- Fleisher, N. (2007). *Quick notes on HPSG. Syntax*.
- Gazdar, G., Klein, E., Pullum, G. K., & Sag, I. A. (1985). *Generalized phrase structure grammar*. Massachusetts: Harvard University Press.
- Geeksforgeeks. (2019). *Game development with unity: Introduction*. Retrieved December 23, 2019, from <https://www.geeksforgeeks.org/game-development-with-unity-introduction/>

- Gruber, J. S. (2008). *Lexical structures in syntax and semantics* (25th ed.). Amsterdam: North-Holland Pub. Co.
- Gunji, T. (1986). "Generalized phrase structure grammar" by G. Gazdar, E. Klein, G. K. Pullum, and I. A. Sag. *New Generation Computing*, 4 (2), 217–219.
- Hart, J. (2008). *The art of the storyboard: A filmmaker's introduction* (2nd ed.). Massachusetts: Focal Press.
- Hesse, H. (1922). *Siddhartha*. Berlin: S. Fischer Verlag.
- Jackendoff, R. (1972). *Semantic interpretation in generative grammar*. Cambridge: The MIT Press.
- Jackendoff, R. (2003). *Foundations of language: Brain, meaning, grammar, evolution* (1st ed.). Oxford: Oxford University Press.
- Jantke, K. P., & Knauf, R. (2005). Didactic design through storyboarding : Standard concepts for standard tools. *4th International Symposium on Information and Communication Technologies*, 205, 20–25.
- Jensenius, A. R. (2018). *Definitions: Motion, action, gesture*. Retrieved December 23, 2019, from <http://www.arj.no/2012/11/01/motion-action-gesture/>
- Joshi, A. K., & Schabes, Y. (1997). Tree-Adjoining grammars. In *handbook of formal languages* (69–123). Berlin: Springer.
- Judaeus, P. (1981). *Philo*. Massachusetts: Harvard University Press.
- Jung, C. G. (1964). *Man and his symbols*. Massachusetts: Anchor Press.
- Jung, C. G. (2003). *Psychology of the unconscious*. New York: Dover Publications.

- Jung, C. G. (2014a). *Collected works of c.g. jung, volume 14: Mysterium coniunctionis*. New Jersey: Princeton University Press.
- Jung, C. G. (2014b). *Collected works of c.g. jung, volume 9 (part 1): Archetypes and the collective unconscious*. New Jersey: Princeton University Press.
- Kaplan, R. M., & Bresnan, J. (1982). Lexical-Functional grammar: A formal system for grammatical representation. In *The mental representation of grammatical relations*, (173–281). Massachusetts: MIT Press.
- Kiliçaslan, Y., Güner, E. S., & Yildirim, S. (2009). Learning-based pronoun resolution for Turkish with a comparative evaluation. *Computer Speech and Language*, 23 (3), 311–331.
- Kilicaslan, Y., & Tuna, G. (2014). An nlp-based approach for improving human-robot interaction. *Journal of Artificial Intelligence and Soft Computing Research*, 3 (3), 189–200.
- Kiliçaslan, Y., Uçar, Ö., Uçar, E., & Güner, E. S. (2008). Visualization of Turkish for autistic and mentally retarded children. *Proceedings of the 2nd International Conference on Pervasive Computing Technologies for Healthcare 2008, PervasiveHealth*, 1 (1), 144–147.
- Kılıçaslan, Y., & Tuna, G. (2015). Linking FrameNet and natural languages via thematic role structures. *International Journal of Languages, Literature and Linguistics*, 1 (3), 158–163.
- König, E., & van der Auwera, J. (1994). *The germanic languages* (1st ed.). Abingdon: Routledge.

