View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Sabanci University Research Database

BUILDING A LEXICAL FUNCTIONAL GRAMMAR FOR TURKISH

by

ÖZLEM ÇETİNOĞLU

Submitted to the Graduate School of Engineering and Natural Sciences
in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

Sabancı University

June 2009

BUILDING A LEXICAL FUNCTIONAL GRAMMAR FOR TURKISH

APPROVED BY

Prof. Dr. Kemal Oflazer (Thesis Supervisor)	.....
Assoc. Prof. Dr. Aslı Göksel	.....
Assoc. Prof. Dr. Berrin Yanıkoğlu	.....
Prof. Dr. Cem Say	.....
Assoc. Prof. Dr. Hüsnü Yenigün	.....

DATE OF APPROVAL: .....

© Özlem Çetinoğlu 2009
All Rights Reserved

to my parents

Acknowledgments

I would like to express my deepest gratitude to my supervisor Prof. Dr. Kemal Oflazer, for his guidance and advices not only for this research but also for academic life in general.

I thank Prof. Dr. Cem Say, Assoc. Prof. Dr. Berrin Yanıkoğlu, Assoc. Prof. Dr. Hüsnü Yenigün and Assoc. Prof. Dr. Aslı Göksel for their kind attendance to the thesis committee and for their valuable contributions.

My sincere gratitude goes to Prof. Dr. Miriam Butt for patiently answering my questions and for her enormous help in linguistics. I would also like to thank all ParGram members for the insightful discussions and comments. I learned a lot from them.

Önsel Armağan and Tuba Gümüş helped me in integrating their work into my system. I thank them for their efforts.

Thanks to everyone at FENS 2014 for creating a fun place to work. I'm grateful first to İlknur and then Ayça, Alisher, Reyyan, Süveyda, Ayşegül, Serdar, Burak, Hanife, Murat Can. I enjoyed all the time we had spent together.

Annette Hautli proofread the manuscipt but beyond that I'm indebted to her for the moral support for the last couple of weeks of thesis-writing. Thanks for the warmth you brought to rainy and windy Dublin.

I would like to thank my best friends Bahar Öztok and Bahar Güçver for their invaluable support through all these years. Evrim Güngör printed the drafts of the thesis. Thanks for being there for me when I need you.

Finally, I am forever indebted to my parents for their endless love, support, and patience. I am so lucky to have them.

This work is financially supported by TUBİTAK by grant 105E021.

BUILDING A LEXICAL FUNCTIONAL GRAMMAR FOR TURKISH

Özlem Çetinoğlu

Electronics Engineering and Computer Science

Ph.D. Thesis, 2009

Thesis Supervisor: Prof. Dr. Kemal Oflazer

Keywords: Natural Language Processing, Syntax, Parsing, Lexical Functional Grammar

Abstract

Large-scale, deep grammars with structurally rich output are basic resources for complex tools in human-computer interaction and also for exploring the linguistic phenomena of a language. In this thesis, we introduce a large scale grammar for Turkish implemented in the Lexical Functional Grammar formalism.

Developing a large scale grammar requires that several issues be solved, both linguistically and computationally. As the language to be dealt with is Turkish, rich morphological structures play an important role in constructing the basis of the representation. We follow an approach based on building units that are larger than a morpheme but smaller than a word, in encoding rules of the grammar to explain the linguistic phenomena in a more formal and accurate way.

Our implementation covers rules ranging from basic constituents such as adjective, adverbial, or prepositional phrases to more complex types with derivations such as sentential complements, sentential adjuncts, and relative clauses. The noun phrase subgrammar is the core of the system. Other important rules deal with several types of sentence structures, free word order, and coordination. Also, a date-time grammar developed earlier is integrated into our system.

Some of the frequently occuring phenomena, such as causatives, passives, noun-verb compounds, and non-canonical objects, are also important from a theoretical perspective. We first examine their linguistic representation and then analyze the details of different types of causatives and non-canonical objects by conducting several tests. We then provide their implementation.

To evaluate our grammar we have experimented with real world data. Results show that we have a reasonably high coverage in noun phrases (85.5%). We have also integrated our system into a tool called LingBrowser.

TÜRKÇE İÇİN SÖZCÜKSEL İŞLEVSEL GRAMER GELİŞTİRİLMESİ

Özlem Çetinoğlu

Elektronik Mühendisliği ve Bilgisayar Bilimi

Doktora Tezi, 2009

Tez Danışmanı: Prof. Dr. Kemal Oflazer

Anahtar Sözcükler: Doğal Dil İşleme, Sözdizim, Çözümleme, Sözcüksel İşlevsel Gramer

Özet

Zengin yapısal gösterimli sonuçlar sunan büyük ölçekli derin gramerler, bir dilin dilbilimsel olaylarını araştırmak için olduğu kadar bilgisayar insan etkileşimindeki karmaşık araçlar için de temel kaynaklardandır. Bu tezde, Türkçe için Sözcüksel İşlevsel Gramer kuramı içinde gerçeklenmiş büyük ölçekli bir gramer sunuyoruz.

Büyük ölçekli bir gramer geliştirmek hem dilbilim hem de bilgisayar bilimleri açısından çözülmesi gereken bir çok konuyu beraberinde getirir. Çalışılan dil Türkçe olduğunda, zengin biçimbilimsel yapılar, gösterimin temelini oluşturmakta önemli bir rol oynar. Gramerimizi geliştirirken, dil olaylarını formel ve doğru bir şekilde ifade edebilmek amacıyla, biçimbirimlerden büyük ancak sözcüklerden de küçük yapıtaşları kullandık.

Gerçeklediğimiz sistemde kurallar, sıfat, zarf, edat öbekleri gibi temel bileşenlerden, isim-fiiller, zarf-fiiller, sıfat-fiiller gibi daha karmaşık türemiş yapılara kadar geniş bir alanı kapsamaktadır. İsim öbegi alt grameri sistemin esas bileşenidir. Cümle çeşitleri, serbest sözcük dizilişi, bağlaç öbekleri gramerimizin çözümlediği diğer önemli yapılardır. Ayrıca daha önce geliştirilmiş bir tarih-zaman çözümleyicisi de sistemimize eklenmiştir.

Etken yapılar, edilgen yapılar, isim ve fiilden oluşan fiiller, ve ismin belirtme halini almayan nesneler gibi sıklıkla karşılaştığımız dil olayları, teorik açıdan da önemlidirler. Bu yapıların önce dilbilmsel gösterimleri incelenmiş, sonra çeşitli testler yapılarak etken yapılar ve ismin belirtme halini almayan nesnelerin farklı türleri ayrıntısıyla çözümlenmiştir. Daha sonra çözümlerin gerçeklenme detayları sunulmuştur.

Gramerin değerlendirilmesi için gerçek metin belgeleri üzerinde testler yapılmıştır. Sonuçlar isim öbeklerinde %85.5 oranında başarım olduğunu göstermektedir. Sistemimiz, LingBrowser adlı araca da eklenmiştir.

Contents

Acknowledgments **v**

Abstract **vi**

Özet **vii**

List of Abbreviations **xiii**

1 INTRODUCTION **1**

1.1 Outline of the Thesis 3

2 LEXICAL FUNCTIONAL GRAMMAR **4**

2.1 Overview of Lexical Functional Grammar 4

2.1.1 Constituent Structure 5

2.1.2 Functional Structure 5

2.1.3 Mapping from Constituent Structure to Functional Structure . . 7

2.2 XLE and its Architecture 11

3 TURKISH **17**

3.1 Morphology 17

3.2 Syntax 20

3.3 Inflectional Groups 24

3.3.1 Inflectional Groups as Lexical Units 24

3.3.2 Inflectional Groups and Lexical Integrity 28

3.4 Other Grammars for Turkish 35

4 LEXICAL FUNCTIONAL GRAMMAR ANALYSES OF VARIOUS TURKISH LINGUISTIC PHENOMENA **38**

4.1 General Overview of Rules . . . 39

4.1.1 Noun Phrases . . . 39

4.1.2 Adjective, Adverbial, and Postpositional Phrases . . . 43

4.1.3 Sentential Complements, Sentential Adjuncts, and Relative Clauses 49

4.1.4 Sentences and Free Word Order . . . 57

4.1.5 Coordination . . . 61

4.1.6 The Date-Time Grammar . . . 63

4.2 Causatives . . . 64

4.2.1 Causatives: Monoclausal or Biclausal? . . . 65

4.2.2 Implementation in Lexical Functional Grammar . . . 75

4.2.3 Double Causatives . . . 83

4.3 Passives . . . 87

4.3.1 Impersonal and Double Passives . . . 87

4.3.2 Implementation in Lexical Functional Grammar . . . 90

4.3.3 Passivization of Causatives . . . 92

4.4 Noun-Verb Compound Verbs . . . 95

4.5 Non-canonical Objects . . . 97

4.5.1 Non-Canonical Object Marking in Turkish . . . 98

4.5.2 Object Tests . . . 101

4.5.3 Analysis and Implementation . . . 117

5 EVALUATION **123**

5.1 Test Suites . . . 124

5.2 Optimality Theory Marks . . . 130

5.3 LingBrowser . . . 131

6 SUMMARY AND CONCLUSION **133**

6.1 Future Work . . . 137

A Morphological Tags **140**

B ParGram Sentences **143**

B.1 Spring 2006 Meeting 143

B.2 Fall 2006 Meeting 145

B.3 Spring 2007 Meeting 146

B.4 Fall 2007 Meeting 148

B.5 Spring 2008 Meeting 150

B.6 Fall 2008 Meeting 152

C Sentence Test Suite **155**

C.1 Main Text 155

C.2 Sentences 156

D Noun Phrase Test Suite **158**

D.1 Simple Noun Phrases 158

D.2 Relative Clauses 160

D.3 Sentential Complements 161

D.4 Coordination 162

D.5 XLE Parses of Selected Phrases 163

BIBLIOGRAPHY **186**

List of Figures

2.1 Components of the XLE architecture 12

3.1 Dependency relations of a sentence from the Turkish Treebank 25
3.2 Dependency relations of the phrase *kentin en canlı yeri* 26
3.3 Dependency relations of the phrase *kentin en canlısı* 26

4.1 C-structure of the NP *benim tarih dersimin kitabının yeni basımı* 'the new edition of my history course's book' 44
4.2 F-structure of the NP *benim tarih dersimin kitabının yeni basımı* 'the new edition of my history course's book' 45

List of Tables

4.1 Types of noun phrase rules 42
4.2 Types of adjective phrase rules 46
4.3 Types of adverbial phrase rules 48

5.1 The coverage for ParGram sentences 124
5.2 Statistics about test sentences 125
5.3 Types of phrases used in the noun phrase test 128

List of Abbreviations

A	Adjective
ADV	Adverb
ADVP	Adverb phrase
AP	Adjective phrase
CAT	Category
CCG	Combinatory categorial grammar
COMP	Close complement
CONJ	Conjunction
CP	Complement phrase
DERIV	Derivation
DET	Determiner
DS	Derivational suffix
IG	Inflectional group
LFG	Lexical functional grammar
N	Noun
NLP	Natural language processing
NP	Noun phrase
NUM	Number
OBJ	Object
OBJ-TH	Thematic object
OBL	Oblique object
OBL-AG	Oblique agent
OT	Optimality theory

PART	Participle
PERS	Person
POSS	Possessive
POSTPP	Postposition phrase
PRED	Predicate
PROP	Proper noun
RG	Relational grammar
S	Sentence
SPEC	Specifier
SUBJ	Subject
V	Verb
VP	Verb phrase
XCOMP	Open complement
XFST	Xerox finite state tools

Chapter 1

INTRODUCTION

Natural language processing (NLP) is a subfield of computer science that deals with the research and development of computationally effective methods for analyzing and synthesizing human languages. The applications we commonly use in our daily life such as word processors, spelling correctors, and search engines already benefit from NLP techniques.

High quality machine translation, human computer interaction in a natural dialogue, or question answering systems require that computers make deeper analyses that go beyond superficial aspects. Such deep analyses are made possible by developing linguistically motivated grammars.

Such grammars have a key role in revealing the semantics of sentences in a language. Parsing a sentence with a grammar describes how words come together and form constituents for a grammatical sentence, and determines the structural role of each constituent within the sentence. There are mainly two approaches in parsing: shallow parsing uses simple grammars coupled with statistical approaches to automatically produce bracketed structures and deep parsing targets linguistically motivated, rich output, that is, provides semantic information as well as syntactic structure.

The value of a large scale deep grammar is not just to be a primary resource for many NLP applications. It is also necessary to understand, define and represent the

linguistic phenomena of the language in question in more formal ways. In this thesis, we aim to build a large scale grammar for Turkish with various computational aspects in mind, but without leaving aside the interesting linguistic problems to be solved. One of the distinguishing aspects of this work is the implementation of the grammar by employing parsing units smaller than words but larger than morphemes. This approach allows to incorporate the complex morphology and the syntactic relations mediated by morphological units in a manageable way and to handle lexical representations of very productive derivations.

Our grammar is implemented using the Lexical Functional Grammar (LFG) formalism [Kaplan and Bresnan, 1982], a well-established unification-based theory. LFG is a widely used theory with many contributors working on various languages from different language families. The different experiences of these contributors are shared through the ParGram(Parallel Grammars) project [Butt et al., 1999]. The resulting grammars are used in several projects such as statistical machine learning, syntax/semantics interface, and translations based on parallel grammars[1]. Recently, a search engine company, Powerset[2], bases its indexing technology on parsing the web documents by using English LFG grammar.

The Turkish LFG grammar is part of the ParGram project. The project aims to develop large scale grammars for a range of languages (Arabic, Chinese, English, French, Georgian, German, Hungarian, Japanese, Malagasy, Norwegian, Urdu, and Welsh) within the LFG framework. Despite the differences between the languages involved, the aim is to produce parallel syntactic analyses with the assumption that although word order, surface representation, or constituent hierarchy may differ, the function of constituents are the same for equivalent sentences among languages. As a result of this assumption, a new grammar developed within ParGram benefits from sharing the linguistic know-how on some well studied topics. Semi-annual ParGram meetings help the grammar writers keep the grammars parallel and discuss solutions for problematic cases.

[1] http://www2.parc.com/isl/groups/nltt/homepage.html#activities
[2] www.powerset.com

1.1 Outline of the Thesis

The organization of this thesis is as follows:

Chapter 2 introduces the Lexical Functional Grammar formalism and how it represents syntactic structures. It also describes the architecture of the software system (XLE) by summarizing each of its components.

Chapter 3 gives some basic information about Turkish morphology and syntax, focusing mainly on phenomena implemented in the LFG grammar.

Chapter 4 examines the details of the grammar. First, it discusses the basic components of the grammar. Then it investigates linguistic phenomena such as causatives, passives, and non-canonical objects in detail, and provides implementational details.

Chapter 5 describes the evaluation of our grammar with a series of experiments. It also describes a prototype integration of our grammar into LingBrowser [Armağan, 2008], an intelligent browser that provides users with linguistic information.

Chapter 6 closes the thesis with an extensive summary and future work.

In this thesis we simplify the linguistic representation in various examples so as to highlight only the relevant aspects under discussion. Thus we may not display all the syntactic or semantic structure all the time. For the cited examples, glosses and judgement marks are taken with no modification. We use '*' to indicate ungrammaticality, and '?' and '??' to indicate variability. Appendix A lists morphological abbreviations that we use to indicate Turkish morphological features.

Chapter 2

LEXICAL FUNCTIONAL GRAMMAR

The foundations of Lexical Functional Grammar (LFG) are motivated by linguistic, computational, and psycholinguistic considerations. LFG was introduced by Joan Bresnan and Ronald Kaplan who published two important papers that explain the theory in detail, define the model and the concepts, and compare the differences with existing approaches [Kaplan and Bresnan, 1982; Bresnan and Kaplan, 1982]. XLE [Maxwell and Kaplan, 1996] was developed to help grammarians write grammars in the LFG formalism. It facilitates implementing large scale grammars for several languages from several sites.

This chapter explains the XLE architecture by giving examples from the current Turkish grammar and gives a brief introduction to LFG, focusing on the features used during the grammar implementation. For further details on LFG, the reader is referred to a collection of comprehensive LFG literature [Sells, 1985; Dalrymple et al., 1995; Butt et al., 1999; Bresnan, 2001; Dalrymple, 2001; Falk, 2001].

2.1 Overview of Lexical Functional Grammar

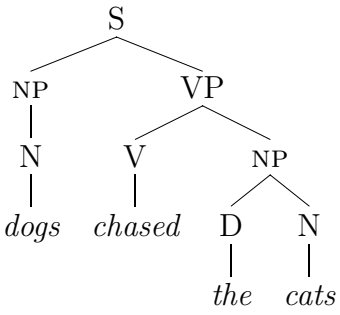
LFG is a theory representing the structure of natural language utterances in two parallel levels: the constituent structure (c-structure) and the functional structure (f-structure).

The c-structure defines the order and grouping of constituents, whereas the f-structure defines functional roles of these constituents. Therefore c-structures are rather language specific, whereas the corresponding f-structures in different languages are expected to be crosslinguistically parallel.

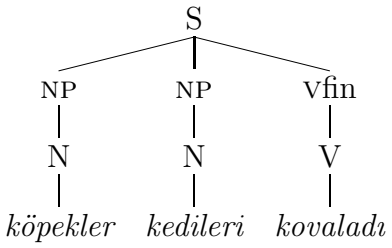
2.1.1 Constituent Structure

Constituent structures have the form of context-free phrase structure trees. (1) and (2) give the c-structures of the English sentence *Dogs chased the cats.* and its Turkish counterpart *köpekler kedileri kovaladı*, respectively. In English a basic sentence consists of a noun phrase and a verb phrase. The noun phrase is the subject of the verb. If the verb is transitive, the verb phrase consists of the verb itself, followed by a noun phrase which is the object of the verb. On the contrary, the c-structure of the Turkish sentence is flat to allow varying word order.[1]

(1)



(2)



2.1.2 Functional Structure

Functional structures are in the form of attribute value matrices. Attributes can be features, such as tense and gender, or functions, such as subject and object. Values corresponding to these attributes can be

- atomic symbols (e.g., value `past` of TENSE in (3))

[1]Depending on the discourse context Turkish allows all six possible Subject-Object-Verb orders with minimal formal constraints.

- semantic forms (e.g., value ‘chase⟨dog, cat⟩’ of PRED in (3))
- subsidiary f-structures (e.g., the f-structure corresponding to SUBJ in (3))

(3) and (4) give the simplified f-structures for the sentences used in c-structure examples (1) and (2). Both f-structures demonstrate that the verb *chase/kovala* is a two place predicate where *dog/köpek* fills in the SUBJect and *cat/kedi* fills in the OBJect position of the verb. There are also additional features in the f-structure, e.g., the TENSE of the verb, or the CASE of the nouns. Note that, although the functional values are the same for these simple sentences, the f-structures have some differences, e.g., the objects have different CASE values.

(3)
$$\begin{bmatrix} \text{PRED} & \text{‘chase⟨dog,cat⟩’} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{‘cat’} \\ \text{SPEC} & \begin{bmatrix} \text{DET} & \begin{bmatrix} \text{PRED} & \text{‘the’} \end{bmatrix} \end{bmatrix} \\ \text{CASE} & \text{obl} \\ \text{NUM pl, PERS 3} \end{bmatrix} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{‘dog’} \\ \text{CASE} & \text{nom} \\ \text{NUM pl, PERS 3} \end{bmatrix} \\ \text{TENSE} & \text{past} \end{bmatrix}$$

(4)
$$\begin{bmatrix} \text{PRED} & \text{‘kovala⟨köpek,kedi⟩’} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{‘kedi’} \\ \text{CASE} & \text{acc} \\ \text{NUM pl, PERS 3} \end{bmatrix} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{‘köpek’} \\ \text{CASE} & \text{nom} \\ \text{NUM pl, PERS 3} \end{bmatrix} \\ \text{TENSE} & \text{past} \end{bmatrix}$$

There are three conditions that an f-structure should satisfy in order to be well-formed:

- **Uniqueness Condition:** Each attribute should have a unique value. The example in (5a) is not well-formed since the CASE feature of a noun cannot be `nom` and `acc` at the same time.
- **Completeness Condition:** An f-structure has to explicitly contain the functions that the value of its PRED feature subcategorizes for. In (5b), the f-structure of the sentence *Mary saw.* is incomplete due to a missing object.

- **Coherence Condition:** All functional attributes represented in the f-structure should be the arguments of the PRED feature on the same f-structure level. (5c) exemplifies an incoherent case. The sentence *Mary slept cats.* has the intransitive verb *sleep*, nevertheless contains an ungoverned OBJect in the corresponding f-structure.

(5) a.
$$\begin{bmatrix} \text{PRED} & \text{'cat'} \\ \text{CASE} & \text{acc, nom} \\ \text{NUM} & \text{PL, PERS 3} \end{bmatrix}$$

b.
$$\begin{bmatrix} \text{PRED} & \text{'see}\langle\text{Mary, }\rangle\text{'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'Mary'} \\ \text{CASE} & \text{nom} \\ \text{NUM} & \text{sg, PERS 3} \end{bmatrix} \\ \text{TENSE} & \text{past} \end{bmatrix}$$

c.
$$\begin{bmatrix} \text{PRED} & \text{'sleep}\langle\text{Mary}\rangle\text{'} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'cat'} \\ \text{CASE} & \text{obl} \\ \text{NUM} & \text{pl, PERS 3} \end{bmatrix} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'Mary'} \\ \text{CASE} & \text{nom} \\ \text{NUM} & \text{sg, PERS 3} \end{bmatrix} \\ \text{TENSE} & \text{past} \end{bmatrix}$$

The relation between a c-structure and its corresponding f-structure is set by using a mapping function, which is discussed in the following.

2.1.3 Mapping from Constituent Structure to Functional Structure

The information to construct the c-structures and f-structures is encoded, in annotated phrase structure rules. (6) gives the rules to parse the Turkish example in (2).

(6) a. S → NP (↑ SUBJ) = ↓ NP (↑ OBJ) = ↓ Vfin ↑ = ↓

b. NP → N

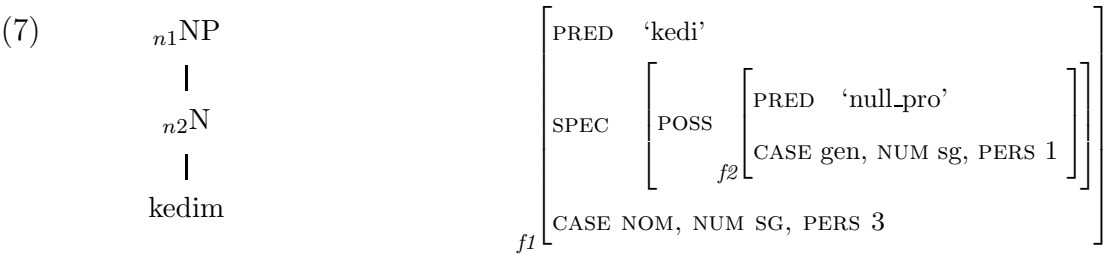
c. Vfin → V

In the LFG notation, ↑ and ↓ are metavariables representing the f-structure of the mother node and the f-structure of the node itself, respectively. In (6a), the equation (↑ SUBJ) = ↓ means that the attribute SUBJ of the mother node's f-structure (here, the f-structure of S) has the f-structure of the current node (here, NP) as its value. ↑ = ↓

states that the f-structure of the node itself (here, Vfin) unifies with the f-structure of its mother node (here S). That is, all information encoded in the f-structure of Vfin goes into the f-structure of S. Note that there are no annotations in (6b) and (6c). This is because, in the general convention, each nonterminal in the right hand side of the phrase structure rule is associated with $\uparrow = \downarrow$ unless indicated otherwise.

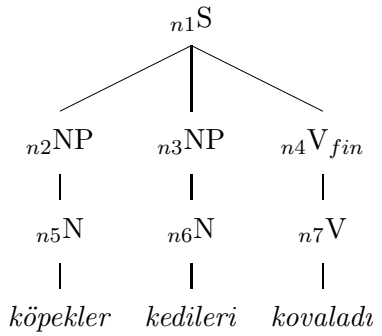
The correspondence or mapping relation from c-structure to f-structure is called ϕ projection. This projection function is many-to-one and into, that is, more than one c-structure node can correspond to the same f-structure and there can be f-structures that have no corresponding c-structure node. (7) shows the NP *kedim* 'my cat' which is parsed with the rule NP $\rightarrow$ N. The possessive marker is a suffix in Turkish, hence there is no explicit node in the c-structure. But in the f-structure representation, it has a separate f-structure. Both *n1* and *n2* map to *f1* (ϕ(n1)=f1, ϕ(n2)=f1) and there is no corresponding c-structure node for *f2*.

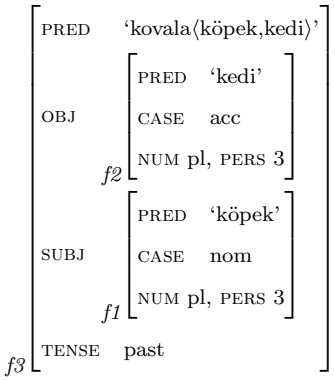
(7)



(8) depicts the mapping between c-structure and f-structure of *köpekler kedileri kovaladı* 'Dogs chased the cats.', where nodes of c-structures and outer and inner f-structures are labeled to highlight the correspondence. The noun *köpekler*, hence its category N (labeled *n5*), is represented with *f1* and *kedileri* corresponds to the f-structure *f2*. Due to the equation $\uparrow = \downarrow$ in the rule NP $\rightarrow$ N, *f1* is also the f-structure for *n2* and similarly *n3* maps to *f2*. By following the equation $\uparrow = \downarrow$ in (6a) and (6c), the f-stucture of the verb becomes the outermost f-structure of the sentence, namely representing the nodes *n1*, *n4* and *n7*. Again, from the constraints ($\uparrow$ SUBJ) = $\downarrow$ and ($\uparrow$ OBJ) = $\downarrow$ of (6a), *f1* which represents *n2* becomes the SUBJect of *f3*. *f2* which corresponds to *n3* is placed as the OBJect.

(8)





The mappings are also shown with the set of equations given in (9)

(9) ϕ(n1)=f3 ϕ(n4)=f3 ϕ(n7)=f3
ϕ(n2)=f1 ϕ(n5)=f1
ϕ(n3)=f2 ϕ(n6)=f2

LFG employs several descriptional instruments to facilitate the construction and representation of f-structures. Here we present two of them that are used in implementing the Turkish LFG grammar.

Functional Uncertainty

Consider the English sentences in (10) which have nonlocal dependencies. For all the sentences, *the girl* fills the gap, but its syntactic function changes in each sentence, depending on the structure of the complement phrase.

(10) a. the girl$_i$ Mary saw $__i$
b. the girl$_i$ John claimed Mary saw $__i$
c. the girl$_i$ Tom said John claimed Mary saw $__i$

For (10a), the empty OBJect of the complement phrase is filled by the NP *the girl*. In LFG notation, the rule given in (11) would parse the whole phrase where the NP on the right hand side covers *the girl* and CP covers *Mary saw*.

(11) NP → NP CP
↑ = ↓ ↑ = (↓ OBJ)

If we want to parse (10b) and (10c), we need to insert the constraints (12a) and (12b) respectively, instead of ↑ = (↓ OBJ). Adding two more constraints covers the local dependencies in (10b) and (10c) but it is not possible to enumerate all disjunctive constraints to cover unbounded local dependencies.

(12) a. ↑ = (↓ COMP OBJ)
b. ↑ = (↓ COMP COMP OBJ)

To solve this problem, Kaplan and Zaenen [1989] proposed functional uncertainty equations by extending the notation and allowing regular expressions in place of simple attributes within f-structure constraints. Instead of writing separate rules for each sentence, the single constraint ↓ = (↑ COMP* OBJ) can capture all possibilities. The Kleene star * allows COMP to be repeated zero or more times. With this notation, phenomena requiring multiple disjunctive enumeration can be described with a simple expression.

Restriction Operator

Restriction enables modifying f-structures in terms of features. Kaplan and Wedekind [1993] introduced the restriction operator '\', that allows to restrict out some features from the existing f-structure. For instance, ↑ \CASE denotes an f-structure identical to ↑ except that it does not have the CASE feature. The restriction operator can be used to eliminate some features from the existing f-structure, or to change the value of a feature during unification. As an example, we present the rule ↑ \CASE = ↓ \CASE\PERS with the constraint (↑ CASE)= acc. According to this rule, the f-structures of the mother node and current node are unified. However, the CASE features are excluded during this unification. According to the given constraint, `acc` is assigned to the CASE feature of the mother node's f-structure. The PERS feature of the current node is also excluded

during the unification and there is no other assignment for this feature for the mother node. If the rule is applied to (13a), we get the f-structure in (13b).

(13) a.
$$\begin{bmatrix} \text{PRED} & \text{'kedi'} \\ \text{CASE} & \text{nom} \\ \text{NUM} & \text{sg} \\ \text{PERS} & 3 \end{bmatrix}$$

b.
$$\begin{bmatrix} \text{PRED} & \text{'kedi'} \\ \text{CASE} & \text{acc} \\ \text{NUM} & \text{sg} \end{bmatrix}$$

2.2 XLE and its Architecture

XLE [Maxwell and Kaplan, 1996] (formerly known as Xerox Linguistic Environment) is a grammar development platform that facilitates the integration of various modules, such as tokenizers, finite-state morphological analyzers, and lexicons in order to build wide-coverage, deep, constraint-based LFG grammars. Figure 2.1 shows the components of the XLE architecture. In this section, we briefly explain each of these components and give examples from the implemented Turkish LFG grammar for clarification.

Tokenizer

The first component of the XLE pipeline, as in any string processing system, is the tokenizer. It splits input text into tokens. Our sample sentence *köpekler kedileri kovaladı* 'Dogs chased the cats.' gets the tokenization shown in (14). It is possible to include multiple tokenizers in this step. Depending on the implementation of the further steps, it is possible to design the tokenizer in a way that it analyzes multiple words as a single token, i.e., multiword expressions. The current version of the Turkish LFG grammar uses the default XLE tokenizer only.

(14) `köpekler @ kedileri @ kovaladı @`

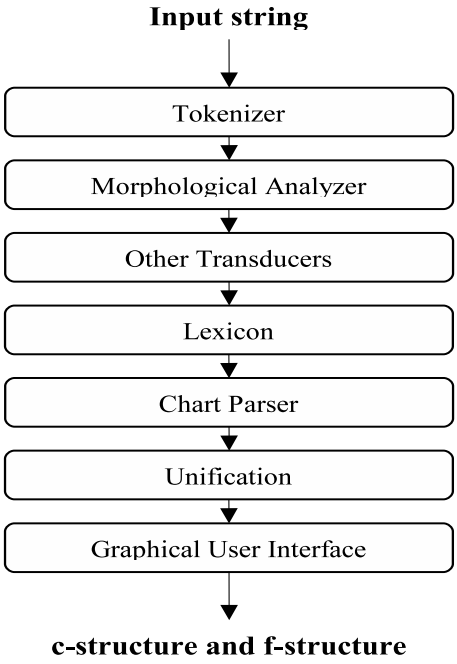


Figure 2.1: Components of the XLE architecture

Morphological Analyzer

The input to the morphological analyzer is a tokenized string like (14). XLE is designed to facilitate the usage of morphological analyzers built by Xerox Finite State Tools (XFST) [Beesley and Karttunen, 2003; Kaplan et al., 2004b]. As the Turkish morphological analyzer [Oflazer, 1994] is built within LFG, it can be easily integrated into the system. (15) gives the output of the analyzer for the noun *kedileri*. Note that all possible morphological analyses are produced as the output. The representation used by the Turkish morphological analyzer is discussed in Section 3.1.

(15) a. `kedi+Noun+A3pl+Pnon+Acc`

b. `kedi+Noun+A3pl+P3sg+Nom`

c. `kedi+Noun+A3sg+P3pl+Nom`

d. `kedi+Noun+A3pl+P3pl+Nom`

Other Transducers

XLE allows to use multiple transducers in a very flexible way. With the help of a configuration file, it is possible to cascade the transducers or use them in a parallel among other configurations [Kaplan and Newman, 1997]. XLE also allows the construction of text-based transducers usually used for adding or overriding the analyses of the primary morphological analyzer. For instance, if *seskaydedici* 'voice recorder' is an unknown word for the Turkish morphological analyzer, we could include it as a new entry in our text-based transducer with no need to change the morphological analyzer.

The current Turkish LFG grammar uses transducers to analyze multiword expressions, especially date and time expressions. (16) gives the input and output of one of these transducers [Gümüş, 2007]. The input is the morphological analyzer output of the expression *2 Ekim 2008* 'October 2nd, 2008' [2] and the output is the multiword stem followed by the appropriate tags.

(16) Input: `2+Num+Card Ekim+Noun+Prop+A3sg+Pnon+Nom 2008+Num+Card`
Output: `2 ekim 2008+Noun+DateTime+A3sg+Pnon+Nom`

Lexicon

XLE enables the grammar writer to enter lexical entries in more than one way. In the basic form, a lexical entry for *kedileri* would be in the form given in (17). The headword, which is the surface representation, is followed by the category of the word and an * denoting that the information is not coming from the morphological unit. Then the set of attribute value pairs defining the word is listed. Note that the information encoded in these pairs forms the f-structure of *kedileri* in (4). This method is not applicable to large-scale grammars since the surface form of each lexical item should be listed separately, but it can still be used to cover alternative analyses the morphological analyzer does not output.

[2]There is more than one analysis of *Ekim*, but only the relevant sense is given as the input to the transducer in the example.

(17) kedileri N * (↑ PRED) = 'kedi'
(↑ CASE) = acc
(↑ NUM) = pl
(↑ PERS) = 3.

Instead of listing every single lexical entry, each tag in the morphological analyzer is assigned a separate entry in the lexicon. After that, rules that parse the morphological output are encoded. The entries for tags are called sublexical entries and the rules that parse these sublexical entries are called sublexical rules. (18) shows the sublexical entries required to parse `kedi+Noun+A3pl+Pnon+Acc`. Each tag has its headword and category and this time an XLE tag in the third column denoting that the information is coming from the morphological analyzer. In the last column, instead of assigning attribute value pairs explicitly, we prefer templates that take the values as arguments and assign them to the attributes. Templates, starting with an @ sign, allow generalizations and facilitate modularity.

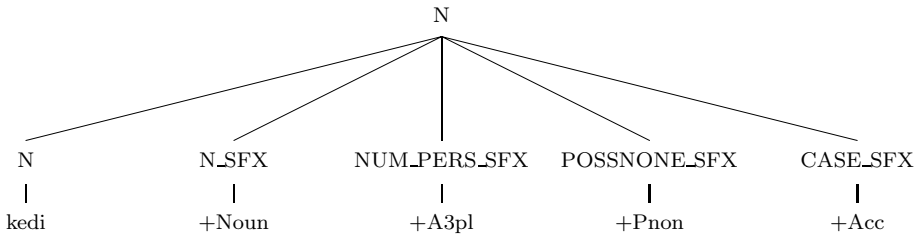
(18)	kedi	N	XLE @(NOUN kedi).
	+Noun	N_SFX	XLE.
	+A3pl	NUM_PERS_SFX	XLE @(NUM pl) @(PERS 3).
	+Pnon	POSSNONE_SFX	XLE.
	+Acc	CASE_SFX	XLE @(CASE acc).

Sublexical rules function in the same way as the usual phrase structure rules in LFG. Categories of the suffixes correspond to variables on the right hand side of the sublexical rules with a _BASE tag added to each of them. To be able to parse the morphological output in (15a), the sublexical rule in (19) should be encoded as well as the sublexical entries in (18).

(19) N → N_BASE
N_SFX_BASE
NUM_PERS_SFX_BASE
POSSNONE_SFX_BASE
CASE_SFX_BASE.

Just like a usual phrase structure rule, sublexical rules construct phrase structure trees (in this case it is morphological information), but they are not explicitly displayed in the c-structure representation. To get this information, XLE enables the user to switch to the expanded display mode to view the sublexical information. The sublexical tree of the noun *kedileri* is given in (20).

(20)



The tags corresponding to the suffixes of the morphological analyzer are easily enumerable but stems cannot be enumerated that easily. Thanks to the XLE facilities, not all stems are necessarily listed as entries in the lexicon. It is possible to define a generic rule that places the variable '-unknown' as the headword of a lexical entry and lists the possible categories of the unknown word by using templates. The argument '%stem' of the templates is a variable that matches the same value '-unknown' takes. For instance, the rule in (21), along with the sublexical rules defining adjectives and nouns, will catch adjectives and nouns which are parsed by the morphological analyzer but do not have explicit headwords in the lexicon. Consider a case where the adjective *iyi* 'good' is parsed as `iyi+Adj` by the morphological analyzer but there is no lexical entry for *iyi* in the lexicon. The tag `+Adj` has a sublexical entry and there is a sublexical rule for parsing the morphological analyses of adjectives. In this case, '-unknown' matches `iyi` and provides the adjective stem required for the adjective sublexical rule.

(21) -unknown A XLE @(ADJ %stem);
N XLE @(NOUN %stem).

Chart Parser and Unification

XLE uses an efficient parser based on three important ideas to improve the performance. The first key point to consider is the interface between the phrasal and functional constraints [Maxwell and Kaplan, 1993]. Instead of interleaving the phrasal and functional constraints, first the phrasal constraints are processed and then the results are used to facilitate the processing of functional constraints in a more effective way.

The second idea is using packed feature structures constructed by "contexted unification" [Maxwell and Kaplan, 1991]. For instance, depending on the context, the noun *ata* might be interpreted either as 'to the horse' or as 'ancestor' which will correspond to two different f-structures in LFG. In the contexted feature representation, XLE will produce the packed structure in (22) by merging the two f-structures into one and labeling the alternatives.

(22)
$$\begin{bmatrix} \text{PRED} & \begin{bmatrix} \langle \text{a:1} \rangle \text{ 'at'} \\ \langle \text{a:2} \rangle \text{ 'ata'} \end{bmatrix} \\ \text{CASE} & \begin{bmatrix} \langle \text{a:1} \rangle \text{ dat} \\ \langle \text{a:2} \rangle \text{ nom} \end{bmatrix} \\ \text{NUM SG, PERS 3} & \end{bmatrix}$$

The last key idea to improve efficiency is the lazy contexted copying during unification [Maxwell and Kaplan, 1996]. XLE employs a bottom up approach in unifying the contexted feature structures. Instead of copying up the whole daughter feature structures, lazy copying links are used and structures are expanded only when necessary. All nodes include Boolean expressions of bad analyses. Daughter structures that satisfy those bad analyses with inconsistent feature values do not pass their information up in the tree and therefore limit the solution space of the mother node.

Chapter 3

TURKISH

In this chapter, we present an overview of Turkish morphology and syntax with special emphasis on the concepts that we will refer to when we describe our grammar. We then continue with the definition of the inflectional groups and discuss the effects of using them in our grammar.

3.1 Morphology

The most important aspect of Turkish morphology is its agglutinative nature where sequences of inflectional and derivational morphemes attach to a root in a predefined order [Oflazer, 1994]. Surface realizations of the morphemes are determined by various morphophonemic rules such as vowel harmony and alternations of voiced/voiceless consonants. Therefore it is possible to encounter several allomorphs of a morpheme. With the exception of loanwords, Turkish morphotactics is quite regular yet complicated, especially when derivation is involved. Multiple derivations are frequent and the number of word forms one can generate from a nominal or verbal root is essentially infinite. (23) gives a simple example that demonstrates the morphemes of an inflected noun in their surface realization in (23a) and the lexical representation of the surface form in (23b).

(23) a. kedi-ler-imiz-de

b. kedi-lAr-HmHz-DA

In the lexical representation, A stands for the back and unrounded vowels {a,e}, D stands for the dental consonants {d,t}, and H stands for the high vowels {ı,i,u,ü}. Therefore, depending on the morphophonemic rules, the lexical morpheme -DA is realized as one of the four possible allomorphs {da, de, ta, te} on the surface level.

Oflazer [1994] uses this two-level representation [Koskenniemi, 1983] in implementing a Turkish morphological analyzer which is built using the Xerox Finite State Tools [Beesley and Karttunen, 2003]. The surface forms are mapped onto their lexical forms by using the encoded morphophonemic rules. They are then transformed into a sequence of tags representing each morpheme with the help of a finite state transducer. The morphological output for the noun *kedilerimizde* 'in our cats' in (23) is given in (24).

(24) kedi -lAr -HmHz -DA
`kedi+Noun +A3pl +P1pl +Loc`

If there is a derivation in the analyzed word, the morphological output contains the tag `^DB` denoting the derivational boundary. We call the sequence of inflectional morphemes between each derivational boundary *inflectional groups* (IGs hereafter). If we represent the morphological information in Turkish in the general form of `^DB`s representing derivational boundaries and $\mathtt{m}_i$s representing morphemes, then the IGs will be grouped as in (25).

(25) $\underbrace{\mathtt{root+m}_1\mathtt{+m}_2\mathtt{+}\cdots\ \mathtt{m}_i\mathtt{\hat{}DB}}_{\mathtt{IG}_1}\ \underbrace{\mathtt{+m}_{i+1}\mathtt{+}\cdots\mathtt{\hat{}DB+}\cdots\mathtt{\hat{}DB+}\cdots\mathtt{+m}_k}_{\mathtt{IG}_2\ \cdots\ \mathtt{IG}_n}$

$\mathtt{IG}_1$ includes the root, $\mathtt{IG}_2 \cdots \mathtt{IG}_n$ each include a tag representing the semantics of the derivation as well as the part of speech information and inflectional tags. A given word may have multiple such representations depending on any morphological ambiguity brought about by alternative segmentations of the word, and by ambiguous interpretations of morphemes.

For instance, the morphological analysis of the derived modifier interpretation of `uzaklaştırılacak` '(the one) that will be sent away' (lit., '(the one) that will be made to be far') would be:[1]

```
uzak+Adj^DB+Verb+Become^DB+Verb+Caus^DB+Verb+Pass+Pos
      ^DB+Adj+FutPart+Pnon
```

The five IGs in this word are:

1. `uzak+Adj`
2. `+Verb+Become`
3. `+Verb+Caus`
4. `+Verb+Pass+Pos`
5. `+Adj+FutPart+Pnon`

The first IG indicates that the root is a simple adjective meaning 'far'. The second IG indicates a derivation into a verb whose semantics is 'to become' the preceding adjective (here the adjective is 'far', so the verb is equivalent to 'to move away' in English). The third IG indicates that a causative verb (equivalent to 'to send away' in English) is derived from the previous verb. The fourth IG indicates the derivation of a passive verb with positive polarity from the previous verb. Finally the last IG represents a derivation into a future participle which will function as a modifier in the sentence.

The given example is not an extreme case in terms of the number of IGs per word. Eryiğit and Oflazer [2006] state that Turkish words found in a typical text average about 3-4 morphemes including the stem, with an average of about 1.2 derivations per word. Given that certain noninflecting function words such as conjuctions, determiners, etc. are rather frequent, this number is rather close to 2 for inflecting word classes. Statistics from the Turkish Treebank [Oflazer et al., 2003] show that for sentences ranging between 2 and 40 words (with an average of about 8 words), the number of IGs range from 2 to

[1]The other interpretation is 's/he will be sent away'

55 IGs (with an average of 10 IGs per sentence).

3.2 Syntax

Turkish is considered to be a free word order language with Subject-Object-Verb as the main order. There are some restrictions on the constituent order in the main sentence level and more restrictions in the clausal level. A constituent that is to be emphasized is generally placed immediately in front of the verb.