- Kracht, M. (2002). Referent systems and relational grammar. *Journal of Logic, Language and Information*, 11(2), 251–286.
- Langacker, R. W. (2008). *Cognitive grammar: An introduction*. Oxford: Oxford University Press.
- Mai, L., Le, H., Niu, Y., & Liu, F. (2011). Rule of thirds detection from photograph. *Proceedings - 2011 IEEE International Symposium on Multimedia, ISM 2011*, 91–96.
- Makharoblidze, T. (2018). On Georgian preverbs. *Open Linguistics*, 4 (1), 163–183.
- Mani, I. (2001). *Automatic summarization*. Amsterdam: John Benjamins Publishing Company.
- Manning, C. D. (1994). *Ergativity: Argument structure and grammatical relations*. PhD Thesis, Stanford University, Stanford.
- Manning, C.D., Raghavan, P. (2008). *Introduction to Information Retrieval*. Cambridge: Cambridge University Press.
- Margolis, M. L. (1909). The Greek preverb and its Hebrew-Aramaic equivalent. *The University of Chicago Press Journals*, 26 (1), 33–61.
- Microsoft Corporation. (2019a). *C# programming guide*.
- Microsoft Corporation. (2019b). *What is .NET?* Retrieved December 25, 2019, from https://dotnet.microsoft.com/learn/dotnet/what-is-dotnet?WT.mc_id=Educationaldotnet-c9-scottha
- Morenberg, M. (2009). *Doing grammar* (4th ed.). Oxford: Oxford University Press.

- Müldür, F. (2016). Noam Chomsky’de üretici dilbilgisi: Derin yapı ve yüzey yapı ayrımı. *Uludağ University Faculty of Arts and Sciences Journal of Philosophy*, (27), 59–74.
- Myss, C. (2019). *The four archetypes of survival*. Retrieved December 23, 2019, from <https://www.myss.com/free-resources/sacred-contracts-and-your-archetypes/appendix-the-four-archetypes-of-survival/>
- Payne, T. E. (2006). Grammatical relations. In *Exploring language structure: A student’s guide* (Student ed.). (210–236). Cambridge: Cambridge University Press.
- Pearson Education. (2019). *Longman dictionary of contemporary English*. Retrieved May 20, 2019, from <http://www.ldoceonline.com/dictionary/maze>
- Perlmutter, D. M. (1986). *Studies in relational grammar 1*. Chicago: University of Chicago Press.
- Perlmutter, D. M., & Postal, P. M. (1986). Some proposed laws of basic clause structure. In *Studies in relational grammar 1* (81–129). Chicago: University of Chicago Press.
- Pollard, C., & Sag, I. A. (1994). *Head-Driven phrase structure grammar*. Chicago: University of Chicago Press.
- Quiles, C., & Lopez-Menchero, F. (2017). *A grammar of Indo-European*. Badajoz: Academia Prisca.
- Radford, A. (1997). *Syntactic theory and the structure of English*. Cambridge: Cambridge University Press.
- Reiter, E., & Dale, R. (1997). Building applied natural language generation systems. *Natural Language Engineering*, 3(1), 57–87.

- Rhys, C. S. (2000). Chinese coverbs and argument. *Lingusitica Atlantica*, 22, 69–87.
- Sanfilippo, A. (1990). *Grammatical relations , Thematic roles and verb semantics*. PhD Thesis, University of Edinburgh, Edinburgh.
- SQLite. (2019). *About SQLite*. Retrieved December 23, 2019, from <https://www.sqlite.org/about.html>
- Traffis, C. (2020). *Transitive and intransitive verbs—What’s the difference?*. Retrieved December 23, 2019, from <https://www.grammarly.com/blog/transitive-and-intransitive-verbs/>
- Van Eynde, F., Schuurman, I., & Schelkens, N. (1998). *Computational linguistics in the Netherlands 1998: Selected Papers from the Ninth Clin Meeting*. Leiden: Rodopi.
- Wikipedia. (2019a). *Bayeux tapestry*. Retrieved December 23, 2019 from https://en.wikipedia.org/wiki/Bayeux_Tapestry
- Wikipedia. (2019b). *Google translate*. Retrieved December 23, 2019, from https://en.wikipedia.org/wiki/Google_Translate

APPENDICES

APPENDIX A Word Types

APPENDIX A.1 Verbs

All languages have verbs. Verbs are the type of words, which express an event, a situation, an occurrence, or a state. Without a verb, there can be no sentence. Verbs are distinguished from other parts of speech because of their possible inflection in tense, aspect, and mood, which is not possible for other parts of speech in most languages.

Tense–aspect–mood is a group of grammatical categories that covers the expression of location in time, the fabric of time – a single block of time, continuous flow of time, or repetitive occurrence, and mood degree of necessity, obligation, probability, ability (Bybee, Perkins, & Pagliuca, 1994). In addition, the voice may also be added.

In English, there are three tenses, which are past, present, and future. They have four aspects, which are simple, progressive, perfect, and perfect progressive. Moreover, there are three moods, which are indicative, subjunctive, imperative. If we talk about voice, there are two types, which are passive and active.