It is possible to drop subjects of sentences and possessive pronouns of noun phrases depending on the discourse context, since the information in the dependent is also repeated in the head. The verb in (26a) has an agreement marker denoting the person. Similarly, the modified noun in (26b) has a person marker (P1sg) denoting the possessor.

(26) a. (ben) uyu-du-m
(I.Nom) sleep-Past-1sg
'I slept.'

b. (benim) kedi-m
(my) cat-P1sg.Nom
'my cat'

Turkish is a head-final language, that is, dependents are placed before heads, as in (27a), but it also allows scrambling in some exceptional cases like the pronominal possessive noun phrases as in (27b).

(27) a. beyaz kedi / *kedi beyaz
white cat.Nom / cat.Nom white
'white cat'

b. benim kedi-m / kedi-m benim
my cat-P1sg.Nom / cat-P1sg.Nom my
'my cat'

The case of a noun phrase determines its grammatical function in the sentence.

In general, the subject is in the nominative case and the object is in nominative or accusative case, as in (28), depending on its specificity [Enç, 1991]. Note that, in this work we assign case to all nouns and derived nominals. When a case marker is not overtly present, we say that the word has 'nominative' case, without implying any further grammatical role or information. Thus a noun with no explicit case marking (hence marked with nominative case in morphology), can function as an indefinite direct object. In this case, we call such an object as 'having a nominative (morphological) case'.

The nominative object is restricted to immediate preverbal position.[2] There is also a group of verbs where the object can bear cases other than nominative/accusative as in (29).

(28) a. köpek kedi kovaladı
dog.Nom cat.Nom chase.Past.3sg
'The dog chased cats (The dog did cat chasing).'

b. köpek kedi-yi kovaladı
dog.Nom cat-Acc chase.Past.3sg
'The dog chased the cat.'

(29) kedi köpek-ten korktu
cat.Nom dog-Abl fear.Past.3sg
'The cat feared the dog.'

Causatives

Causatives in Turkish are constructed morphologically with the minor exceptions of lexical causatives. There are two productive causative morphemes: *-DHr* and *-t*.[3] More than one causative suffix can be attached to the verb. Double causatives are used

[2]There are some exceptions to this rule. In the sentence *yapayım sana yemek* 'Let me cook for you', the nominative object *yemek* comes after the verb *yapayım*. Kemal Oflazer (p.c.) attributes this example to Sarah Kennely. Aslı Göksel(p.c.) gives another example: *ekmek ben hiç yemem* 'I never eat bread.' The nominative object precedes the nominative subject.

[3]There are 3 other morphemes which are not productive and apply to a very small subset of the verbal roots.

frequently, triple causatives are also encountered but further ones are not applicable. Sample morphological analyses of the single and double causative of the verb *uyu* 'sleep' are given in (30).

(30)

uyu-du	`uyu+Verb+Pos+Past+A3sg`
uyu-**t**-tu	`uyu+Verb^DB+Verb+Caus+Pos+Past+A3sg`
uyu-**t**-**tur**-du	`uyu+Verb^DB+Verb+Caus^DB+Verb+Caus+Pos+Past+A3sg`

(31) and (32) exemplify causativizations of an intransitive verb and a transitive verb respectively. The nominative subject *kedi* 'cat' becomes accusative when causativized in (31b). Double causativization of intransitives is similar to single causativization of transitives (compare 31c with 32b). Nominative *çocuk* becomes dative and *kediyi* preserves its case.

(31) a. kedi uyu-du
cat.Nom sleep-Past.3sg
'The cat slept.'

b. çocuk kedi-yi uyu-t-tu
child.Nom cat-Acc sleep-Caus-Past.3sg
'The child made the cat sleep.'

c. anne çocuğ-a kedi-yi uyu-t-tur-du
mother.Nom child-Dat cat-Acc sleep-Caus-Caus-Past.3sg
'The mother made the child make the cat sleep.'

If the verb is transitive, as in (32a), the nominative subject *köpek* 'dog' becomes dative and the accusative object *kediyi* 'cat' preserves its case ((32b)). Double causativization of transitives has some fuzzy meaning. It is certain that somebody else is involved in the causation hierarchy but its ranking is ambiguous. Furthermore, one cannot place that person explicitly in the sentence. (32c) gives both interpretations.

(32) a. köpek kedi-yi kovala-dı
dog.Nom cat-Acc chase-Past.3sg
'The dog chased the cat.'

b. çocuk köpeğ-e kedi-yi kovala-t-tı
child.Nom dog-Dat cat-Acc chase-Caus-Past.3sg
'The child made the dog chase the cat.'

c. çocuk köpeğ-e kedi-yi kovala-t-tır-dı
child.Nom dog-Dat cat-Acc chase-Caus-Caus-Past.3sg
'The child made the dog make someone chase the cat.'
'The child made someone make the dog chase the cat.'

When a verb subcategorizes for an object with a case marker other than accusative, the causativization patterns differ from the verbs with canonical objects. The nominative *kedi* 'cat' becomes accusative and *köpekten* 'from the dog' preserves its case.

(33) a. kedi köpek-ten kork-tu
cat.Nom dog-Abl fear-Past.3sg
'The cat feared the dog.'

b. çocuk kedi-yi köpek-ten kork-ut-tu
child.Nom cat-Acc dog-Abl fear-Caus-Past.3sg
'The child made the cat fear the dog'

Passives

The passive construction is also a morphological process in Turkish. The passive morphemes are *-Hl* and *-Hn*. (34) gives a basic example on passivization of a transitive verb. The direct object in the accusative case becomes the subject in the nominative case after causativization. The verb agrees with the subject.

(34) a. köpek ben-i kovala-dı
dog.Nom cat-Acc chase-Past.3sg
'The dog chased me.'

b. ben (köpek tarafından) kovala-n-dı-m
I.Nom (dog.Nom by) chase-Pass-Past-1sg
'I was chased (by the dog).'

Again, the verbs with different case-marked objects have different behaviors in passivization than the nominative/accusative ones.

(35) a. kedi köpek-ten kork-tu
cat.Nom dog-Abl fear-Past.3sg
'The cat feared the dog.'

b. köpek-ten kork-ul-du
dog-Abl fear-Pass-Past.3sg
'The dog was feared.'

The behavior of non-canonical objects under certain linguistic phenomena is examined thoroughly in Section 4.5.

3.3 Inflectional Groups

Due to the agglutinative nature of the language, the syntax of Turkish has a strong connection with the morphology. Derivational processes occur morphologically, thus units smaller than words affect the syntax. In this section we explain how and why we use inflectional groups in our system.

3.3.1 Inflectional Groups as Lexical Units

In order to help clarify how IGs are involved in syntactic relations, a sentence from the Turkish Treebank [Oflazer et al., 2003] is given in Figure 3.1.[4] Morpheme boundaries are represented by the '-' sign and morphemes in dashed boxes define one IG. A solid box denotes a word boundary. If there is only one IG in the word, no dashed boxes are used. As the example indicates, IGs may consist of one or more morphemes. Each column underneath the boxes represents the morphological output tags of an IG corresponding to that column. For this example, there are three words where derivation took place,

[4]The sentence is slightly simplified for demonstrative purposes. It is the main clause of a conditional sentence in the treebank, the if-clause is omitted for space limitations.

the vertical dashed lines represent the derivational boundaries in the morphological outputs. Arrowed arcs show the dependencies from the dependant to the head and labels above the arcs denote the type of the dependencies. There are also implicit arcs from a left IG to its right IG, labeled with a DERIV in the treebank, but they are not represented in the figure. Note that dependencies are between IGs, not words.

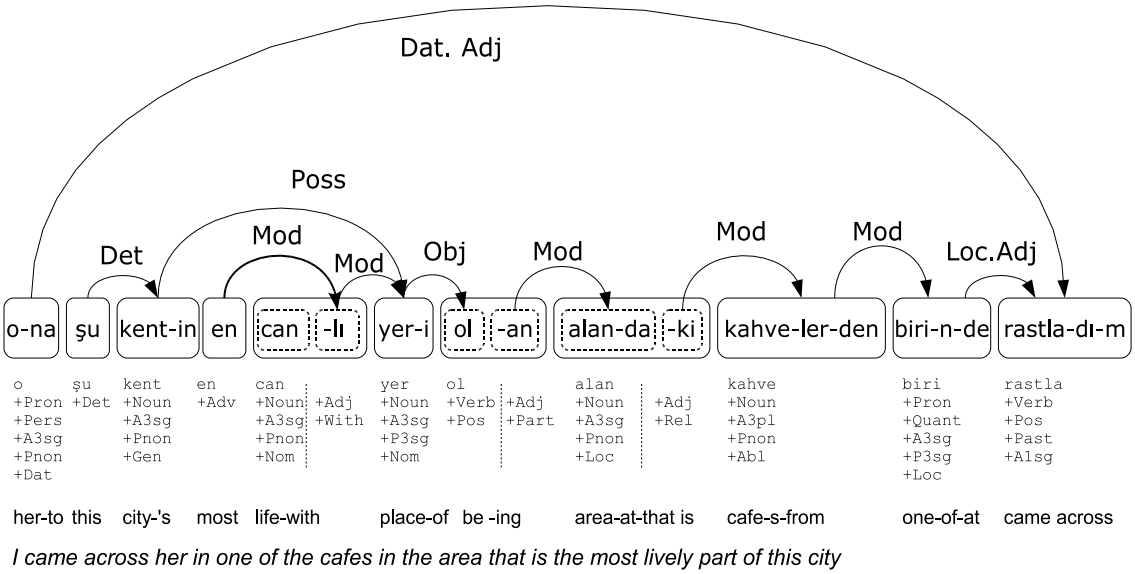


Figure 3.1: Dependency relations of a sentence from the Turkish Treebank

We focus on a shorter phrase taken from the big example in Figure 3.1, to explain in detail why dependencies are between IGs instead of words. Figure 3.2 depicts the relations of the phrase *kentin en canlı yeri* 'the most lively place of the city'.

Here, *en* 'most' modifies *canlı* 'lively' (literally 'with life') and not *can* 'life'. It is the derived adjective *canlı*, again not *can*, that modifies the noun *yer* 'place'. The genitive noun *kentin* 'city's' specifies the derived phrase *en canlı yeri* 'the most lively place'. The morpheme *-i* of the noun *yeri* is the possessive marker. To emphasize the use of IGs, the phrase in Figure 3.3 is introduced which is similar to the phrase in Figure 3.2 but contains one more derivation. The noun *canlı* is derived from the adjective *canlı* with no explicit derivational morpheme. The noun *kentin* now specifies the derived noun, hence the possessive marker *-sı* is attached to *canlı* instead of *yer* (Figure 3.2).

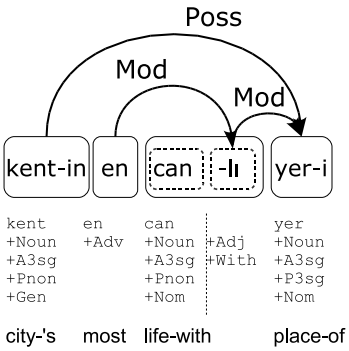


Figure 3.2: Dependency relations of the phrase *kentin en canlı yeri*

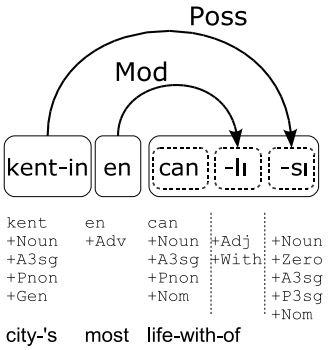


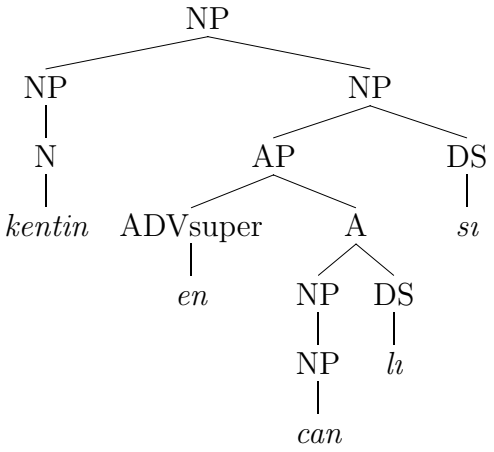
Figure 3.3: Dependency relations of the phrase *kentin en canlısı*

(36) shows the corresponding f-structure for the NP *kentin en canlısı* 'the most lively one of the city'. The semantics of the derivational suffix *-li* is shown as 'li⟨↑ OBJ⟩'. First, the f-structure of noun *can* 'life' is placed as the OBJ of the derivational suffix. Supporting the dependency representation in Figure 3.3, the f-structure of the adverb *en* is placed as the adjunct of *li⟨can⟩*, that is, the adjective *canlı*. Zero derivation of an adjective to a noun, as exemplified in the given phrase, indicates that there is a generic person modified by the adjective in question. In terms of f-structure representation this corresponds to a new PRED `'null-pro'` with the adjective as the ADJUNCT of the new structure which is shown as the outermost matrix in (36). The derived noun behaves essentially like a lexical noun and can be specified by another noun, here by *kentin* 'city's'.

(36)
$$\begin{bmatrix}
\text{PRED} & \text{'null-pro'} \\
\text{ADJUNCT} & \begin{bmatrix}
\text{PRED} & \text{'li}\langle\text{can}\rangle\text{'} \\
\text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'can'} \\ \text{CASE nom, NUM sg, PERS 3} \end{bmatrix} \\
\text{ADJUNCT} & \begin{bmatrix} \text{PRED} & \text{'en'} \end{bmatrix} \\
\text{ATYPE attributive, DEGREE superlative}
\end{bmatrix} \\
\text{SPEC} & \begin{bmatrix} \text{POSS} & \begin{bmatrix} \text{PRED} & \text{'kent'} \\ \text{CASE gen, NUM sg, PERS 3} \end{bmatrix} \end{bmatrix} \\
\text{CASE NOM, NUM SG, PERS 3}
\end{bmatrix}$$

The effect of using IGs as the representative units can be explicitly seen in the c-structure where each IG corresponds to a separate node, as in (37).

(37)



Within the tree representation, each IG corresponds to a separate node. Thus, the LFG grammar rules constructing the c-structures are encoded using IGs as units of parsing. If an IG contains the root morpheme of a word, then the node corresponding to that IG is named as one of the syntactic category symbols. The rest of the IGs are given the node name DS to indicate derivational suffix.

Note that in (37), the node representing the surface morpheme *-sı* seems to be carrying an inflectional suffix rather than a derivational one. This is because the derivation from an adjective to a noun does not have a surface morpheme and the possessive suffix is attached to the derived noun.

3.3.2 Inflectional Groups and Lexical Integrity

The representation of derivational suffixes in Turkish has been the most discussed subject since the beginning of the grammar development within the ParGram project. Basically, the IG approach goes against the Lexical Integrity Principle [Bresnan and Mugane, 2006] of the LFG theory:

> Every lexical head is a morphologically complete word formed out of different elements and by different principles from syntactic phrases.

However, in our approach, lexical heads might not be morphologically complete words but derivational suffixes, causing the words to be separated into several nodes in c-structures. For instance, in (37), the noun *canlısı* is represented with three different nodes although it is a single word.

There are five lexical integrity tests employed by Bresnan and Mchombo [1995] to decide whether the words constructed by derivational suffixes are lexicalized or not. Once these tests are applied to derived words in Turkish, it can be observed that there are certain suffixes which do not obey the standard definition of suffixes although they are attached to words orthographically. The most distinctive results come from tests on phrasal recursivity. In this section we briefly give the definitions and examples from Bresnan and Mchombo [1995] and then provide the Turkish examples with our comments.

Extraction

Bresnan and Mchombo [1995] give the definition of extraction as follows and examplify the test with sentences in (38).

> Constituents of words cannot be extracted by syntactic operations, such as relativization, clefting or topicalization, which leave visible gaps in structure.

(38) a. They've been [American history] teachers for years.

b. ∗American history, which they've been _ teachers for years ...

c. ∗American history, which they've been it teachers for years, ...

Although the examples do not attempt to extract the constituents of a word, the definition also holds for Turkish, as it is not possible to extract the stem of a derived word by using syntactic operations.

Conjoinability

The paper distinguishes between the behavior of syntactic and morphological constituents by stating that "while syntactic categories can be conjoined by syntactic conjunctions, stems and affixes normally cannot". It supports this claim with (39) and (40).

(39) a. Mary outran and outswam Bill.

b. ∗Mary outran and -swam Bill.

(40) a. John's joyfulness and cheeriness kept us going.

b. ∗John's joyful, and cheeriness kept us going.

Bresnan and Mchombo [1995] state that examples like *outswam* and *joyfulness* are lexicalized. The paper also indicates that there are counterexamples and explain their behavior with the help of phonological words. (41) gives examples of a conjoinable suffix in Turkish.

(41) a. ev-de-ki ve araba-da-ki
house-Loc-Rel and car-Loc-Rel
'in the house and in the car'

b. [ev-de ve araba-da]-ki
[house-Loc and car-Loc]-Rel

c. [ev ve araba]-da-ki
[house.Nom and car]-Loc-Rel

The usage in (41c) is more common than the usage in (41b) and the example may be more related to suspended affixation [Kabak, 2007] than conjoinability.[5] The next example is more convincing. The derivational affix *-ken* which derives an adverb with the meaning of 'while' can always be conjoined as given in (42b).

(42) a. ev-den gel-ir-ken ve okul-a gid-er-ken
house-Abl come-Aor-While and school-Dat go-Pres-While
'while coming from the house and going to the school'

b. [ev-den gel-ir ve okul-a gid-er]-ken
[house-Abl come-Aor and school-Dat go-Aor]-While

Another conjoinable derivational suffix is given in (43). In this case the usage in (43b) is much more common than the one in (43a). The suffix *-DHr* is used to form copular sentences from adjective phrases, noun phrases, or postpositional phrases.

(43) a. genç-tir ve güzel-dir
young-Cop and beautiful-Cop
'S/he is young and beautiful.'

b. [genç ve güzel]-dir
[young and beautiful]-Cop

Also, there are cases where the derivational suffix cannot be conjoined as exemplified by the suffix *-(y)An* which derives a participle from a sentence in (44).

(44) a. ev-den gel-en ve okul-a gid-en çocuk
house-Abl come-Prespart and school-Dat go-Prespart child.Nom
'the child who comes from the house and who goes to the school'

b. *[ev-den gel ve okul-a gid]en çocuk
[house-Abl come and school-Dat go]-Prespart child.Nom

[5]Suspended affixation is defined in Section 4.1.5.

Gapping

Bresnan and Mchombo takes Simpson's observations [1983; 1991] as the third test: "gapping or ellipsis can apply to syntactic, but not morphological, constituents".

(45) a. John outran Bill and Mary, Patrick.

b. ∗John outran Bill and Mary -swam Patrick. [Simpson, 1991]

(46) a. John liked the play and Mary, the movie.

b. ∗John liked the play, and Mary dis- it. [Simpson, 1991]

There is no derivational suffix in Turkish that we can apply gapping to.

Inbound Anaphoric Islands

The fourth test claims that "while phrases can contain anaphoric and deictic uses of syntactically independent pronouns, derived words and compounds cannot". A supporting example from Postal [1969] is given in (47).

(47) a. McCarthyite

b. ∗himite [Postal, 1969]

In Turkish, there are examples for both supporting and opposing this argument. (48) shows an ungrammatical case, but phrases in (49)-(52) are quite possible. In usage, a native speaker will understand the meaning of the first example although it is ungrammatical. Note that the suffix *-lHk* has two interpretations. The *-lHk* we use in (52) derives an adjective from a noun. The other interpretation derives a noun from an adjective and has the meaning of the suffix *-ness* in English.[6]

[6]In showing the surface suffix boundaries we follow the output of the Turkish morphological analyzer[Oflazer, 1994].

(48) a. [kedi]-li
cat-With
'with a cat'

b. ∗[o]-nlu
s/he-With
'with her/him'

(49) a. [kedi]-siz
it-Without
'without a cat'

b. [o]-nsuz
it-Without
'without it'

(50) a. [kız]-cağız
girl-Dim.Nom
'poor girl'

b. [o]-ncağız
she-Dim.Nom
'poor she'

(51) a. [Ali'-de]-ki
Ali-Loc-Rel
'the one at Ali'

b. [o-nda]-ki
he-Loc-Rel
'the one at him'

(52) a. bu ceket tam [babam]lık
this jacket.Nom just father-Pos-Fitfor
'this jacket is just right for my father. (e.g. fits well or his style)'

b. bu ceket tam [on]luk
this jacket.Nom just he-Fitfor
'This jacket is just right for him.'

Phrasal Recursivity

Bresnan and Mchombo [1995] state that "word-internal constituents generally differ from word-external phrases in disallowing the arbitrarily deep embedding of syntactic phrasal modifiers" and give the example in (53).

(53) a. [happy]-ness

b. ∗[quite happi]-ness

c. ∗[more happy [than sad]]-ness

This test is the one most similar to our basic concerns. We have adopted the IG-based approach to correctly identify the dependency relations among the phrases

and thus obtain the bracketing of the phrases as given in (54) - (57), which verifies phrasal recursivity. But not all of the derivational suffixes can satisfy this condition, a counterexample is given in (58).

(54) a. evde-ki
house-Loc-Rel
'in the house'

b. [bu ev-de]-ki
[this house-Loc]-Rel
'in this house'

c. [sen-in ev-in-den daha güzel ev-de]-ki
[you-Gen ev-Poss-Abl more beautiful house-Loc]-Rel
'in the house which is more beautiful than your house'

(55) a. gel-en adam
come-Prespart man.Nom
'the man who comes'

b. [geç gel]-en adam
[late come]-Prespart man.Nom
'the man who comes late'

(56) a. elbise-li
dress-With
'with a dress'

b. [mavi elbise]-li
[blue dress]-With
'with a blue dress'

(57) a. perde-lik kumaş
curtain-Fitfor fabric.Nom
'fabric for curtains'

b. [kısa perde]-lik kumaş
[short curtain]-Fitfor fabric.Nom
'fabric for short curtains'

(58) a. mutlu-luk
happy-Ness.Nom
'happiness'

b. *[cok mutlu]luk
[very happy]-Ness.Nom
'[very happy]ness'

c. *[sen-den daha mutlu]-luk
[you-Abl more happy]-Ness.Nom
'[happier than you]ness'

The paper points out some possible syntactic phrases that can be derived, mentioning Afrikaans, English and Japanese examples. They follow Spencer's analyses [Spencer, 1988, 1991] by claiming that such phrases are lexicalized. In Turkish, however, none of the phrases that undergo derivations in the given examples above are lexicalized.

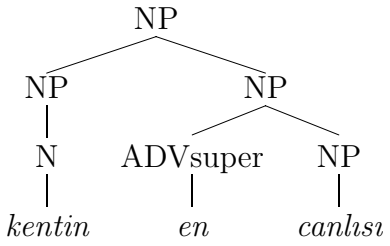
Lieber's [1988; 1992] approach is similar to ours in the way that it allows phrasal recursion within lexical categories, in violation of the lexical integrity principle. According to the authors, one of the problems of this approach is that Lieber would also try to syntactically construct examples like (59). These problematic cases are grammatical sentences in Turkish, because every sentence can be used as a noun phrase, hence the authors' argument is not applicable to Turkish.

(59) a. ??the Prince of Wales and the woman that he married syndrome,
b. ??an ate too much and smoked a post-prandial cigar headache,
c. ??who's the manager, proprietor, or CEO wink

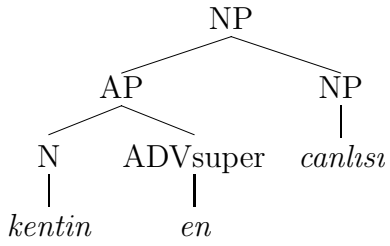
Conclusion

In summary, most Turkish suffixes have phrasal scope. Without the IG approach, one would end up with c-structures that do not reflect the linguistic intuitions. Consider the phrase *mavi elbiseli* 'with a blue dress' in (56b). If we attached the suffix *-li* to the stem *elbise* without considering the phrasal scope, the adjective *mavi* would seem to modify the derived adjective *elbiseli*. Similarly, the c-structure in (60) would be the representation of the phrase in Figure 3.3, instead of (37), p.27 if the IG represantation had not been preferred.

(60)



(61)



Another proposed alternative was implementing the approach in Bresnan and Mugane [2006]. (61) gives the c-structure of Figure 3.3 according to this approach. In any of these alternatives, the lexical integrity is preserved but the c-structure does not reflect the actual relations between the phrases. There is both information loss and misconception about the phrase structures of the language. For instance, in (60) the adverb *en* seems to modify the derived noun *canlısı* although adverbs cannot modify noun phrases in Turkish. Further, in (61) an NP and an ADV seem to construct an AP, and again, it is not one of the generalizations of Turkish grammar. Thus, we claim that our approach fits better the computational treatment of Turkish syntax.

3.4 Other Grammars for Turkish

Güngördü and Oflazer [1995] describe a rather extensive grammar for Turkish using the LFG formalism. Although this grammar had a good coverage and handled phenomena such as free-constituent order, the underlying implementation was based on pseudo-unification. But most crucially, it employed a rather standard approach to represent lexical units: words with multiple nested derivations were represented with complex nested feature structures where linguistically relevant information could be embedded at unpredictable depths which made access to them in rules extremely complex and unwieldy.

Bozşahin [2002] has concerns similar to ours on the scope of derivational morphemes. He argues that inflectional morphemes also have phrasal scope and the most appropriate way to handle these scope relations (both for inflections and derivations) is to prefer morphosyntactic rules instead of syntactic rules. Therefore he employs morphemes

overtly as lexical units in a CCG framework to account for a variety of linguistic phenomena. The implementation aims to solve the problematic cases rather than to extend coverage. The drawback is that morphotactics is explicitly raised to the level of the sentence grammar, hence the categorial lexicon accounted for both constituent order and the morpheme order with no distinction.

Oflazer's dependency parser [Oflazer, 2003] is based on an extended finite state approach where the dependency relations are established between IGs. The rules of the grammar are defined in terms of regular expressions that form a composed finite state transducer. There is also a syntactic filtering component to filter the overparses, again implemented as a finite state transducer. The input sentence is first morphologically analyzed and converted into an IG representation. Then the parser and filter components are applied to the IG representation iteratively. Each iteration sets head and dependent relations between the IGs, until a fixed point is reached, i.e., there are no more dependency relations added in an iteration. Parses are then ranked according to the total link length. He also provides lenient filtering for robustness and allows the system to output partial dependency structures when there is no full parse.

Çakıcı [2005], uses relations between IG-based representations encoded within the Turkish Treebank [Oflazer et al., 2003] to automatically induce a CCG grammar lexicon for Turkish. She uses the dependencies in the treebank except that coordination is left for future work. The version of the Turkish Treebank that is used, does not contain dependency information for relative clauses. Labels that represent such dependencies are manually added in order to extract the information in long distance dependencies. It is the earliest attempt for Turkish to automatically build a large coverage and linguistically expressive grammar by using a treebank.

Another work that investigates the use of IGs is Eryiğit, Nivre, and Oflazer's [2008] dependency parsing experiments. They conduct tests on a probabilistic parser and a classifier-based parser with words or IGs as parsing units. They also test the effects of adding morphological information as features and lexicalization. In all possible test cases, taking IGs as the parsing unit outperforms word-based parsing. The best score

is achieved when the classifier based parser is run with parameters combining IG based representation, morphological information, and lexicalization.

Chapter 4

LEXICAL FUNCTIONAL GRAMMAR ANALYSES OF VARIOUS TURKISH LINGUISTIC PHENOMENA

This chapter summarizes all rules in the grammar in general and mainly focuses on how inflectional groups are used in derivational linguistic phenomena by giving example sentences and their corresponding f-structures. The derivational suffix attached to the verb may change the function of the sentence containing the verb as a whole, as in infinitives and participles, or may modify the function of verb arguments in the derived structure in a valency alternating case like causativization and passivization.

We only briefly mention the rules that are comparatively straightforward either in terms of linguistics or in terms of implementation, and explain the more interesting cases in detail. Section 4.1 gives a general overview of the rules in the grammar. The following sections first analyze a linguistic phenomenon and then explain the LFG implementation. In Section 4.2 we focus on causatives. We discuss their clausal representation by conducting tests, and then present their implementation. Section 4.3 investigates different types of passives and provides their f-structures. An analysis for noun verb compound verbs is proposed in Section 4.4. Finally, Section 4.5 groups non-canonical objects into subsets and observes their behavior under causativization, passivization, and raising. We present our analyses and implementation in Section 4.5.3.

4.1 General Overview of Rules

This section presents on overview of the set of rules that make up the majority of the grammar. Our grammar comprises an extensive set of rules to handle noun phrases. After an overview of recent relevant work on Turkish noun phrases, we give a parsing example with a sample noun phrase and its rule and list the other types of noun phrase rules in Section 4.1.1. In addition, we present an overview of adjective, adverbial, and postpositional rules in Section 4.1.2. Sentential complements, sentential adjuncts, and relative clauses are all constructed by morphological derivations. Section 4.1.3 goes into detail with these derivations by using examples and presenting the LFG analyses. We present the main sentence rule and discuss the problems we encountered in implementing free word order (Section 4.1.4) and coordination (Section 4.1.5). The section concludes with a description of the date-time grammar (Section 4.1.6).

4.1.1 Noun Phrases

A noun phrase is any sequence of words that can function as a subject, or as some kind of an complement such as an object, a subject complement, the complement of a postposition [Göksel and Kerslake, 2005]. The case and referentiality plays an important role in determining the argumenthood of noun phrases. Recently there has been extensive work on Turkish that examines the case and referentiality features [Öztürk, 2005; Arslan Kechriotis, 2006].

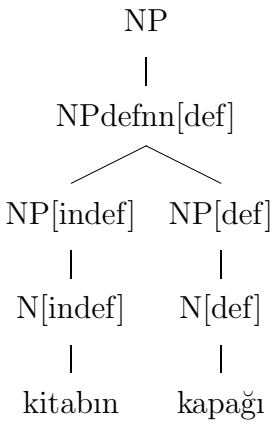
Öztürk [2005] claims that case and referentiality are strongly correlated and they are assigned by the same functional projection, since there is no Determiner Phrase (DP) layer in terms of Minimalist Program to assign referentiality separately from case. Arslan Kechriotis [2006] takes a contrary position and argues that Turkish employs DP despite the lack of an overt determiner system. She compares (morphologically) nominative NPs with no determiner with nominative [bir NP] constructions and concludes that there are syntactic differences between them. This finding is, again, contrary

to Öztürk's analyses. Observing that referential nominals are DPs and non-referential nominals are NPs, she also discusses the position and function of these phrases such as the behavior as subject and object, position with respect to adverbials, and position and case marking in embedded clauses and under relativization. The related work on Turkish noun phrases provides us important analyses on explaining the behavior of different noun phrases within the sentence.

In our approach, we take the determiners as the modifiers of noun phrases, unlike the Minimalist Program which takes determiners as the heads of DPs and nouns as the complements. This section only deals with the construction of several types of noun phrases. The role of noun phrases within the sentence is discussed in Sections 4.1.4, 4.2, and 4.5.

Our grammar covers a wide range of different types of noun phrases, including indefinite and definite noun compounds, possessives, pronouns, proper nouns, derived noun phrases, NPs modified by adjectives, determiners, numbers, measure phrases, postpositions, and combinations of these. In indefinite noun compounds, an NP in nominative case modifies the head NP and the modifying NP functions as MODifier in the LFG representation. In definite noun compounds, an NP in genitive case modifies the head NP, and this time the modifying NP functions as a possessive specifier, namely SPEC POSS. (62) and (63) give the c-structure and the f-structure for the simple definite noun compound *kitabın kapağı* 'book's cover'.

(62)



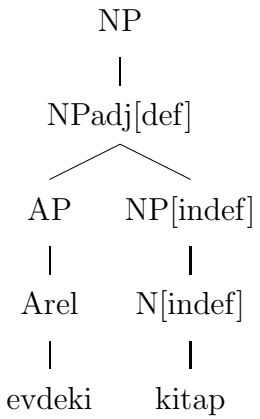
(63)

$$\begin{bmatrix} \text{PRED} & \text{'kapak'} \\ \text{SPEC} & \begin{bmatrix} \text{POSS} & \begin{bmatrix} \text{PRED} & \text{'kitap'} \\ \text{CASE} & \text{gen, NUM sg, PERS 3} \end{bmatrix} \end{bmatrix} \\ \text{CASE} & \text{NOM, NUM SG, PERS 3} \end{bmatrix}$$

The definiteness feature of nouns is stored in the c-structure by using complex

categories, i.e, categories that can take arguments, to be able to modify its value during unification. For noun phrases, the value of the argument is either `def` or `indef`. An example which makes use of this property is given in (64). The head of the NP is *kitap* 'book' which is indefinite as a single noun but the whole phrase *evdeki kitap* 'the book at the house' is definite. During parsing, the f-structure of the head unifies with the f-structure of the whole phrase. Having a feature value pair [DEF -] in the f-structure of *kitap* 'book' would result in an unwanted [DEF -] in the final f-structure. Instead, we do not carry the argument `indef` of the NP up the tree and assign the correct value `def` to the argument of the complex category NPadj.

(64)



The rule for the noun phrase *evdeki kitap* 'the book at the house' is given in (65). NPadj is composed of an AP followed by an NP. The NP is the head of the NPadj ($\uparrow = \downarrow$), and AP is the ADJUNCT in the resulting f-structure (($\uparrow$ ADJUNCT) $= \downarrow$). There are three disjuncts in the rule, each representing a generalization on NP types. Only NPs falling into one of these disjuncts can be modified by adjectives derived by *-ki*.

(65)

NPadj[_var] →	AP	{ NP[indef]	NP[_var]	NPvalid }
	($\uparrow$ ADJUNCT) $= \downarrow$	$\uparrow = \downarrow$	$\uparrow = \downarrow$	$\uparrow = \downarrow$
		_var=def	_var=def	_var=def
			($\uparrow$ SPEC DEF)	($\downarrow$ SPEC POSS)

The first type deals with indefinite NPs; in this case the final NPadj is definite (`_var=def`). Our example phrase falls into this group. The second type deals with

Rule Name	Description
NPdet[_defvar]	determiner-modified NPs, e.g., *bu kitap* 'this book'
NPnum[_defvar]	number-modified NPs, e.g., *iki kitap* 'two books'
NPadj[_defvar]	adjective-modified phrases, e.g., *mavi kitap* 'blue book'
NPmeas[_defvar]	measure phrases, e.g., *büyük bir kutu kitap* 'a big box of books'
NPpostp[_defvar]	postposition-modified NPs, e.g., *kitaba ait kapak* 'cover belonging to the book'
NPnn[_defvar]	indefinite noun compounds, e.g., *kitap kapağı* 'book cover'
NPposs[_defvar]	covert possessive NPs, e.g., *kitap kapağım* 'my book cover'
NPdefnn[_defvar]	definite noun compounds, e.g, *kitabın kapağı* 'book's cover'
NPpron[_defvar]	possessive NPs, e.g. *benim kedim* 'my cat'
PRON	pronouns, e.g., *ben* 'I'
PROP	proper names, e.g., *Ahmet*
PROPloc	proper location names, e.g., *İstanbul*
N[_defvar]	basic nouns, e.g., *kitap* 'book', *kitabım* 'my book'
NPpart	sentential complement, infinitives, e.g., *gitmek* 'to go'
NPderiv	NPs derived from adjectives or numbers, e.g., *ikide* 'at two'

Table 4.1: Types of noun phrase rules

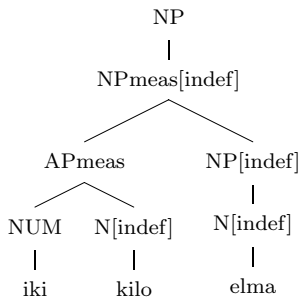
definite NPs with a determiner, e.g., *evdeki bu kitap* 'this book at the house'. And finally, the third type is used for valid NPs with a possessor, where NPvalid represents the set of definite possessive NPs, definite nouns, or nouns derived from adjectives.[1] The phrase *evdeki kitaplarım* 'my books at the house' is an example for the third set. The NP grammar is composed of rules that follow the basic rule structure of (65). We summarize these rules in Table 4.1.

The actual noun f-structures also carry semantic information about nouns (e.g., `common, proper, count, mass, measure`). This information is crucial for parsing some phrases. The morphological analyzer outputs some semantic information such as `proper`, but most of the semantic details are manually encoded in the lexicon. For instance, measure nouns have a semantic marker in the lexicon and measure phrases

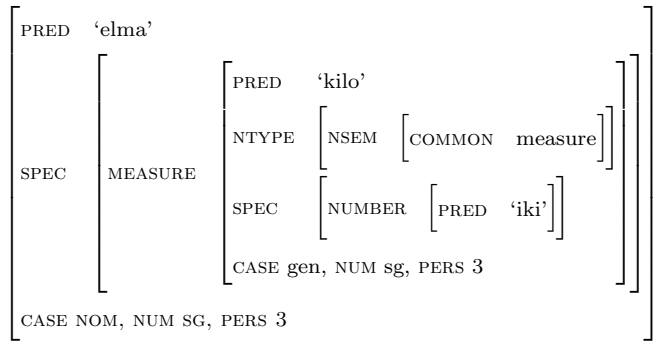
[1] NPvalid is defined as `NPvalid = { NPposs[def] | N[def] | NPderiv }.` in the grammar.

have a separate rule in the grammar. (66) and (67) show the c-structure and f-structure of the phrase *iki kilo elma* 'two kilos of apple'. The marker `measure` placed in the f-structure of *kilo* enables the phrase to be parsed by the rules APmeas (for measure APs) and NPmeas (for measure NPs).

(66)



(67)



We conclude this section with the structures for a relatively complex NP, giving the actual XLE output of the phrase instead of simplified representative structures. Figure 4.1 and Figure 4.2 illustrate the c-structure and f-structure of the NP *benim tarih dersimin kitabının yeni basımı* 'the new edition of my history course's book'. All parts of speech have type information (e.g., ATYPE, NTYPE, PRON-TYPE) in their f-structures and there is also the CHECK feature that keeps information on well-formedness which we usually omit in simplified structures.

4.1.2 Adjective, Adverbial, and Postpositional Phrases

Similar to noun phrases, adjective, adverbial, and postpositional phrases are essential components of a wide coverage grammar. This section summarizes the basic rules of those phrases. Deverbal constructions of adjectives and adverbs are discussed separately in Section 4.1.3.

Adjective Phrases

The adjective phrase grammar includes rules for basic, comparative and superlative adjectival phrases such as *mutlu* 'happy', *daha mutlu* 'happier', *en mutlu* 'the happiest'.

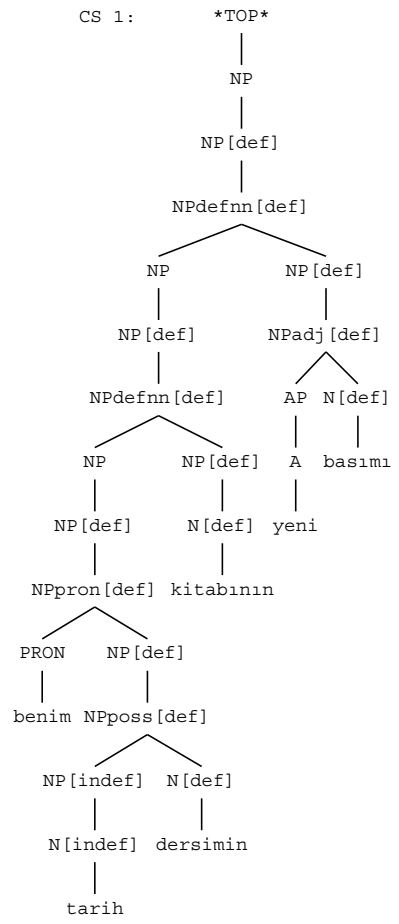
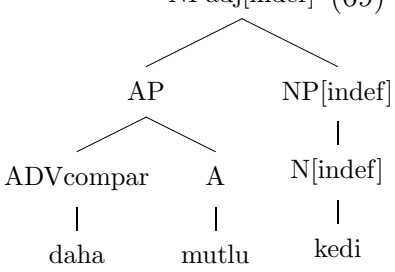


Figure 4.1: C-structure of the NP *benim tarih dersimin kitabının yeni basımı* 'the new edition of my history course's book'

The degree of the adjective is also represented in the f-structure, with values `positive`, `comparative`, and `superlative` respectively. (68) and (69) give the c-structure and f-structure for the AP *daha mutlu kedi* 'happier cat'.

(68)



(69)

$$\begin{bmatrix} \text{PRED} & \text{'kedi'} \\ \text{ADJUNCT} & \begin{bmatrix} \text{PRED} & \text{'mutlu'} \\ \text{ADJUNCT} & \begin{bmatrix} \text{PRED} & \text{'daha'} \\ \text{DEGREE} & \text{positive} \end{bmatrix} \\ \text{DEGREE} & \text{comparative} \end{bmatrix} \\ \text{CASE NOM, NUM SG, PERS 3} \end{bmatrix}$$

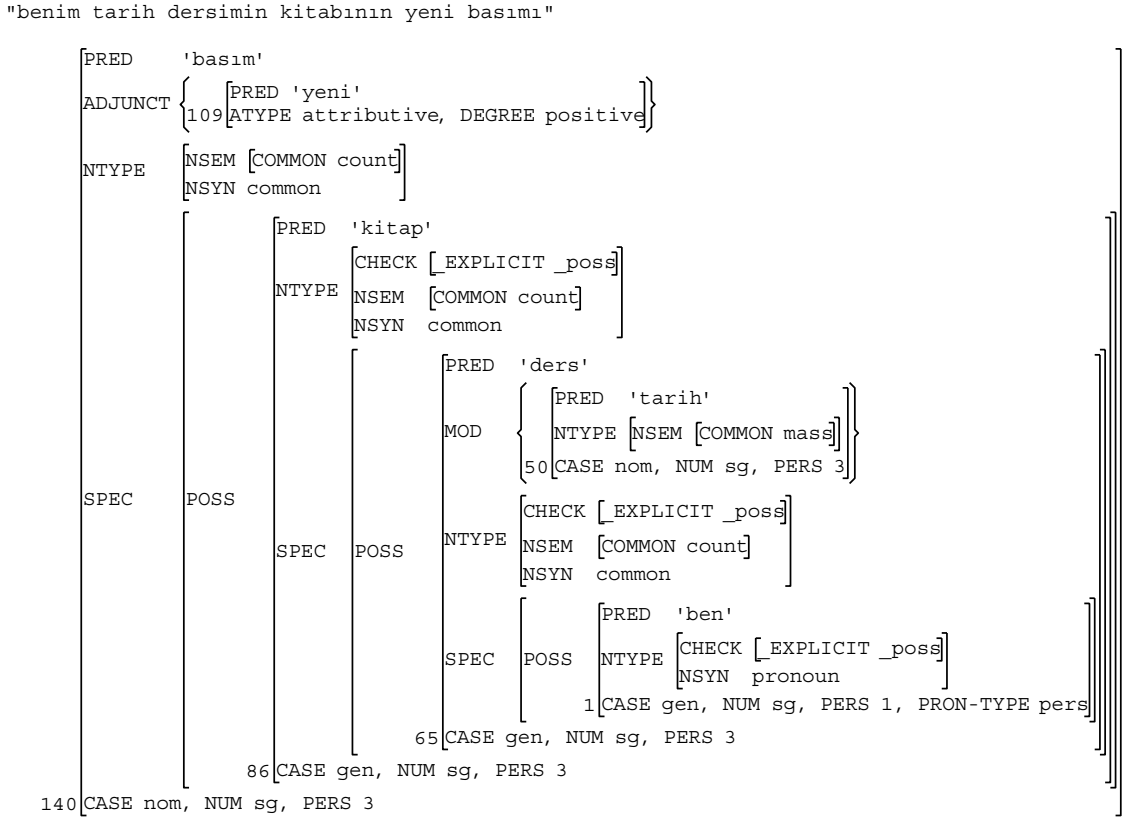


Figure 4.2: F-structure of the NP *benim tarih dersimin kitabının yeni basımı* 'the new edition of my history course's book'

There is a group of phrases that requires special treatment due to their semantics. Although the phrase *iki fincan* 'two cups' should be a noun phrase as *fincan* is a noun, it is parsed as an adjective phrase APcont (container adjective phrase), so that the container phrase can modify a mass noun, e.g. *iki fincan kahve* 'two cups of coffee'. We follow exactly the same approach for the measurement phrases and treat them as adjective phrases as well.

Derived adjectives are handled by encoding two types of rules. If the derivational suffix has phrasal scope it has a separate rule. If the adjective suffix is attached to simple words, for instance *-CH* '-ist' in e.g. *merkez-ci* 'centralist', *barış-çı* 'pacifist', then the generic rule Aderiv is used. Table 4.2 summarizes the rules in the adjective phrase grammar. APpart which covers relative clauses is explained in Section 4.1.3.

Rule Name	Description
A	basic adjectives, e.g., *mavi* 'blue'
APcont	container APs, e.g., *iki fincan* 'two cups (of)'
APmeas	measure APs, e.g., *iki kitap* 'two books'
Aderiv	derived adjectives with no phrasal scope, e.g., *milliyetci* 'nationalist'
Awith, Arel	derived adjectives with phrasal scope, e.g., *beyaz elbiseli* 'with a white dress'
APpart	participles, e.g., *uyuyan* 'sleeping'

Table 4.2: Types of adjective phrase rules

Adverbial Phrases

The part of the grammar that handles adverbial phrases consists of rules for parsing simple, comparative, and superlative adverbs, adverbs modifying other adverbs, e.g. *az* 'less', *çok* 'more', derived adverbs, e.g. *sakince* 'calmly', and adverbs formed by duplicating adjectives, e.g, *sakin sakin* 'calmly, lit. calm calm'. There is also a special constituent focus rule[2] for adverbs like *bile* 'even', *dA* 'too', *falan/filan* 'etc.'. They attach these adverbs after every possible phrase. For the basic sentence in (70a), the sentences in (70b) - (70c) represent all possible placements of the adverb *bile* 'even'.

(70) a. Zeynep sabah yumurta-sı-nı ye-di
Zeynep.Nom morning.Nom egg-Poss-Acc eat-Past.3sg
'Zeynep ate her egg in the morning.'

b. Zeynep bile sabah yumurtasını yedi
Zeynep.Nom even morning.Nom egg-Poss-Acc eat-Past.3sg
'Even Zeynep ate her egg in the morning.'

c. Zeynep sabah bile yumurtasını yedi
Zeynep.Nom morning.Nom even egg-Poss-Acc eat-Past.3sg
'Zeynep ate her egg even in the morning.'

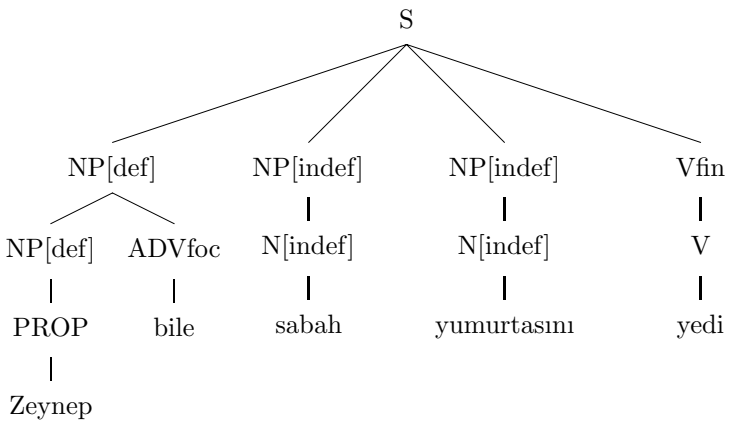
d. Zeynep sabah yumurtasını bile yedi
Zeynep.Nom morning.Nom egg-Poss-Acc even eat-Past.3sg
'Zeynep ate even her egg in the morning.'

[2]very similar to the one used in the ParGram English grammar

e. Zeynep sabah yumurtasını yedi bile
Zeynep.Nom morning.Nom egg-Poss-Acc eat-Past.3sg even
'Zeynep even ate her egg in the morning.'

The c-structure and f-structure of (70b) is given in (71) and (72) respectively.

(71)



(72)

$$\begin{bmatrix} \text{PRED} & \text{'ye⟨Zeynep, yumurta⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'Zeynep'} \\ \text{ADJUNCT} & \begin{bmatrix} \text{PRED} & \text{'bile'} \end{bmatrix} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'yumurta'} \\ \text{SPEC} & \begin{bmatrix} \text{POSS} & \begin{bmatrix} \text{PRED} & \text{'null_pro'} \\ \text{NUM sg, PERS 3} \end{bmatrix} \end{bmatrix} \\ \text{CASE} & \text{acc} \end{bmatrix} \\ \text{ADJUNCT} & \begin{bmatrix} \text{PRED} & \text{'sabah'} \\ \text{CASE} & \text{nom} \end{bmatrix} \end{bmatrix}$$

Table 4.3 summarizes the rules in the grammar for adverbial phrases. ADVsub which covers subordinate clauses is explained in Section 4.1.3.

Rule Name	Description
ADV	basic adverbs, e.g., *erken* 'early'
ADVcompar	comparative adverb *daha* 'more'
ADVsuper	superlative adverb *en* 'most'
ADVderiv	derived adverbs, e.g., *sakince* 'calmly'
Adup	adverbs derived by duplicating adjectives, e.g., *sakin sakin* 'calmly'
ADVsub	subordinate clauses, e.g., *uyurken* 'while sleeping'
ADVmodADV	adverbs modifying adverbs, e.g., *çok* 'very'
ADVfoc	constituent focusing adverbs, *bile* 'even'

Table 4.3: Types of adverbial phrase rules

Postpositional Phrases

The postposition rule is straightforward, the only crucial information, that is, the case marker of the NP that the postposition subcategorizes for, comes from the morphological analyzer. The analysis for *ait* 'belonging to' is `ait+Postp+Dat`. `+Dat` indicates that the NP should be dative, hence the dative marked *Ali'ye* 'to Ali' can function as the OBJect of *ait*. The f-structure of the postpositional phrase *Ali'ye ait* 'belonging to Ali' is illustrated in (73). Whether the resulting postposion phrase (POSTPP) modifies an NP, e.g., *Ali'ye ait kitap* 'the book belonging to Ali', or serves as an adverbial phrase, e.g., *yemekten sonra* 'after the dinner', is determined by semantic markers.

(73)
$$\begin{bmatrix} \text{PRED} & \text{'ait}\langle\text{Ali}\rangle\text{'} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'Ali'} \\ \text{CASE dat, NUM sg, PERS 3} & \end{bmatrix} \\ \text{CASE NOM, NUM SG, PERS 3} & \end{bmatrix}$$

There is also a handful of words that behave as postpositions although they are nouns. They cannot be taken as simple lexicalized postpositions neither by the morphology nor by the syntax due to agreement in person during the phrase construction. *yüzünden* 'because of', as one of the members of the set, has the alternations in (74a) and (74b) for 1st and 3rd person singular. The lemma (here, *yüz*) and the case (here,

ablative) of the noun acting as postposition are handcoded in the grammar. Other information can be generalized: agreement in person and number with the exception of nominative case in 3rd person nouns (cf. (74c)).

(74) a. ben-im yüz-üm-den
I-Gen because.of-P1sg-Abl
'because of me'

b. on-un yüz-ün-den
he/she/it-Gen because.of-P2sg-Abl
'because of him/her/it'

c. kedi yüz-ün-den
cat.Nom because.of-P2sg-Abl
'because of the cat'

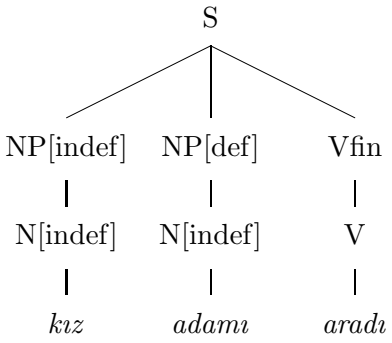
The very few postpositions originating from other categories (*başka* 'other than, lit. other', *diye* 'in the way of, lit. say-Opt', *nazaran* 'as compared to, lit. by glance') are lexicalized in our morphological analyzer and are handled by the standand postposition rule.

4.1.3 Sentential Complements, Sentential Adjuncts, and Relative Clauses

In Turkish, sentential complements and adjuncts are marked by productive verbal derivations into nominals (infinitives, participles) or adverbials. Relative clauses with subject and non-subject (object or adjunct) gaps are formed by participles which function as adjectivals modifying a head noun. (75) shows a simple sentence that will be used throughout the following examples. Its c- and f-structure are given in (76a) and (76b), respectively.

(75) kız adam-ı ara-dı
girl.Nom man-Acc call-Past.3sg
'The girl called the man.'

(76) a.



b.
$$\begin{bmatrix} \text{PRED} & \text{'ara⟨kız,adam⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'kız'} \\ \text{CASE} & \text{nom} \\ \text{NUM sg, PERS} & 3 \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'adam'} \\ \text{CASE} & \text{acc} \\ \text{NUM sg, PERS} & 3 \end{bmatrix} \\ \text{TENSE} & \text{past} \end{bmatrix}$$

Sentential Complements

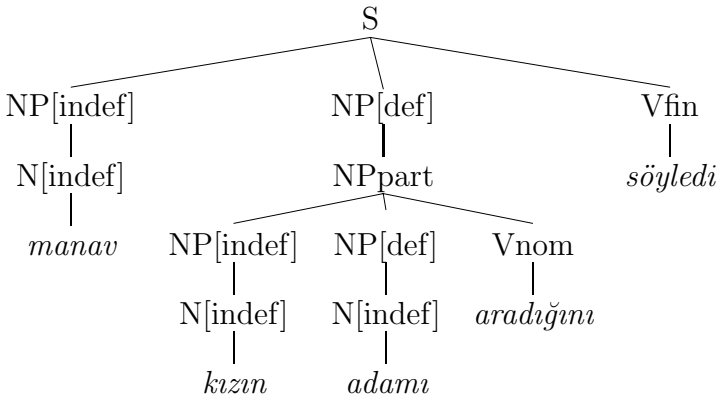
In (77), we give a past participle form as the head of a sentential complement. This complement functions as an object for the verb *söyledi* 'said'. It is derived from (75).

(77) manav kız-ın adam-ı ara-dığı-nı söyle-di
grocer.Nom girl-Gen man-Acc call-PastPart-Acc say-Past.3sg
'The grocer said that the girl called the man.'

Once the grammar encounters such a sentential complement, the verb with its empty arguments (here, SUBJ and OBJ) and the participle IG with its nominal features, e.g., CASE, construct the derivation. Later, the constituents of the sentential complement fill in those empty arguments as in a normal sentence.

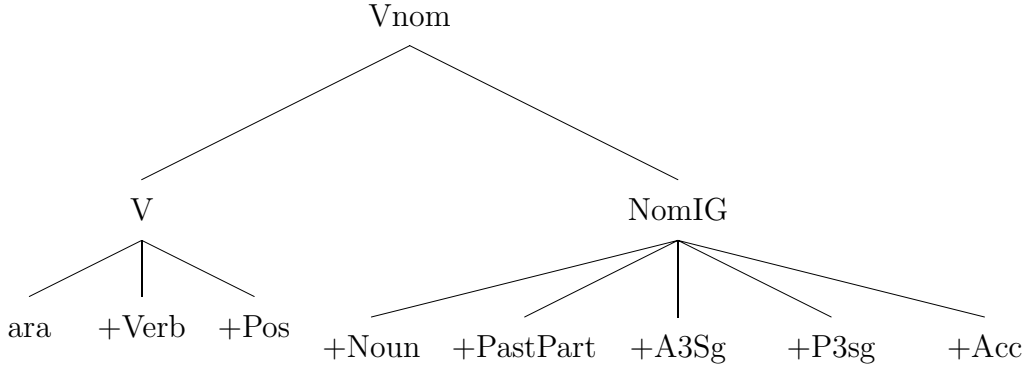
(78) gives the c-structure of the sentence in (77). Note that the participle IG including the derivational morpheme is attached to the base verb in the node Vnom, unlike placement of the IG in (37), p.27, which is a separate node in the tree. This is necessitated by the free constituent order: the NP *adamı kızın aradığını* is valid, as well as the NPs with other permutations of the constituents within the participle phrase. Representing the IG on the sublexical level never causes loss of information that we discussed in Section 3.3. In participle derivation, there cannot be nested subtrees where one of the nodes modifies the inner nodes of the head —here, the verb— thanks to the characteristics of the derivational suffixes of this kind.

(78)



The IG is part of the sublexical tree of Vnom, which is invisible in the standard c-structure representation. (79) unfolds the leaves of the sublexical tree. The subcategorization information is carried in the root *ara* and the nominal features come from the IG part.

(79)



The resulting f-structure is for a noun phrase, which is now the object of the matrix verb *söyledi* 'said' in (77). The final f-structure for the whole sentence is shown in (80). Since the participle IG has the complete set of syntactic features of a noun, no new rules are needed to incorporate the derived f-structure to the rest of the grammar, that is, the derived phrase can be used as if it is a simple NP.

(80)
$$\begin{bmatrix}
\text{PRED} & \text{`söyle⟨manav, ara⟩'} \\
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`manav'} \\ \text{CASE nom, NUM sg, PERS 3} \end{bmatrix} \\
\text{OBJ} & \begin{bmatrix}
\text{PRED} & \text{`ara⟨kız, adam⟩'} \\
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`kız'} \\ \text{CASE gen, NUM sg, PERS 3} \end{bmatrix} \\
\text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`adam'} \\ \text{CASE acc, NUM sg, PERS 3} \end{bmatrix} \\
\text{CHECK} & \begin{bmatrix} \text{PART} & \text{pastpart} \end{bmatrix} \\
\text{CASE acc, NUM sg, PERS 3, CLAUSE-TYPE nom}
\end{bmatrix} \\
\text{TENSE} & \text{past}
\end{bmatrix}$$

The f-structure and c-structure similarities of sentences in (75) and (77) can easily be observed. In both cases, the structures of (77), in a way, encapsulate the structures of (75). The structures of the basic sentence and the derived sentential complement have many features in common. We can observe the same similarity in the grammar rules too. In a very simplified representation, the sentence has the rule in (81a) and the sentential complement is parsed by (81b).[3]

(81)

a. S →	NP[_var]	NP[_var]	Vfin
	(↑ SUBJ) = ↓	(↑ OBJ) = ↓	↑ = ↓
	(↓ CASE)= nom		(↓ CASE)
b. NPpart →	NP[_var]	NP[_var]	Vnom
	(↑ SUBJ) = ↓	(↑ OBJ) = ↓	↑ = ↓
	(↓ CASE)= gen		

[3]Note that the rules are oversimplified to focus on the similarities and distinguish the major differences. It is possible to have non-genitive subjects in the sentential complement as given in (1)

(1) yol-dan bir araba geç-tiği-ni gör-dü-m
road-Abl a car.Nom went.by-PastPart-Acc see-Past-1sg

'I saw that a car went by on the road.' [Kornfilt, 2002]

The rules also have disjuncts and constraints to handle such cases.

Basically the rules differ in the construction of the verb and some minor constraints, e.g., the case of the subject. To understand whether the parsed sentence is a complete sentence or not, the finite verb requirement is checked. Since the requirement is met by the existence of the TENSE feature, (77) is parsed as a complete sentence. There is no TENSE feature in the participle, hence it is not a complete sentence. Indeed the sentential complement also includes temporal information as the `pastpart` value of PART feature, in the OBJect's f-structure, denoting an event in the past.

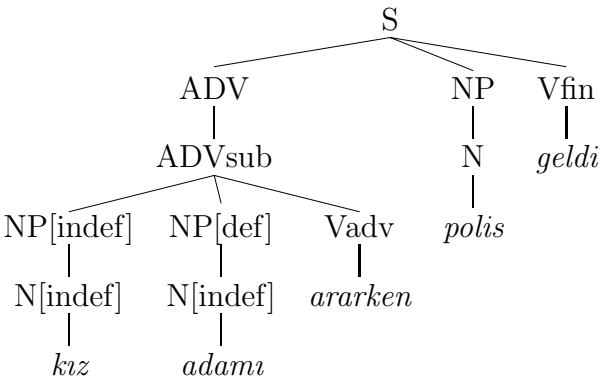
Sentential Adjuncts

Another verbal derivation that follows the same mechanism is the construction of sentential adjuncts. A sentential adjunct example which derives (75) into an adverb is given in (82).

(82) kız adam-ı ara-r-ken polis gel-di
girl.Nom man-Acc call-Aor-While police.Nom come-Past.3sg
'The police came while the girl called the man.'

The c-structure construction of the adverbial clause in (83) is similar to the sentential complement c-structure in (78). Again, Vadv of the adverbial clause is constructed first. The ADVsub rule is similar to the basic S rule in (81a) with a Vadv instead of a Vfin.

(83)



The f-structure for this sentence is shown in (84). Similar to the nominalized clause, which functions as an OBJ in (80), the derived ADJUNCT contains the verb's SUBJect and OBJect as well as the features of the adverb such as ADJUNCT-TYPE. The CHECK feature is important for controlling the SUBJect of the adverbial clause.

(84)
$$\begin{bmatrix} \text{PRED} & \text{'gel⟨polis⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'polis'} \\ \text{CASE nom, NUM sg, PERS 3} \end{bmatrix} \\ \text{ADJUNCT} & \begin{bmatrix} \text{PRED} & \text{'ara⟨kız, adam⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'kız'} \\ \text{CASE nom, NUM sg, PERS 3} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'adam'} \\ \text{CASE acc, NUM sg, PERS 3} \end{bmatrix} \\ \text{CHECK} & \begin{bmatrix} \text{SUB} & \text{while} \end{bmatrix} \\ \text{ADJUNCT-TYPE} & \text{sub} \end{bmatrix} \\ \text{TENSE} & \text{past} \end{bmatrix}$$

Deverbal adverbs can be divided into two groups according to subject control: one group, namely *-(y)AlH* 'since having *verb*ed', *-(y)HncA* 'when (s/he) *verb*s', *-ken* 'while (s/he is) *verb*ing', *-[mA]dAn* 'without having *verb*ed', *-[DHk]çA* 'as long as (s/he) *verb*s', allows different subjects for the adverbial clause and the main sentence. In the other group, namely *-(y)Hp* 'after having *verb*ed' and *-(y)ArAk* 'by *verb*ing', the subject of the matrix verb is also the subject of the inner clause. *-CAsHnA* 'as if (s/he is) *verb*ing' belongs to both of the groups depending on the tense of the verb. If the verb is in aorist tense, then the subjects of the matrix verb and the inner clause should match, but if the verb is in narrative tense, then the subjects might differ.

Relative Clauses

Relative clauses in Turkish are gapped sentences which function as modifiers of nominal heads. Turkish relative clauses have been previously studied [Güngördü and Engdahl,

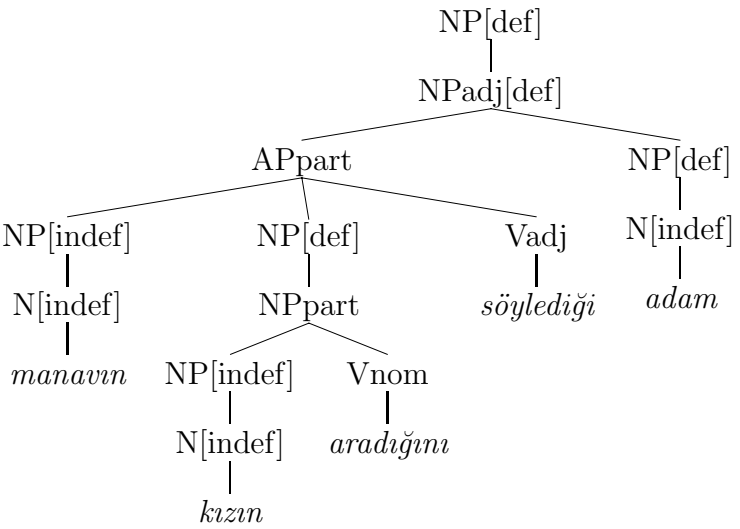
1998; Barker et al., 1990] and found to pose interesting issues for linguistic and computational modeling. Our aim here is not to address this problem in its generality but show with a simple example, how IGs that encode derived forms, handle the mechanics of generating f-structures for such cases.

We basically follow Kaplan and Zaenen's [1989] functional uncertainty approach in handling long distance dependencies. Once we derive the participle phrase we unify it with the appropriate argument of the verb using rules based on functional uncertainty. (85) shows a relative clause where a participle form is used as a modifier of a head noun, *adam* in this case.

(85) manav-ın kız-ın $[\,]_i$ ara-dığı-nı söyle-diği adam$_i$
grocer-Gen girl-Gen obj.gap call-PastPart-Acc say-PastPart man.Nom
'the man the grocer said the girl called'

The rule parsing the relative clause is similar to the other verbal derivation rules. This time, we replace Vfin of the basic sentence rule with Vadj. The c-structure of the sentence in (85) is given in (86). The sentential NP denoted as NPpart in the tree is treated like any regular NP by the rule that parses the participle AP. NPpart has an implicit gap but empty nodes are not allowed in LFG c-structures. The verb *ara* 'call' of NPpart subcategorizes for a subject and an object, and the f-structure of NPpart, hence all the f-structures encapsulating it, would be incomplete with a missing object.

(86)



There is an NPadj rule given in (87) for filling in the gaps like the one inside NPpart. By default, it treats the adjective phrase that modifies a noun phrase as the ADJUNCT of that NP, i.e, (↑ ADJUNCT) = ↓. Additionally, the constraint (↓ OBJ+) = ↑ of the APpart rule states that the mother node of the participle adjective unifies with the current node's function that is composed of at least one OBJect. The f-structure of the participle adjective's mother node (↑, here NPadj) is the f-structure of the head NP by the constraint ↑ = ↓ of NP. Therefore, the rule covers all possible gaps in the path starting with head noun's ADJUNCT OBJ and can continue with infinitely many OBJs.

(87)

NPadj[_var] →	APpart	NP[_var2]
	(↑ ADJUNCT) = ↓	↑ = ↓
	(↓ OBJ+) = ↑	_var = def

The resulting f-structure can be examined more easily from (88). At the innermost level, the NP *kızın aradığını* 'that the girl called' is parsed with a gap object. It then functions as the OBJect of the outer adjectival phrase *manavın kızın aradığını söylediği* 'that the grocer said the girl called'. The participle adjective modifies the head NP *adam* 'man', hence functions as the ADJUNCT of the topmost level f-structure. The gap in the derived form, the object here, is then unified with the head word *adam* as marked with co-indexation in (88). As a result, *adam* unifies with its ADJUNCT's OBJect's OBJect.

(88)
$$\begin{bmatrix}
\text{PRED} & \text{'adam'}\ \boxed{1} \\
\text{ADJUNCT} & \begin{bmatrix}
\text{PRED} & \text{'söyle}\langle\text{manav, ara}\rangle\text{'} \\
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'manav'} \\ \text{CASE gen, NUM sg, PERS 3} \end{bmatrix} \\
\text{OBJ} & \begin{bmatrix}
\text{PRED} & \text{'ara}\langle\text{kız, adam}\rangle\text{'} \\
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'kız'} \\ \text{CASE gen, NUM sg, PERS 3} \end{bmatrix} \\
\text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'adam'}\ \boxed{1} \end{bmatrix} \\
\text{CHECK} & \begin{bmatrix} \text{PART} & \text{pastpart} \end{bmatrix} \\
\text{CASE acc, NUM sg, PERS 3, CLAUSE-TYPE nom}
\end{bmatrix} \\
\text{CHECK} & \begin{bmatrix} \text{PART} & \text{pastpart} \end{bmatrix} \\
\text{ADJUNCT-TYPE relative}
\end{bmatrix}
\end{bmatrix}$$

The example sentence (85) includes (77) as a relative clause with the object extracted, hence the similarity in the f-structures can be observed easily. The ADJUNCT in (88) is almost the same as the whole f-structure of (80), differing only in TENSE and ADJUNCT-TYPE features.

4.1.4 Sentences and Free Word Order

A simplified rule to parse a sentence has been given in (81a). The actual sentence rule is very similar to this simple rule with additional constituents on the right hand side, such as adverbial phrases, postpositional phrases, NPs functioning as adverbs. The most complex part of the rule is Vfin that represents a finite verb. Vfin can be a simple or a derived verb, a noun-verb compound, or can have one of valency alternating suffixes. There is a meta sentence rule which checks if the verb is finite, controls whether subcategorization frames are filled and assigns PRO. No matter how complicated the verb formation is, all sentences are parsed with the same rule.

Copular sentences, on the other hand, have a special rule. When the copular suffix

-DHr is attached to an NP, AP, or POSTPP, the morphological output is parallel to a regular verb, hence sentences containing such copular verbs are parsed with the standard sentence rule.[4] However it is also possible to construct copular sentences by using NPs, APs, or POSTPPs as the predicate without any explicit derivation. (89a) and (89b) give two copular sentences with and without the copular suffix, respectively. The special copular sentence rule covers cases like (89b) to assure that f-structures are identical. Moreover, the representation of the past tense of copular verbs is parallel to that of regular verbs, but the future tense is a construction with the light verb *ol-* 'be'. (89c) and (89d) give two copular sentences in the past and future tense, respectively.

(89) a. kedi mutlu-dur
cat.Nom happy-Cop.3sg
'The cat is happy.'

b. kedi mutlu
cat.Nom happy
'The cat is happy.'

c. kedi mutlu-ydu
cat.Nom happy-Past.3sg
'The cat was happy.'

d. kedi mutlu ol-acak
cat.Nom happy be-Fut.3sg
'The cat will be happy.'

In the implementation, we pay attention to the parallelism of the structures of different sentence types represented in (89). The value of the PRED in the f-structure is '`ol⟨(↑ SUBJ), (↑ XCOMP-PRED)⟩`' where the XCOMP-PRED contains '`pred⟨(↑ SUBJ)⟩`'. `pred` is the predicate of the sentence. (90)-(92) illustrate the f-structures of (89b)-(89d). The differences in the f-structures are their TENSE values. Also note that the value of VTYPE in (92) is `main` instead of `copular`.

[4] i.e., the extended version of (81a).

(90)
$$\begin{bmatrix} \text{PRED} & \text{`ol⟨mutlu, kedi⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`kedi'} \\ \text{CASE} & \text{nom} \end{bmatrix}\boxed{1} \\ \text{XCOMP-PRED} & \begin{bmatrix} \text{PRED} & \text{`mutlu⟨kedi⟩'} \\ \text{SUBJ} & \boxed{1} \\ \text{ATYPE} & \text{predicative} \end{bmatrix} \\ \text{TENSE PRES, VTYPE COPULAR} \end{bmatrix}$$

(91)
$$\begin{bmatrix} \text{PRED} & \text{`ol⟨mutlu, kedi⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`kedi'} \\ \text{CASE} & \text{nom} \end{bmatrix}\boxed{1} \\ \text{XCOMP-PRED} & \begin{bmatrix} \text{PRED} & \text{`mutlu⟨kedi⟩'} \\ \text{SUBJ} & \boxed{1} \\ \text{ATYPE} & \text{predicative} \end{bmatrix} \\ \text{TENSE PAST, VTYPE COPULAR} \end{bmatrix}$$

(92)
$$\begin{bmatrix} \text{PRED} & \text{`ol⟨mutlu, kedi⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`kedi'} \\ \text{CASE} & \text{nom} \end{bmatrix}\boxed{1} \\ \text{XCOMP-PRED} & \begin{bmatrix} \text{PRED} & \text{`mutlu⟨kedi⟩'} \\ \text{SUBJ} & \boxed{1} \\ \text{ATYPE} & \text{predicative} \end{bmatrix} \\ \text{TENSE FUT, VTYPE MAIN} \end{bmatrix}$$

Although Turkish is known to be a free word order language, there are still some restrictions on the word order, especially in the constituent order of subordinate clauses. The nominative object is restricted to immediate preverbal position, but accusative objects can move freely.[5] Still, the usage of some adverbs restrict the position of direct objects. (93) exemplifies the different placement of the adverb *hızlı* 'fast' in sentences with direct or indirect objects. (93d) is not grammatical if we want the adverb to modify the verb. This restriction comes from the semantics of the adverb, as *hızlı* is both an adjective and an adverb, and in (93d) it modifies *kitabı* 'book' instead of the verb *read*. If the adverb has no adjective interpretation, it can be placed in a prenominal position and it still modifies the verb as given in (93e).

(93) a. ben kitab-ı hızlı oku-r-um
I.Nom book-Acc fast read-Aor-1sg
'I read the book fast.'

b. *ben kitap hızlı oku-r-um
I.Nom book.Nom fast read-Aor-1sg
'I read books fast.'

[5]As exemplified in Footnote 2, p.21, there are exceptions to this rule.

c. ben hızlı kitap okurum
I.Nom fast book.Nom read-Aor-1sg
'I read books fast.'

d. *ben hızlı kitabı oku-r-um
I.Nom fast book-Acc read-Aor-1sg
'I read the book fast. (intended)'

e. ben sabahleyin kitab-ı oku-r-um
I.Nom in.the.morning book-Acc read-Aor-1sg
'I (will) read the book in the morning.'

Our implementation allows the constituents of sentential complements move freely within the participle. But there is also a possibility that the constituents of the sentential complement interfere with the constituents of the main sentence, as in (94a). As can be observed from the subtree NPpart in (78), p.51, the whole participle phrase is parsed at once and then used in the main sentence level. Hence, it is not possible to parse non-contiguous chunks of the participle in our approach. Note that the other non-contiguous possibilities, such as (94b) and (94c) are not grammatical.

(94) a. manav **adam-ı** **ara-dığı-nı** söyle-di **kız-ın**
grocer.Nom man-Acc call-PastPart-Acc say-Past.3sg girl-Gen
'The grocer said that the girl called the man.'

b. ***kız-ın** manav **adam-ı** **ara-dığı-nı** söyle-di
girl-Gen grocer.Nom man-Acc call-PastPart-Acc say-Past.3sg
'The grocer said that the girl called the man.'

c. ***adam-ı** **ara-dığı-nı** manav **kız-ın** söyle-di
grocer.Nom man-Acc call-PastPart-Acc say-Past.3sg girl-Gen
'The grocer said that the girl called the man.'

In general, question sentences are constructed by simply omitting the target of the question and inserting the question word into its place, as exemplified in (95a) and (95b). But there is an exception for this generalization; although (95c) is grammatical, (95d) is not.

(95) a. kitab-ı ben oku-du-m
book-Acc I.Nom read-Past-1sg
'I read the book.'

kitab-ı kim oku-du
book-Acc who.Nom read-Past.3sg
'Who read the book?'

b. ben kitap oku-du-m
I.Nom book.Nom read-Past-1sg
'I read books.'

kim kitap oku-du
who.Nom book.Nom read-Past.3sg
'Who read books?'

c. kitap/kitabı oku-du-m ben
book.Nom/Acc read-Past-1sg I.Nom
'I read books/the book.'

d. *kitap/kitabı oku-du kim
book.Nom/Acc read-Past.3sg who.Nom
'Who read books/the book?'

Question sentences like (95a) and (95b) are parsed with the standard sentence rule. The major difference is the value of the feature CLAUSE-TYPE. It is `decl` for declarative sentences but `int` for questions. The grammar also contains rules to parse interrogative sentences.

4.1.5 Coordination

Coordination is an important issue to be solved especially in a computational approach, as the number of possible interpretations of the coordination increases by the number of constituents involved in the coordination. Hence many ambiguous cases occur. Efforts of ParGram members brought up a common set of rules which facilitate the implementation of coordinated structures in XLE. In simple coordination, coordination is a set consisting the f-structure of each conjunct [Kaplan and Maxwell, 1988]. The standard coordination rule is given in (96) where CAT represents any category such as N, NP, S, etc. There are at least two conjuncts of the same category, and they are conjoined by a CONJunction. Between the first conjunct and the conjunction, one or more conjuncts can follow, separated by commas. The mother node is the same category as the daughter nodes.

(96) $\text{SCCOORD(CAT)} = \underset{\uparrow\in\downarrow}{\text{CAT}}\ (\underset{\uparrow\in\downarrow}{[\text{COMMA CAT}]+}\ (\text{COMMA}))\ \underset{\uparrow=\downarrow}{\text{CONJ}}\ \underset{\uparrow\in\downarrow}{\text{CAT}}$

(97) gives the f-structure of the phrase *adam ve kadın* 'the man and the woman'. Some of the attributes are nondistributive across the members of the set, instead they have their own attribute value pairs in the set itself. For instance, PERS is a nondistributive attribute, so that two singular nouns can form a coordinate structure which is plural. The outermost f-structure does not have a PRED, but the coordinator is represented in COORD-FORM. `<s` inside the f-structure of *kadın* indicates that *adam* precedes *kadın* in the coordination structure.

(97)
$$\left[\left\{\left[\begin{array}{l}\text{PRED}\quad \text{'adam'}\\ \text{CASE NOM, NUM SG, PERS 3}\end{array}\right],\left[\begin{array}{l}\text{PRED}\quad \text{'kadın'}\\ \langle \text{S}\quad \left[\ \text{'adam'}\right]\\ \text{CASE NOM, NUM SG, PERS 3}\end{array}\right]\right\}\right.$$
$$\left.\begin{array}{l}\text{CASE NOM, COORD +_, COORD-FORM VE, NUM PL, PERS 3}\end{array}\right]$$

As well as standard coordination, Turkish has interesting coordination structures using *suspended affixation* [Kabak, 2007], in which the inflectional features of the last element in a coordination have phrasal scope, that is, all other coordinated constituents have certain default features which are then 'overridden' by the features of the last element in the coordination. A very simple case of such suspended affixation is exemplified in (98a) and (98b). Note that although this is not due to the derivational morphology that we mentioned in Section 3.3, it is due to a more general nature of morphology in which affixes can have phrasal scope.

(98) a. kız adam ve kadın-ı ara-dı
girl.Nom man.Nom and woman-Acc call-Past.3sg
'The girl called the man and the woman.'

b. kız [adam ve kadın]-ı ara-dı
girl.Nom [man.Nom and woman]-Acc call-Past.3sg
'The girl called the man and the woman.'

The f-structure of *adam ve kadını* in (98b) is given in (99). For Turkish, CASE is also one of the nondistributive attributes. The standard coordination rule is modified so that the case of the coordination is the case of the last conjunct if the previous conjuncts are in nominative case. In (99), the CASE of the coordination is `acc` although *adam* has CASE `nom`.

(99)
$$\begin{bmatrix} \left\{ \left[\begin{array}{ll} \text{PRED} & \text{'adam'} \\ \multicolumn{2}{l}{\text{CASE NOM, NUM SG, PERS 3}} \end{array} \right] \right. \\ \left. \left[\begin{array}{lll} \text{PRED} & \text{'kadın'} & \\ \langle \text{S} & \left[\text{'adam'} \right] & \\ \multicolumn{3}{l}{\text{CASE ACC, NUM SG, PERS 3}} \end{array} \right] \right\} \\ \text{CASE ACC, COORD } +_, \text{ COORD-FORM VE, NUM PL, PERS 3} \end{bmatrix}$$

Although it is possible to parse basic coordinated phrases with or without suspended affixation in the current implementation, the grammar lacks a wide coverage of coordinated structures especially for verbal coordination where one or more arguments are shared by the coordinated verbs.

4.1.6 The Date-Time Grammar

Tuba Gümüş at Istanbul Technical University has implemented a date-time grammar for Turkish [Gümüş, 2007], based on our grammar. Her work covers point-in-time expressions, particularly clock-time expressions (*saat 2'de* 'at 2 o'clock', *gecenin üçünde* 'at three (oclock) at night'), days of the week (*Salıları* 'on Tuesdays', *Cuma günü* 'on Friday'), calender dates (*9 Mart 2007* '9th March 2007', *Ekim 19'da* 'on October 19th'), seasons (*yazın* 'in summer', *kış mevsiminde* 'in winter'), and some general phrases (*şimdi* 'now', *dün sabah* 'yesterday morning').

The core of the developed grammar uses our NP rules, hence the implementational approach is parallel to ours. Also the features and templates are based on our version for the sake of consistency. Gümüş added new rules to parse temporal phrases that

are not covered by the NP rules (e.g. a nominative N modifying an N for *dün sabah*). For the date expressions, finite state transducers are introduced. She also semantically marked certain types of words as being temporal with more specific information such as date, clock-time, day, or season.

We then integrated this date-time grammar into our system. The integration process brings about some ambiguity which is solved by introducing OT-marks[6] that help to rank the parser outputs.

4.2 Causatives

Crosslinguistically, causatives can give rise to either biclausal or monoclausal structures and they can be formed either periphrastically or morphologically. In Turkish, causatives are formed morphologically and a natural assumption would be that these morphological formations are monoclausal structures. However, as discussions with respect to morphologically-formed causatives in Japanese [Matsumoto, 1998] have shown, morphological causatives can also give rise to biclausal structures as well.

Previous work on Turkish causatives [Gibson and Özkaragöz, 1981; Aissen and Hankamer, 1980; Knecht, 1986] has been formulated within Relational Grammar (RG) and has arrived at differing conclusions with respect to the monoclausality (*clause union* in terms of RG) of the construction. Knecht [1986] has supported the ideas of Aissen and Hankamer [1980] on a monoclausal structure, whereas Gibson and Özkaragöz [1981] have argued that a biclausal approach is more appropriate. Knecht [1986] gives different RG-based explanations for the evidence Gibson and Özkaragöz [1981] proposed in favor of biclausality.

In this section we reexamine the structural representation of causatives by applying several language dependent tests to decide whether the causative constructions are indeed monoclausal, that is, with a single predicate, or biclausal, that is, with an

[6]The discussion on OT-Marks is given in Section 5.2

embedded clause. The basic data with respect to causative formation in Turkish is provided in Section 3.2. We introduce the possible tests that can be applied to decide whether the causatives are monoclausal or biclausal in Section 4.2.1, with subsections that discuss these tests in more detail. Concluding that the majority of the tests points towards a monoclausal status of Turkish causatives, we present the analysis and implementation in our LFG grammar in Section 4.2.2. We then continue with the explanation and implementation of double causatives in Section 4.2.3.

Most of the research in this section is done in collaboration with Miriam Butt and published in Çetinoğlu et al. [2008].

4.2.1 Causatives: Monoclausal or Biclausal?

There are several language dependent tests to decide whether the causative constructions are monoclausal or biclausal. Butt [2003] uses object agreement, anaphora, and control for Urdu and also gives examples of clitic climbing for French [Rosen, 1989] and cooccurrence of negative polarity items for Korean [Choi, 2002]. Matsumoto [1998] and Yokota [2001] use subject honorification, passivization, pronominal binding, control and adjunct interpretation for Japanese. Yokota [2001] also tests the double-*o* constraint, and *shika-na(i)* 'only-Neg' construction for functional monoclausality. Among these possible tests, five are applicable to Turkish: Passivization, Reflexive Binding, Control, Adjunct Interpretation, and Negative Polarity Items.

For all the tests, the sample sentence is first used in the causative and then in a 'tell' construction to compare and contrast the mono/biclausality of causatives with a clearly biclausal construction [cf. Butt 1995].

Passivization

In the passivization test, the behavior of the object of the base verb is observed when the base verb is first causativized and then passivized. The object of the base verb

can be the subject of the passivized causativized sentence, which indicates that the causative construction is monoclausal. (100a) and (100b) give the base sentence and causativized sentence respectively.

(100) a. süt-ü bütün çocuk-lar-a iç-ir-di
milk-Acc all child-Pl-Dat drink-Caus-Past.3sg
'(S/he) made all children drink the milk.'

b. süt bütün çocuk-lar-a içir-il-di
milk.Nom all child-Pl-Dat drink-Caus-Pass-Past.3sg
'All children were made to drink milk.'

süt 'milk', which is the object of the base verb *iç* 'drink' and also the object of the causativized verb *içir* 'make drink', is the subject of the passivized causativized verb. There is no clausal barrier that prevents the innermost object behave as a subject through the causativization and passivization processes.

The difference can be observed by comparing the causative construction with a 'tell' construction where the 'drink milk' clause is embedded by the 'tell' matrix verb in an infinitive in (101). Here, the embedded object cannot become the subject in the passive version in (101b). Instead, a different construction is used in which the entire infinitive 'drink the milk' functions as the subject of the construction as in (101c). *süt* 'milk' is still the object of the sentence constructing the NP.

(101) a. bütün çocuk-lar-a süt-ü iç-me-leri-ni söyle-di
all child-Pl-Dat milk-Acc drink-Inf-Poss-Acc tell-Past.3sg
'(S/he) told all children to drink the milk.'

b. *süt bütün çocuk-lar-a iç-me-leri söyle-n-di
milk.Nom all child-Pl-Dat drink-Inf-Poss.Nom tell-Pass-Past.3sg
'All children were told to drink the milk.'

c. bütün çocuk-lar-a süt-ü iç-me-leri söyle-n-di
all child-Pl-Dat milk-Acc drink-Inf-Poss.Nom tell-Pass-Past.3sg
'All children were told to drink the milk.'

In sum, data from passivization exhibits a clear difference between causatives and an embedded infinitive as in the 'tell' construction. In particular, in the causative, the "embedded" object can be passivized, indicating that it is in fact an object argument of a monoclausal, albeit complex predication.

Reflexive Binding

Reflexive binding is a further possible test for monoclausality, as reflexives crosslinguistically tend to be clause-bound. However, this test is also tricky, since it may not refer to syntactic boundaries, but operate on semantic grounds. With respect to Japanese, according to Matsumoto [1998] it depends on whether the causative is permissive or coercive; he concludes that the former is biclausal and the latter is monoclausal, but Yokota [2001] claims that regardless of the type of the causative, binding the reflexive pronoun to the causer or the causee is possible.

The sentence in (102a) is similar to Japanese example in (Yokota 2001:7). As can be seen, the reflexive pronoun *kendi* 'self' in Turkish can be bound to both the subject of the base verb, here *Arda*, and the subject of the causativized verb, here *Ali*. We give the tell construction as comparison in (102b). Again, the reflexive pronoun can be bound to both of the subjects.

(102) a. Ali_i Arda'-ya_j $\text{kendi-ni}_{i/j}$ savun-dur-du
Ali.Nom Arda-Dat him(self)-Acc defend-Caus-Past.3sg
'Ali made Arda defend him(self).'

b. Ali_i Arda'-ya_j $\text{kendi-ni}_{i/j}$ savun-ma-sı-nı söyle-di
Ali.Nom Arda-Dat him(self)-Acc defend-Inf-Poss-Acc tell-Past.3sg
'Ali told Arda to defend him(self).'