The number of their valency arguments categorizes verbs. Valency refers to the argument-taking potential of verbs. According to this reference system, verbs can be transitive, intransitive, ditransitive, copular, and avalent (Jackendoff, 2003; Morenberg, 2009).

Avalent verbs do not take any noun (subject or object). These type verbs are not available in English, but it can be seen in Spanish. For example, in (1), example sentences, other than English, do not take any subject or object. (1.b) is Spanish, (1.c) is Turkish and which are translations of (1.a).

- a) It rains. [English]
- b) Llueve. [Spanish] (1)
- c) Yağıyor. [Turkish]

Transitive verbs are followed by a noun or a noun phrase. These nouns should be direct objects. Intransitive verbs do not take any direct object. Ditransitive verbs take two nouns such as in (2). Copular verbs connect the subject with a complement. An example of these verbs, *seem*, *be*, and *become* can be given such as in (3).

Mehmet gave **Ayşe**_(N1) a **book**_(N2). (2)

This book **seems** funny. (3)

APPENDIX A.2 Nouns

The word “noun” is from the Latin word “nomen” and it means literally “name”. Therefore, it can be said that all named entities or entity groups such as living creatures, places, objects, ideas, etc. are nouns.

In many languages, nouns have a grammatical gender classifier. Grammatical gender classification includes masculine, feminine, neuter, animate, and inanimate. Sometimes, this classification is provided by using an affix, and sometimes, it is provided by making an agreement with a noun, determiner, or adjective. Other languages do not indicate gender.

The origin of gender systems is certainly not a well-understood phenomenon. A reason lies in the heterogeneous nature of gender markers whose origin is known. Sources for gender markers are numerous, and of different types (Aikhenvald, 2003).

In Latin, *ignis* means “fire” and it is masculine, differently, *flamma* means “flame” and it is feminine. In different circumstances, in German, *hund* means “dog” and when we need to express any dog, *hund* should be used. If when she-dog is expressed, there needs “-in” suffix and *hündin* should be used. Except for those, the word *arbeit* is

always feminine and it means “job” whereas also “job” is used German, which is then masculine. And also, German has neuter cases such as *auto*, which means “car”.

APPENDIX A.3 Pronouns

Pronouns are used as nouns. They are divided into seven classes: personal, reflexive, possessive, demonstrative, relative, interrogative, and indefinite. They are pronouns in origin, although some of them are processed in other word categories such as determiner. As in nouns, pronouns may have gender discrimination. For example, to indicate a noun, “he”, “she”, or “it” is used in English. The same rule applies in German and French.

Again, as in nouns, some languages such as Turkic, Austronesian, etc. have not any gendered situation. For example, in Turkish, there is only one type of third singular personal pronoun, which is *o*, although English has three which are *he*, *she*, *it*.

APPENDIX A.4 Determiners

Determiners serve to express the reference of that noun or a noun phrase in the context. The determiners can be articles, quantifiers, demonstrative adjectives, and possessive adjectives.

Articles are used for grammatical definiteness of the noun. Articles are used in different situations according to languages. English articles only state the definiteness or indefiniteness, German and French articles specify genders. In English, when there is an unspecified noun in context, indefinite articles “a” or “an” are used, but when a specified noun is indicated, definite article “the” is used. There is no gender specification in English.

Contrarily, in German, *der* is used for masculine, *die* is used for feminine and *das* is used for neuter nouns. In French, there are two cases, which are feminine *la* and masculine *le*, but for the plural noun, *les* is used regardless of gender.

In Turkish, there is no word for articles. In (4) and (5), it can be seen how the article mechanism works. As we can see in (4), there is no need to use an article but when the specific object is mentioned as in (5), a suffix needs for the definite article.

Et ye-di-m.
 Meat eat-past tense suffix-1st singular pronoun suffix
 I ate **a** meat. (4)

Et-i ye-di-m.
 Meat-the eat-past tense suffix-1st singular pronoun suffix
 I ate **the** meat (5)

English does not use an article before a proper noun such as “Mehmet” but some languages use. For example, in Portuguese, “o Mehmet” should be used.

Quantifiers are used to indicate the quantity of the noun. Quantifiers are only used to specify an indefinite number of nouns. If number values such as twelve, ten, etc. are used, that word becomes an adjective.

Demonstrative adjectives are used for distinguishing a noun from others. There are two of these type determiners in English, which are *this* and *that*. Nevertheless, some languages have more than two excluding plural form. In Table 1, we can see the differences.

Table Appendix A.1 Demonstrative adjectives distances in Japanese, English, and Turkish

Japanese	Turkish	English	Distance
Kono ringo	Bu elma	This apple	Next to
Sono ringo	Şu elma	That apple	Near
Ano ringo	O elma	-	Far

Possessive adjectives show the owner of the noun. Some languages use a possessive adjective before the noun, some use after the noun and some of them do not use any word but affix, as we can see in (6) as bold marked.

- a) **my** cat [English]
 - b) katten **min** [Norwegian]
 - c) kedi **-m** [Turkish]
- (6)

APPENDIX A.5 Auxiliary Verbs

Auxiliary verbs combine the main verb. It reinforces the effects of the main verb in terms of tense, aspect, and mood. In addition, they help to build question sentences and passive voices. Almost all languages have an auxiliary mechanism. Auxiliary verbs are never used alone.

Although there are approximately fifteen auxiliary verbs, which are divided into two classes that are primary and modal in English, German has three that *sein*, *haben*, and *werden*. Turkish has mainly two, which are *et-mek* and *ol-mak*. Note that, in Turkish, *-mek* or *-mak* are infinitive suffixes. To see an example, look at (7).

- Mutlu **ol**.
 - a) Happy be
Be happy.
 - Beni rahatsız **et-me**.
 - b) Me disturb do-not
Do not disturb me.
- (7)

APPENDIX A.6 Adjectives

In many languages, there is one lexical category whose primary function is to modify nouns. This category is generally labeled adjective. Adjectives can sometimes

be a single word as in (8.a), sometimes a word phrase as in (8.b) and sometimes multiple as in (8.c).

- a) **beautiful** girl
 - b) the **larger than expected** table
 - c) **fast, furious** driver
- (8)

In Proto-Indo-European languages, the noun could be modified in three different ways; with another noun, as in *stone wall*; with a noun in genitive, as in *the father's house*; or with an adjective, as in *paternal love* (Quiles & Lopez-Menchero, 2017).

APPENDIX A.7 Adverbs

Adverbs are words that modify a verb, adjective, determiner, clause, preposition, or sentence. They are generally answering questions such as *how*, *in what way*, *when*, *where*, and *to what extent*. They can be independent as in (9.a) or multiple as in (9.b), prefixes of verbal stems – originally independent but usually merged with them – and also a nexus between a noun and a verb as in (9.c), expressing a non-grammatical relationship, normally put behind, but sometimes coming before the word or sentence as in (9.d).

- a) He talks **slowly**.
 - b) He talks **very slowly**.
 - c) This one **always** smells.
 - d) **Especially**, this one is correct.
- (9)

APPENDIX A.8 Interjections

Interjections are exclamations of instant feelings such as to curse, to greet, to respond, etc. They change very quickly due to daily experiences. This category is closer to the phonological level. Many of them may not be directly translated into written language, hence the closest equivalents are tried to be extracted. These sounds

are written in different ways according to the language structure. Think about your expressions under any circumstances and try to write on it. Some of the close examples can be seen in (10).

- | | | |
|----|----------------------------------|------|
| a) | Hey! Buraya bak. [Turkish] | |
| | Hey! Look over here. | |
| | | (10) |
| b) | Berk! C'est dégoûtant.. [French] | |
| | Ugh! This is disgusting. | |

As we can see in (10), the sound *hey* used for calling is the same in English and in Turkish, but the sound that expresses disgust is written differently in French and English.

APPENDIX A.9 Clitics

Clitics refer to a form which resembles a word, but which cannot stand on its own as a normal utterance, being phonologically dependent upon a neighboring word (its host) in a construction (Crystal, 2008). According to language some of them can be placed before the host and this is called proclitic. Some of them can be placed after the host and are called enclitic. Moreover, other types can be placed middle of the stem of the host and are named endoclitics. Most known clitic in English is 's in "it's" for "it has" or "it is".

APPENDIX A.10 Contractions

Contraction is the process or result of phonologically reducing a linguistic form so that it comes to be attached to an adjacent linguistic form, or fusing a sequence of forms so that they appear as a single form (Crystal, 2008). Differently from clitic, contractions are the convenience of speech and the reflection of the writing. In English, contractions are used commonly such as *haven't*. Some examples can be seen in Table 2.