We go one step further and apply Yokota's [2001] test to see whether there is a distinction between permissive and coercive meanings. (103) introduces the adverb *forcibly* for the coercive meaning. The behavior of *kendi* both for the causative and the tell constructions remains the same for the coercive case.

(103) a. Ali$_i$ zorla Arda'-ya$_j$ kendi-ni$_{i/j}$ savun-dur-du
Ali.Nom forcibly Arda-Dat him(self)-Acc defend-Caus-Past.3sg
'Ali forcibly made Arda defend him(self).'

b. Ali$_i$ zorla Arda'-ya$_j$ kendi-ni$_{i/j}$ savun-ma-sı-nı
Ali.Nom forcibly Arda-Dat him(self)-Acc defend-Inf-Poss-Acc
söyle-di
tell-Past.3sg
'Ali forcibly told Arda to defend him(self).'

Given that the reflexive could be sensitive to logical subjects, rather than synactic subjects [Mohanan, 1994], this test is thus inconclusive with respect to monoclausality in Turkish.

Control

Syntactic control is a well-established crosslinguistic test for subjecthood. It has been used for both Urdu and Japanese causatives. In Urdu, control clauses differ with respect to morphological causatives versus the biclausal 'tell' construction, clearly indicating that causatives are monoclausal [Butt, 2003]. In Japanese, however, the situation is more complex. Matsumoto [1998] uses this test also as an evidence for different types of causatives. Yokota [2001] again argues that this distinction is not applicable for Japanese causatives reanalyzing the examples given in [Matsumoto, 1998]. A similarly complex situation holds in Turkish. (104a) is parallel to examples in Matsumoto [1998].

(104) a. Can$_i$ çocuğ-a$_j$ [PRO$_{i/j}$ televizyon seyred-er-ken] çorap-lar-ı
Can.Nom child-Dat television watch-Aor-While sock-Pl-Acc
giy-dir-di
wear-Caus-Past.3sg
'Can made the child put on the socks while watching TV.'

b. Can$_i$ çocuğ-a$_j$ [PRO$_j$ televizyon seyred-er-ken] çorap-lar-ı
Can.Nom child-Dat television watch-Aor-While sock-Pl-Acc
giy-me-si-ni söyle-di
wear-Inf-Poss-Acc tell-Past.3sg
'Can told the child to put on the socks, while watching TV.'

Turkish patterns similarly to Japanese [Yokota, 2001]. In causative sentences, as in (104a), subject of the control clause can be controlled either by the subject of the base verb or by the agent (logical subject) of the causativized verb. In (104b), on the contrary, the subject of the control clause is controlled by the matrix object only.

Notice that this pattern is independent of word order. Since word order is free in Turkish, the adverbial control clause can be placed in several positions within the sentence. (106) gives all possible placements of the adverbial clause *televizyon seyrederken* 'watching TV' for the sentence in (105). Some of the placements are biased towards *Can* but in all the arrangements either *Can* or the child can be watching TV.

(105) Can çocuğ-a çorap-lar-ı giy-dir-di.
Can child-Dat sock-Pl-Acc wear-Caus-Past.3sg
'Can made the child put on the socks'

(106) a. [PRO$_{i/?j}$ televizyon seyrederken] Can$_i$ çocuğa$_j$ çorapları giydirdi.

b. Can$_i$ [PRO$_{i/j}$ televizyon seyrederken] çocuğa$_j$ çorapları giydirdi.

c. Can$_i$ çocuğa$_j$ [PRO$_{i/j}$ televizyon seyrederken] çorapları giydirdi.

d. Can$_i$ çocuğa$_j$ çorapları [PRO$_{i/j}$ televizyon seyrederken] giydirdi.

e. Can$_i$ çocuğa$_j$ çorapları giydirdi [PRO$_{i/?j}$ televizyon seyrederken].

This word order test is also applied to the biclausal tell construction. (107) gives the basic tell sentence, and items of (108) give the possible phrase ordering. For the tell constructions there is no ambiguity. The subject watching TV is either *Can* ((108a-b,e)) or the child ((108c-d)) unlike the ambiguous cases in causatives. If the adverb is close to the inner clause to be a part of it, then it is the child who is watching TV. Otherwise, the adverb is attached to the verb in the main clause.

(107) Can çocuğ-a$_i$ [PRO$_i$ çorap-lar-ı giy-me-si-ni] söyle-di
Can.Nom child-Dat sock-Pl-Acc wear-Inf-Poss-Acc tell-Past.3sg
'Can told the child to put on the socks.'

(108) a. [PRO$_i$ televizyon seyrederken] Can$_i$ çocuğa$_j$ çorapları giymesini söyledi.

b. Can$_i$ [PRO$_i$ televizyon seyrederken] çocuğa$_j$ çorapları giymesini söyledi.

c. Can$_i$ çocuğa$_j$ [PRO$_j$ televizyon seyrederken] çorapları giymesini söyledi.

d. Can$_i$ çocuğa$_j$ çorapları [PRO$_j$ televizyon seyrederken] giymesini söyledi.

e. Can$_i$ çocuğa$_j$ çorapları giymesini söyledi [PRO$_i$ televizyon seyrederken].

We take it to be significant that the causative and the biclausal 'tell' construction do not pattern in parallel, but show differences.

Adjunct Interpretation

Matsumoto [1998] and Yokota [2001] give examples of adjunct interpretation in discussion of mono/biclausality of Japanese causatives. Whether manner adverbs are interpreted with respect to the base verb or the causativized verb, or both can give us an idea of the structure of the causatives. In (109) the adverb is interpreted with respect to the causer (mother), not the causee (baby), which is taken to be clear evidence for monoclausality.

(109) anne bebeğ-i isteksizce uyu-t-tu
mother.Nom baby-Acc reluctantly sleep-Caus-Past.3sg
'The mother reluctantly made the baby sleep'

If we want to say that the baby is sleeping reluctantly we cannot use an adverb to express it. Instead, we can use an adjective as in (110).

(110) anne isteksiz bebeğ-i uyu-t-tu
mother.Nom reluctant baby-Acc sleep-Caus-Past.3sg
'The mother made the reluctant baby sleep.'

Now let us compare the causative data with that of the biclausal 'tell' construction in (111). As can be seen, there are more interpretive possibilities, as the adverb 'reluctantly' can apply either within the matrix clause (the mother was reluctant) or the

embedded clause (the sleeping of the baby was reluctant). We thus again have a clear contrast between the causative and a biclausal construction.

(111) a. anne bebeğ-e isteksizce uyu-ma-sı-nı söyle-di
mother.Nom baby-Dat reluctantly sleep-Inf-Poss-Acc tell-Past.3sg
'The mother told the baby to sleep reluctantly.'
'The mother reluctantly told the baby to sleep.'

b. anne bebeğ-e uyu-ma-sı-nı isteksizce söyle-di
mother baby-Dat sleep-Inf-Poss-Acc reluctantly tell-Past.3sg
'The mother reluctantly told the baby to sleep.'

Recall that in Japanese coercive and permissive causatives patterned differently. In Turkish, the coercive reading generally is the default interpretation for causatives. However, with respect to some verbs, the permissive meaning is more frequent than the coercive one. An example is provided in (112) and (113) checks on the adjunct interpretation in this sentence.

(112) bisiklet-in-i kullan-dır-ır mı-sın?
bicycle-P2sg-Acc use-Caus-Aor Ques-2sg
'Would you let (me) ride your bicycle?'

(113) bisiklet-in-i ban-a sessizce kullan-dır-dı
bicycle-P3sg-Acc I-Dat quietly use-Caus-Past.3sg
'He let me quietly ride his bicycle.'
'He quietly let me ride his bicycle.'

Unlike with the coercive causative, a permissive reading thus seems to allow an ambiguous interpretation along the lines of a biclausal. The second interpretation in (113) is more probable, but both are possible. However, if we use a different adverb, the ambiguity vanishes. Consider (114a) and (114b) with the adverbs 'forcibly' and 'reluctantly', respectively.

(114) a. bisiklet-in-i ban-a zorla kullan-dır-dı
bicycle-P3sg-Acc I-Dat forcibly use-Caus-Past.3sg
'He forcibly let me ride his bicycle.'

b. bisiklet-in-i ban-a isteksizce kullan-dır-dı
bicycle-P3sg-Acc I-Dat reluctantly use-Caus-Past.3sg
'He reluctantly let me ride his bicycle.'

These examples show that the interplay between adverbial meaning, lexical semantics and context is rather complex and that the data must be treated with care. However, the central contrast in (109) vs. (111) would seem to indicate that causatives differ from biclausal structures.

Negative Polarity Items

We now turn to the last test and one that has been proven to be quite robust as a test for monoclausality, namely negative polarity [cf. Choi 2005]. The scope of a negative polarity item tends to be clause-bound. In Turkish this plays out as follows: the pronoun *hiç kimse* 'anybody' in conjunction with the negative suffix *-mA* means *nobody* ((115)).

(115) a. hiç kimse kestane yedi mi
anybody.Nom chestnut.Nom eat.Past Ques
'Did anybody eat chestnuts?'

b. hiç kimse kestane yemedi
anybody.Nom chestnut.Nom eat.Neg.Past
'Nobody ate chestnuts.'

c. *hiç kimse kestane yedi
anybody.Nom chestnut.Nom eat-Past.3sg
'Anybody ate chestnuts.'

(116) gives a causative sentence with *hiç kimse*. The negative pronoun and the negative suffix should be in the same clause therefore this example favors monoclausal constructions.

(116) hiç kimse Cem'-e kestane ye-dir-me-di
anybody.Nom Cem-Dat chestnut-Nom eat-Caus-Neg-Past.3sg
'Nobody let Cem eat chestnuts.'
'Nobody fed Cem with chestnuts.'

We can see the difference better by using the same items in a tell construction as in (117). In (117a) *hiç kimse* and the negative marker on the verb are in the same clause, so the sentence is grammatical, but (117b) exemplifies an ungrammatical sentence where *hiç kimse* is used in the matrix verb and *-mA* negates the verb of the inner clause.

(117) a. hiç kimse Cem'-e kestane ye-me-si-ni söyle-me-di
anybody.Nom Cem-Dat chestnut-Nom eat-Inf-Poss-Acc tell-Neg-Past.3sg
'Nobody told Cem to eat chestnuts'

b. *hiç kimse Cem'-e kestane ye-me-me-si-ni söyle-di
anybody.Nom Cem-Dat chestnut-Nom eat-Neg-Inf-Poss-Acc tell-Past.3sg
'Nobody told Cem not to eat chestnuts'

If we use *hiç kimse* in another role within the sentence, as in (118), the same pattern as in (117) is observed. This is the expected result, the person who is told to (here *hiç kimse*) is not a part of the embedded clause.

(118) a. Cem hiç kimse-ye kestane ye-me-si-ni söyle-me-di
Cem.Nom anybody-Dat chestnut-Nom eat-Inf-Poss-Acc tell-Neg-Past.3sg
'Cem told nobody to eat chestnuts'

b. *Cem hiç kimse-ye kestane ye-me-me-si-ni söyle-di
Cem.Nom anybody-Dat chestnut-Nom eat-Neg-Inf-Poss-Acc tell-Past.3sg
'Cem told nobody not to eat chestnuts'

So in order to test the behavior of *anybody* as a part of the embedded clause, (119) is introduced. In (119a) negation is in the matrix sentence but *hiç kimse* is in the embedded clause, therefore it is ungrammatical as expected.[7] Satisfying the *same clause* rule, (119b) is grammatical. (119c) is also grammatical; once the inner clause has both the negation and negative polarity item we can negate the matrix verb as well.

(119) a. *Cem Ayşe'ye hiç kimse-yi öp-me-si-ni söyle-me-di
Cem.Nom Ayşe-Dat anybody-Acc kiss-Inf-Poss-Acc tell-Neg-Past.3sg
'Cem didn't tell Ayşe to kiss nobody.'

[7]Actually (119a) is grammatical when we interpret it as 'Cem didn't tell Ayşe 'go and kiss that person' it is Ayşe who decided to kiss'. But we think this is not what we are looking for in the tests.

b. Cem Ayşe'ye hiç kimse-yi öp-me-me-si-ni söyle-di
Cem.Nom Ayşe-Dat anybody-Acc kiss-Neg-Inf-Poss-Acc tell-Past.3sg
'Cem told Ayşe to kiss nobody.'

c. Cem Ayşe'ye hiç kimse-yi öp-me-me-si-ni söyle-me-di
Cem.Nom anybody-Dat chestnut-Nom kiss-Neg-Inf-Poss-Acc tell-Past.3sg
'Cem didn't tell Ayşe to kiss nobody.'

The following examples give the causative forms parallel to (118) and (119). Especially (121) provides a good evidence in contrasting the biclausal construction in (119a).

(120) Cem hiç kimse-ye kestane ye-dir-me-di
Cem.Nom nobody-Dat chestnut.Nom eat-Caus-Neg-Past.3sg
'Cem let nobody eat chestnuts.'
'Cem didn't feed anybody with chestnuts.'

(121) Cem Ayşe'ye hiç kimse-yi öp-tür-me-di
Cem.Nom Ayşe-Dat nobody-Acc kiss-Caus-Neg-Past.3sg
'Cem didn't let Ayşe kiss anybody.'

Thus, the interaction of causatives with negative polarity again demonstrates that causatives do not pattern along the lines of a biclausal construction.

Though, there is a drawback of the monoclausal representation when the causative sentence is negative. In her thesis [1993] Göksel uses examples like (122) to indicate the scope of negation and to prove that [[V+neg]caus] is semantically possible in Turkish. Our implementation only represents one of the interpretations. We believe the answer to represent both of the interpretations could only be found by considering the representation of negation as well.

(122) Ali Cem'-e kestane ye-dir-me-di
Ali Cem-Dat chesnut-Nom eat-Caus-Neg-Past.3sg
'Ali didn't let Cem eat chesnuts'
'Ali didn't feed Cem with chesnuts'

Summary

The results of the tests are mixed: some of the tests completely favor monoclausality, whereas some others provide counterexamples for representation with a single predicate. The Passivization test clearly shows the distinction between the causative structures and the biclausal 'tell' constructions, and favors monoclausality. The Reflexive Binding test supports biclausal structure but this might be due to logical subjects. The Control test seems to give evidence for biclausality if we only consider the causative example but a comparison with the tell construction clearly demonstrates a distinction. The Adjunct Interpretation, on the other hand, favors monoclausal structures in almost all cases but there are few ambiguous interpretations. This problem again, targets the semantic interpretation of the adjuncts rather than the clausal structure, hence does not completely negate our conclusions. Another test that clearly favors monoclausality is the use of Negative Polarity Items. The result of these observations leads us to assume a monoclausal structure.

4.2.2 Implementation in Lexical Functional Grammar

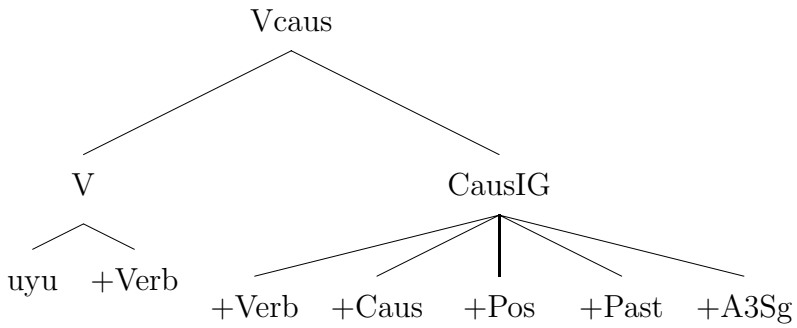
Having established that Turkish causatives are best analyzed as monoclausal, we now turn to their representation. Modeling a monoclausal structure in which two predicates (in our case the main verb and the causative morphology) merge to a predicate as a single unit is tricky because the analysis involves argument structure merger. Within LFG, argument structure merger can be effected in various ways. In terms of our actual implementation, we use the Restriction Operator [Kaplan and Wedekind, 1993] and base ourselves on the approach suggested by [Butt and King, 2006] for Urdu causatives.

As can be seen in our analysis in (131), the complex causative predication is represented as a monoclausal structure, that is, as a flat f-structure with no embeddings. The way we arrive at this analysis is complex and works as follows. For one, we assume a base f-structure as in (130), which is combined with the predicative information of the causative morpheme. That is, there are two morphemes containing the predicative

information of a causativized verb: the verb stem and the causative suffix. These two predicates are merged to form the new complex predicate by substituting in the argument structure of the verb stem into one of the arguments of the causative morpheme.

(123) illustrates the sublexical tree representation of a causativized verb. The morphological output `uyu+Verb^DB+Verb+Caus+Pos+Past+A3sg` of *uyuttu* 'made sleep' splits into two IGs by the derivational boundary ^DB .

(123)



The lexical entries for the intransitive verb *uyu* and the causative suffix are given in (124a) and (124b) respectively. The second argument `%PRED2 of` the causative suffix is a local variable that will be filled in by the predicate of the base verb. As can be seen in (124b), the causative suffix has a two place predicate where the first argument is the causer and the second argument is the event that is caused. The verb stem in our case has only one argument ((124a)). When this information is substituted in for `%PRED2` in (124b), the number of arguments of the base verb is preserved. However, the nature of the arguments themselves are altered.

(124) a. (↑ PRED) = 'uyu<(↑ SUBJ)>'

b. (↑ PRED) = 'caus<(↑ SUBJ), %PRED2>'

(125) gives the semantic representations of the main verb and its causativized form, and the mapping of arguments. For intransitive verbs as in (124a), the SUBJect of the base verb becomes the OBJect of the merged structure.

(125) uyu⟨SUBJ⟩ caus⟨SUBJ, uyu⟨OBJ⟩⟩

The XLE code snippet in (126) is the part of the causative rule that handles intransitive verbs. The equation ↑=↓ under CausIG states that the f-structure of CausIG (↓, current node) is unified with the f-structure of Vcaus (↑, mother node), therefore all the features of CausIG, including its predicate is passed to Vcaus. The equation under V also unifies the current node with the mother node but this time some of the features are restricted out not to carry those features to the mother node and instead to construct the merged structure. The predicate of the mother node is '`caus⟨SUBJ, %PRED2⟩`' coming from the causative morpheme. Therefore the PRED features should be excluded during the unification. The subject of the main verb (↓) is the object of the complex predicate (↑) ((↓ SUBJ) = (↑ OBJ)) and the new subject of the complex predicate will be filled in by a phrase other than the arguments of the main verb. Therefore SUBJ and OBJ should also be excluded from the equation ↑=↓, i.e., ↑ \PRED\SUBJ\OBJ= ↓ \PRED\SUBJ\OBJ. The constraint (↓ PRED) = (↑ PRED ARG2) states that the second argument of PRED of the mother node, here Vcaus, is the PRED of the current node, here V.

(126)

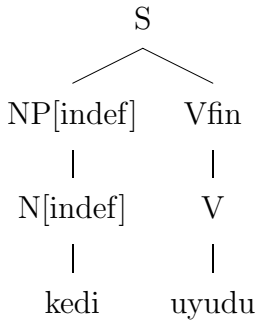
Vcaus →	V	CausIG
	↑ \PRED\SUBJ\OBJ= ↓ \PRED\SUBJ\OBJ	↑ = ↓
	(↓ SUBJ) = (↑ OBJ)	
	(↓ PRED) = (↑ PRED ARG2)	

We revisit (31) to give its implementation. C-structures corresponding to the base sentence in (127a) and its causativized form in (127b) are given in (128) and (129) respectively. In accordance with our analysis of the basic sentences, causatives also have a flat structure in order to account for the possibility of free word order.

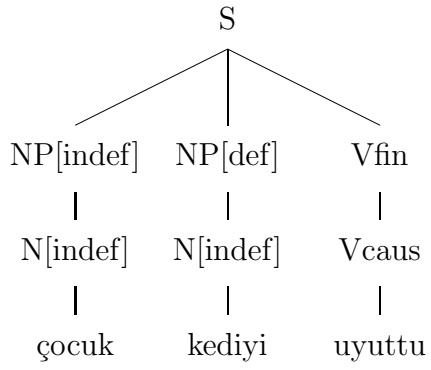
(127) a. kedi uyu-du
cat.Nom sleep-Past.3sg
'The cat slept.'

b. çocuk kedi-yi uyu-t-tu
child.Nom cat-Acc sleep-Caus-Past.3sg
'The child made the cat sleep.'

(128)



(129)



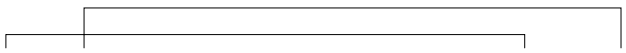
F-structures (130) and (131) show the initial representation of the base sentence and the resulting structure after causativization. The former subject *kedi* 'cat'in nominative case is the object in accusative case when causativized. The subject of the new sentence is *çocuk* 'child'.

(130)
$$\begin{bmatrix} \text{PRED} & \text{'uyu⟨kedi⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'kedi'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{TENSE} & \text{past} \end{bmatrix}$$

(131)
$$\begin{bmatrix} \text{PRED} & \text{'caus⟨çocuk, uyu⟨kedi⟩⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'çocuk'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'kedi'} \\ \text{CASE} & \text{acc} \end{bmatrix} \\ \text{TENSE} & \text{past} \end{bmatrix}$$

When the verb in question is transitive, the lexical entry has a subject and an object argument as exemplified for *kovala* 'chase' in (132a). The merged structure in (132b) reflects the new functions assigned after the causativization process. For transitive verbs, the subject of the base verb becomes the thematic object (OBJ_θ) of the merged structure; the object remains the same.

(132) a. (↑ PRED) = 'kovala<(↑ SUBJ),(↑ OBJ)>'



b. kovala⟨SUBJ, OBJ⟩ caus⟨SUBJ, kovala⟨OBJ-TH, OBJ⟩⟩

(133) gives the disjunction of the XLE causative rule handling transitive verbs.

The mapping in (132b) is encoded by using the restriction operator and constraints. The f-structure of V is unified with the f-structure of the mother node, that is Vcaus, without the SUBJ and OBJ-THfeatures. The PRED is also restricted out to be able to place the merged predicate in the mother f-structure. The SUBJ of the mother node comes from the causativized sentence. The equation (↓ SUBJ)= (↑ OBJ-TH) places the SUBJ of the base verb V(↓) as the OBJ-THof the causativized verb Vcaus(↑). The rule makes sure that the verb is transitive with the existential constraint (↓ OBJ). As in the intransitive rule disjunct, the constraint (↓ PRED) = (↑ PRED ARG2) states that the second argument of the PRED of Vcaus will be filled by the PRED of V.

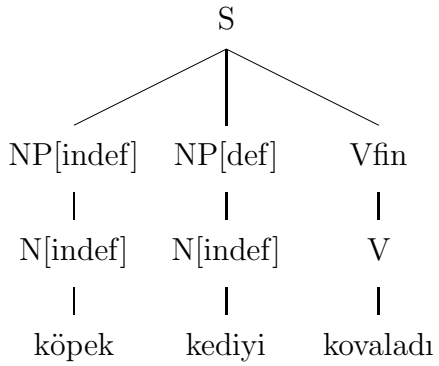
(133) Vcaus → V CausIG
↑ \SUBJ\OBJ-TH\PRED = ↓\SUBJ\OBJ-TH\PRED ↑ = ↓
(↓ SUBJ)= (↑ OBJ-TH)
(↓ OBJ)
(↓ PRED) = (↑ PRED ARG2)

For the transitive verbs, we present the sentences in (32) once more, as (134). The c-structures of transitive verbs have no representational difference from intransitive ones. The c-structures for (134a) and its causativized form (134b) are given in (135) and (136).

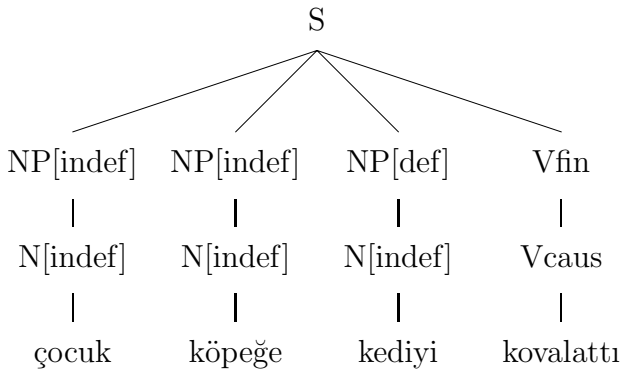
(134) a. köpek kedi-yi kovala-dı
dog.Nom cat-Acc chase-Past.3sg
'The dog chased the cat.'

b. çocuk köpeğ-e kedi-yi kovala-t-tı
child.Nom dog-Dat cat-Acc chase-Caus-Past.3sg
'The child made the dog chase the cat.'

(135)



(136)



(137) and (138) give the corresponding f-structures of (135) and (136), respectively. *kediyi* 'cat (acc)', the object of the first sentence, preserves its case and function whereas the nominative subject *köpek* 'dog' becomes a dative thematic object when the causativization occurs. The subject of the new sentence is *çocuk* 'child'.

(137)
$$\begin{bmatrix} \text{PRED} & \text{'kovala⟨köpek,kedi⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'köpek'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'kedi'} \\ \text{CASE} & \text{acc} \end{bmatrix} \\ \text{TENSE} & \text{past} \end{bmatrix}$$

(138)
$$\begin{bmatrix} \text{PRED} & \text{'caus⟨çocuk, kovala⟨köpek,kedi⟩⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'çocuk'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'kedi'} \\ \text{CASE} & \text{acc} \end{bmatrix} \\ \text{OBJTH} & \begin{bmatrix} \text{PRED} & \text{'köpek'} \\ \text{CASE} & \text{dat} \end{bmatrix} \\ \text{TENSE} & \text{past} \end{bmatrix}$$

Following these implementation basics, we show the f-structures of the example sentences used in the mono/bi-clausality tests. The f-structure of the sentence *Ali Arda'ya*

kendini savundurdu 'Ali made Arda defend him(self).' given in (102a) is depicted in (139). The reflexive pronoun *kendi* 'self' is not bound to any of the subjects in the implementation.

(139)
$$\begin{bmatrix} \text{PRED} & \text{'caus⟨Ali, savun⟨Arda,kendi⟩⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'Ali'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'kendi'} \\ \text{CASE} & \text{acc} \end{bmatrix} \\ \text{OBJTH} & \begin{bmatrix} \text{PRED} & \text{'Arda'} \\ \text{CASE} & \text{dat} \end{bmatrix} \\ \text{TENSE} & \text{past} \end{bmatrix}$$

The control test sentence (104a) *Can çocuğa televizyon seyrederken çorapları giydirdi.* 'Can made the child put on the socks while watching TV.' has the f-structure in (140). The subject of the inner clause is `'null_pro'` and there is no co-indexation, neither with *Can* nor with *çocuk* 'child' in the implementation.

(140)
$$\begin{bmatrix} \text{PRED} & \text{'caus⟨Can, giy⟨çocuk, çorap⟩⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'Can'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'çorap'} \\ \text{CASE} & \text{acc} \end{bmatrix} \\ \text{OBJTH} & \begin{bmatrix} \text{PRED} & \text{'çocuk'} \\ \text{CASE} & \text{dat} \end{bmatrix} \\ \text{ADJUNCT} & \begin{bmatrix} \text{PRED} & \text{'seyret⟨null_pro, televizyon⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'null_pro'} \\ \text{CASE} & \text{nom} \end{bmatrix} \quad \text{OBJ} \quad \begin{bmatrix} \text{PRED} & \text{'televizyon'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{ADJUNCT-TYPE} & \text{sub} \end{bmatrix} \\ \text{TENSE} & \text{past} \end{bmatrix}$$

(141) gives the f-structure for *bisikletini bana sessizce kullandırdı* 'He quietly let me ride his bicycle.' given in (113). The second interpretation 'He let me quietly ride his bicycle.' is lost due to the monoclausal representation.

(141)
$$\begin{bmatrix} \text{PRED} & \text{'caus⟨o, kullan⟨ben,bisiklet⟩⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'o'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'bisiklet'} \\ \text{CASE} & \text{acc} \end{bmatrix} \\ \text{OBJTH} & \begin{bmatrix} \text{PRED} & \text{'ben'} \\ \text{CASE} & \text{dat} \end{bmatrix} \\ \text{ADJUNCT} & \begin{bmatrix} \text{PRED} & \text{'sessizce'} \end{bmatrix} \\ \text{TENSE} & \text{past} \end{bmatrix}$$

Finally, the f-structure of *hiç kimse Cem'e kestane yedirmedi* 'Nobody let Cem eat chestnuts.' given in (116) is shown in (142). Negation is represented in the main clause level in the structure, hence the representation covers the second interpretation in (116) as well.

(142)
$$\begin{bmatrix} \text{PRED} & \text{'caus⟨hiç kimse, ye⟨Cem,kestane⟩⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'hiç kimse'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'kestane'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{OBJTH} & \begin{bmatrix} \text{PRED} & \text{'Cem'} \\ \text{CASE} & \text{dat} \end{bmatrix} \\ \text{TENSE PAST, NEG} & + \end{bmatrix}$$

4.2.3 Double Causatives

Double causativization of verbs is frequently used in Turkish, especially if the verb is intransitive. We revisit the examples we have given in Section 4.2, to analyze the representation further and to present the details of our implementation.

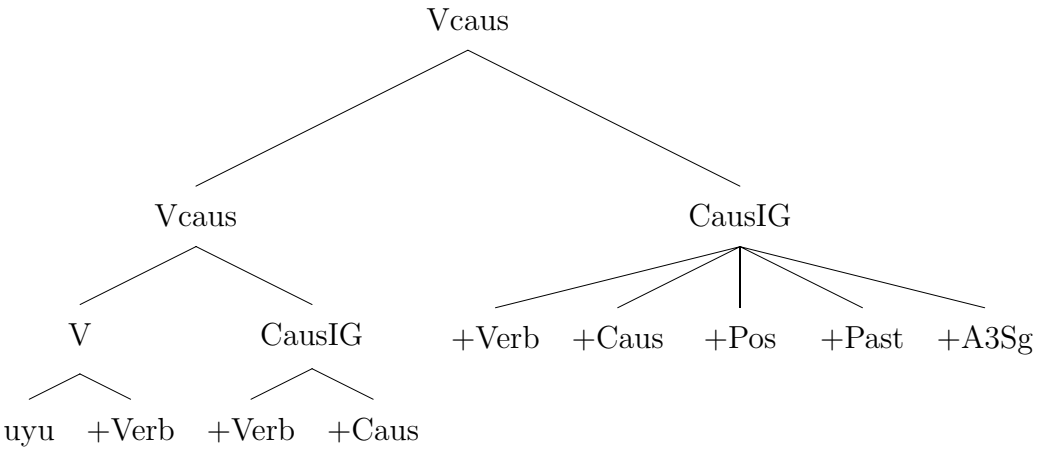
(143a) and (143b) demonstrate the example sentences for the double causativization of intransitives. Once an intransitive verb is causativized, the resulting predicate '`caus⟨SUBJ, pred⟨OBJ⟩⟩`' bears the grammatical functions of a canonical transitive. Therefore it will be parsed without any need for modifications in the grammar rules.

(143) a. çocuk kedi-yi uyu-t-tu
child.Nom cat-Acc sleep-Caus-Past.3sg
'The child made the cat sleep.'

b. anne çocuğ-a kedi-yi uyu-t-tur-du
mother.Nom child-Dat cat-Acc sleep-Caus-Caus-Past.3sg
'The mother made the child make the cat sleep.'

The c-structures of double causativized intransitives are usual flat trees but the sublexical tree of the verb is interesting in that it has a hierarchical structure. (144) depicts the sublexical tree of *uyutturdu* 'made someone make sleep'.

(144)



Causativizing the same verb for the second time ends up with an f-structure parallel to the single causativization of transitives. In (145), we repeat (131) as the f-structure

of the base sentence (143a), and give the f-structure of the causativized sentence (143b) in (146).

(145)
$$\begin{bmatrix} \text{PRED} & \text{‘caus⟨çocuk, uyu⟨kedi⟩⟩’} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{‘çocuk’} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{‘kedi’} \\ \text{CASE} & \text{acc} \end{bmatrix} \\ \text{TENSE} & \text{past} \end{bmatrix}$$

(146)
$$\begin{bmatrix} \text{PRED} & \text{‘caus⟨anne, caus⟨çocuk, uyu⟨kedi⟩⟩⟩’} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{‘anne’} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{‘kedi’} \\ \text{CASE} & \text{acc} \end{bmatrix} \\ \text{OBJ-TH} & \begin{bmatrix} \text{PRED} & \text{‘çocuk’} \\ \text{CASE} & \text{dat} \end{bmatrix} \\ \text{TENSE} & \text{past} \end{bmatrix}$$

Double causativization of transitives, however, is controversial. A single causativization example along with two double causativization examples are given in (147). As exemplified in (147b), it is not considered to be grammatical to overtly state both of the intermediaries between the agent and the theme of the event. Unlike (147b), the sentence in (147c) is grammatical when one of the intermediaries is covert. But then, the ranking is ambiguous although it is certain that somebody else is involved in the causation hierarchy. We give both possible interpretations in (147c).

(147) a. çocuk köpeğ-e kedi-yi kovala-t-tı
child.Nom dog-Dat cat-Acc chase-Caus-Past.3sg
‘The child made the dog chase the cat.’

b. *çocuğ-a köpeğ-e kedi-yi kovala-t-tır-dı
child-Dat dog-Dat cat-Acc chase-Caus-Caus-Past.3sg
‘S/he made the child make the dog chase the cat.’

c. çocuk köpeğ-e kedi-yi kovala-t-tır-dı
child.Nom dog-Dat cat-Acc chase-Caus-Caus-Past.3sg
‘The child made someone make the dog chase the cat.’
‘The child made the dog make someone chase the cat.’

Dede [1984] explains the ungrammaticality of (147b) with a constraint against two derived datives. The sentence is ungrammatical when the datives are derived from former subjects of the base and single causativized verb. To support her argument,

she provides grammatical examples of causatives with two overt dative noun phrases in (148) where one of the datives is originally dative in the base sentence (cf. also Zimmer 1976, Kornfilt 1997:332, Göksel 1993:216).

(148) a. ben para-yı çocuğ-a ver-di-m
I.Nom money-Acc child-Dat give-Past-1sg
'I gave the money to the child.'

b. ban-a para-yı çocuğ-a ver-dir-di
I-Dat money-Acc child-Dat give-Caus-Past.3sg
'S/he made me give the money to the child.' [Dede, 1984]

Note that (147c) is not the causativized form of (147a). We have to introduce an agent and omit either *çocuk* 'child' or *köpek* 'dog' from the sentence in favor of grammaticality to get the causativized (147a). Instead, the agent *çocuk* of (147a) preserves its function and another intermediary is introduced in (147c). Göksel and Kerslake consider some instances of the second causative as an emphasizer so that there are no more valency alternations in the causativized verb [Göksel, 1993; Göksel and Kerslake, 2005]. According to them, (149a) and (149b) are identical in meaning. We believe (149b) would include an interpretation with an intermediary as well and treat all double causatives the same in our implementation.

(149) a. saç-ım-ı kes-tir-di-m
hair-P1sg-Acc cut-Caus-Past-1sg
'I had my hair cut.'

b. saç-ım-ı kes-tir-t-ti-m
hair-P1sg-Acc cut-Caus-Caus-Past-1sg
'I had my hair cut.' [Göksel and Kerslake 2005:148]

Now let us examine how we can represent double causative sentences like (147c) in LFG. Recall the three place predicate of intransitives after the first causativization (cf. (132b)); this time, with one more argument, the predicate will be four place and the intermediary will be represented by the special symbol NULL to indicate the absence

of a grammatical function. (150) illustrates the two possible predicates of the verb in (147c).

(150) a. caus⟨SUBJ, caus⟨NULL, kovala⟨OBJ-TH, OBJ⟩⟩⟩

b. caus⟨SUBJ, caus⟨OBJ-TH, kovala⟨NULL, OBJ⟩⟩⟩

The grammar needs an additional causative rule to handle double causatives of transitives. The methodology is the same but we introduce another causative predicate with **NULL** as one of the arguments ((151)) and a CHECK feature **_DOUBLE-CAUS** to control second causativization.

(151) (↑ PRED) = 'caus<NULL, %PRED2>'

Additional contraints to (133) are given in boldface in (152) and the rule for the second causative is given in (153). Briefly, we force the inner causative to have an argument structure with a **NULL** if it will have the outer causative, and force the double causative not to have a **NULL** instead of a SUBJ, that is, pick the lexical representation in (124b) instead of (151).

(152)	Vcaus →	V	CausIG
		↑ \SUBJ\OBJ-TH\PRED = ↓\SUBJ\OBJ-TH\PRED	↑ = ↓
		(↓ SUBJ)= (↑ OBJ-TH)	
		(↓ OBJ)	
		(↑ CHECK _DOUBLE-CAUS)	
		(↑ PRED ARG1) = NULL	
		(↓ PRED) = (↑ PRED ARG2)	

(153)	Vdoublecaus →	Vcaus	CausIG
		↑ \SUBJ\PRED = ↓\SUBJ\PRED	↑ = ↓
		(↑ CHECK _DOUBLE-CAUS)= +	
		(↑ PRED ARG1) ∼= NULL	
		(↓ PRED) = (↑ PRED ARG2)	

The simplified f-structure of (147c) is depicted in (154). *kedi* 'cat' is chased by *köpek* 'dog', and *çocuk* 'child' is the agent that starts the causation. The intermediary person between the child and dog is not explicit in the sentence, hence is represented as NULL in the f-structure.

(154)
$$\begin{bmatrix} \text{PRED} & \text{'caus⟨çocuk, caus⟨NULL, kovala⟨köpek,kedi⟩⟩⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'çocuk'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'kedi'} \\ \text{CASE} & \text{acc} \end{bmatrix} \\ \text{OBJTH} & \begin{bmatrix} \text{PRED} & \text{'köpek'} \\ \text{CASE} & \text{dat} \end{bmatrix} \\ \text{TENSE} & \text{past} \end{bmatrix}$$

4.3 Passives

As briefly introduced in Section 3.2, passivization is also realized morphologically in Turkish. We have given a canonical passivization example in (34), p.23 where the accusative object becomes nominative subject when passivized. We now continue with impersonal and double passivization discussed in Section 4.3.1 and then provide our implementation in Section 4.3.2. The analysis and implementation of passivized causatives are presented in Section 4.3.3.

4.3.1 Impersonal and Double Passives

In Turkish, it is possible to passivize intransitives with constituents other than direct object, as in (155) and (156). In those cases, passivization is impersonal, that is, the constituent preserves its function (and also its case marking) and there is no subject in the passivized sentence. Kornfilt [1997] shows such passives are impersonal by stating

the two properties that do not obey the subjecthood rules: the constituent is not in nominative case ((155b) and (156b)) and it does not agree with the verb in person and number ((156b) and (156c)). Still, we can derive a participle from the passivized sentence and extract the constituent in the same way as subject, as in (155c).

(155) a. Ali okul-a git-ti
Ali.Nom school-Dat go-Past.3sg
'Ali went to the school.'

b. okul-a gid-il-di
school-Dat go-Pass-Past.3sg
'The school was gone to.'

c. gid-il-en okul
go-Pass-Prespart school.Nom
'the school that was gone to'

(156) a. Ali san-a git-ti
Ali you-Dat go-Past.3sg
'Ali went to you.'

b. san-a gid-il-di
you-Dat go-Pass-Past.3sg
'You were gone to.'

c. *sen/san-a gid-il-di-n
you.Nom/you-Dat go-Pass-Past-2sg
'You were gone to. (intended meaning)'

When the constituent is used with a transitive verb, instead of an intransitive one as in (157a), the object becomes the subject as expected ((157b)), and the behavior of the constituent in the participle construction changes. We can see the object-like behavior in (157c). The derivation in (157d), which is parallel to (155c) in terms of the participle suffix, is now ungrammatical with a transitive verb.

(157) a. Ali'-yi okul-a götür-dü-m
Ali-Acc school-Dat take-Past-1sg
'I took Ali to the school.'

b. Ali okul-a götür-ül-dü
Ali.Nom school-Dat take-Pass-Past.3sg
'Ali was taken to the school.'

c. Ali'nin götür-ül-düğü okul
Ali-Gen take-Pass-Pastpart school.Nom
'the school that Ali was taken to'

d. *Ali götür-ül-en okul
Ali.Nom take-Pass-PresPart school.Nom
'the school that Ali was taken to' (intended meaning)

(158) illustrates another examples of an impersonal passive. To prevent a confusion that might arise, (158) is only given to show the passivization of an intransitive verb in terms of syntactic and morphological modifications; it does not necessarily mean that (158b) is the passive form of the 1st person singular verb in (158a). In all cases of impersonal passivization, the agent is uncertain, yet can be identified as a group of people, not a single person.

(158) a. ben ev-de uyu-du-m
I.Nom home-Loc sleep-Past-1sg
'I slept at home.'

b. ev-de uyu-n-du
home-Loc sleep-Pass-Past.3sg
'It was slept at home.'

When the 'group' meaning is intended in the sentence, transitive verbs can also be impersonally passivized by using double passivization on transitives. (159) gives two double passivized sentences, both having the meaning that the actions are taken together with a group. It may also contain the generic meaning, as exemplified in (160).[8]

(159) a. film izle-n-il-di
movie.Nom watch-Pass-Pass-Past.3sg
'The movie was watched.'

[8]The single passivization of the sentences in (159) can be assumed to have the same interpretation with the double passivization, but (160) does not have such a parallelism.

b. tatlı-lar ye-n-il-di
Dessert-Pl.Nom eat-Pass-Pass-Past.3sg
'Desserts were eaten.'

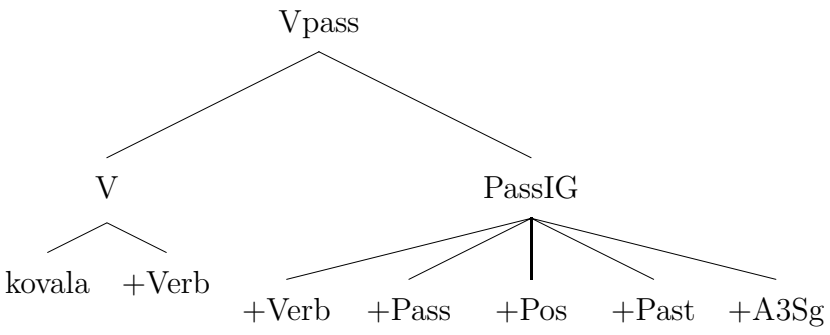
(160) harp-te vur-ul-un-ur
war-Loc shoot-Pass-Pass-Aor.3sg
'One is shot (by one) in war' [Özkaragöz, 1986]

In the following section, we give the implementation of different passivization types.

4.3.2 Implementation in Lexical Functional Grammar

The basic passivization is handled with the standard lexical rule that takes an OBJ and makes it a SUBJ.[9] The passivization information is carried in the morphological tag `+Pass` and is represented as the feature-value pair [PASSIVE +]. The sublexical tree in (161) illustrates the representation of the passivized form of *kovala* 'chase'.

(161)



(162) gives the lexical entry for the same verb, which is modified in order to handle passivization. Now the basic lexical entry, also given in (132a), is the argument of the template @PASS. There are no separate lexical entries for a verb and its passive form in the lexicon.

(162) @(PASS (↑ PRED)='kovala⟨(↑ SUBJ), (↑ OBJ)⟩').