Table Appendix A.2 Contraction examples in different languages

Full-Form	Contraction Form	Language	English Translation
Ne haber?	Naber? (Mostly in slang)	Turkish	What is the news?
Für das	Fürs	German	For the
Le animal	L'animal	French	The animal

APPENDIX A.11 Adpositions

Adpositions are subsumed of prepositions, postpositions, and less encountered circumpositions. As its name implies, it is preposition if it comes before its complement, postposition if it comes after, or circumposition if it comes in the middle of its complement.

In English, there are more prepositions, such as in (11.a), than postpositions, (11.b), but vice versa in Turkish.

- a) The book is **on** the table.
b) I came here five years **ago**. (11)

In Turkish, postpositions can be provided with suffixes or separate words. In other words, almost all postpositions have a suffix form. In (12.a), the separate word form of a postposition can be seen and in (12.b), there is a suffix form. Also, in (12.c), there is a postposition that has not any suffix form.

- Kitap **ile**
a) book with
‘with book’
Kitap **-la**
b) book-with
‘with book’ (12)

c)

van der Auwera, 1994). Various constructions in English might also be analyzed as

‘a week ago’

(13)

APPENDIX A.12 Conjunctions

The main task of conjunctions is to connect words or phrases. In other

Coordinating conjunctions are used to give equal emphasis to a pair of main clauses.

Subordinators connect an independent clause and a dependent clause. As an

As an example of the adverb cases, the sentence “The heatwave was forecast to end

APPENDIX A.13 Coverbs

Coverbs are basically prefixes of verbs. This structure is mostly found in Asian languages and the most well-known language is Chinese. (14) shows an example from Chinese. In this example, *yong* is a coverb of verb *chi* (Rhys, 2000).

Lisi	yong	kuaizi	chi	fan	
Lisi	use	chopsticks	eat	rice	(14)
Lisi eats with chopsticks.					

APPENDIX A.14 Preverbs

A preverb is a prefix or particle preceding the root or stem of a verb, “a convenient term for a prepositional element in a compound verb” (Margolis, 1909). This category is named as a separable verb in Indo-European languages and the mechanism is slightly different. Examples from Georgian (15) and German (16) are given below (Makharoblidze, 2018).

aprens [Georgian]					
a-	pren	-s			(15)
he/she	lets	him/her/it/them	fly		

sie	kommt	sofort	an		
she	comes	immediately	at		(16)
‘She is arriving immediately.’					

APPENDIX A.15 Particles

Particle is a general term for a type of words that are not major parts of speech and are not inflected, especially one that does not readily fit into a standard classification of parts of speech. In German, particles are used to increase discourse quality as in (18). In this example, *doch* literally means “but” and *mal* literally means “once”. On

the other hand, in Afrikaans, a double particle system is used for negation such as in (17).

Hy	is	nie	hunger	nie	
He	is	not	hungry	not	(17)
'He is not hungry.'					

Versuch's	doch	mal	
Try it	but	once	(18)
'Come on, try it.'			

APPENDIX A.16 Classifiers

Measure words denote a unit of measurement and are used with mass nouns (uncountable nouns), and in some cases, also with count nouns such as “three cups of tea”. For English, after measure words usually *of* is used but some other languages do not use such structure such as in German “drei Tassen Tee”. This literally means “three cups tea”.

APPENDIX A.17 Cardinal Numbers

The traditional term retained in some models of grammatical description, referring to the class of numerals one, two, etc. cardinal numbers contrast with the ordinal numbers first, second, etc (Crystal, 2008).

APPENDIX B Full Form of the Developed Phrase Structure Tree

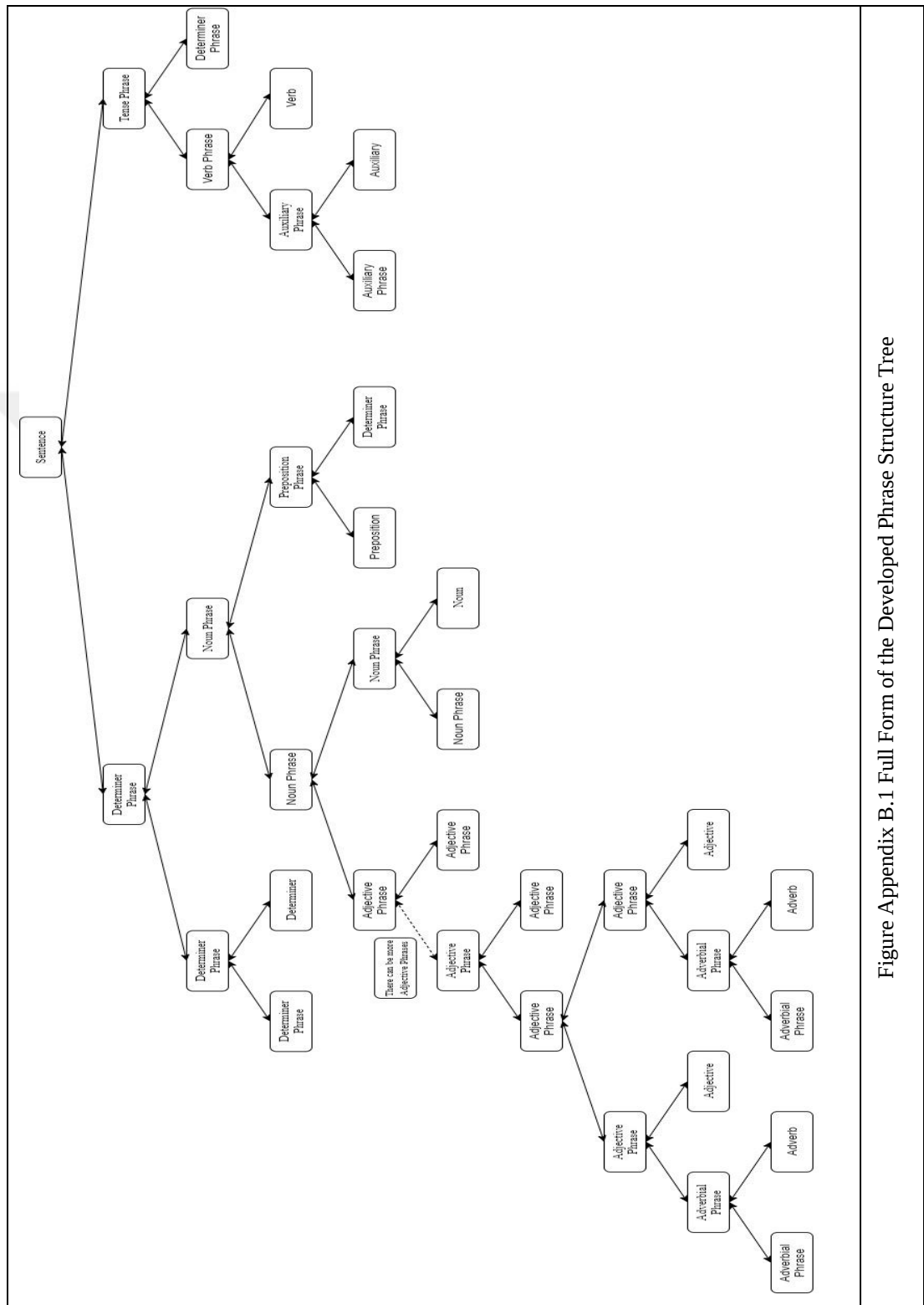


Figure Appendix B.1 Full Form of the Developed Phrase Structure Tree

APPENDIX C Sentence Examples to Test the Developed Method

Some sentences are deliberately written using the wrong grammar to test the developed phrase parser.

1	The girl walked from the room to the kitchen in home.
2	The girl saw the dog.
3	The dog was chasing the cat.
4	The cat ran to the garden.
5	The cat hid under the table.
6	The boy is running.
7	All the school boys are coming.
8	Some students have been running.
9	The eagles helped the fellowship of the ring.
10	The boy is watching the TV.
11	The extremely big surprisingly fast old fat man runs.
12	All the new school boys are working.
13	The extremely big surprisingly fast incredibly old fat man walks.
14	John built a tree house.
15	Colorless green ideas sleep furiously.
16	The cat from the home runs from the garden to the street.
17	The John has come.
18	Mary took a cat to her house.
19	Mary took the cat to her house.
20	Mary took a cat to a house.
21	Mary took a cat the home.
22	Mary took the cat.
23	She found a bird.
24	She found a bird on the tree.
25	She found a bird from the garden.
26	There is a big bench.

27	There is a very big dog.
28	A dog is running.
29	I was watching the TV.
30	Randy is watching a movie on the TV.
31	He is came here.
32	He is was run.
33	The she has come.
34	Some people are coming.
35	No one is coming.
39	John came home early.
40	John came to home early.
41	John came from home early.
42	The new best friends are talking.