[9]We slightly modified the version available at http://www2.parc.com/isl/groups/nltt/xle/doc /notations.html

The template @PASS in (163) leaves its argument SCHEMATA as is, when there is no passivization. The SCHEMATA is the PRED schema of the verb. If the verb is passivized, the passive morpheme inserts an + as the value of the PASSIVE feature. Since the constraint (↑ PASSIVE)=c + is satisfied, the second disjunct of the passive rule is selected during the parse. The OBJect of the PRED becomes SUBJect by the lexical rewrite rule (↑ OBJ)→(↑ SUBJ) and the SUBJect is replaced with an OBLique AGent if a by-phrase is present, otherwise it will only be represented as NULL. The resulting PRED schema is either (↑ PRED)='kovala⟨(↑ OBL-AG), (↑ SUBJ)⟩' or (↑ PRED)='kovala⟨NULL, (↑ SUBJ)⟩', respectively. Then, the governing functions are filled in by the appropriate noun phrases by using the standard sentence rule.

(163) PASS(SCHEMATA)= { SCHEMATA (↑ PASSIVE)=c -
| SCHEMATA
(↑ PASSIVE)=c +
(↑ OBJ)→(↑ SUBJ)
{ (↑ SUBJ)→(↑ OBL-AG)
|(↑ SUBJ)→ NULL}}.

The different outputs of this rule are illustrated in (164) and (165), which provide simplified f-structures of the sentences *köpek beni kovaladı* 'The dog chased me.' (also in (34a)) and *ben kovalandım* 'I was chased.' (also in (34b)) respectively.

(164)
$$\begin{bmatrix} \text{PRED} & \text{'kovala⟨köpek, ben⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'köpek'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'ben'} \\ \text{CASE} & \text{acc} \end{bmatrix} \\ \text{TENSE} & \text{past} \end{bmatrix}$$

(165)
$$\begin{bmatrix} \text{PRED} & \text{'kovala⟨NULL, ben⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'ben'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{TENSE PAST, PASSIVE} & + \end{bmatrix}$$

A similar rule is applied for the implementation of impersonal passivization. (166) gives the f-structure for the impersonally passivized sentence in (158b).

(166)
$$\begin{bmatrix} \text{PRED} & \text{`uyu}\langle\text{NULL}\rangle\text{'} \\ \text{ADJUNCT} & \begin{bmatrix} \text{PRED} & \text{`ev'} \\ \text{CASE} & \text{loc} \end{bmatrix} \\ \text{TENSE PAST, PASSIVE} & + \end{bmatrix}$$

In cases of double passivization, the morphological analyzer produces the analysis of a single passivized verb although there are two passive morphemes in the surface level. *ye-n-di* (eat-Pass-Past.3sg) and *ye-n-il-di* (eat-Pass-Pass-Past.3sg) have the same morphological output `ye+Verb^DB+Verb+Pass+Pos+Past+A3sg`.

In our implementation, we accept the second passivization only as an emphasizer which is not reflected in our representation. Hence, we use the output of the morphological analyzer without any modification. The double passivized sentence in (159b) has the f-structure provided in (167).

(167)
$$\begin{bmatrix} \text{PRED} & \text{`ye}\langle\text{NULL, tatlı}\rangle\text{'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`tatlı'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{TENSE PAST, PASSIVE} & + \end{bmatrix}$$

4.3.3 Passivization of Causatives

Passivization of causatives is straightforward from a theoretical point of view but poses interesting issues in terms of implementation. The nominative subject *kedi* 'cat' in (168a) becomes the accusative object in (168b) when causativized. When the causative sentence in (168b) is passivized the accusative *kedi* 'cat' becomes the nominative subject again in (168c).

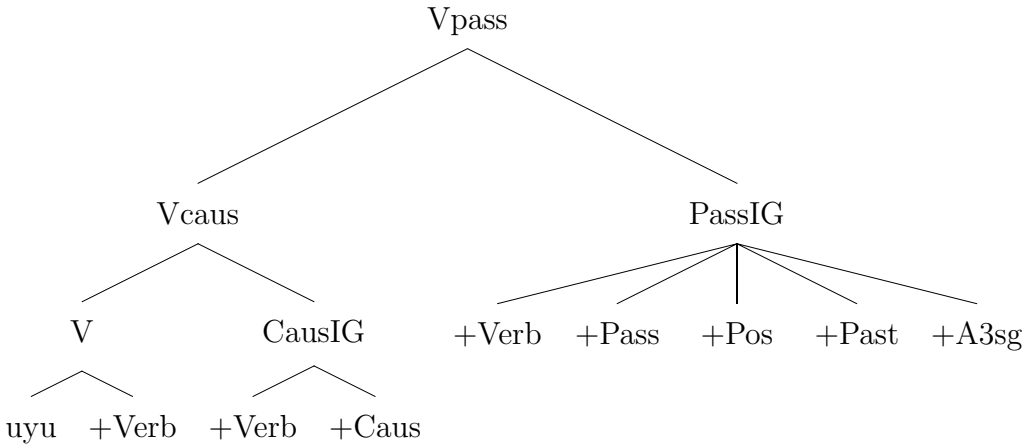
(168) a. kedi uyu-du
cat.Nom sleep-Past.3sg
'The cat slept.'

b. çocuk kedi-yi uyu-t-tu
child.Nom cat-Acc sleep-Caus-Past.3sg
'The child made the cat sleep.'

c. kedi uyu-t-ul-du
cat.Nom sleep-Caus-Pass-Past.3sg
'The cat was made to sleep.'

(169) illustrates the sublexical tree of the verb *uyutuldu* 'was made to sleep' where the hierarchy of the causative and passive morphemes can be observed. According to this sublexical tree, the main verb and the causative morpheme come together to constuct the causative complex predicate which is represented as Vcaus in the tree. Then the IG including the passive morpheme is attached to Vcaus to passivize it.

(169)



Causativization increases the valency of the verb by one. If the verb is intransitive, as in (168), the result is a transitive verb. Therefore, one would expect the passivization of causatives to be like the passivization of transitive verbs. The morpheme `+Caus` carrying the causative information has the modified lexical entry given in (170), similar to other verbs (cf. (162)). Then, the passivization will be handled by the standard passive rule.

(170) @(PASS (↑ PRED) = 'caus⟨(↑ SUBJ), %PRED2⟩').

Hovewer, the implementation does not go in parallel with the linguistic theory. The passive template in (163) is a lexical rule and it is called from the suffix lexicon of the

grammar ((170)). When there is no passivization, the PRED for the causative morpheme is `'caus<(↑ SUBJ), %PRED2>'` as expected. But when there is passivization, only the lexical rewrite rule (↑ SUBJ)→ NULL applies since there is no explicit OBJ in the provided PRED schema of the `+Caus` morpheme. Therefore the resulting PRED schema is (↑ PRED)='caus⟨NULL, %PRED2⟩' and the obligatory rewrite rule (↑ OBJ)→(↑ SUBJ) is lost in this step.

This is why there is a special disjunct devoted to passivized causatives in the causative rule ((171)). It makes use of the fact that the subject of the base verb is also the subject of the passivized causativized verb. There is no risk of allowing [[V+pass]+caus] constructions since they cannot pass the morphology barrier.

(171)

Vcaus →	V		CausIG
	↑ \PRED = ↓ \PRED		↑ = ↓
	(↑ PASSIVE)=c +		
	(↓ PRED) = (↑ PRED ARG2)		

As a result, the f-structure of the passivized causative sentence in (168c) is given in (173).

(172)
$$\begin{bmatrix} \text{PRED} & \text{'caus⟨çocuk, uyu⟨kedi⟩⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'çocuk'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'kedi'} \\ \text{CASE} & \text{acc} \end{bmatrix} \\ \text{TENSE} & \text{past} \end{bmatrix}$$

(173)
$$\begin{bmatrix} \text{PRED} & \text{'caus⟨NULL, uyu⟨kedi⟩⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'kedi'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{TENSE PAST, PASSIVE} & + \end{bmatrix}$$

The passivization of causativized transitives has no difference in implementation. (175) repeats the f-structure of the causativized transitive (174a) and (176) gives the f-structure of its passivized form in (174b).

(174) a. çocuk köpeğ-e kedi-yi kovala-t-tı
child.Nom dog-Dat cat-Acc chase-Caus-Past.3sg
'The child made the dog chase the cat.'

b. kedi (çocuk tarafından) köpeğ-e kovala-t-ıl-dı
cat.Nom child.Nom by dog-Dat chase-Caus-Pass-Past.3sg
'The dog was made to chase the cat (by the child).'

(175)
$$\begin{bmatrix} \text{PRED} & \text{'caus⟨çocuk, kovala⟨köpek,kedi⟩⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'çocuk'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'kedi'} \\ \text{CASE} & \text{acc} \end{bmatrix} \\ \text{OBJTH} & \begin{bmatrix} \text{PRED} & \text{'köpek'} \\ \text{CASE} & \text{dat} \end{bmatrix} \\ \text{TENSE} & \text{past} \end{bmatrix}$$

(176)
$$\begin{bmatrix} \text{PRED} & \text{'caus⟨NULL, kovala⟨köpek,kedi⟩⟩'} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'kedi'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{OBJTH} & \begin{bmatrix} \text{PRED} & \text{'köpek'} \\ \text{CASE} & \text{dat} \end{bmatrix} \\ \text{TENSE PAST, PASSIVE} & + \end{bmatrix}$$

4.4 Noun-Verb Compound Verbs

In Turkish, N-V constructions that act as a single verb are commonly used. Most frequently, the light verbs *et* 'do' and *ol* 'become', followed by *al* 'take', *ver* 'give', *koy* 'put' form the V part of the construction. (177) exemplifies a light verb in use.

(177) Ayşe geçmiş-i yad et-ti
Ayşe.Nom past-Acc remembrance.Nom do-Past.3sg
'Ayşe remembered the past.'

Verbs that are constructed using a noun and a light verb possess the characteristics of complex predicates. Consider the passivization test in (178). We argue that it is *geçmiş* 'past', not *yad* 'remembrance', which functions as object in (177). Supporting our argument, *geçmiş* 'past' becomes the nominative subject when the sentence is passivized in (178).

(178) geçmiş yad ed-il-di
past-Acc remembrance.Nom do-Pass-Past.3sg
'The past was remembered.'

Further evidence for the complex predication is that the noun and the light verb jointly affect the argument structure. (179) shows the effect of using two different light verbs with the same noun. In (179a) the object is in accusative, but in (179b) it is in dative case.

(179) a. çocuk öğretmen-**i** örnek **aldı**
child.Nom teacher-Acc role.model.Nom take-Past.3sg
'The child took the teacher as a role model.'

b. çocuk öğretmen-**e** örnek **oldu**
child.Nom teacher-Dat role.model.Nom become-Past.3sg
'The child became a role model to the teacher.'

In opposition to (179), if we use one light verb and change the noun part of the compound verb, as in (180a) and (180b), then we can observe different case markers in the object. Thus, it is the combination of light verb and noun that determines the case of the object.

(180) a. çocuk öğretmen-**i** **örnek** aldı
child.Nom teacher-Acc role.model.Nom take-Past.3sg
'The child took the teacher as a role model.'

b. çocuk öğretmen-**den** **haber** aldı
child.Nom teacher-Abl news.Nom take-Past.3sg
'The child learned news about the teacher'
'The child learned news from the teacher'

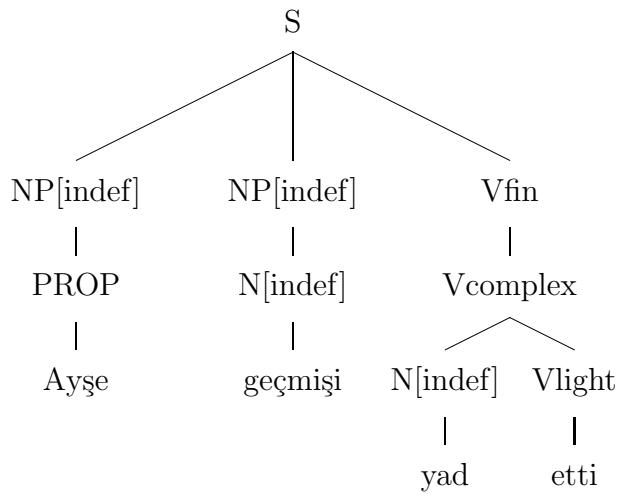
In terms of the LFG representation, we follow the N-V analysis of Butt et al. [2008]. The argument structure of verb and noun is mapped into a merged monoclausal structure. Together with its verb meaning, *et* 'do' has an additional entry in the lexicon as a light verb, given in (181).

(181) (↑ PRED) = 'et<(↑ SUBJ), %PRED2>'

The predicate of the noun that forms a compound verb with the given light verb is placed as the second argument of the complex predicate. The rule handling this

transformation is completely parallel to that of causatives. The c-structure and f-structure of the light verb example in (177) are given in (182) and (183) respectively.

(182)



(183)

$$\begin{bmatrix} \text{PRED} & \text{`et}\langle \text{Ayşe, yad}\langle \text{geçmiş}\rangle\rangle\text{'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`Ayşe'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`geçmiş'} \\ \text{CASE} & \text{acc} \end{bmatrix} \\ \text{TENSE} & \text{past} \end{bmatrix}$$

Compound verbs also act as a single constituent within the sentence. None of the other constituents of the sentence can interfere. Only the question clitic *mH*, e.g, *yad mı etti* 'did s/he remember' and the adverb *dA*, e.g, *yad da etti* 's/he remembered too' can be used in between. The Vcomplex representation in the c-structure makes sure that this property holds.

4.5 Non-canonical Objects

Turkish has a well-known case alternation on objects that correlates with the semantics of specificity [Enç, 1991]. A nonspecific direct object generally bears nominative case

and a specific direct object is marked with the accusative. (184) and (185) exemplify this well-known contrast.

(184) a. Ali bir **piyano** kiralamak istiyor
Ali one piano.Nom to.rent want.Prog.3sg
'Ali wants to rent one (some) piano.' [Enç, 1991]

b. Ali bir **piyano-yu** kiralamak istiyor
Ali one piano-Acc to.rent want.Prog.3sg
'Ali wants to rent a certain piano.' [Enç, 1991]

(185) a. **su** içtim
water.Nom drink.Past.1sg
'I drank water.'

b. **su-yu** içtim
water-Acc drink.Past.1sg
'I drank the water.'

In this section, we survey a less well-known fact that Turkish contains further semantically conditioned case markings. There are at least two identifiable groups (Section 4.5.1). One involves Differential-Object Marking [Aissen, 2003], encoding semantic differences at a clausal level, and in the other one, the non-canonical object marking seems to be conditioned exclusively by the lexical semantics of the verb. In Section 4.5.2, we go through a number of tests involving passivization, causativization and raising in order to get a handle on the distribution and behavior of the non-canonical objects. We present our analysis and its implementational details in Section 4.5.3.

The research in this section is done in collaboration with Miriam Butt and published in Çetinoğlu and Butt [2008].

4.5.1 Non-Canonical Object Marking in Turkish

In addition to the well-known specificity alternation in (184) and (185), an ablative object indicates partitivity when the object is consumable [Dede, 1981; Kornfilt, 1990],

as in (186), or it expresses a membership of a category (e.g. *örnek* 'example', [Göksel and Kerslake, 2005]). As (187) illustrates, the relation does not hold when the object does not belong to any of these groups.

(186) su-dan içtim
water-Abl drink.Past.1sg
'I drank some of the water.'

(187) şişe-den içtim
bottle-Abl drink.Past.1sg
'I drank (something) from the bottle.'

In addition to signaling partitivity, case in Turkish also appears to make distinctions between the degree of *affectedness* of an object. The examples in (188) and (189) illustrate this type of case alternation, which occurs with a group of verbs that also includes *bak* 'look' and *üfle* 'blow on'. Here the dative encodes less affected objects and alternates with the accusative. For example, in (188) the action and, indeed, the verb are the same. However, if an accusative is used, the interpretation is that the child was shot; when a dative is used, the object *child* is less affected and the interpretation is that the child was merely hit.

(188) a. Ali **çocuğ-u** vur-du
Ali.Nom child-Acc hit-Past.3sg
'Ali shot the child.' [Dede 1981:41]

b. Ali **çocuğ-a** vur-du
Ali.Nom child-Dat hit-Past.3sg
'Ali hit the child.' [Dede 1981:41]

(189) a. fare **peynir-i** ye-di
mouse.Nom cheese-Acc eat-Past.3sg
'The mouse ate the cheese.' [Dede 1981:41]

b. fare **peynir-e** dokun-du
mouse.Nom cheese-Dat touch-Past.3sg
'The mouse touched the cheese.' [Dede 1981:41]

In (189) the verbs differ, but the effect of the case alternation is the same: actions affecting an object to differing degrees are encoded via differential case marking.

Alternating case markers due to the affectedness of the object are also found in many other languages (e.g., Scottish Gaelic, Finnish, South Asian languages in general, cf. Butt 2006). For example, Kiparsky [1998] analyzes a Finnish alternation that is very similar to the one in (188) as involving boundedness.

(190) a. Ammu-i-n **karhu-n**
shoot-Past-1sg bear-Acc
'I shot the/a bear.' [Kiparsky 1998:267]

b. Ammu-i-n **karhu-a**
shoot-Past-1sg bear-Part
'I shot at the/a bear (bear is not dead).' [Kiparsky 1998:267]

We leave aside the question of the exact semantics underlying the observed alternations in (188) and (189) and move on to another type of non-canonical case marking on objects found with a large subset of psych verbs. Although all the verbs given in (191) are similar in meaning, only (191a) bears the canonical accusative case. (191b) and a group of verbs such as *nefret et* 'hate', *kork* 'fear', *şüphelen* 'suspect', *iğren* 'be disgusted' have ablative objects and (191c), and another subset of pysch verbs such as *yalvar* 'beg', *kız* 'be angry', *inan* 'believe' have dative objects.

(191) a. Ali **Ayşe'-yi** seviyor
Ali.Nom Ayşe-Acc love.Prog.3sg
'Ali loves Ayşe.'

b. Ali **Ayşe'-den** hoşlanıyor
Ali.Nom Ayşe-Abl like.Prog.3sg
'Ali likes Ayşe.'

c. Ali **Ayşe'-ye** tapıyor
Ali.Nom Ayşe-Dat adore.Prog.3sg
'Ali adores Ayşe.'

There is also another set of verbs which simply take non-canonical objects. These verbs do not have a common semantic property and can have either ablative or dative objects. *bin* 'ride' in (192) and *yardım et* 'help' are from this class.

(192) Hasan at-a bindi
Hasan.Nom horse-Dat ride.Past.3sg
'Hasan rode the horse.'

In our work, we focus on how these non-canonical objects should be analyzed. Given that they are clearly semantically restricted ([+r]), we would expect them to function as OBJ_θ or even OBL in terms of LFG's linking theory [Bresnan and Zaenen, 1990]. A related question is whether these non-canonical objects, when passivized, should be analyzed as subjects. In the next section, we therefore examine data with respect to passivization, causativization and raising.

4.5.2 Object Tests

Both causativization and passivization affect argument structure and thus are potentially good tests to distinguish between types of objects. In addition to these tests we consider the data from raising tests and observe that there are two classes of objects.

Passivization

We have given the passivization of verbs with canonical objects in Section 4.3. In standard LFG analyses (e.g., Bresnan 1982; Sells 1985; Butt et al. 1999), the assumption is that the OBJ, but not OBJ_θ, is realized as the SUBJ of the passive clause (also see the discussion of the status of OBJ in Börjars and Vincent [2008]). This section thus investigates the behavior of the non-canonical objects with respect to passivization.

Recall that in canonically marked clauses, the nominative/accusative object is realized as a standard nominative subject which agrees with the verb under passivization. (193) gives a simple canonical example.

(193) a. ben-i kovala-dı
I-Acc chase-Past.3sg
'S/he chased me.'

b. ben kovala-n-dı-m
I.Nom chase-Pass-Past-1sg
'I was chased.'

In contrast, the ablative partitive object preserves its case under passivization. As Dede [1981] points out, if the ablative were absorbed under passivization with ablative partitives, then the partitive reading would be lost. There is thus a *clausal semantic* reason for the ablative to be preserved.

(194) a. **su** iç-il-di
water.Nom drink-Pass-Past.3sg
'Water was drunk.'

b. **su-dan** iç-il-di
water-Abl drink-Pass-Past3sg
'Some of the water was drunk.'

Given this observation, the next question is the function of the ablative partitives in the passivized sentence. Subjecthood rules given in Kornfilt [1997] are the nominative case and the agreement with the verb in person and number [cf. also Göksel and Kerslake 2005], and (194b) fails with respect to both of them. This is more clearly illustrated by the (semantically somewhat strange) examples in (195) where the verb agreement is 3sg in (195b).

(195) a. **ben** iç-il-di-m
I.Nom drink.Pass-Past.1sg
'I was drunk.'

b. **ben-den** iç-il-di
I-Abl drink.Pass-Past.3sg
'Some of me was drunk.'

However, there are indications, as in (196), that these ablative partitives function as subjects. Kornfilt [1990] points out that these examples involve unaccusative verbs

where the ablative is the sole core argument and is naturally analyzed as a subject (despite the absence of verb agreement). Kornfilt [1990] argues that the ablative objects have the same distribution as canonical objects and proposes a *pro* which receives a phonologically unrealized Structural Case, thus bringing ablatives in line with canonical nominative/accusative objects (Kornfilt abandons the subjecthood criterion of verb agreement with respect to these examples).

(196) a. biz-de bu kitap-tan kal-ma-dı
we-Loc this book-Abl remain-Neg-Past
'We don't have any (copies) of this book left.' [Kornfilt 1990:287]

b. dolap-ta bu sucuk-lar-dan var/yok
cupboard-Loc this sausage-Pl-Abl exist/Neg.exist
'There are/aren't (some/any) of these sausages in the cupboard.' [Kornfilt 1990:287]

In Göksel and Kerslake's [2005] analysis, a type of partitive constructions is composed of an ablative noun phrase as the modifier and the constituent expressing the part as the head. (197) exemplifies the partitive construction *bu kitaptan iki tane* 'two copies of this book'. In such constructions it is possible to omit the head. When the head *iki tane* 'two copies' is omitted, we get the ablative noun phrase in (196a).Thus, the analysis of (196a) is parallel to that of (197). In our analysis, we do not employ such a parallelism. Our approach follows a parallelism between the ablative and nominative/accusative alternations of the same phrase.

(197) biz-de bu kitap-tan iki tane kal-ma-dı
we-Loc this book-Abl two copy remain-Neg-Past
'We don't have two copies of this book left.'

Non-canonical case encoding degree of affectedness/boundedness is also preserved under passivization. When (198a) is passivised the dative object is still dative in (198b) instead of nominative. Again, case absorption would erase the semantic contrast; the sentence would mean 'shot the child' rather than 'hit the child'.

(198) a. Ali çocuğ-a vur-du
Ali.Nom child-Dat hit-Past
'Ali hit the child.'

b. çocuğ-a vur-ul-du
child-Dat hit-Pass-Past
'The child was hit.' [Dede 1981:45]

If we apply a test on both alternatives of *vur*, we can observe that the passivized accusative and dative behave exactly alike with respect to anaphora resolution. This indicates that the passivized dative argument may be functioning as a subject.

(199) a. **çocuk** kendisi tarafından vuruldu
child.Nom self.P3sg by shoot.Pass.Past.3sg
'The child was shot by itself.'

b. **çocuğ-a** kendisi tarafından vuruldu
child-Dat self.P3sg by shoot.Pass.Past.3sg
'The child was hit by itself.'

In psych verbs, the object also preserves its case under passivization as exemplified in (200b). A small group of native speakers also accept the passivization as grammatical when the object becomes nominative as in (200c).

(200) a. san-a tap-tı
you-Dat worship.Past.3sg
'S/he worshipped you.'

b. san-a tapıldı
you-Dat worship.Pass.Past.3sg
'You were worshipped.'

c. ?sen tapıldın
you.Nom worship.Pass.Past.2sg
'You were worshipped.'

Although (200c) is ungrammatical for some speakers, the same data providers find (201) grammatical. In this example, *tapılarak* '(while) being worshipped' is the sentential complement which behaves as an adverb and is constructed by appending an

-(y)ArAk suffix to the verb. The subject of the *while*-clause always matches the subject of the main sentence (presumably via obligatory anaphoric control, cf. Dalrymple 2001). So, it seems that, to be able to construct the matrix sentence, the inner sentence should have a subject, and the verb *tap* 'worship' is forced to be passivized and has a nominative case marker, rather than a dative one.

(201) **öküz** tap-ıl-arak kilise-ye getir-il-di
ox.Nom worship-Pass-ByDoingSo church-Dat bring-Pass-Past.3sg
'The ox, while being worshipped, was brought to the church.'
(Knecht [1986] taken from Özkaragöz [1979])

When the matrix verb is impersonally passivized, *ox* can keep its dative case marker in the embedded clause. On the whole, the evidence from passivization with respect to the psych verbs again seems to indicate that the non-canonical object is indeed functioning as a direct object that is realized as a subject under passivization.

(202) **öküz-e** tap-ıl-arak dans ed-il-di
ox-Dat worship-Pass-ByDoingSo dance make-Pass-Past.3sg
'It was danced while the ox was worshipped.'

Lastly, we turn to the class of verbs like *bin* 'ride', which have dative objects. As shown in (203), case is again preserved under passivization.

(203) a. Hasan at-a bindi
Hasan.Nom horse-Dat ride.Past.3sg
'Hasan rode the horse.'

b. at-a bin-il-di
horse-Dat ride-Pass-Past.3sg
'The horse was ridden.'

However, this data by itself again is not sufficient to establish the potential subjecthood (and hence the precise object status of the non-canonical object), as it is also possible to passivize clauses with an intransitive verb and constituents other than the

direct object, as in (204). In these cases passivization is impersonal, that is, the constituent preserves its function (and also its case marking) and there is no subject in the passivized sentence ((204a) and (204b)).[10]

(204) a. Ali okul-a git-ti
Ali.Nom school-Dat go-Past.3sg
'Ali went to the school.'

b. okul-a gid-il-di
school-Dat go-Pass-Past.3sg
'The school was gone to. (Somebody went to the school)'

But all is not lost as the derivational suffixes *-(y)An* and *-dHğH* help distinguish between subject and non-subject gaps in participles. The suffix *-(y)An* is used in relativizing subjects, constituents expressing the location of the activity indicated by the relative clause, and some possessors [Göksel and Kerslake, 2005].

Thus, if we convert a passivized sentence with neither a location constituent nor a possessor into a participle and extract the constituent we are interested in, we can restrict ourselves to determine whether or not it is functioning as a subject. Consider the data in (205). (205a) represents the base predication. In (205b) and (205c), participles corresponding to the base predication have been formed. In (205b), the suffix *-(y)An* indicates that there is a subject gap, i.e., *köpek* 'dog' is the missing subject of the participle. In (205c), on the other hand, the object *kedi* 'cat' has been extracted and the non-subject suffix *-dHğH* marks this.

(205) a. köpek kedi-yi kovaladı
dog.Nom cat-Acc chase.Past.3sg
'The dog chased the cat.'

b. [ ]$_i$ kedi-yi kovala-**yan** köpek$_i$
cat-Acc chase.PresPart dog.Nom
'The dog that chased the cat.'

[10]The impersonal passive in (204) repeats (155). The detailed discussion on impersonal passives is given in Section 4.3.

c. köpeğ-in [ ]$_i$ kovala-**dığı** kedi$_i$
dog-Gen chase-PastPart.3sg cat.Nom
'The cat that the dog chased.'[11]

So let us try the participle extraction test with the *bin* 'ride' class. When we make a participle out of the passive version in (203) and extract the constituent *at* 'horse', the morphological marking on the participle indicates that the former non-canonical object is now patterning with subjects (cf. (206a) and (206b)). We take this as an indication that these non-canonical objects behave like subjects when they are passivized.

(206) a. bin-il-en at
ride-Pass-PresPart.3sg horse.Nom
'The horse that was ridden.'

b. *bin-il-diği at
ride-Pass-PastPart.3sg horse.Nom
'The horse that was ridden.' (intended meaning)

The data in this section has demonstrated that in all instances of non-canonical object marking, the case was preserved under passivization. Despite this case preservation and the lack of agreement with the verb, a range of tests indicate that these non-canonical objects function as subjects when passivized. Thus, the passivization data so far also suggest that all of the objects could be analyzed as OBJ. In the next section we turn to data from causativization to see whether this analysis can be confirmed or whether our analysis needs to be more differentiated.

Causativization

Both single and double causativization of verbs with canonical objects are discussed in the introductory Section 3.2 and in more detail in Section 4.2. If the verb is intransitive, the subject becomes an accusative object (cf. (31b)). In transitive clauses, the canonical nominative/accusative object preserves its case and function when the verb

[11]The genitive case on *dog* is because it is functioning as the agent/Spec of the participle.

is causativized. The causee (former nominative subject) is marked with the dative (cf. (32b)).

The ablative on partitive objects is similarly preserved under causativization. The causee is again dative, as exemplified in (207b). This is parallel to the canonical causative in (32), indicating that the ablative object patterns with canonical objects.

(207) a. su-dan iç-ti-m
water-Abl drink-Past-1sg
'I drank some of the water.'

b. annem ban-a su-dan iç-ir-di
mother.P1sg I-Dat water-Abl drink-Caus-Past.3sg
'My mother made me drink some of the water.'

Where a dative object signals low affectedness, we encounter a difficulty because Turkish has a general constraint which disprefers two dative-marked objects in a clause. However, if one of the datives is an indirect object, then two datives in a clause are allowed, as in (208).

(208) Babam-a çocuklar-a masal anlat-tır-dı-m
father.P1sg-Dat child.Pl-Dat story.Nom tell-Caus-Past-1sg
'I had my father tell stories to the children.' [Göksel 1993:216]

The pattern with causatives of dative less affected objects is complex in that it allows for an alternative realization of both the causee and the object. Each can be realized with a dative or an accusative, depending on whichever is compatible with an affectedness/boundedness reading. Consider *bak* 'look' in (209), which takes a dative object in the base predication. In (209a), the causee is in the dative, but in (209b), the causee is accusative and *kapı* 'door' (which is not affected) is dative.

(209) a. **hizmetçi-ye** çocuğ-u bak-tır-dı-k
maid-Dat child-Acc look-Caus-Past-1pl
'We made the maid look after the child.' [Dede 1981:43]

b. **herkes-i** kapı-ya bak-tır-dı-m
everybody-Acc door-Dat look-Caus-Past-1sg
'I made everybody look at the door.' [Dede 1981:43]

A similar pattern can be observed in (210) with the shoot/hit alternation.[12] When the child is less affected (hit rather than shot), it appears in the dative.

(210) a. Ahmet Ali'-**ye** çocuğ-**u** vur-dur-du
Ahmet.Nom Ali-Dat child-Acc shoot-Caus-Past.3sg
'Ahmet made Ali shoot the child.'

b. Ahmet Ali'-**yi** çocuğ-**a** vur-dur-du
Ahmet.Nom Ali-Acc child-Dat hit-Caus-Past.3sg
'Ahmet made Ali hit the child.'

Knecht [1986] gives another interesting example which allows two causativization patterns for a verb with a non-canonical object. The verb *hohla* 'blow on' subcategorizes for a dative object. Most of the native speakers prefer to keep *ayna* 'mirror' in the dative case, and convert *Ufuk* into accusative when causativized (211c). But it is also acceptable to transform the non-canonical object of the main verb into the accusative object of the causative verb, demonstrating the alternative possibilities in verbs with no clearly affected object (211b).

(211) a. Ufuk ayna-**ya** hohla-dı
Ufuk.Nom mirror-Dat blow.on-Past.3sg
'Ufuk blew on the mirror.'

b. Ufuk'-**a** ayna-**yı** hohla-t-tı-m
Ufuk-Dat mirror-Acc blow.on-Caus-Past-1sg
'I made Ufuk blow on the mirror.'

c. Ufuk'-**u** ayna-**ya** hohla-t-tı-m
Ufuk-Acc mirror-Dat blow.on-Caus-Past-1sg
'I made Ufuk blow on the mirror.'

[12]Note that an "affectedness" alternation in causatives has also been documented in Romance, Bantu and South Asian languages [Alsina and Joshi, 1991; Alsina, 1997; Butt, 1998].

The fact that causatives of non-canonical dative objects do not allow two datives in the clause indicates that both the causee and the non-canonical object should be analyzed as objects — the causee cannot be analyzed as an indirect object, otherwise two datives in a clause should be licit, as in (208). Furthermore, modulo the double-dative constraint, the non-canonical objects pattern like canonical transitives in terms of causativization.

We now turn to the pattern with psych verbs and verbs of the *bin* 'ride' type. Both with ablative and dative objects of psych verbs, the case is preserved under causativization. However, the causee (former nominative subject) is accusative rather than dative, as shown in (212) and (213).

(212) a. kedi köpek-ten kork-tu
cat.Nom dog-Abl fear-Past.3sg
'The cat feared the dog.'

b. çocuk **kedi-yi** köpek-ten kork-ut-tu
child.Nom cat-Acc dog-Abl fear-Caus-Past.3sg
'The child made the cat fear the dog.'

(213) a. Ali ateş-e tap-tı
Ali.Nom fire-Dat worship-Past.3sg
'Ali worshipped the fire.'

b. baba-sı **Ali'-yi** ateş-e tap-tır-dı
father-P3sg Ali-Acc fire-Dat worship-Caus-Past.3sg
'His father made Ali worship the fire.'

The same pattern holds for the *bin* 'ride' type. As shown in (214), the case of the object is preserved under causativization, and again, the causee must be accusative.

(214) a. Hasan at-a bin-di
Hasan.Nom horse-Dat ride-Past.3sg
'Hasan rode the horse.'

b. baba-sı **Hasan'-ı** at-a bin-dir-di
father-P3sg Hasan-Acc horse-Dat ride-Caus-Past.3sg
'His father made Hasan ride the horse.'

The evidence from causativization thus partitions the data into two sets: those which allow for a dative causee in parallel to canonical transitive clauses and those which require an accusative causee, deviating from the canonical pattern. Under the assumption that causatives always need to include an OBJ in the subcategorization frame, we suggest that the data from causativization can be understood as follows: ablative partitives and affectedness alternation involve "real" objects, i.e., OBJ. However, psych verbs and other non-canonical case marking verbs subcategorize for OBJ_θ. That is, when a clause with a partitive or less affected object is causativized, then the causee is realized as a dative OBJ_θ (or the causee as an OBJ and the affected object as an OBJ_θ in the case of the alternative possibilities in examples as in (209) or (211)) because there is already an OBJ in the clause. On the other hand, when a psych verb or *bin* 'ride' type verb is causativized, there is only a lexically determined OBJ_θ in the clause and so the causee is linked to an OBJ.

Passives of Causatives

In order to test this hypothesis, we examine the behavior of the causativized clauses with non-canonical objects when these in turn are passivized. As a benchmark, the passivization of a causativized canonical verb is given in (215). Note that the translation in (215b) might be misleading. In the Turkish sentence, *kedi* 'cat' is the subject whereas in the English sentence *dog* is the subject.[13]

(215) a. çocuk köpeğ-e kedi-yi kovala-t-tı
child.Nom dog-Dat cat-Acc chase-Caus-Past.3sg
'The child made the dog chase the cat.'

b. kedi (çocuk tarafından) köpeğ-e kovala-t-ıl-dı
cat.Nom child.Nom by dog-Dat chase-Caus-Pass-Past.3sg
'The dog was made to chase the cat (by the child).'

The ablative partitives again pattern canonically in that the causee remains dative. However, the ablative case is preserved and the subject is non-nominative. That is, the

[13]The example in (215) is also given in (174). Its f-structure analysis can be found in (176).

ablative object of the main verb seems to be the one linked to the OBJ in the causative version and it is this argument which is subject to passivization in (216b). Again, the English translation might be misleading.

(216) a. anne-m ban-a su-dan iç-ir-di
mother-P1sg I-Dat water-Abl drink-Caus-Past.3sg
'My mother made me drink some of the water.'

b. ban-a su-dan iç-ir-il-di
I-Dat water-Abl drink-Caus-Pass-Past.3sg
'I was made to drink some of the water.'

The dative less affected objects pattern like the ablatives. The verb *vur* 'shoot', which represents the canonical part of the affectedness alternation has the behavior given in (217). Both (217a) and (217b) have two readings caused by free word order. Note that *çocuk* 'child' is the subject of the first interpretation in (217b) despite the English translation.

(217) a. Ahmet Ali'**ye** çocuğ-**u** vur-dur-du
Ahmet Ali-Dat child-Acc shoot-Caus-Past.3sg
'Ahmet made Ali shoot the child.'
'Ahmet made the child hit Ali'

b. çocuk Ali'**ye** vur-dur-ul-du
child.Nom Ali-Dat shoot-Caus-Past.3sg
'Ali was made to shoot the child.'
'The child was made to hit Ali.'

The next example uses the verb *vur* 'hit' which represents the non-canonical part of the affectedness alternation. Similar to (217), (218) is also ambiguous. The second interpretation of (218b) would be more frequent than the first one among native speakers, though both are quite grammatical.[14]

[14]Note that these examples are somewhat artifical in daily usage although they are grammatical. A native speaker would prefer using a periphrastic causative verb, as in (2b). Then, the causative is a biclausal structure, the whole sentence in (2a) is nominalized so the dative case marker of the inner clause is preserved. *sebep ol* 'cause' is a N-V complex predicate and takes a dative object.

(218) a. Ahmet Ali'-**yi** çocuğ-**a** vur-dur-du
Ahmet Ali-Acc child-Dat hit-Caus-Past.3sg
'Ahmet made Ali hit the child.'
'Ahmet made the child shoot Ali. '

b. Ali çocuğ-**a** vur-dur-ul-du
Ali.Nom child-Dat hit-Caus-Pass-Past.3sg
'Ali was made to hit the child.'
'The child was made to shoot Ali.'

So in order to avoid ambiguity, we introduce an example with an inanimate object (219). The resulting sentence in (219b) is parallel to the *worship* example and in compliance with our findings.

(219) a. Ahmet Ali'-**yi** kapı-ya vur-dur-du
Ahmet Ali-Acc door-Dat hit-Caus-Past.3sg
'Ahmet made Ali hit the door.'

b. Ali kapı-ya vur-dur-ul-du
Ahmet Ali-Acc door-Dat hit-Caus-Pass-Past.3sg
'Ali was made to hit the door.'

The psych verbs and *bin* 'ride' type verbs again exhibit a different pattern. Examples of a psych verb with an ablative object ((220)), a psych verb with a dative object ((221)), and *bin* 'ride' with the dative object ((222)) are provided below. In every example the accusative causee in the causativized sentences becomes nominative under passivization. This is consistent with our analysis of the accusative causee having been linked to OBJ in the causative and then being available for standard passivization whereby a canonical OBJ is realized as a nominative SUBJ.

(2) a. Ali **çocuğ-a** vur-du
Ali.Nom child-Dat hit-Past
'Ali hit the child.'

b. Ahmet [Ali-'nin **çocuğ-a** vur-ma-sı]-na sebep ol-du
Ahmet.Nom Ali-Gen child-Dat hit-Inf-Poss-Dat cause.Nom become-Past.3sg
'Ahmet caused Ali hit the child.'

(220) a. çocuk kedi-yi köpek-ten kork-ut-tu
child.Nom cat-Acc dog-Abl fear-Caus-Past.3sg
'The child made the cat fear the dog.'

b. kedi köpek-ten kork-ut-ul-du
cat.Nom dog-Abl fear-Caus-Pass-Past.3sg
'The cat was made to fear the dog.'

(221) a. babası Ali'yi ateş-e taptırdı
father.P3sg Ali-Acc fire-Dat worship.Caus.Past.3sg
'His father made Ali worship the fire.'

b. Ali ateş-e taptırıldı
Ali.Nom fire-Dat worship.Caus.Pass.Past.3sg
'Ali was made to worship the fire.'

(222) a. babası Hasan'-ı at-a bindirdi
father.P3sg Hasan-Acc horse-Dat ride.Caus.Past.3sg
'His father made Hasan ride the horse.'

b. Hasan at-a bindirildi
Hasan.Nom horse-Dat ride.Caus.Pass.Past.3sg
'Hasan was made to ride the horse.'

In sum, the data from passivized causatives are consistent with our analysis made on the basis of the data with respect to simple causatives and passives. Ablative partitive and dative less affected objects behave in parallel to canonical objects, strengthening our claim that they are OBJ. For the sentences in (220)–(222), the result of the passivization is as expected: causativization introduces OBJs with an accusative case to these sentences, and passivization makes these OBJs nominative SUBJs. Hence the psych verbs and the *bin* 'ride' type of verbs with non-canonical objects can be analyzed as subcategorizing for OBJ_θs in their basic form.

Raising

Raising is another possible test for subject status. That is, one could take a passivized version of the clauses with non-canonical objects and see if the passivized object is

able to be raised out of the clause, as a normal subject would. However, it turns out that verbs like *görün* 'seem' and *inan* 'believe', which are equivalent to raising verbs in other languages, display a quite complex set of syntactic properties [a.o., Mulder 1976; Kornfilt 1977; Moore 1998] in Turkish.

When the lexical item *gibi* 'like' is used, agreement markers can appear on both the matrix and the embedded verb. Since this provides information about subject status and is thus potentially interesting for our investigation, we only provide examples with *gibi*, as in (223). Note that the agreement marker of the matrix verb is optional.

(223) biz san-a süt iç-ti-k gibi görün-dü-k
we.Nom you-Dat milk drink-Past.1pl like seem-Past-1pl
'We seemed to you to have drunk milk.' [Mulder 1976:(26b)]

The *biz* 'we' here is nominative and is clearly the subject of the matrix verb *görün* 'seem' ; as evidenced by verb agreement, it is also the subject of the embedded verb.

In (224), we have taken our benchmark transitive clause, passivized it and then embedded it in a raising construction. As can be seen, the embedded subject is raised to be the matrix nominative subject which agrees with the raising verb. Interestingly, this subject (*biz* 'we') may or may not agree with the embedded verb.

(224) a. biz sana kovala-n-dı-k gibi görün-dü-k
we.Nom you.Dat chase-Pass-Past-1pl like seem-Past-1pl
'We seemed to you to have been chased.'

b. biz sana kovala-n-dı gibi görün-dü-k
we.Nom you.Dat chase-Pass-Past.3sg like seem-Past-1pl
'We seemed to you to have been chased.'

Now let us examine what happens with respect to clauses with non-canonical objects. First, we take the examples of semantic case alternation. As can be seen from the alternation in (225), the case is again preserved in order to be able to preserve the semantic distinction of partitivity.

(225) a. su iç-il-di gibi görün-dü
water.Nom drink-Pass-Past.3sg like seem-Past.3sg
'It seemed that water was drunk.'

b. su-dan iç-il-di gibi görün-dü
water-Abl drink-Pass-Past.3sg like seem-Past.3sg
'It seemed that some of the water was drunk.'

The same is true for the affectedness alternation, where a nominative on *child* in (226a) would result in the reading that the child was shot, rather than hit (cf. [Kornfilt 1977]). This can be seen in (226b), which is ambiguous. In the second reading, the subject has been pro-dropped and is interpreted as a third person pronoun. Actually, (226a) also has a second reading parallel to that of (226b).

(226) a. çocuğ-a vur-ul-du gibi görün-dü
child-Dat hit-Pass-Past.3sg like seem-Past.3sg
'It seemed that the child was hit.'
'It seemed to the child that s/he was shot.'

b. ban-a vur-ul-du gibi görün-dü
I-Dat hit-Pass-Past.3sg like seem-Past.3sg
'It seemed that I was hit.'
'It seemed to me that s/he was shot.'

So, again it seems that in these cases the non-canonical object is acting as a direct object which can be raised out of a clause after passivization, though preserving its case marking for reasons of semantic contrast.

The pattern with respect to the psych verbs and the *bin* 'ride' type again differs. We illustrate this only with respect to the verb *kork* 'fear'(all the other verbs behave the same way as this one). As can be seen from (227a) vs. (227b), *biz* 'we' can marginally be raised; however it is not the subject of the embedded verb, as it cannot agree with that. Furthermore, as illustrated by (227c), one cannot raise *biz* 'we' while preserving its non-canonical case marking. *biz* 'we' can appear with the non-canonical case marking, but then only as part of the embedded clause, as in (227d) (cf. [Kornfilt, 1977] on a

discussion of the significance of word order in such examples) and the verb *görün* 'seem' must be interpreted as having an impersonal subject.

(227) a. *biz sana kork-ul-duk gibi görün-dük
we.Nom you.Dat fear-Pass-Past.1pl like seem-Past.1pl
'We seemed to you to have been feared.'

b. ?biz sana kork-ul-du gibi görün-dük
we.Nom you.Dat fear-Pass-Past.3sg like seem-Past.1pl
'We seemed to you to have been feared.'

c. *biz-den sana kork-ul-du gibi görün-dü
we-Abl you.Dat fear-Pass-Past.3sg like seem-Past.3sg
'It seemed to you that we were feared.'

d. sana [biz-den kork-ul-du] gibi görün-dü
you.Dat we-Abl fear-Pass-Past.3sg like seem-Past.3sg
'It seemed to you that we were feared.'

To summarize, the raising data confirms the patterns observed with respect to causativization and passivization: the non-canonical objects in Turkish can be grouped into two types. On the one hand, the non-canonical marking is used to express a semantic case alternation at clausal level and here the object can be analyzed as an OBJ. On the other hand, the non-canonical case marking is tied to the inherent lexical semantics of particular verbs, such as psych verbs and verbs such as *bin* 'ride', and in this case, the object can be analyzed as an OBJ_θ.

4.5.3 Analysis and Implementation

Given the empirical considerations made above, we conclude that the instances of Differential Object Marking (DOM), namely the ablative partitives and the affectedness alternation should be analyzed as involving OBJ. On the other hand, the cases of lexically specified non-canonical case marking involving dative and ablative arguments should be analyzed as inherently semantically-restricted objects, i.e., as OBJ_θ. We show how this analysis plays out in the actual implementation with respect to passivization

and causativization thereby further confirming the formal validity of our analysis.

Passivization

The subpart of the passive lexical rule dealing with canonical verbs is given in detail in Section 4.3.2. Another subpart of the passive lexical rule deals with psych verbs and *bin* 'ride' type verbs. For these, we posit the subcategorization frame `pred<SUBJ, OBJ-TH>` and add a disjunction to the standard passive lexical rule to encode that an OBJ-THbecomes SUBJ ((↑ OBJ-TH) → (↑ SUBJ)) when there are no OBJ available in the clause. The result is illustrated in (228) and (229), which give the simplified f-structures of the sentences in *Hasan ata bindi* 'Hasan rode the horse.' in (203a) and *ata binildi* 'The horse was ridden.' in (203b), respectively.

(228)
$$\begin{bmatrix} \text{PRED} & \text{'bin⟨Hasan, at⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'Hasan'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{OBJ-TH} & \begin{bmatrix} \text{PRED} & \text{'at'} \\ \text{CASE} & \text{dat} \end{bmatrix} \\ \text{TENSE} & \text{past} \end{bmatrix}$$

(229)
$$\begin{bmatrix} \text{PRED} & \text{'bin⟨NULL, at⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'at'} \\ \text{CASE} & \text{dat} \end{bmatrix} \\ \text{TENSE PAST, PASSIVE} & + \end{bmatrix}$$

Finally, the partitivity and affectedness relations are controlled via CHECK features, which are generally used within ParGram to enforce well-formedness constraints. Thus, for example, if a verb of consumption has a consumable object, it is allowed to have an ablative object in the basic sentence and an ablative subject in its passive form.[15] (230) shows the f-structure analysis of (186). The passivized sentence (194b) has the f-structure in (231).

[15]Ideally, this kind of information should be encoded and checked at the level of the representation of world knowledge.

(230)
$$\begin{bmatrix} \text{PRED} & \text{`iç⟨ben, su⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`ben'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`su'} \\ \text{CASE} & \text{abl} \\ \text{CHECK} & \text{consumable} \end{bmatrix} \\ \text{TENSE} & \text{past} \end{bmatrix}$$

(231)
$$\begin{bmatrix} \text{PRED} & \text{`iç⟨NULL, su⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`su'} \\ \text{CASE} & \text{abl} \\ \text{CHECK} & \text{consumable} \end{bmatrix} \\ \text{TENSE PAST, PASSIVE} & + \end{bmatrix}$$

Causativization

For the implementation of causatives with non-canonical objects, we follow the approach explained in Section 4.2.2. The standard rule that if the core predication already contains an OBJ, then the causee (former SUBJ) is realized as a dative OBJ_θ, applies to ablative partitives and the affectedness alternation.

Both the base version *ben sudan içtim* 'I drank some of the water.' and causativized version *annem bana sudan içirdi* 'My mother made me drink some of the water.' of the partitive example in (207) are represented by the f-structures (232) and (233), respectively.

(232)
$$\begin{bmatrix} \text{PRED} & \text{`iç⟨ben, su⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`ben'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`su'} \\ \text{CASE} & \text{abl} \\ \text{CHECK} & \text{consumable} \end{bmatrix} \\ \text{TENSE} & \text{past} \end{bmatrix}$$

(233)
$$\begin{bmatrix} \text{PRED} & \text{`caus⟨anne, iç⟨ben,su⟩⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`anne'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`su'} \\ \text{CASE} & \text{abl} \\ \text{CHECK} & \text{consumable} \end{bmatrix} \\ \text{OBJ-TH} & \begin{bmatrix} \text{PRED} & \text{`ben'} \\ \text{CASE} & \text{dat} \end{bmatrix} \\ \text{TENSE} & \text{past} \end{bmatrix}$$

If the core predication does not contain an OBJ, then the causee has to be real-

ized as an accusative OBJ. Psych verbs and *bin* 'ride' type verbs subcategorize for an OBJ-THinstead of an OBJ, therefore the SUBJ of the base verb becomes the OBJ after causativization. (234) illustrates this mapping for the psych verb *kork* 'fear'.

(234) kork⟨SUBJ, OBJ-TH⟩ caus⟨SUBJ, kork⟨OBJ, OBJ-TH⟩⟩

The code snippet required to implement the mapping in (234) is shown in (235). Since OBJ-THhas no change during the causativization process, there are no constraints for this argument in the rule. Actually, there is no separate disjunction for the verbs with non-canonical objects in the causative rule, the rule in (235) is identical to (126). The implementation for intransitive verbs is used to parse the verbs subcategorizing for a subject and a thematic object, too.

(235) Vcaus → V CAUSIG
↑ \PRED\SUBJ\OBJ= ↓ \PRED\SUBJ\OBJ ↑ = ↓
(↓ SUBJ) = (↑ OBJ)
(↓ PRED) = (↑ PRED ARG2)

Finally, we give the structures of the non-canonical objects. (236) and (237) depict f-structures of *kedi köpekten korktu* 'the cat feared the dog' and *çocuk kediyi köpekten korkuttu* 'the child made the cat fear the dog', given in (212).

(236)
$$\begin{bmatrix} \text{PRED} & \text{'kork⟨kedi, köpek⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'kedi'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{OBJ-TH} & \begin{bmatrix} \text{PRED} & \text{'köpek'} \\ \text{CASE} & \text{abl} \end{bmatrix} \\ \text{TENSE} & \text{past} \end{bmatrix}$$

(237)
$$\begin{bmatrix} \text{PRED} & \text{'caus⟨çocuk, kork⟨kedi, köpek⟩⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'çocuk'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'kedi'} \\ \text{CASE} & \text{acc} \end{bmatrix} \\ \text{OBJ-TH} & \begin{bmatrix} \text{PRED} & \text{'köpek'} \\ \text{CASE} & \text{abl} \end{bmatrix} \\ \text{TENSE} & \text{past} \end{bmatrix}$$

We treat *bin* 'ride' class verbs in the same manner. (238) and (239) are the implementations for *Hasan ata bindi* 'Hasan rode the horse' and *babası Hasan'ı ata bindirdi* 'His father made Hasan ride the horse', given in (214).

(238)
$$\begin{bmatrix} \text{PRED} & \text{'bin⟨Hasan, at⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'Hasan'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{OBJ-TH} & \begin{bmatrix} \text{PRED} & \text{'at'} \\ \text{CASE} & \text{dat} \end{bmatrix} \\ \text{TENSE} & \text{past} \end{bmatrix}$$

(239)
$$\begin{bmatrix} \text{PRED} & \text{'caus⟨baba, bin⟨Hasan, at⟩⟩'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'baba'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'Hasan'} \\ \text{CASE} & \text{acc} \end{bmatrix} \\ \text{OBJ-TH} & \begin{bmatrix} \text{PRED} & \text{'at'} \\ \text{CASE} & \text{dat} \end{bmatrix} \\ \text{TENSE} & \text{past} \end{bmatrix}$$

Our partitioning of non-canonical objects in Turkish into two distinct sets, one which subcategorizes for OBJ but with special case marking that is motivated by clausal semantic factors, and one which subcategorizes for an OBJ_θ due to inherent lexical semantic factors, thus allows for a straightforward implementation.

Summary

In this section we analyzed objects that bear cases other than the canonical nominative/accusative case in Turkish. With a set of examples, we observed the possible alternation scenarios and divided the non-canonical objects into subsets. Some verbs have ablative objects when the object is consumable and only part of the object is affected from the action. Degree of affectedness or boundedness causes alternation in object cases for another set of verbs as well. Most of the psych verbs subcategorize for either dative or ablative objects, as do a small subset of verbs with no common semantics.

When the sentences including non-canonical objects are passivized, all of the objects preserve their case. Although Turkish has nominative subjects in general, there are indications that non-canonical objects might turn into subjects. On the other hand,

data from causativization points to two distinct groups. Objects with partitivity or affectedness/boundedness alternations behave the same as canonical objects, with the difference that they preserve their non-canonical case in order to keep the semantic information coded by them. Objects of psych verbs and the *bin* 'ride' type behave as if they do not already contain an OBJ, as the accusative causee fills that role. We thus analyze these non-canonical objects as OBJ_θ.

Chapter 5

EVALUATION

Testing is one of the crucial steps of developing an accurate large-scale grammar. The initial attempts of testing our grammar started with a set of manually constructed test files. During the development of the grammar, we built a small test set each time we introduced a group of rules to parse certain types of phrases. There are basically four test files: noun phrases, basic and complex sentence structures, participles, and copular sentences. We have a total of 318 phrases/sentences in those files, with 76 additional phrases for the date-time grammar [Gümüş, 2007]. After major modifications in the grammar, these files are tested again in order to detect any possible bugs. Section 5.1 gives information about the more structured tests conducted.

Outputting all possible parses of a phrase is the major goal of our hand written grammar but highly ambiguous cases cause an exhaustive number of parses when the phrases get more complex. Hence, getting the optimal results is another crucial step in building a large scale grammar. In Section 5.2 we explain our attempts to rank the more possible parses higher than less possible ones. Finally we describe the integration of our system into a tool called LingBrowser [Armağan, 2008] in Section 5.3.

5.1 Test Suites

We divide the test suites into two groups. The first group consists of manually constructed test files. They are used to test specific linguistic phenomena and are usually introduced after those phenomena are implemented. ParGram sentences also fall into the first group. The second group consists of real world examples. We conducted a test on sentences and another on noun phrases where the test files are extracted from fiction novels in both of the cases.

ParGram Sentences

A set of sentences called ParGram sentences was distributed to the attendants before the semi-annual ParGram meetings. These are important for testing the coverage of the grammar on different and possibly problematic linguistic phenomena, as well as testing the parallelism among the participating grammars. Table 5.1 gives the total number of sentences, the number of relevant sentences, that is, the number of sentences that have a counterpart in Turkish in terms of parallel linguistic structure, and then the number of sentences parsed successfully. Each test suite has a set of basic test sentences followed by a set of more complex structures. Appendix B gives the sentences covered in ParGram meetings.

Meetings	total # of sent.	# of relevant sent.	# of parsed sent.
March 2006	23	20	13
September 2006	18	17	15
March 2007	18	17	17
August 2007	18	17	16
March 2008	21	19	19
September 2008	20	20	19

Table 5.1: The coverage for ParGram sentences

# of sentences	43
# of words	301
# of unique words	245
# of morphological analyses	636
# of unique morphological analyses	482

Table 5.2: Statistics about test sentences

Sentence Test Suite

One of the two important tests we conducted includes a test set of complex sentences. Unlike manually constructed test sets or ParGram sentences, this testfile is completely taken from running text. We used file 00007121.txt from METU Corpus [Say et al., 2002] which contains an excerpt from the fiction book *Öykümü Kim Anlatacak* 'Who will Tell My Story' [İşigüzel, 1994]. We took the first four paragraphs of the text and prepared an XLE test file by removing punctuation marks and placing one sentence per line. Table 5.2 shows the basic statistics concerning the test file.

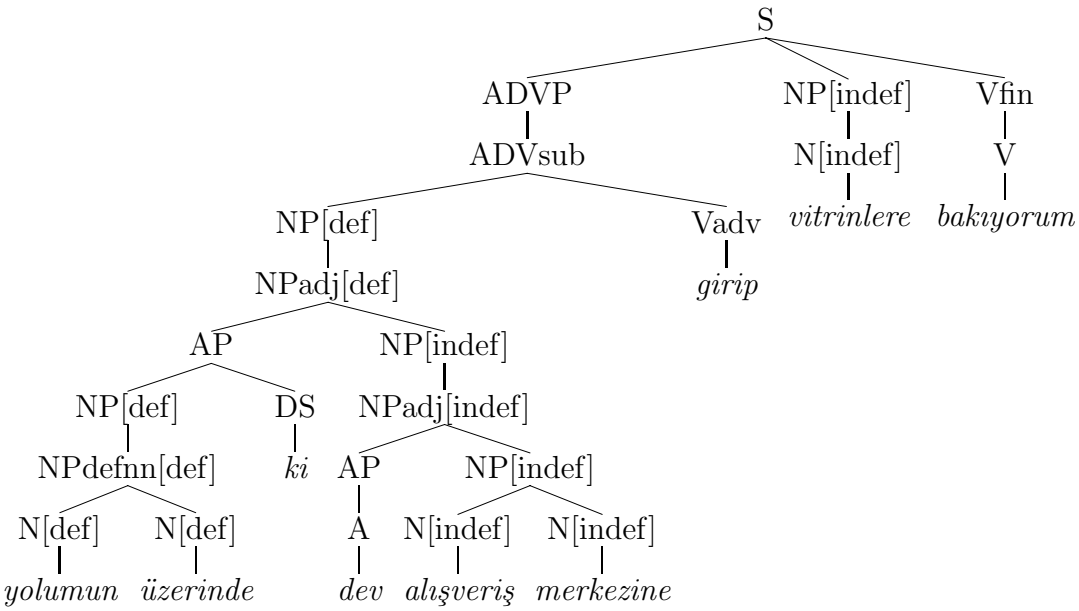
The shortest sentence contains a single word and the longest sentence contains 27 words. The average sentence length is 7 words. In terms of IGs, the shortest sentence has only one IG and the longest sentence has 35 IGs. The average number of IGs per sentence is 8.83. The number of morphemes in Table 5.2 and the number of IGs per sentence indicate that the sentences are more complex than the word counts indicate.

Of the 43 sentences, 29 are parsed in the first attempt. Later, the number is increased to 33 after the addition of some new rules. The remaining sentences get no parse. (240) is one of the parsed sentences. Its c-structure and f-structure are given in (241) and (242), respectively. The complete set of sentences is given in Appendix C.

(240) yol-um-un üzeri-nde-ki dev alışveriş merkez-i-ne
way-P1sg-Gen on-Loc-Rel huge shopping center-P3sg-Dat
gir-ip vitrin-ler-e bak-ıyor-um
enter-AfterDoingSo shopwindow-Pl-Dat look.at-Prog-1sg
'I look at the shop windows by entering the huge shopping center on my way.'

In the deep NP subtree in (241), first the phrase *alışveriş merkezi* 'shopping center' is constructed and then it is modified by the adjective *dev* 'huge'. On the left of the subtree the NP *yolumun üzeri* 'on my way' is constructed and derived into an AP by adding the derivational suffix *-ki*. This derived AP is the ADJUNCT of the phrase *dev alışveriş merkezi* 'huge shopping center', in which *merkez* 'center' is the head and, *dev* 'huge' and *alışveriş* 'shopping' modify it. This complex NP forms the adverbial sentence with the verb *gir-* 'enter' which is derived into an adverb. In the topmost level the main sentence consist of three nodes: ADVP for the adverbial sentence, NP for *vitrinlere* 'to the shop windows' and V_{fin} for *bakıyorum* 'I look at'.

(241)

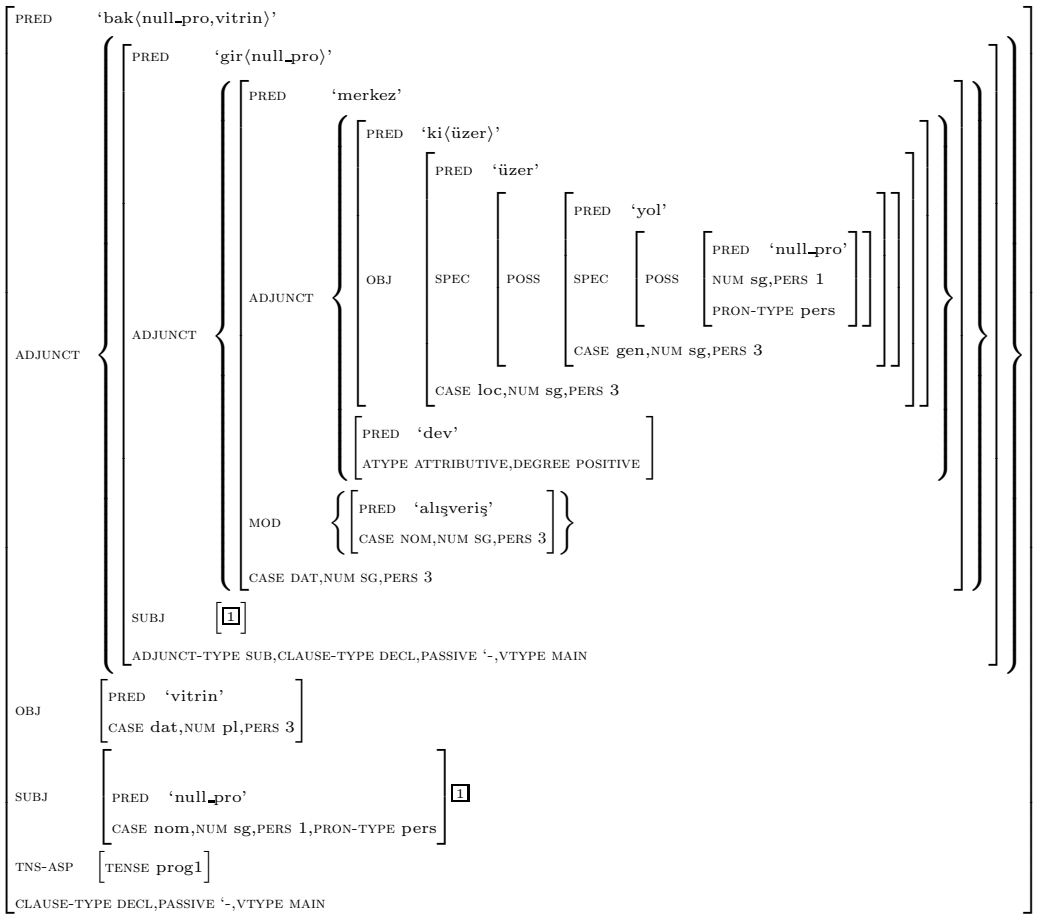


In (242), we can see the f-structures of the nodes represented in the c-structure. The five innermost nested f-structures represent the phrase *yolumun üzerinde* 'on my way'. *yol* 'way' is specified by the 1st person possesive marker and *üzerinde* 'on' is specified by *yol* 'way'. This f-structure is the OBJect of the suffix *-ki*. The derived adjective *yolumun üzerindeki* 'on my way' modifies the NP *dev alışveriş merkezi* 'huge shopping center'. This complex NP is the ADJUNCT of the verb *gir* 'enter'. The sentence is derived into an adverb by attaching the suffix *-ip* to its verb.

The PREDicate of the main sentence is the main verb *bak* 'look' and the information comes from the Vfin node. The ADVP node in the c-structure functions as the ADJUNCT

of the outermost matrix and the NP node is the OBJect. The SUBJect (i.e., *ben* 'I') is not explicit in the sentence but is formed by using the person marker in the verb. The pro-dropped subject of the main sentence is also the subject of the adverbial sentence. This is given by numbered square indicators.

(242)



Noun Phrase Test Suite

The second important test measures the coverage of noun phrases. We randomly picked file 00033224.txt [Duman, 1997] and file 00129176.txt [Peksoy, 2000] from the literature section of METU Corpus [Say et al., 2002]. Then the noun phrases in these files were manually extracted and divided into four groups. Table 5.3 gives the number of phrases in each subset of the test NPs. The complete list of phrases is given in Appendix D.

Type	Number
Simple Noun Phrases	194
Relative Clauses	48
Sentential Complements	36
Coordination	19
Total	297

Table 5.3: Types of phrases used in the noun phrase test

Simple Noun Phrases:

The set of simple noun phrases is composed of simple nouns, derived nouns, indefinite and definite noun compounds, adjective-modified NPs, pronouns and alike. Since these simple noun phrases are the base constituents of more complex noun phrases, the success rate is high in this set. 182 out of 194 phrases have a correct parse (93.8% success). 12 phrases get no parse. Some of the parsed phrases are given in (243). The complete set of simple noun phrases and the parses of (243) can be found in Appendices D.1 and D.5, respectively.

(243) a. *alarm sistemi falan* 'alarm system etc.'

b. *altıncı katın düğmesine* 'to the button of the sixth floor'

c. *arkadaşımızın doğum gününe* 'to our friend's birthday'

d. *aşağı kattaki ana vezneye* 'main pay desk at the lower floor'

e. *bir dakika bile* 'even a moment'

f. *biraz mahçup bir eda* 'a bit of an embarrassed expression'

g. *bütün eller* 'all hands'

h. *şu siyah uzun saçlı olanı* 'that one with long black hair'

i. *Tuğba'nın bu aşırı güvenine* 'to this over confidence of Tuğba'

Relative Clauses:

The group of relative clauses is important in that the rules parsing these phrases are parallel to the rules parsing sentences. Hence this subtest also gives us some idea

about the coverage of the sentences. 37 out of 48 phrases have a correct parse, and the rest gets no parse. Some of the parsed phrases are given in (244). The complete set of relative clauses and the parses of (244) can be found in Appendices D.2 and D.5, respectively.

(244) a. *bitip tükenmek bilmeyen bir yol* 'a never ending road'

b. *elindeki kitabı kapatan öğretmenin* 'of the teacher who closes the book in her hands '

c. *gözleriyle çevreyi araştıran Candan* 'Candan who is exploring the around with her eyes'

d. *mağazanın camlarının arkasını çepeçevre saran çelik perde* 'the steel panel that covers all of the rear sides of the windows of the store'

Sentential Complements:

Similar to relative clauses, sentential complements are indicators of sentence coverage as well as the noun phrase coverage. 30 out of 36 phrases get a correct parse and 6 phrases get no parse. Some of the parsed phrases are given in (245). The complete set of sentential complements and the parses of (245) can be found in Appendices D.3 and D.5, respectively.

(245) a. *daha erken gelebilmem* 'that I can come earlier'

b. *Mina'yı sevmemen* 'that you do not like Mina'

c. *sanatçının sahneye çıkışından* 'from the artist's getting to the stage'

d. *bir kentin ortasında yitmek* 'to get lost in the middle of a city'

Coordinated Noun Phrases:

Coordination has the lowest success rate among all kinds of noun phrases. The coordination rules do not cover different types of coordinated noun phrases. As a consequence, only 5 out of 19 phrases are parsed in this subset. The complete set of coordinated noun phrases can be found in Appendix D.4.

There are 297 phrases in total. The total number of phrases with a correct parse is 254 which means that our grammar can parse 85.5% of the test phrases. The remaining 43 phrases do not get an parse. A very important observation is that when our grammar provides multiple parses for a given input string, all parses have plausable interpretations. However, the system is not able to handle partial parses. These observations also hold for sentences test suite, that is, the grammar only outputs the correct parses. Once the system fails in parsing some constituents of the input, then parsing fails completely.

5.2 Optimality Theory Marks

Both the sentence and the NP tests show the strong and weak points of our grammar. One of the problems we encountered is highly ambiguous output. The source of the ambiguity in the parser outputs might be at the morphological or syntactic level. The sentence in (246), which is taken from the sentence test suite, gives an idea on how the ambiguity in words or in syntactic constructions affect the ambiguity of the whole sentence. The English translation of the sentence gives the intended meaning, but in the actual implementation this is just one of the eight different outputs. *kimse* is both a pronoun meaning 'anybody' and a noun meaning 'person'. For the word *bana*, the morphological analyzer gives the pronoun *me* in accusative and an infrequent noun root *ban*[1] in accusative. Moreover, the determiner *bu* 'this' may specify either the NP *kötü büyü* 'bad spell' or the NP *kötü büyüyü bozacak sihirli sözcük* 'the magical word that will break the bad spell'.

(246) kimse ban-a bu kötü büyü-yü boz-acak sihir-li sözcüğ-ü
nobody I-Acc this bad spell-Acc break-FutPart magic-With word-Acc
fısılda-ya-ma-dı
whisper-Able-Neg-Past.3sg
'Nobody was able to whisper me the magical word that will break this bad spell'

[1]An Ottoman title used for Crotian princes

In Lexical Functional Grammar, one widely used solution for the ambiguity problem is applying the Optimality Theory [Prince and Smolensky, 2004] by using Optimality Theory Marks (OT-marks) [Frank et al., 2001]. With the help of the OT-marks it is possible to mark the rules that cause a phrase to have different parses and to rank those rules in a user-defined order. For instance, the NP *kitap kapağı* has two interpretations in Turkish: more frequently used 'book cover', with the f-structure given in (247) and less frequently used 'his/her book cover', with the f-structure given in (248).

(247)
$$\begin{bmatrix} \text{PRED} & \text{'kapak'} \\ \text{MOD} & \begin{bmatrix} \text{PRED} & \text{'kitap'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{CASE} & \text{nom} \end{bmatrix}$$

(248)
$$\begin{bmatrix} \text{PRED} & \text{'kapak'} \\ \text{MOD} & \begin{bmatrix} \text{PRED} & \text{'kitap'} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{SPEC} & \begin{bmatrix} \text{POSS} & \begin{bmatrix} \text{PRED} & \text{'null_pro'} \\ \text{NUM sg, PERS 3} \\ \text{PRON-TYPE pers} \end{bmatrix} \end{bmatrix} \\ \text{CASE} & \text{nom} \end{bmatrix}$$

We assign OT-Marks, namely `NP-nn` and `NP-poss` respectively, to rules parsing these phrases and then give precedence to `NP-nn` over `NP-poss` in the OT-Mark ranking. XLE gives `1+1` results instead of `2` as the output. Only the preferred solution is displayed unless the user chooses to see unoptimal solutions in the output window.

This simple rule highly facilitates appropriate ranking since it applies to one of the very basic NP types that is frequently used in constructing more complex phrases. We also use OT-Marks to rank the temporal interpretation of NPs higher and to prefer lexicalized parses over derived ones when the morphological analyzer outputs both alternatives.

5.3 LingBrowser

Önsel Armağan at Sabancı University developed an NLP based hypertext browser that aims at helping advanced users acquire linguistic information on Turkish in an efficient

and user-friendly environment [Armağan, 2008]. It uses linguistic resources such as the Turkish morphological analyzer [Oflazer, 1994], Turkish WordNet [Bilgin et al., 2004], and TELL [Oflazer and Inkelas, 2006] to provide information like morphological segmentation and features, alignments of lexical and surface morphemes along with the explanation of any allomorph, segmental structure, pronunciation and stress information, meanings of roots, and advanced search in terms of linguistic information in the source text.

LingBrowser is designed in a modular way that enables the integration of new components. Our parser is integrated to LingBrower to parse arbitrary sentences and noun phrases. Paul Meuer from University of Bergen has developed XLE-Web,[2] a software that enables uploading the grammars to a server so that users can access the system via an online user interface. Extensions to LingBrowser set the communication between the user and XLE-Web. The user can choose a sentence and one of the menu options is to parse the sentence with the LFG grammar.

[2]http://maximos.aksis.uib.no/Aksis-wiki/XLE-Web

Chapter 6

SUMMARY AND CONCLUSION

In this thesis we presented our work on developing a large scale grammar for Turkish implemented in the Lexical Functional Grammar formalism. The grammar developed so far addresses many important linguistic aspects ranging from free constituent order, subject and non-subject extractions, all kinds of subordinate clauses mediated by derivational morphology, valency changing alternations, and has a very wide coverage NP subgrammar.

One of the tenets of our approach is the use of *inflectional groups* (IGs) as parsing units. IGs represent the inflectional properties of segments of a complex word structure separated by derivational boundaries (`^DB`) . An IG is typically larger than a morpheme but smaller than a word (except when the word has no derivational morphology in which case the IG corresponds to the word). It turns out that it is the IGs that actually define syntactic relations between words. A grammar for Turkish that is based on words as units would have to refer to information encoded at arbitrary positions in words, making the task of the grammar writer much harder. However, treating morphemes as units in the grammar level implies that the grammar will have to know about morphotactics making either the morphological analyzer redundant, or repeating the information in the morphological analyzer at the grammar level which is not very desirable. IGs bring a certain form of normalization to the lexical representation of a language like Turkish, so that units that the grammar rules refer to are simple enough to allow easy access to

the information encoded in complex word structures.

We developed a wide coverage noun phrase subgrammar with rules covering indefinite and definite noun compounds, possessives, pronouns, proper names, derived noun phrases, NPs modified by adjectives, determiners, numbers, measure phrases, postpositions, and combinations of these. Adjectives, postpositions, and adverbs also have their own rule sets. Sentential complements, sentential adjuncts and relative clauses present interesting challenges both in terms of linguistic analysis and in terms of implementation. All these are morphologically constructed by derivational suffixes attached to the verb. For the relative clauses, we employed functional uncertainty equations [Kaplan and Zaenen, 1989] to solve long distance dependencies. The rules parsing sentential derivations are parallel to the rules parsing sentences. We implemented free word order in sentences in addition to copular sentences, interrogative, negative sentences, and sentence level coordination. A date-time grammar developed by Gümüş [2007] was integrated into our system and improved our sentence coverage by parsing temporal phrases successfully.

We implemented sentence level coordination and coordination in noun phrases but there is still room for improvement in handling various types of coordinated phrases, especially in verb phrases where arguments of the verb are shared. Apart from these common types of coordination, Turkish employs suspended affixation [Kabak, 2007] where only the last conjunct of the coordinated phrase explicitly gets the inflectional features although these features scope over all the conjuncts. Parsing simple phrases with suspended affixation is implemented but a comprehensive solution to cover more complex phrases should be developed.

We thoroughly examined the representation and implementation of causatives and carried out a number of language specific tests to understand whether there is one combined clause (monoclausal) or two clauses with one embedded in another (biclausal) in causative constructions. The passivization, adjunct interpretation, and negative polarity item tests supports monoclausality, whereas reflexive binding and control tests have some counterexamples that favor biclausal structures. The result of these obser-

vations led us to assume a monoclausal structure. We then implemented our proposed analysis as complex predicates [Butt and King, 2006] by taking advantage of the Restriction Operator [Kaplan and Wedekind, 1993]. We provided details of our rules and illustrated our results with sample c-structures and f-structures. We also included the implementation of double causatives into our grammar.

We discussed impersonal passives, double passives, and passives of causatives as well as basic passivization. We followed the standard approach used in the ParGram grammars to implement basic passivization and extended this approach to other types of passives. Compound verbs which are composed of a noun and a light verb, e.g., *yardım etmek* 'help, lit. help do', were also treated as complex predicates and implemented in a similar fashion as causatives.

Another extensive study within this thesis covers non-canonical objects. We investigated verbs that subcategorize for an object with case markers other than the canonical accusative case. These verbs were divided into four subsets: ablative partitives, affectedness alternations, psych verbs, and a small subset of verbs with no common semantics. We revisited causativization, passivization, and passivization of causatives, this time to observe behaviour of non-canonical objects, and also tested non-canonical objects under raising constructions. Given these empirical considerations, we concluded that the ablative partitives and the affectedness alternation are parallel to canonical objects and should subcategorize for a SUBJ and an OBJ. The objects of psych verbs and the small subset of verbs with no common semantics should be analyzed as inherently semantically-restricted objects, i.e., as OBJ_θ. We again provided implementational details and sample c-structures and f-structures.

The ParGram sentences were helpful in testing the coverage of the grammar on linguistically challenging topics and the qualitative evaluation of c-structures and f-structures. Moreover, they are crucial in keeping our grammar parallel to other grammars. The sentence test suite and the noun phrase test suite provide important data for grammar evaluation, since they are directly taken from running text. We used three separate files from the METU Corpus [Say et al., 2002] for these two tests, all of which

are excerpts from stories. In the sentence test suite, 33 out of 43 sentences have correct parses and 10 sentences get no parse. The noun phrase test suite has 297 phrases which are divided into four groups. 182 out of 194 simple noun phrases, 37 out of 48 relative clauses, 30 out of 36 participles, and 5 out of 19 coordinated NPs get correct parses. The remaining phrases do not have any parser output. The percentage of successful parses is 93.8% in simple noun phrases and 85.5% in total. The drop is mainly caused by coordinated noun phrases. The tests not only show that our grammar has a high coverage in noun phrases but also informs us on the sentence coverage since relative clauses and sentential complements have parallel rules to sentence parsing rules. The results of the tests conducted also address a major drawback: highly ambiguous output. Although we attempted to rank the outputs by using OT-Marks, the results were not satisfactory. We see it as an important avenue for future work.

In summary;

- we employed parsing units that we call inflectional groups in building our grammar. This choice enables us to handle the very productive derivational morphology in Turkish in a rather principled way and has made the grammar more or less oblivious to morphological complexity. We presented the architecture of our grammar earlier in Çetinoğlu and Oflazer [2006] and the updated version in Çetinoğlu and Oflazer [2009].

- we built a wide coverage grammar with rules parsing an extensive set of noun phrases, adjectival, adverbial, postpositional phrases, sentential complements, adjuncts, relative clauses, basic sentence types, basic coordinated phrases.

- we integrated a date-time grammar [Gümüş, 2007] into our system and improved the coverage on temporal adjuncts.

- we thoroughly examined some of the linguistic phenomena, such as causativization, passivization, light verbs, and non-canonical objects. We proposed solutions on how they can be represented structurally and how we can implement them

within the LFG architecture. We presented our findings in Çetinoğlu et al. [2008] for causatives and in Çetinoğlu and Butt [2008] for non-canonical objects.

- we produced linguistically motivated, deep, and rich outputs which are useful for semantics both linguistically and computationally.

- we tested our grammar coverage on sentences and noun phrases with real world data. We correctly parsed 33 out of 43 sentences in the sentences test suite and 254 out of 297 (85.5%) phrases in the noun phrases test suite.

- we integrated our system as the syntactic component into LingBrowser [Armağan, 2008] which provides end users with linguistic information on Turkish, such as morphological structures, glosses, pronunciation and stress representations.

6.1 Future Work

We presented an LFG based Turkish grammar which covers many aspects of the language and outputs rich and structured parses. It is, though, still at the beginning of the development when compared to large scale robust grammars which can parse nearly every sentence. To extend the coverage, coordination should be revisited. Coordination structures are frequent in real world data, and present challenges for efficient implementation and ranking optimal solutions. More complex sentence structures, punctuation, multiword expressions, and named entities are among the most important topics that should follow.

The robustness of a grammar is measured by its capability of parsing real world data. The current grammar is capable of giving accurate outputs for the phrases it can parse, but fails to give an output for many others. We observe that we can parse many of the constituents though we cannot find a parse for the complete sentence. A good way of handling this problem is to use a fragment rule that will parse the phrase as a set of fragments [Butt et al., 1999]. XLE is configurable in a way that the fragment rule can be used when no valid parses are available.

As the grammar coverage is extended, the complexity of the parseable sentences increases, resulting in many possible parses. We attacked the problem by introducing OT-Marks for ranking the most probable outputs higher. The next step should be to enrich the grammar with more OT-Marks. This can be achieved with the help of linguistic heuristics, and statistical information. OT-Marks are also a key to robustness by allowing parses with common mistakes in written data or daily speech although they are not strictly grammatical [Frank et al., 2001]. In addition, XLE facilitates integrating statistical methods into the system to output the most probable one among correct parses [Kaplan et al., 2004a]. Guidance on preparing the statistical input to train the system can be obtained from previous work [Riezler et al., 2002; Riezler and Vasserman, 2004].

So far, all proposed solutions and most of the future work that will improve the grammar are based on manual work. Obviously, this means years of effort by advanced developers with linguistic expertise. Alternatively, already existing resources can be used as tools to improve the grammar in a more efficient way. One of the best available resources for our needs is the Turkish Treebank [Oflazer et al., 2003]. Actually, we used the treebank to retrieve the most frequently used subcategorization frames of verbs[1] and to import this data to the verb lexicon of our grammar. But it is just a minor attempt as compared to other ways to benefit from the well-structured data the treebank employs.

Cahill et al. [2008] show that it is possible to automatically induce wide-coverage, robust, deep LFG grammars from the Penn-II Treebank [Marcus et al., 1994] for English. The idea is to annotate the treebank with f-structure equations and extract a parser from this annotated treebank. This parser can be used to parse unseen data and to output annotated trees which then can be converted to f-structures by collecting and resolving the annotations on the nodes of the tree. Evaluation on gold standards prove that the results are competitive with the results from hand written grammars. The approach is also successfully applied to languages from different language families and with varying amount of resources such as Arabic [Tounsi et al., 2009], Chinese [Burke

[1]We thank Reyyan Yeniterzi and Süveyda Yeniterzi for helping with this.

et al., 2004; Guo et al., 2007], French [Schluter and van Genabith, 2008], German [Cahill et al., 2005], Japanese [Oya and van Genabith, 2007], and Spanish [O'Donovan et al., 2005; Chrupała and van Genabith, 2006]. We believe it is an interesting research topic to apply this framework to a morphologically rich language with a dependency treebank encoding relations between IGs instead of words.

Appendix A

Morphological Tags

+A1pl	1st person plural
+A1sg	1st person singular
+A2pl	2nd person plural
+A2sg	2nd person singular
+A3pl	3rd person plural
+A3sg	3rd person singular
+Abl	Ablative
+Able	Able to verb
+Acc	Accusative/Objective
+Acquire	To acquire the noun in the stem
+Adj	Adjective
+Adv	Adverb
+AfterDoingSo	After having *verb*ed
+Agt	Involved in some way with the stem
+Aor	Aorist tense
+As	As long as (s/he) *verb*s
+AsIf	As if (s/he is) *verb*ing
+Become	To become like the noun or adj in the stem
+ByDoingSo	By *verb*ing
+Card	Cardinal number

+Caus	Causative
+Cop	Copular
+Dat	Dative
+DemonsP	Demonstrative pronoun
+Det	Determiner
+Dim	Diminutive
+Dup	Duplicative
+FitFor	Fits for that noun
+Fut	Future tense
+FutPart	Future participle
+Gen	Genitive
+Imp	Imperative
+Inf	Infinitive
+Ins	Instrumental
+Interj	Interjection
+Loc	Locative
+Ly	As in slow → slowly
+Neg	Negative polarity
+Ness	As in red → redness
+Nom	Nominative
+Noun	Noun
+Num	Number
+Opt	Optative, let me/him/her *verb*
+Ord	Ordinal number
+P1pl	1st person plural
+P1sg	1st person singular
+P2pl	2nd person plural
+P2sg	2nd person singular
+P3pl	3rd person plural
+P3sg	3rd person singular

+Pass	Passive
+Past	Past tense
+PastPart	Past participle
+Percent	Percentage number
+PersP	Personal pronoun
+Pnon	Pronoun (no overt agreement)
+Pos	Positive polarity
+Postp	Postposition
+Pres	Present tense
+PresPart	Present particple
+Prog	Present continuous
+Pron	Pronoun
+Prop	Proper noun
+Punc	Punctuation
+QuantP	Quantifying pronoun
+Ques	Question clitic
+QuesP	Question pronoun
+Recip	Reciprocal
+Reflex	Reflexive
+ReflexP	Reflexive pronoun
+Rel	Relativization
+SinceDoingSo	Since having *verb*ed
+Verb	Verb
+When	When (s/he) *verb*s
+While	While (s/he is) *verb*ing
+With	With that noun
+Without	Without that noun
+WithoutHavingDoneSo	Without having *verb*ed
+Zero	A derivation with a 0 morpheme
^DB	Derivational boundary

Appendix B

ParGram Sentences

B.1 Spring 2006 Meeting

(1) Tamar
Tamar
'Tamar'

(2) o
she/he/it
'she'

(3) köşedeki şu dayanıksız kutu
corner.Loc.Rel that flimsy box.Nom
'that flimsy box in the corner'

(4) kızlar zıpladı
girl.Pl.Nom jump.Past.3sg
'girls jumped'

(5) kızlar zıpladılar
girl.Pl jump.Past.3pl
'girls jumped'

(6) kızlar oğlanların kutuyu gördüğünü söyledi
girl.Pl.Nom boy.Pl.Gen box.Acc see.PastPart.Sg.Acc say.Past.3sg
'The girls said that the boys saw the box.'

(7) kızlar oğlanların kutuyu gördüğünü söylediler
girl.Pl.Nom boy.Pl.Gen box.Acc see.PastPart.Sg.Acc say.Past.3pl
'The girls said that the boys saw the box.'

(8) kızlar oğlanların kutuyu gördüklerini söyledi
girl.Pl.Nom boy.Pl.Gen box.Acc see.PastPart.Pl.Acc say.Past.3pl
'The girls said that the boys saw the box.'

(9) kızlar oğlanların kutuyu gördüklerini söylediler
girl.Pl.Nom boy.Pl.Gen box.Acc see.PastPart.Pl.Acc say.Past.3pl
'The girls said that the boys saw the box.'

(10) dayanıksız kutu
flimsy box.Nom
'the flimsy box'

(11) kutu dayanıksız
box.Nom flimsy
'The box is flimsy.'

(12) kutu dayanıksızdır
box.Nom flimsy.Cop
'The box is flimsy.'

(13) kırık kutu
broken box.Nom
'the broken box'

(14) Akide zıplayabiliyor
Akaki.Nom jump.Able.Prog.3sg
'Akaki is able to jump.'

(15) Yağmurun yağması olası
rain.Gen rain.Inf.Poss probable
'It is likely to rain.'

(16) Akide'nin zıpladığı doğru
Akaki.Gen jump.PastPart.Poss true
'It is true that Akaki jumped.'

(17) Akide sıfatlardan korkuyor
Akaki.Nom adjective.Pl.Abl fear.Prog.3sg
'Akaki is afraid of adjectives.'

(18) Akide Tamar'ın sıfatları sevdiğinin farkında
Akaki.Nom Tamar.Gen adjective.Pl.Acc like.PastPart.Gen fark.Poss.Loc
'Akaki is aware that Tamar likes adjectives.'

B.2 Fall 2006 Meeting

(19) usta geldi
plumber.Nom come.Past.3sg
'The plumber came.'

(20) usta lavaboyu onardı
plumber.Nom sink.Acc fix.Past.3sg
'The plumber fixed the sink.'

(21) usta duşu onarmadı
plumber.Nom shower.Acc fix.Neg.Past.3sg
'The plumber did not fix the shower.'

(22) usta kaloriferi onardı mı
plumber.Nom heating.Acc fix.Past.3sg Ques
'Did the plumber fix the heating?'

(23) arabayı onarın
car.Acc fix.Imp.2pl
'Fix [= 2Pl] the car.'

(24) bisikleti kim onardı
bike.Acc who fix.Past.3sg
'Who fixed the bike?'

(25) iyi bir usta
good a plumber.Nom
'a good plumber'

(26) usta canayakındır
plumber.Nom friendly.Cop
'The plumber is friendly.'

(27) köpek bahçededir
dog.Nom garden.Loc.Cop
'The dog is in the garden.'

(28) Ebeveynlerimin yeni arabasındaki deri koltuklar
parent.Pl.Gen new car.Poss.Loc.Rel leather seat.Pl.Nom
'the leather(-)seats in my parents' new car'

(29) çok daha büyük
very more big
'very much bigger'

(30) saatte elli kilometrelik hız sınırı
hour.Loc fifty kilometre.For speed.Nom limit.Poss
'a speed limit of fifty kilometres an hour'

(31) saatte elli kilometrelik hız sınırı
hour.Loc fifty kilometre.For speed.Nom limit.Poss
'a speed limit of fifty kilometres an hour'

(32) şişesi iki liraya şarap
bottle.Poss two pound.Dat wine.Nom
'wine for two pounds a bottle'

(33) onun şarap içmesini engelledi
he.Gen wine.Nom drink.Inf.Poss.Acc prevent.Past.3sg
'She prevented him from drinking the wine.'

(34) onların hepsi balık yedi
they.Gen all fish.Nom eat.Past.3sg
'They had all eaten fish and chips.'

B.3 Spring 2007 Meeting

(35) köpekler kedileri kovalar
dog.Pl.Nom cat.Pl.Acc chase.Aor.3sg
'Dogs chase cats.'

(36) köpekler kedileri kovalamaz.
dog.Pl.Nom cat.Pl.Acc chase.Neg.Aor.3sg
'Dogs do not chase cats.'

(37) çocuklar köpekleri ve kedileri kovalar
child.Pl.Nom dog.Pl.Acc and cat.Pl.Acc chase.Aor.3sg
'Children chase dogs and cats.'

(38) çocuklar okula gidiyor
child.Pl.Nom school.Dat go.Prog.3sg
'The children go to school.'

(39) Peter'in horladığını biliyorum
Peter.Gen snore.PastPart.Acc know.Prog.1sg
'I know that Peter snores.'

(40) 3 Şubat 2007'de Meryem güldü
3 February 2007.Loc Mary laugh.Past.3sg
'On February 3, 2007, Mary laughed.'

(41) sabahın 3.00'ünde Meryem güldü
morning.Gen 3.Poss.Loc Mary.Nom laugh.Past.3sg
'At 3:00 in the morning, Mary laughed.'

(42) sabah saat 3'te Meryem güldü
morning.Nom hour.Nom 3.Loc Mary.Nom laugh.Past.3sg
'At 3:00 in the morning, Mary laughed.'

(43) Fransa, Paris'te Meryem güldü
France.Nom, Paris.Loc Mary.Nom laugh.Past.3sg
'In Paris, France, Mary laughed.'

(44) gelemem
come.Able.Neg.1sg
'I cannot come.'

(45) gelmemezlik yapamam
come.Neg.NotState make.Able.Neg.1sg
'I cannot not come.'

(46) gelmezlik yapamam
come.NotState make.Able.Neg.1sg
'I cannot not come.'

(47) kimse gelmedi
anybody.Nom come.Neg.Past.3sg
'Anybody didn't come.'

(48) birazdan yağmur yağmaya başlayacak
soon rain.Nom come-down.Inf.Dat start.Fut.3sg
'Soon it will start to rain.'

(49) yıkan
wash.Reflex.Imp.2sg
'Wash yourself.'

(50) Zeynep elbise dikindi
Zeynep.Nom dress.Nom sew.Reflex.Past.3sg
'Zeynep sewed a dress for herself.'

(51) büyük bir elma
big a apple.Nom
'a big apple'

(52) büyük bir kutu elma
big a box.Nom apple.Nom
'a big box of apples'

(53) balığın yirmi alt türü
fish.Gen twenty sub.Nom type.Poss.
'twenty subtypes of fishes.'

(54) şu restoranda şarap yirmi liradır
that restaurant.Loc wine.Nom twenty lira.Cop
'Wine is 20 euro in that restaurant.'

B.4 Fall 2007 Meeting

(55) kız iki çocuk gördü
girl.Nom two boy.Nom see.Past.3sg
'The girl saw two boys.'

(56) ağlamadılar
cry.Neg.Past.3pl
'They did not cry.'

(57) masadaki yeni kitaplar
table.Loc.Rel new book.Pl.Nom
'the new books on the table'

(58) Can öğretmendir
John.Nom teacher.Cop
'John is a teacher.'

(59) öğrencilerin hepsi Japon
student.Pl.Gen all.Poss Japanese
'All of the students are Japanese.'

(60) öğrencilerin bildiriyi yazması
student.Pl.Gen paper.Acc write.Inf.Poss
'the students' writing the paper'

(61) İngilizce eğitimi
English.Nom study.P3sg
'the study of English'

(62) öğrencilerin İngilizce eğitimi
student.Pl.Gen English.Nom study.Poss
'the students' study of English'

(63) okullardaki eğitim
school.Pl.Loc.Rel study.Nom
'the study in schools'

(64) okullardaki eğitim süreci
school.Pl.Loc.Rel study.Nom process.P3sg
'the studying in schools'

(65) okullar hakkında eğitim
school.Pl.Nom account.Poss.Loc study.Nom
'the study of/about schools'

(66) okullar hakkında eğitim süreci
school.Pl.Nom account.Poss.Loc study.Nom process.P3sg
'the studying of schools'

(67) gezdiğim ülke
visit.PastPart.P1sg country.Nom
'the country I visited'

(68) gezilecek ülke
visit.Pass.FutPart country.Nom
'the country to visit'

(69) ülkeyi gezen kişi
country.Acc visit.PresPart person.Nom
'the person (who) visited (the country)'

(70) ayakta duran kız uzun
foot.Loc stand.PresPart girl.Nom tall
'the girl who is standing is tall.'

(71) kızı gören çocuk onu beğendi
girl.Acc see.PresPart boy.Nom she.Acc like.Past.3sg
'The boy who saw the girl liked her.'

(72) Peter elmaları yetiştirir Kari de yer
Peter.Nom apple.Pl.Acc grow.Aor Kari.Nom too eat.Aor
' Peter grows and Kari eats apples.'

B.5 Spring 2008 Meeting

(73) bazı çocuklar güldü
some kid.Pl laugh.Past.3sg
'Some kids laughed.'

(74) çocukların bazısı güldü
kid.Pl.Gen some.Poss laugh.Past.3sg
'Some of the kids laughed.'

(75) birisi güldü
someone.Poss laugh.Past.3sg
'Someone laughed.'

(76) bazıları güldü
some.Poss laugh.Past.3sg
'Some laughed.'

(77) dört çocuk güldü
Four kid.Nom laugh.Past.3sg
'Four kids laughed.'

(78) iklimden başka faktörler
climate.Abl other factor.Pl
'other factors than (the) climate'

(79) öyle bir gürültü var ki kimse uyuyamıyor
such a noise.Nom existing that anybody sleep.Able.Neg.Prog.3sg
'There is such a noise that nobody can sleep.'

(80) vitamin ve mineraller gibi katkı maddeleri
vitamin.Nom and mineral.Pl.Nom as additive substance.Pl.Poss
'such additives as vitamins and minerals'

(81) yol yorgunu kovboy
road.Nom tired.Poss cowboy.Nom
''

(82) Ali uyudu.
Ali sleep.Past.3sg
'Ali slept.'

(83) annesi Ali'yi uyuttu
mom.P3sg Ali.Acc sleep.Caus.Past.3sg
'His mom made Ali sleep.'

(84) Ali uyutuldu.
Ali.Nom sleep.Caus.Pass.Past.3sg
'Ali was made to sleep.'

(85) annesi Ali'ye muzu yedirdi
mom.P3sg Ali.Dat banana.Acc eat.Caus.Past.3sg
'His mom made Ali eat the banana.'

(86) Ali'ye muz yedirildi
Ali.Dat banana.Nom eat.Caus.Pass.Past.3sg
'Ali was made to eat the banana.'

(87) muz yendi.
banana.Nom eat.Pass.Past.3sg
'A banana was eaten.'

(88) evde uyundu.
home.Loc sleep.Pass.Past.3sg
'It was slept at home.'

(89) annesi Ali'nin yenmesine izin verdi
mom.P3sg Ali.Gen eat.Pass.Inf.Dat permission.Nom give.Past.3sg
'His mother let Ali be eaten.'

(90) annesi Ali'nin kurtlar tarafından yenmesine izin
mom.P3sg Ali.Gen wolf.Pl by eat.Pass.Inf.Dat permission.Nom
verdi
give.Past.3sg
'His mother let Ali be eaten by the wolves.'

(91) Annesi Ali'nin yemesine izin verdi
Mom.P3sg Ali.Gen eat.Inf.Dat permission give.Past.3sg
'His mother let Ali eat'

B.6 Fall 2008 Meeting

(92) kızlar gitti
girl.Pl go.Past.3sg
'The girls left.'

(93) kızlar gitmedi
girl.Pl go.Neg.Past.3sg
'The girls did not leave.'

(94) hiçbir kız gitmedi
no girl.Nom go.Neg.Past.3sg
'No girl left.'

(95) erkekler Meryem'i gördü
boy.Pl Mary.Acc see.Past.3sg
'The boys saw Mary.'

(96) erkekler ve kızlar gitti
boy.Pl and girl.Pl go.Past.3sg
'The boys and girls left.'

(97) erkekler şarkı söyledi ve dans etti
boy.Pl song sing.Past and dance make.Past.3sg
'The boys sang and danced.'

(98) kek Meryem tarafından yendi
cake.Nom Mary.Nom by eat.Pass.Past.3sg
'The cake was eaten by Mary.'

(99) her zaman hava hakkında konuşurlar
all time.Nom weather.Nom about talk.Aor.3Pl
'They always talk about the weather.'

(100) o bir büyükelçi gibi davranır
she.Nom an ambassador.Nom as act.Aor.3sg
'She acts as an ambassador.'

(101) zeki kız
smart girl.Nom
'the smart girl'

(102) beş erkek
five boy.Nom
'five boys'

(103) 5 erkek
5 boy.Nom
'5 boys'

(104) beşinci erkek
five.Ord boy.Nom
'the fifth boy'

(105) erkeklerin beşi
boy.Pl.Gen five.Poss
'five of the boys'

(106) Can'ın gül-me-si
John.Gen laugh-Inf-Poss
'John's laughing'

(107) keki piş-ir-mek
cake.Acc bake-Caus-Inf
'baking the cake'

(108) Can'ın keki piş-ir-me-si
John.Gen cake.Acc bake-Caus-Inf-Poss
'John's baking the cake'

(109) yıkım
destruction.Nom
'the destruction'

(110) Roma'nın şehri yık-ma-sı
Rome.Gen city.Acc destroy-Inf-Poss
'Rome's destruction of the city'

(111) Meryem öğretmen-dir
Mary.Nom teacher-Cop
'Mary is a teacher.'

Appendix C

Sentence Test Suite

C.1 Main Text

Şebnem İşigüzel, *Öykümü Kim Anlatacak* 'Who will Tell My Story', p. 11-12

"Sonra ben öyle çok ağlayıp geceler boyunca telefon bekledim ki...

Kimse bana bu kötü büyüyü bozacak sihirli sözcüğü fısıldayamadı. Ben boğazında yara izi olmayan 'Suskun Adam'la mutlu olacağımı biliyordum. Onun eşi olabilirdim, çocuklarını doğurabilirdim, birbirimize, hiç bağırmadan, sonsuz güven ve mutluluk sunarak yaşayabilirdik. Ama o benim gibi düşünmedi. Benden kaçtı. Kaçtıkça daha da büyüdü, bir tutku oldu. Bu tutku zamanla bana acı vermeye başladı. Okulu ve işi bıraktım. Ağırlaşan ve giderek ölüme yaklaşan bir hastadan farksızdım. Çevremdekiler bana yardım edemiyorlardı. Bir gece uyandım. Giyinip dışarı çıktım. Hava soğuktu. Yürümeye başladım. Bu hoşuma gitti. Ben yürüdükçe gökyüzünün rengi de değişiyordu. Önce koyu bir griydi, martıların kirli tüylerine benzer bir renk almaya başlamıştı ki boğazında yara izi olmayan 'Suskun Adam'ın benim için neden bir tutkuya dönüştüğünü düşünmeye başladım. Yoksa her şey gibi onu da ben mi yaratmıştım? Bildiğim tek şey vardı: Ben ona yakındım. Sanki çok uzun yıllar onunla birlikte yaşamış, birlikte düşler görmüştüm. Psikologa bu yüzden gittim. Terapiler sonuç vermeyince iş hipnozla, geçmişte, çocukluğumda ya da onunla birlikteyken takıldığım noktayı bulmaya, belleğimden kazımaya kaldı. Ama doktorum bilinç bandımı geriye çok hızlı sardı ve ben

bir önceki yaşamıma gittim. Bir liman kentinde çocuklarıyla kaçmaya çalışan bir kadın. Kentin Müslümanların eline geçme olasılığı var. Muhteşem bir kent. Hangi yüzyılda, nerede ve kim olarak yaşadım?

Daha önce yaşamış olduğumu öğrenmek, bana, rengi beğenilmediği ya da solduğu için boyanılan bir kumaş parçasıymışım duygusu veriyor.

Kendime çiçek, taze meyve ve bir sürü renkli dergi alıyorum. Yolumun üzerindeki dev alışveriş merkezine girip vitrinlere bakıyorum. Rahatlıyorum. Çalışmamak güzel bir duygu. Bütün gün gezip dolaşıyorum. Bol bol uyuyup okuyorum. Salı ve Cuma günleri kütüphane günüm. Perşembeleri uzun yürüyüşler ve ziyaretler yapıyorum. Çarşamba, cumartesi, pazartesi psikologa gidiyorum. Bugün pazar, ama ben psikologa gitmek istiyorum. Randevu almaya bile gerek duymuyorum. Doktoruma sadece derin uykuların bana iyi geldiğini söylüyorum. Küçük seskaydedicim yine yanımda."

C.2 Sentences

```
1  sonra ben öyle çok ağlayıp geceler boyunca telefon bekledim ki
2  kimse bana bu kötü büyüyü bozacak sihirli sözcüğü fısıldayamadı
3  ben boğazında yara izi olmayan suskun adamla mutlu olacağımı
   biliyordum
4  onun eşi olabilirdim, çocuklarını doğurabilirdim, birbirimize,
   hiç bağırmadan, sonsuz güven ve mutluluk sunarak yaşayabilirdik
5  ama o benim gibi düşünmedi
6  benden kaçtı
7  kaçtıkça daha da büyüdü, bir tutku oldu
8  bu tutku zamanla bana acı vermeye başladı
9  okulu ve işi bıraktım
10 ağırlaşan ve giderek ölüme yaklaşan bir hastadan farksızdım
11 çevremdekiler bana yardım edemiyorlardı
12 bir gece uyandım
13 giyinip dışarı çıktım
14 hava soğuktu
15 yürümeye başladım
16 bu hoşuma gitti
17 ben yürüdükçe gökyüzünün rengi de değişiyordu
```

önce koyu bir griydi, martıların kirli tüylerine benzer bir renk almaya başlamıştı ki boğazında yara izi olmayan suskun adamın benim için neden bir tutkuya dönüştüğünü düşünmeye başladım
yoksa her şey gibi onu da ben mi yaratmıştım
bildiğim tek şey vardı
ben ona yakındım
sanki çok uzun yıllar onunla birlikte yaşamış, birlikte düşler görmüştüm
psikologa bu yüzden gittim
terapiler sonuç vermeyince iş hipnozla, geçmişte, çocukluğumda ya da onunla birlikteyken takıldığım noktayı bulmaya, belleğimden kazımaya kaldı
ama doktorum bilinç bandımı geriye çok hızlı sardı ve ben bir önceki yaşamıma gittim
bir liman kentinde çocuklarıyla kaçmaya çalışan bir kadın
kentin müslümanların eline geçme olasılığı var
muhteşem bir kent
hangi yüzyılda, nerede ve kim olarak yaşadım
daha önce yaşamış olduğumu öğrenmek, bana, rengi beğenilmediği ya da solduğu için boyanılan bir kumaş parçasıymışım duygusu veriyor
kendime çiçek, taze meyve ve bir sürü renkli dergi alıyorum
yolumun üzerindeki dev alışveriş merkezine girip vitrinlere bakıyorum
rahatlıyorum
çalışmamak güzel bir duygu

bütün gün gezip dolaşıyorum
bol bol uyuyup okuyorum
salı ve cuma günleri kütüphane günüm
perşembeleri uzun yürüyüşler ve ziyaretler yapıyorum
çarşamba, cumartesi, pazartesi psikologa gidiyorum
bugün pazar, ama ben psikologa gitmek istiyorum
randevu almaya bile gerek duymuyorum
doktoruma sadece derin uykuların bana iyi geldiğini söylüyorum
küçük seskaydedicim yine yanımda

Appendix D

Noun Phrase Test Suite

D.1 Simple Noun Phrases

Amerika'da
Candan
Candan'ın gözyaşları
Candan'ın sesi
Mina'nın bir repliğini
Mina'nın kişiliği üzerinde
Mozart'ın bir ezgisi
Tuğba'nın arkasındaki kapıya
Tuğba'nın bu aşırı güvenine
Tuğba'yı
alarm sistemi falan
altıncı katta
altıncı katın düğmesine
altıncı katın veznesinin
alışveriş gezegeninde
arkadaşım
arkadaşımızın doğum gününe
arkadaşının kolunu
ayak uçlarımı
ayaklarımdan
aynı tafralı eda
aşağı kattaki ana vezneye
bazıları
beni
beyaz bir çığlık
bir Alman dostum
bir alt katta
doğum günü hediyesi
düşselliğimizin eğlenceli yanı
evi
eşyalar
fiziksel yapısı
gazeteler
gelecek cevabı
gergin bir yüzle
gerçek bir üzüntüyle
gruplar
gümüş bir ırmak
gündüz düşlerinden
gündüz düşlerinin yazgısına
günü
gürültü
hangi camları
havaya
hemen merdivenlerin yanındaki kapı
herkes
hocam
iki arkadaş
iki genç kız
ilk haftalarda
kadının sesi
kadının titrek sesi
kanepeye
karşı taraftan

bir arkadaşım
bir bekleyişin şarkısı
bir bekçi falan
bir belediye arabası
bir bulut
bir cami
bir dakika bile
bir konser
bir konserde
bir müzik kutusuna
bir pazar
bir resim
bir yerden
bir şarkıyı
biraz mahcup bir eda
birileri
birisi
bu mağazada
bu parlak fikirlerin
bu tür toplantılara
buralarda
bütün eller
büyüleyici bir gerginlik
cuma günleri
ders
ders bitimine
dershane
dershanenin giriş kapısında
dershaneye
derslerin yoğunluğu nedeniyle
dinleyiciler
dinleyicilere
dostlukları
polisler
rakamların yolunu
sahnede
sahneye de
salona
salonda
salondaki tek mum
salonun elektriği
sanatçıların
sesindeki sıkkın ifadeyi
sesini
karşıdan gelen şu kızı
karşısındakinin gözlerinin içine
kendi notlarına yakın notlar
kendi oyun alanı
kendisi
kimi okullarda
kitaplarının üstüne
kocaman mağaza
kolundaki saate
koskoca mağazada
mağazamızın yüzde indirimi
mağazanın güvenliği
merdivenin son basamaklarında
merdivenlerden
merdivenleri
merdivenlerin yanındaki kapıdan
muavinlerle
mum
mumlar
mumların arasında
mumların belli belirsiz hışırtısı
mumların hışırtısı
ne kadar anlamlı bir konuşma
o büyülü ses
o telaş içinde
o tiyatrocu kızla
odanın içinde
okulu
oraya
otobüsün içinde
oyuncak ayılar
parfüm poşetini
parfüm reyonuna
zayıf bir kişilik
zayıf oyunculuk
zemin kata
zemin kattaki vezne
İngilizce sözcükleri
çalgıları
çağırma düğmesine
çelik duvarlarla
çelik duvarın üzerinde
çevremde
çevremizdeki binalara

seyirciler
sinek ilacı
sisler arasında
siyahlarla birlikte bir protesto yürüyüşünde
su sesi
söz konusu
söz konusu ezginin öyküsü

sözleri üzerine
sınıf başkan yardımcısı
tam tersini
tatsız gerçek
telefon
teybin düğmesine

teyp kasetleri
tezgahların arkasına
tezgahtar bayan
tezgahın üstünde
tuvaletin kapısını
tuvaletin yerini
upuzun bir yol
vitrin camlarını
vitrin falan
yanında da
yapılacak bir şey
yürekliliğim
yüzlerce mum

çıkış kapısına
çıplak çalgıların
ön planda
önceki kararlılığım

önümüzdeki derse
özel bir gün
öğrenciler

öğrencilere
öğretmen
öğretmenlerin peşinden
şarkı dinleme saatimi
şarkıcı kadınla
şarkıcı kadınla ilgili bir gazete haberini
şarkılarına
şarkının ikinci kısmı
şarkının ilk bölümü
şarkının içinde
şarkının içine
şarkının serüvenini de
şarkıyı
şaşkınlığını
şu müthiş zekanı
şu parçayı
şu siyah uzun saçlı olanı
şu çayı
şu öğrenci kalabalığından

D.2 Relative Clauses

Alper'in söylediği bir şeye
alanı dolduran on binlerce şarkıcı
anlattığım konuyu
az önce Tuğba ile Candan'ı getiren asansör
az önce durdukları reyonun tezgahına
bitip tükenmek bilmeyen bir yol
elindeki kitabı kapatan öğretmenin
elindeki klasörlerden Tuğba'ya ait olanı
en iyi arkadaşı şule'nin doğum günü partisine gideceğini
gözleriyle çevreyi araştıran Candan
hemen merdivenlerin yanındaki kapı

heyecandan tizleşmiş bir sesle
hiç görmediğim sokakları
insanlarla konuşurken de kendini sahnede sanıp kaptıran biri
insanın üzerine giden nefis biri
istediği markayı
kalan merdivenleri
karşı taraftan hızla gelen bir delikanlıdan aldığı omuz
karşıdan gelen şu kızı
kasanın bulunduğu kısımda
kentin hiç bilmediğim köşelerine
mağazanın camlarının arkasını çepeçevre saran çelik perde
neler yapabileceğimizi
olacak şey
otobüste, dolmuşta, sokak ortasında kurulan düşler
parfümeri reyonuyla ilgilenen hoş görünümlü bayan
parfümü içine koyduğu poşete
partiyi yapana
sahnede görünmeyen bir tip
sessiz sessiz ağlayan arkadaşına
siyahlarla birlikte bir protesto yürüyüşünde çekilmiş bir resim
son dakikada çıkan bu işten
soracağınız bir şey
söylediği sözün anlamını
uzandığım kanepeden
uzayıp giden bir yolu
veznede sıra bekleyen iki müşteri
vitrinde teşhir edilen malların arkasını
vitrine açılan bir giriş
yalnızca karşısındakinin gözlerinin içine bakan nefis biri
yalnızca uzayıp giden bir yol
yapılacak bir şey
yazdığı iki fişten birini
yere damlayan mum
yere saçılan kağıtları
yürümek yerine dalgalanan öyle tül gibi bir kadını
Şule'lere götürmek için hazırladığı kek paketini de
Şule'nin o çok beğendiği parfümü

D.3 Sentential Complements

Mina'yı sevmemen
Tuğba ile aynı dershaneyi seçmeleri

Tuğba ile ortak hediye almaya
Tuğba'nın kurtulmak için hiçbir hareket yapmamış olmasına
Tuğba'nın uzaklaşmasını
alarmın çalışmamasına
bir insanın gündüz düşleri içinde yaşamasının tehlikeli olabileceğini
bir kentin ortasında yitmek
bir şeyler ümit etmek
bitip tükenmek bilmeyen bir yol
bu gece doğum gününde parfümü götürmek
bunun eğlenceli bir yanı olduğunu
buradan çıkmanın bir yolunu
cüzdanımın kaybolması
daha erken gelebilmem
en azından masraf açısından yük olmamaya
en iyi arkadaşı Şule'nin doğum günü partisine gideceğini
gündüz düşü kurma işiyle
kapanma saati
kendi notlarına yakın notlar almasına
mumların söndürülmesi
onu satın almayı
onun iriyarı biri olduğunu
ortalarda dolaşmasına, koşuşturmasına, alabildiğine eğlenmesine
parfüm almak
sanatçının sahneye çıkışından
size dağıttığım testleri
sürekli bir yeniliğin tam ortasında olmak
vitrini hazırlamak
yitip gitmekten de
yitmek
yürümek
zekamın müthiş olduğunu
Şule'lere götürmek için hazırladığı kek paketini de
çayı ısıtmalarını
şarkının bittiğinin

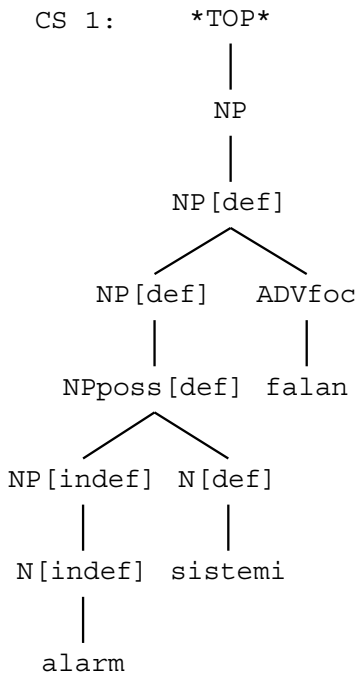
D.4 Coordination

arkadaşları, kişilikleri
elindeki kitapları ve büyükçe bir poşeti
fiziksel olarak zayıf, ince, çelimsiz bir tip
hem içeri girilmemesi, hem de dışarı çıkılmaması
hiç görmediğim sokakları, insanları

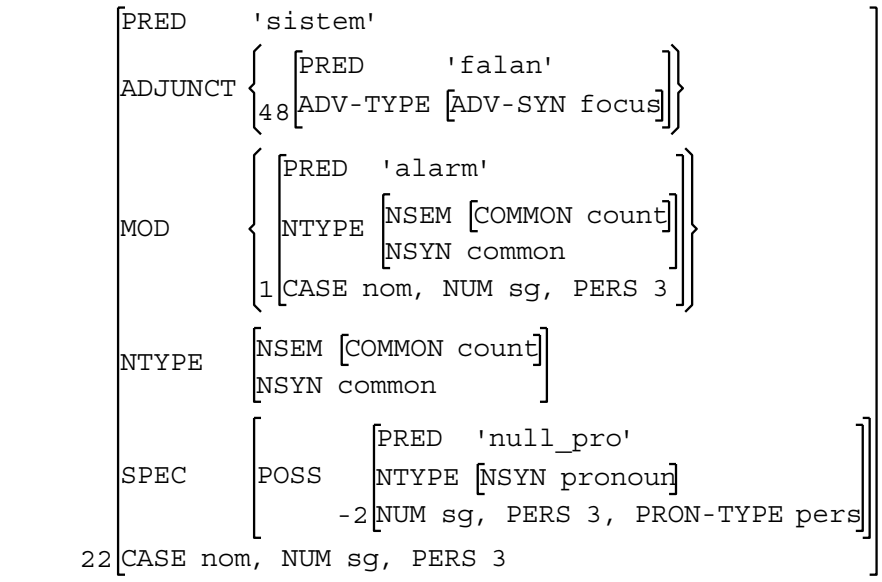
iriyarı, sarışın, ama sahnede görünmeyen bir tip
kek, pasta, çörek gibi bir şeyler
kitapları ve poşetleri
kitaplarını ve poşetini
kitaplarını, defterlerini
kırık, acı, ama dirençli bir bekleyişin
müthiş bir aydınlık ve inanılmaz bir sessizlik
ortalarda dolaşmasına, koşuşturmasına, alabildiğine eğlenmesine
otobüste, dolmuşta, sokak ortasında
otobüste, dolmuşta, sokak ortasında kurulan düşler
saatin akreple yelkovanı
uzun boylu, incecik bir kız
vitrine açılan bir giriş, bir kapı
yorgun ve isteksiz ayaklarını

D.5 XLE Parses of Selected Phrases

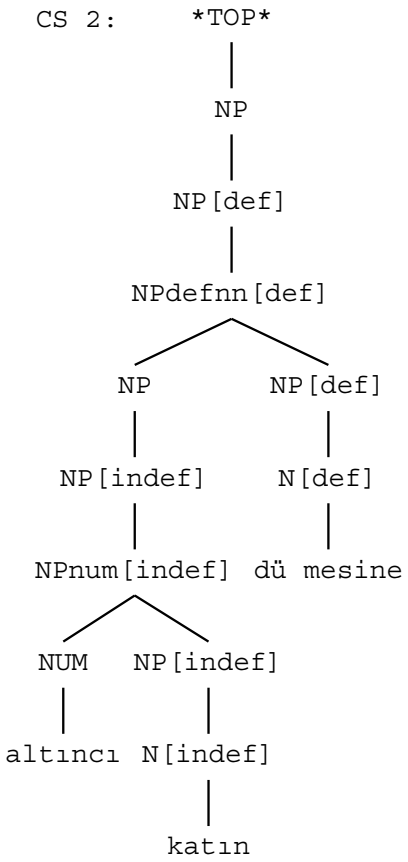
(112) alarm sistemi falan
alarm.Nom system.P3sg etc
'alarm system etc.'



"alarm sistemi falan"



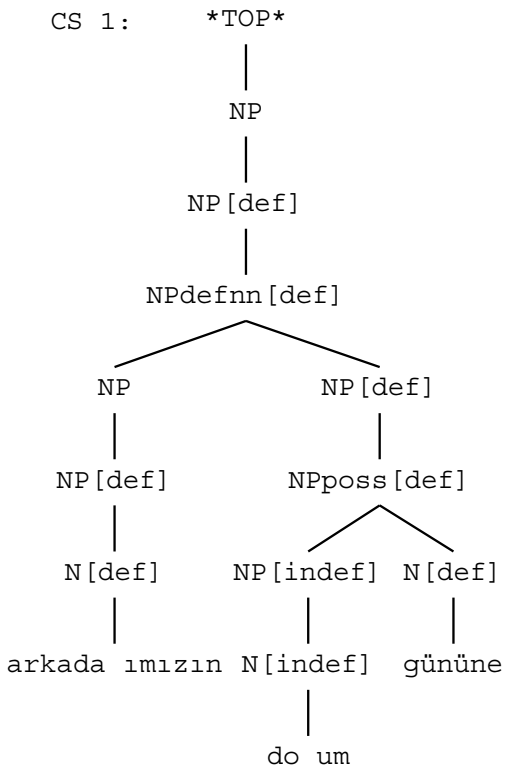
(113) altıncı katın düğmesine
sixth floor.Gen button.P3sg.Dat
'to the button of the sixth floor'



"altıncı katın dü mesine"

$$\begin{bmatrix} \text{PRED} & \text{'dü me'} \\ \text{NTYPE} & \begin{bmatrix} \text{NSEM} & [\text{COMMON count}] \\ \text{NSYN} & \text{common} \end{bmatrix} \\ \text{SPEC} & \begin{bmatrix} \text{POSS} & \begin{bmatrix} \text{PRED} & \text{'kat'} \\ \text{NTYPE} & \begin{bmatrix} \text{CHECK} & [_\text{EXPLICIT} \ _\text{poss}] \\ \text{NSEM} & [\text{COMMON count}] \\ \text{NSYN} & \text{common} \end{bmatrix} \\ \text{SPEC} & \begin{bmatrix} \text{NUMBER} & 1\begin{bmatrix} \text{PRED} & \text{'altıncı'} \\ \text{NUMBER-TYPE} & \text{ord} \end{bmatrix} \end{bmatrix} \\ & 63[\text{CASE gen, NUM sg, PERS 3}] \end{bmatrix} \\ & 124[\text{CASE dat, NUM sg, PERS 3}] \end{bmatrix} \end{bmatrix}$$

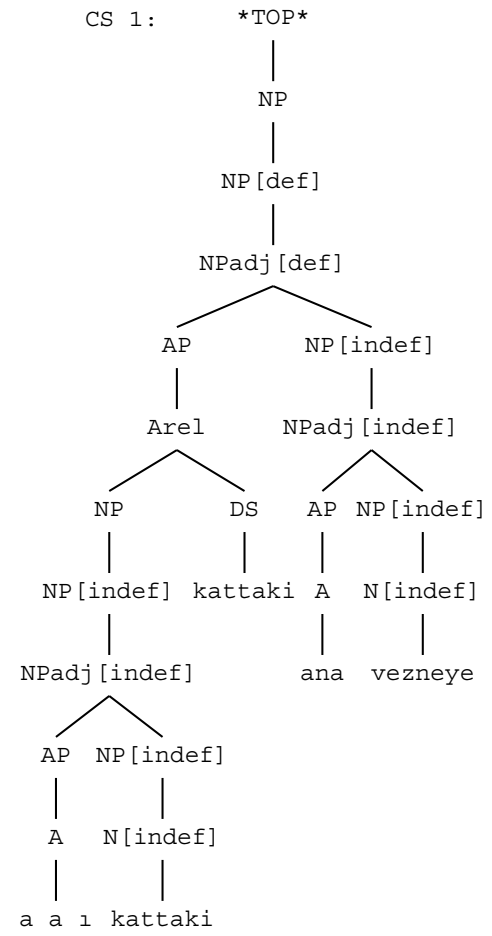
(114) arkadaşımızın doğum gününe
friend.P2pl.Gen birth.Nom day.P3sg.Dat
'to our friend's birthday'



"arkada ımızın do um gününe"

$$\begin{bmatrix}
\text{PRED} & \text{'gün'} \\
\text{MOD} & \left\{ \begin{bmatrix} \text{PRED} & \text{'do um'} \\ \text{NTYPE} & \begin{bmatrix} \text{NSEM} & [\text{COMMON count}] \\ \text{NSYN} & \text{common} \end{bmatrix} \end{bmatrix} \right\} \\
 & 22\begin{bmatrix} \text{CASE nom, NUM sg, PERS 3} \end{bmatrix} \\
\text{NTYPE} & [\text{NSEM } [\text{TIME } +]] \\
\text{SPEC} & \begin{bmatrix} \text{POSS} & \begin{bmatrix} \text{PRED} & \text{'arkada '} \\ \text{NTYPE} & \begin{bmatrix} \text{CHECK} & [_\text{EXPLICIT } _\text{poss}] \\ \text{NSEM} & [\text{COMMON count}] \\ \text{NSYN} & \text{common} \end{bmatrix} \\ \text{SPEC} & \begin{bmatrix} \text{POSS} & \begin{bmatrix} \text{PRED} & \text{'null_pro'} \\ \text{NTYPE} & [\text{NSYN pronoun}] \end{bmatrix} \\ -1 & [\text{NUM pl, PERS 1, PRON-TYPE pers}] \end{bmatrix} \end{bmatrix} \\ 1 & [\text{CASE gen, NUM sg, PERS 3}] \end{bmatrix} \\
61 & [\text{CASE dat, NUM sg, PERS 3}]
\end{bmatrix}$$

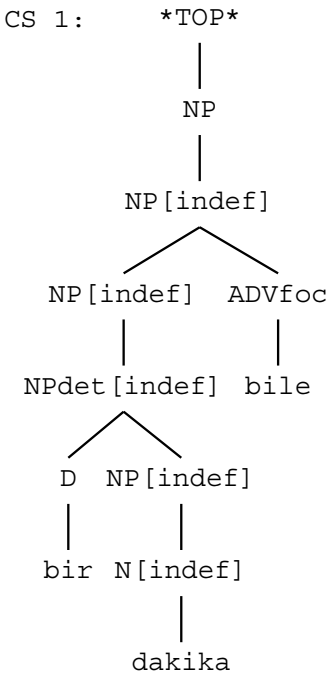
(115) aşağı kattaki ana vezneye
lower floor.Loc.Rel main pay.desk.Dat
'main pay desk at the lower floor'



"a a ı kattaki ana vezneye"

$$\begin{bmatrix}
\text{PRED} & \text{'vezne'} \\
\text{ADJUNCT} & \left\{ \begin{bmatrix}
\text{PRED} & \text{'ki<[28:kat]>'} \\
\text{OBJ} & \begin{bmatrix}
\text{PRED} & \text{'kat'} \\
\text{ADJUNCT} & \left\{ {}_{1}\begin{bmatrix} \text{PRED 'a a ı'} \\ \text{ATYPE attributive, DEGREE positive} \end{bmatrix} \right\} \\
\text{NTYPE} & \begin{bmatrix} \text{NSEM} & [\text{COMMON count}] \\ \text{NSYN} & \text{common} \end{bmatrix} \\
{}_{28}\text{CASE loc, NUM sg, PERS 3} & {}_{28}
\end{bmatrix}
\end{bmatrix}, \right. \\
& \left. \begin{bmatrix}
\text{PRED 'ana'} \\
\text{ATYPE attributive, DEGREE positive} \\
{}_{55}\langle \text{s} \quad ([28{:}ki]) \rangle
\end{bmatrix} \right\} \\
\text{NTYPE} & \begin{bmatrix} \text{NSEM} & [\text{COMMON count}] \\ \text{NSYN} & \text{common} \end{bmatrix} \\
{}_{103}\text{CASE dat, NUM sg, PERS 3} &
\end{bmatrix}$$

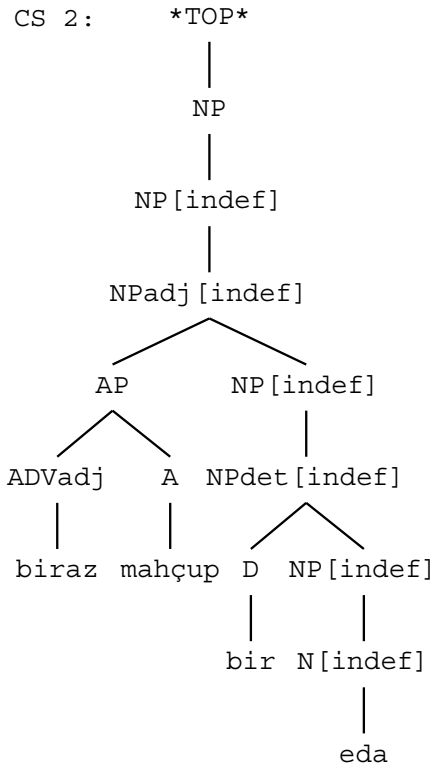
(116) bir dakika bile
a moment.Nom even
'even a moment'



"bir dakika bile"

$$\begin{bmatrix} \text{PRED} & \text{'dakika'} \\ \text{ADJUNCT} & \left\{ {}_{49}\begin{bmatrix} \text{PRED} & \text{'bile'} \\ \text{ADV-TYPE} & \begin{bmatrix} \text{ADV-SYN} & \text{focus} \end{bmatrix} \end{bmatrix} \right\} \\ \text{NTYPE} & \begin{bmatrix} \text{NSEM} & \begin{bmatrix} \text{TIME} & + \end{bmatrix} \end{bmatrix} \\ \text{SPEC} & {}_{1}\begin{bmatrix} \text{DET} & \begin{bmatrix} \text{PRED} & \text{'bir'} \\ \text{DET-TYPE} & \text{indef} \end{bmatrix} \end{bmatrix} \\ {}_{34}\begin{bmatrix} \text{CASE nom, NUM sg, PERS 3} \end{bmatrix} \end{bmatrix}$$

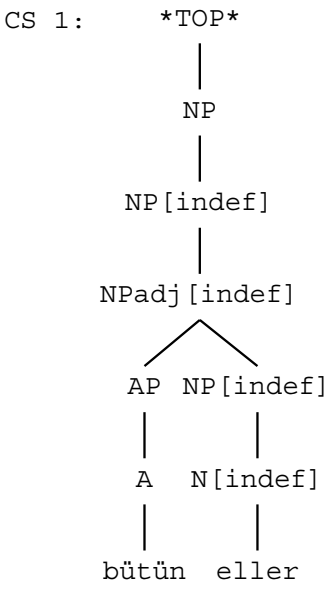
(117) biraz mahçup bir eda
a.bit embarrassed a expression.Nom
'a bit of an embarrassed expression'



"biraz mahçup bir eda"

$$\begin{bmatrix} \text{PRED} & \text{'eda'} \\ \text{ADJUNCT} & \left\{ \begin{bmatrix} \text{PRED} & \text{'mahçup'} \\ \text{ADJUNCT} & \left\{ \begin{bmatrix} \text{PRED} & \text{'biraz'} \\ \text{ADJUNCT-TYPE} & \text{degree, DEGREE positive} \end{bmatrix}_{1} \right\} \end{bmatrix} \right. \\ & \left. \begin{bmatrix} \text{ATYPE} & \text{attributive, DEGREE positive} \end{bmatrix}_{15} \right\} \\ \text{NTYPE} & \begin{bmatrix} \text{NSEM} & \begin{bmatrix} \text{COMMON} & \text{count} \end{bmatrix} \\ \text{NSYN} & \text{common} \end{bmatrix} \\ \text{SPEC} & \begin{bmatrix} \text{DET} & \begin{bmatrix} \text{PRED} & \text{'bir'} \\ \text{DET-TYPE} & \text{indef} \end{bmatrix} \end{bmatrix}_{26} \\ \begin{bmatrix} \text{CASE} & \text{nom, NUM sg, PERS 3} \end{bmatrix}_{59} \end{bmatrix}$$

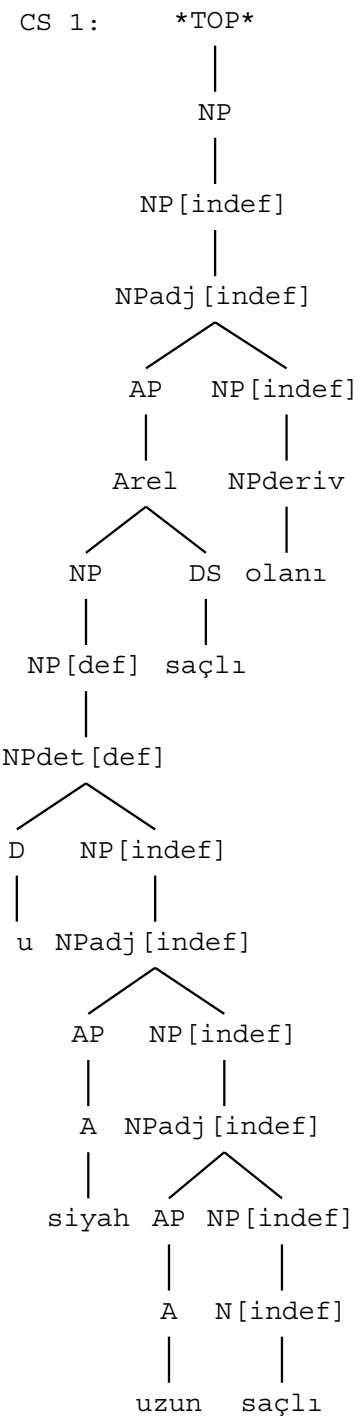
(118) bütün eller
all hand.Pl
'all hands'



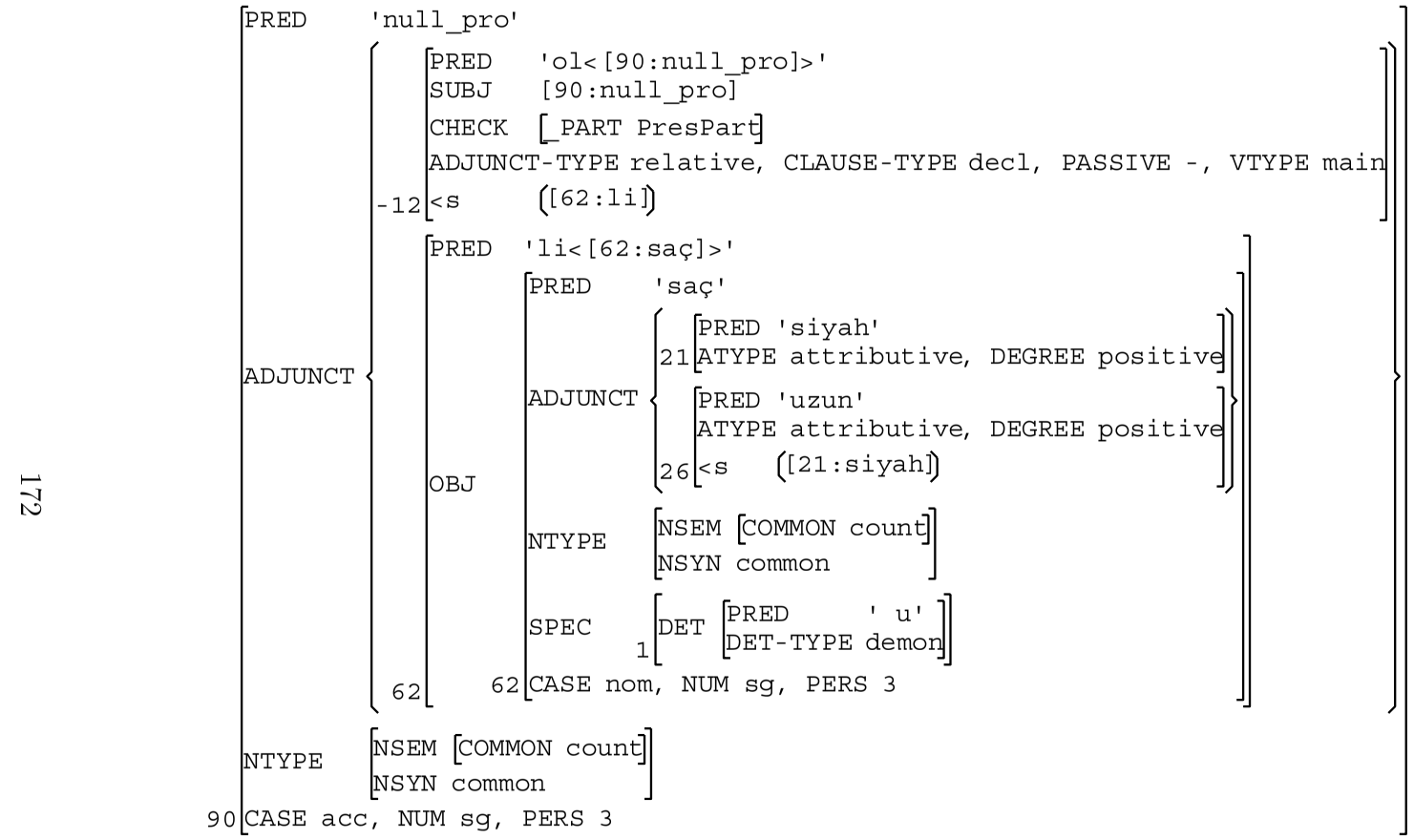
"bütün eller"

$$\begin{bmatrix} \text{PRED} & \text{'el'} \\ \text{ADJUNCT} & \left\{ 1 \begin{bmatrix} \text{PRED} & \text{'bütün'} \\ \text{ATYPE} & \text{attributive, DEGREE positive} \end{bmatrix} \right\} \\ \text{NTYPE} & \begin{bmatrix} \text{NSEM} & [\text{COMMON count}] \\ \text{NSYN} & \text{common} \end{bmatrix} \\ 12 & \begin{bmatrix} \text{CASE} & \text{nom, NUM pl, PERS 3} \end{bmatrix} \end{bmatrix}$$

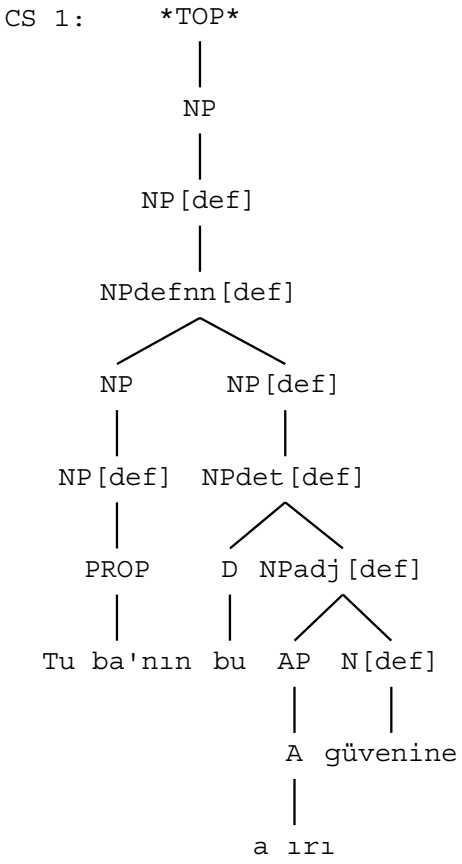
(119) şu siyah uzun saçlı olanı
that black long hair.With be.PresPart.ZeroDeriv.Acc
'that one with long black hair'



" u siyah uzun saçlı olanı"



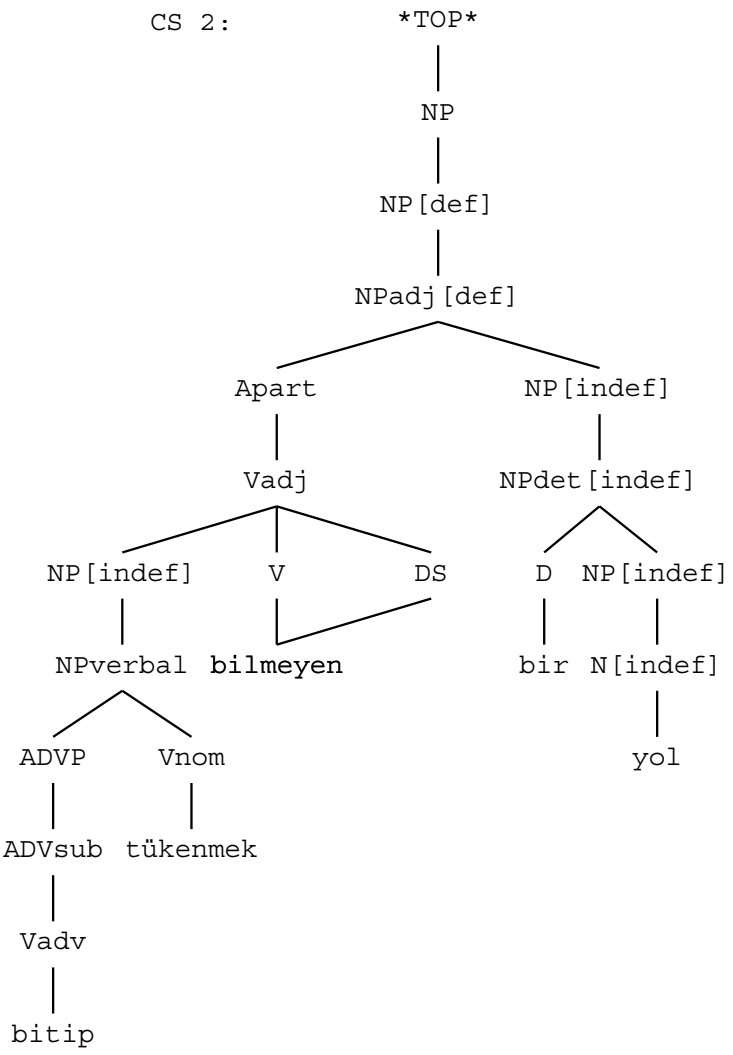
(120) Tuğba'nın bu aşırı güvenine
Tuğba.Gen this over confidence.P3sg.Dat
'to this over confidence of Tuğba'



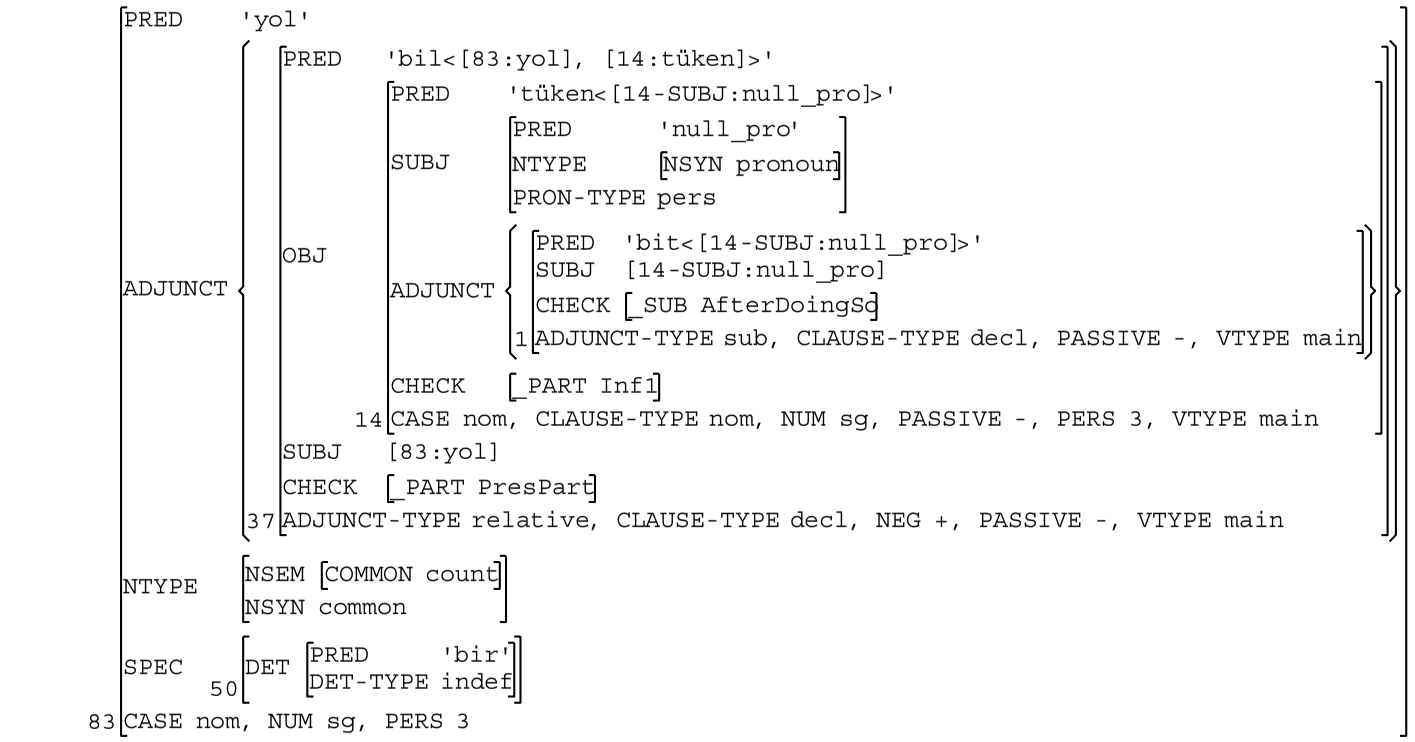
"Tu ba'nın bu a ırı güvenine"

$$\begin{bmatrix}
\text{PRED} & \text{'güven'} \\
\text{ADJUNCT} & \left\{ {}_{54}\begin{bmatrix}\text{PRED 'a ırı'} \\ \text{ATYPE attributive, DEGREE positive}\end{bmatrix} \right\} \\
\text{NTYPE} & \begin{bmatrix}\text{NSEM } \begin{bmatrix}\text{COMMON count}\end{bmatrix} \\ \text{NSYN common}\end{bmatrix} \\
\text{SPEC} & \begin{bmatrix}
\text{DET} & \begin{bmatrix}\text{PRED 'bu'} \\ \text{DET-TYPE demon}\end{bmatrix} \\
\text{POSS} & \begin{bmatrix}
\text{PRED} & \text{'Tu ba'} \\
\text{NTYPE} & \begin{bmatrix}\text{CHECK } \begin{bmatrix}\text{_EXPLICIT _poss}\end{bmatrix} \\ \text{NSEM } \begin{bmatrix}\text{PROPER } \begin{bmatrix}\text{PROPER-TYPE name}\end{bmatrix}\end{bmatrix} \\ \text{NSYN proper}\end{bmatrix} \\
{}_{1}\text{CASE gen, NUM sg, PERS 3}
\end{bmatrix}
\end{bmatrix}_{34} \\
{}_{67}\text{CASE dat, NUM sg, PERS 3}
\end{bmatrix}$$

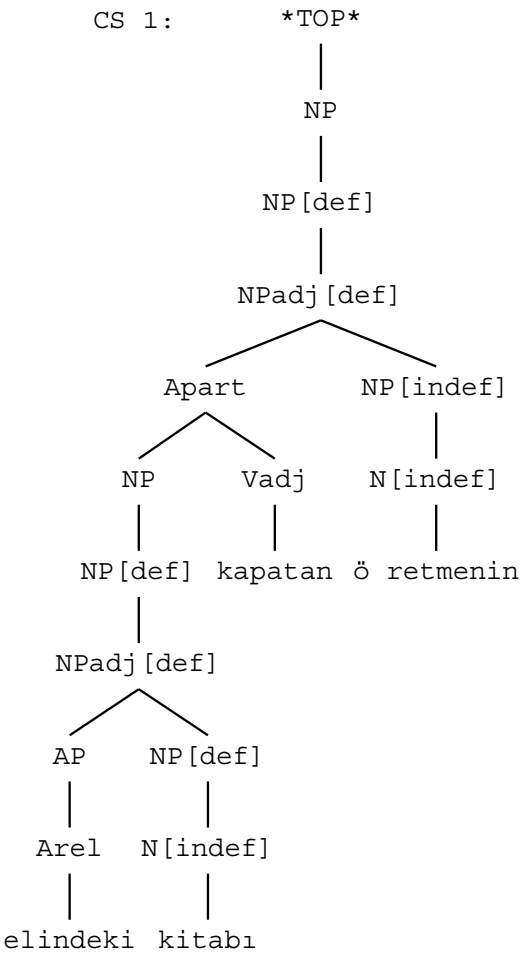
(121) bitip tükenmek bilmeyen bir yol
finish.AfterDoingSo exhaust.Inf know.Neg.Prespart a road.Nom
'a never ending road'



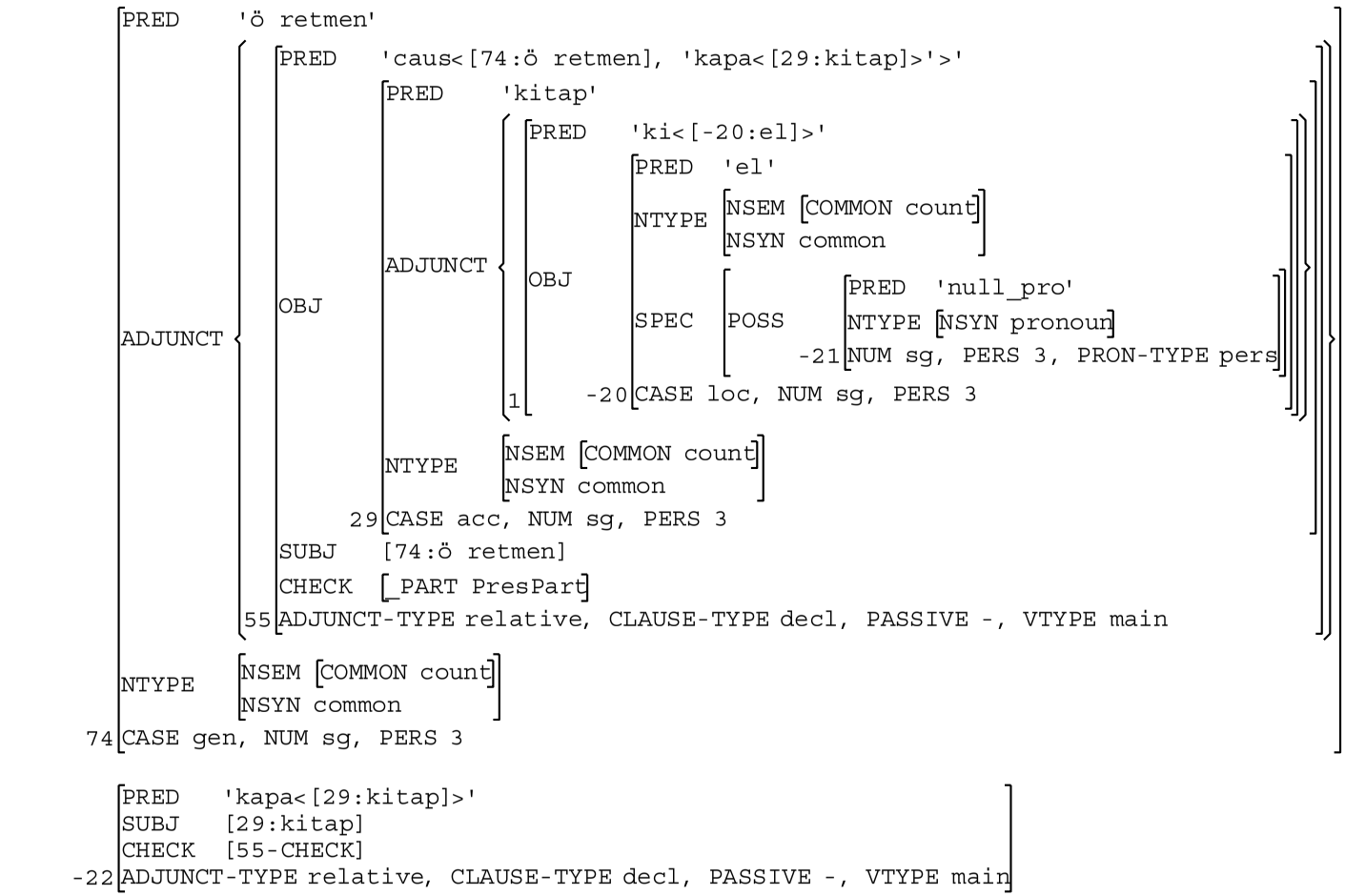
"bitip tükenmek bilmeyen bir yol"



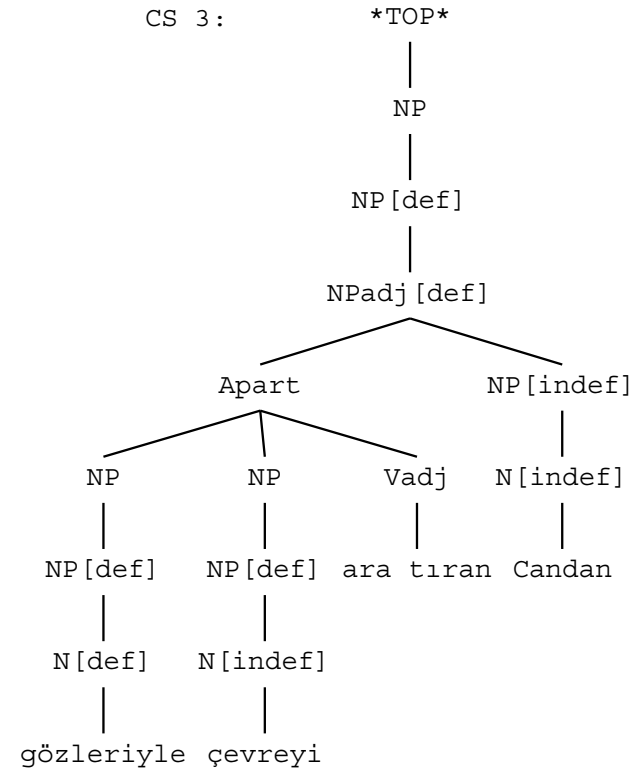
(122) elindeki kitabı kapatan öğretmenin
hand.P3sg.Loc.Rel book.Acc close.PresPart teacher.Gen
'of the teacher who closes the book in her hands'



"elindeki kitabı kapatan ö retmenin"



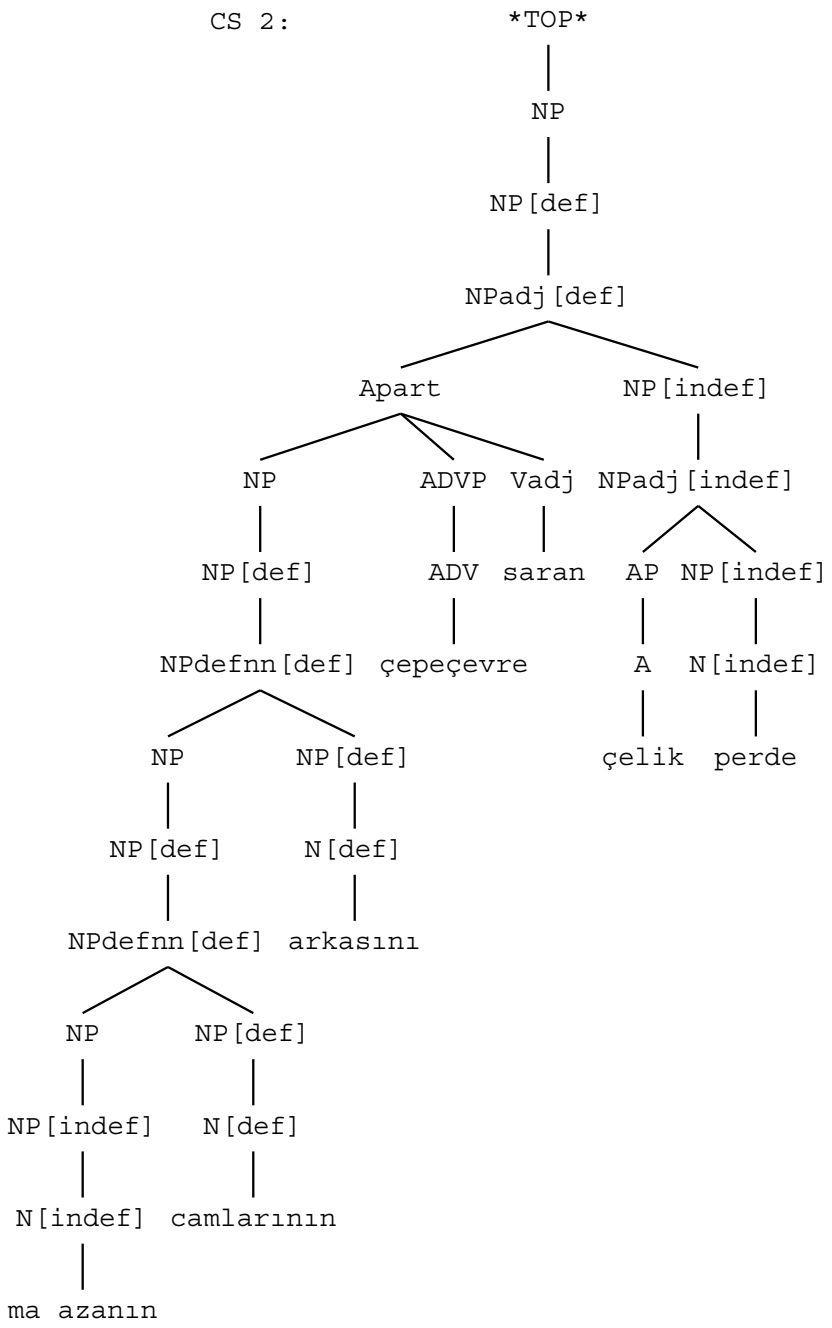
(123) gözleriyle çevreyi araştıran Candan
eye.Pl.P3sg.With around.Acc explore.PresPart Candan.Nom
'Candan who is exploring the around with her eyes'



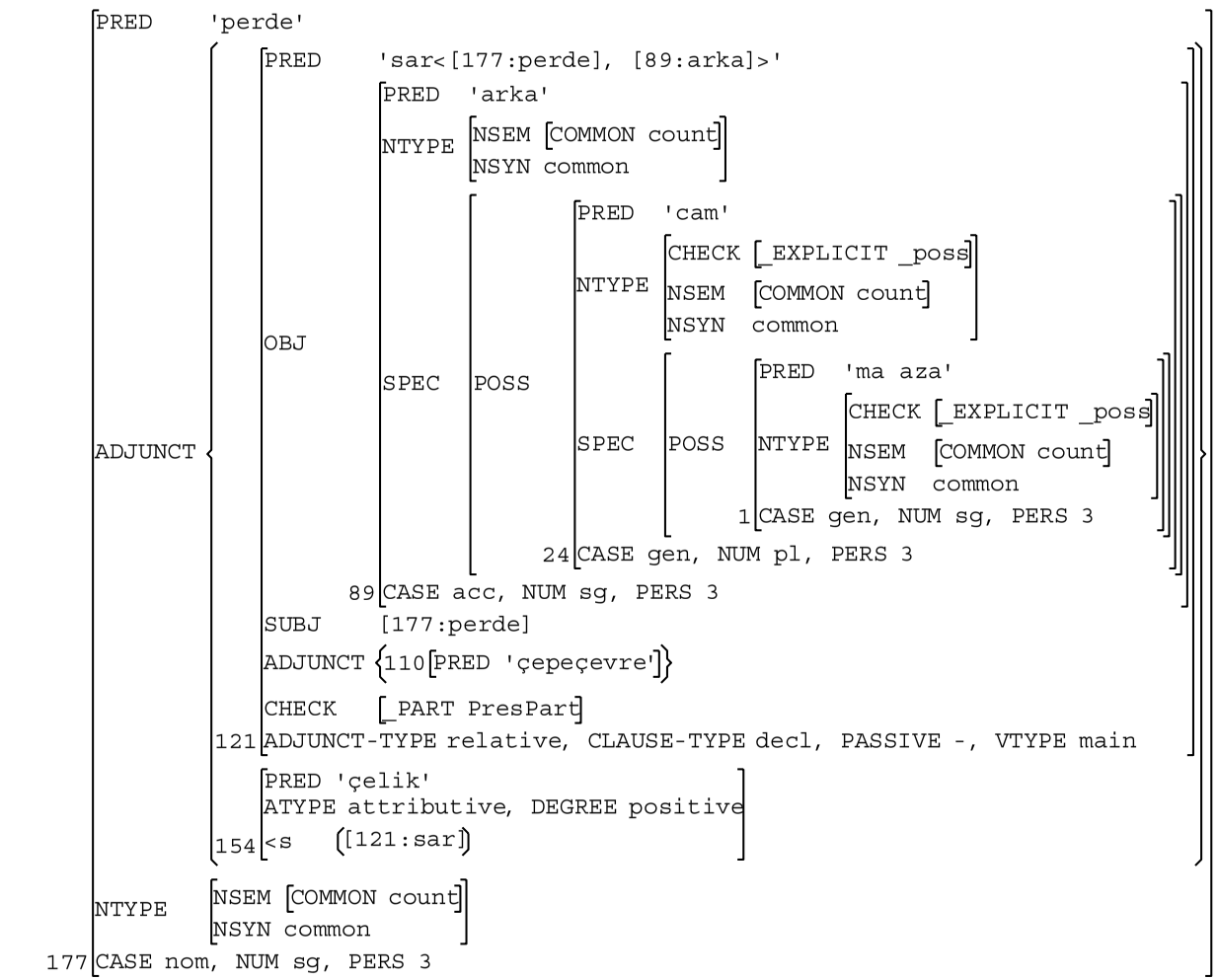
"gözleriyle çevreyi ara tıran Candan"

```
[PRED     'Candan'
 ADJUNCT {[PRED     'ara tır<[128:Candan], [65:çevre]>'
           OBJ      [PRED   'çevre'
                     NTYPE [NSEM [COMMON count]
                            NSYN common]]
                 65[CASE acc, NUM sg, PERS 3]
           SUBJ     [128:Candan]
           ADJUNCT {[PRED   'göz'
                     NTYPE [NSEM [COMMON count]
                            NSYN common]]
                    [SPEC  [POSS [PRED   'null_pro'
                                  NTYPE [NSYN pronoun]]
                           -9[NUM sg, PERS 3, PRON-TYPE pers]]]
                    1[CASE inst, NUM pl, PERS 3]]
           CHECK    [_PART PresPart]
          86[ADJUNCT-TYPE relative, CLAUSE-TYPE decl, PASSIVE -, VTYPE main]}
 NTYPE    [NSEM [PROPER [PROPER-TYPE name]]
           NSYN proper]
128[CASE nom, NUM sg, PERS 3]]
```

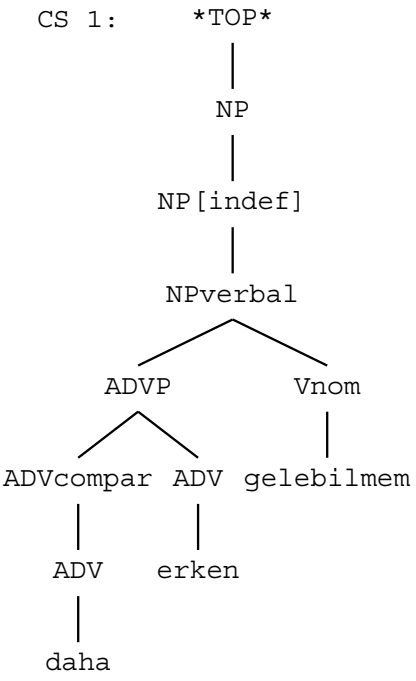
(124) mağazanın camlarının arkasını çepeçevre saran çelik
store.gen window.Pl.P3sg.Gen rear.side.P3sg all.around cover.PresPart steel
perde
panel.Nom
'the steel panel that covers all of the rear sides of the windows of the store'



"ma azanın camlarının arkasını çepeçevre saran çelik perde"



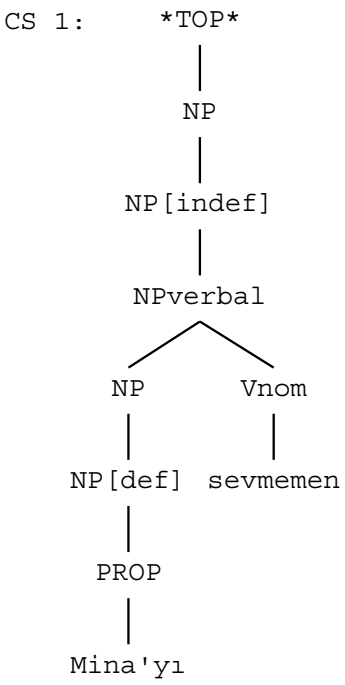
(125) daha erken gelebilmem
more early come.Able.Inf.P1sg.Nom
'that I can come earlier'



"daha erken gelebilmem"

```
[PRED    'yabil<[85-SUBJ:null_pro], [-3:gel]>'
 SUBJ    [PRED  'null_pro'
          NTYPE [NSYN pronoun]
          NUM sg, PERS 1, PRON-TYPE pers]
 XCOMP   [PRED 'gel<[85-SUBJ:null_pro]>'
          SUBJ [85-SUBJ:null_pro]
       -3 CLAUSE-TYPE decl, PASSIVE -, VTYPE main]
 ADJUNCT {[PRED    'erken'
           ADJUNCT {[PRED 'daha'
                  1 ADJUNCT-TYPE degree, DEGREE positive]}
          18 DEGREE  comparative]}
 CHECK   [_PART Inf2]
85 CASE nom, CLAUSE-TYPE nom, NUM sg, PASSIVE -, PERS 3, VTYPE main]
```

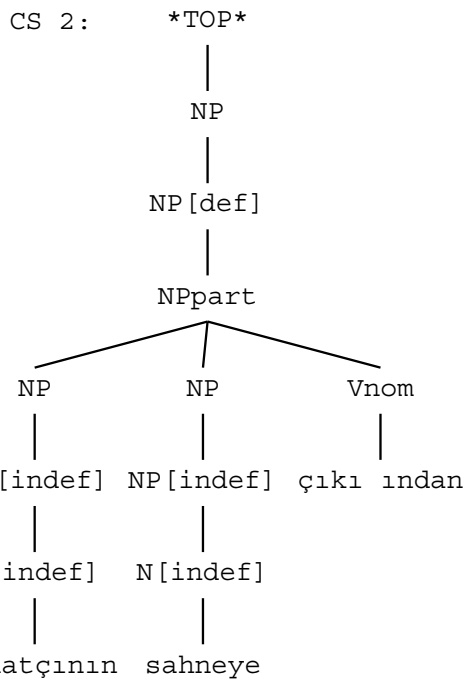
(126) Mina'yı sevmemen
Mina.Acc like.Neg.Inf.P2sg.Nom
'that you do not like Mina'



"Mina'yı sevmemen"

$$\begin{bmatrix} \text{PRED} & \text{'sev<[32-SUBJ:null_pro], [1:Mina]>'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'null_pro'} \\ \text{NTYPE} & \begin{bmatrix} \text{NSYN} & \text{pronoun} \end{bmatrix} \\ \multicolumn{2}{l}{\text{NUM sg, PERS 2, PRON-TYPE pers}} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'Mina'} \\ \text{NTYPE} & \begin{bmatrix} \text{NSEM} & \begin{bmatrix} \text{PROPER} & \begin{bmatrix} \text{PROPER-TYPE} & \text{name} \end{bmatrix} \end{bmatrix} \\ \text{NSYN} & \text{proper} \end{bmatrix} \\ 1 \begin{bmatrix} \multicolumn{2}{l}{\text{CASE acc, NUM sg, PERS 3}} \end{bmatrix} \end{bmatrix} \\ \text{CHECK} & \begin{bmatrix} \text{_PART} & \text{Inf2} \end{bmatrix} \\ 32 \begin{bmatrix} \multicolumn{2}{l}{\text{CASE nom, CLAUSE-TYPE nom, NEG +, NUM sg, PASSIVE -, PERS 3, VTYPE main}} \end{bmatrix} \end{bmatrix}$$

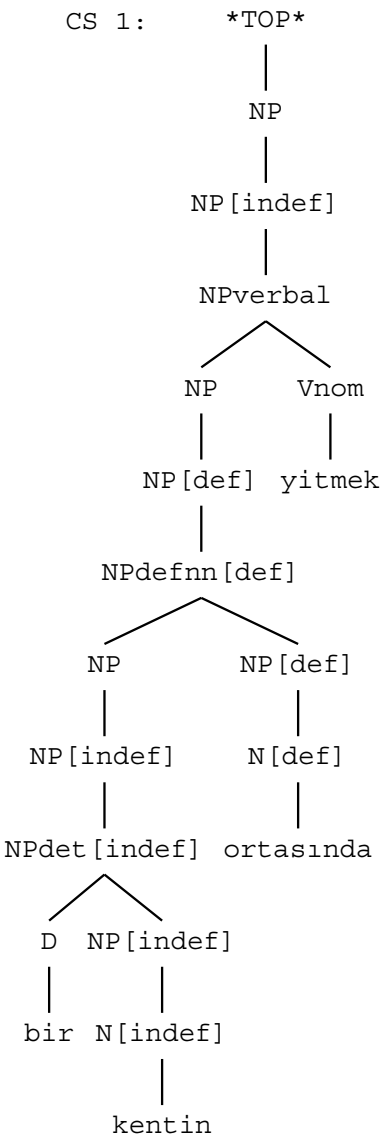
(127) sanatçının sahneye çıkışından
artist.Gen stage.Dat get.Inf.P3sg.Abl
'from the artist's getting to the stage'



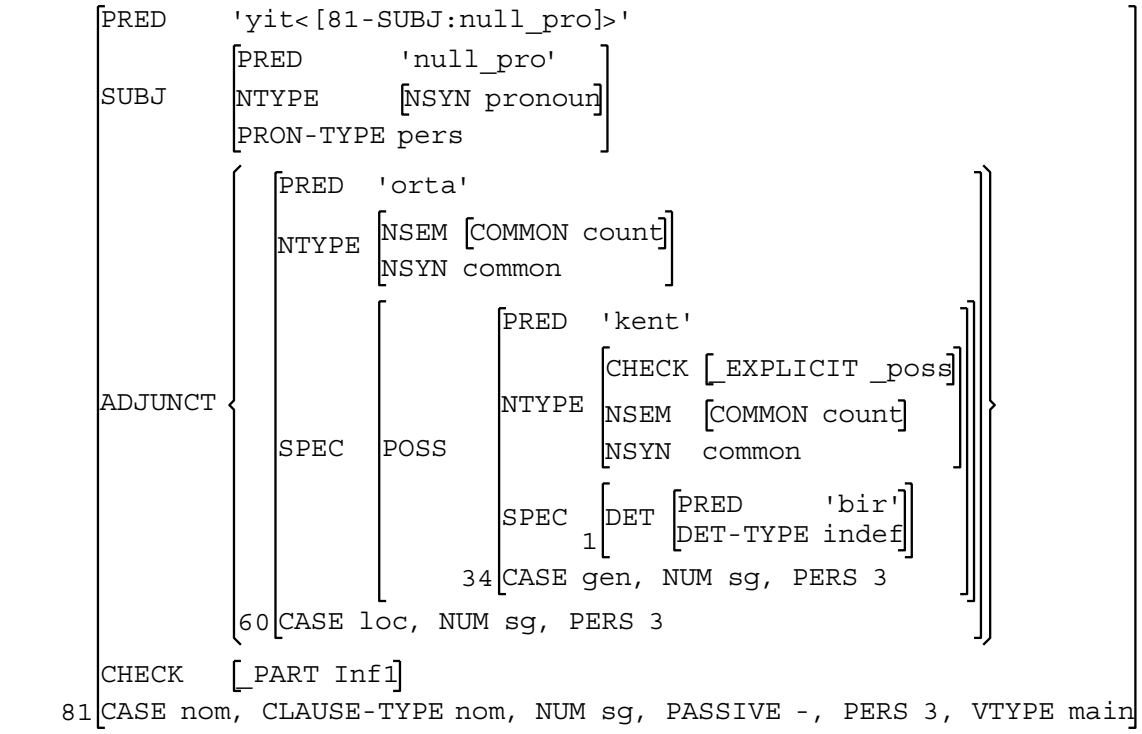
"sanatçının sahneye çıkı ından"

$$\begin{bmatrix}
\text{PRED} & \text{'çık<[1:sanat]>'} \\
\text{SUBJ} & \begin{bmatrix}
\text{PRED} & \text{'sanat'} \\
\text{CHECK} & [\text{_EXPLICIT _subj}] \\
\text{DERIV} & \left\{ \text{-4} \begin{bmatrix} \text{DERIV-FORM ci, DERIV-SEM agt} \end{bmatrix} \right\} \\
\text{NTYPE} & \begin{bmatrix} \text{NSEM} & [\text{COMMON count}] \\ \text{NSYN} & \text{common} \end{bmatrix} \\
1 & \text{CASE gen, NUM sg, PERS 3}
\end{bmatrix} \\
\text{ADJUNCT} & \left\{ \begin{bmatrix}
\text{PRED} & \text{'sahne'} \\
\text{NTYPE} & \begin{bmatrix} \text{NSEM} & [\text{COMMON count}] \\ \text{NSYN} & \text{common} \end{bmatrix} \\
27 & \text{CASE dat, NUM sg, PERS 3}
\end{bmatrix} \right\} \\
\text{CHECK} & [\text{_PART Inf3}] \\
48 & \text{CASE abl, CLAUSE-TYPE nom, NUM sg, PASSIVE -, PERS 3, VTYPE main}
\end{bmatrix}$$

(128) bir kentin ortasında yitmek
a city.Gen middle.P3sg.Loc get.lost.Inf
'to get lost in the middle of a city'



"bir kentin ortasında yitmek"



Bibliography

Judith Aissen. Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory*, 21:435–483, 2003.

Judith Aissen and Jorge Hankamer. Lexical extension and grammatical transformations. In *Proceedings of BLS*, volume 6, Berkeley, 1980. Berkeley Linguistic Society.

Alex Alsina. A theory of complex predicates: evidence from causatives in Bantu and Romance. In Alex Alsina, Joan Bresnan, and Peter Sells, editors, *Complex Predicates*, pages 203–246. CSLI Publications, Stanford, 1997.

Alex Alsina and Smita Joshi. Parameters in causative constructions. In Lise M. Dobrin, Lynn Nichols, and Rosa M. Rodriguez, editors, *Papers from the 27th Regional Meeting of the Chicago Linguistic Society (CLS)*, pages 1–16, 1991.

Önsel Armağan. Lingbrowser: A NLP based browser for linguistic information. Master's thesis, Sabancı University Department of Computer Science, 2008.

Zekiye Ceyda Arslan Kechriotis. *Case as an Uninterpretable Feature*. PhD thesis, Boğaziçi University, 2006.

Chris Barker, Jorge Hankamer, and John Moore. Wa and Ga in Turkish. In Katarzyna Dziwirek, Patrick Farrell, and Errapel Meijas-Bikandi, editors, *Grammatical Relations*. CSLI, 1990.

Kenneth Beesley and Lauri Karttunen. *Finite State Morphology*. CSLI Publications, Stanford, CA, 2003.

Orhan Bilgin, Özlem Çetinoğlu, and Kemal Oflazer. Building a Wordnet for Turkish. *Romanian Journal of Information Science and Technology*, 7(1-2):163–172, 2004.

Kersit Börjars and Nigel Vincent. Objects and OBJ. In Miriam Butt and Tracy Holloway King, editors, *On-Line Proceedings of the LFG08 Conference*, Sydney, 2008. CSLI Publications.

Cem Bozşahin. The combinatory morphemic lexicon. *Computational Linguistics*, 28 (2):145–186, 2002. URL `http://www.aclweb.org/anthology/J02-2002.pdf`.

Joan Bresnan. *Lexical-Functional Syntax*. Blackwell Publishers, Oxford, 2001.

Joan Bresnan. The passive in lexical theory. In Joan Bresnan, editor, *The Mental Representation of Grammatical Relations*, pages 3–86. MIT Press, Cambridge, 1982.

Joan Bresnan and Ronald M. Kaplan. Introduction: grammars as mental representations of language. In Joan Bresnan, editor, *The Mental Representation of Grammatical Relations*, pages xvii–iii. The MIT Press, Cambridge, MA, 1982.

Joan Bresnan and Sam A. Mchombo. The lexical integrity principle: evidence from Bantu. *Natural Language and Linguistic Theory*, 13(2):181–254, 1995.

Joan Bresnan and John Mugane. Agentive nominalizations in Gikuyu and the theory of mixed categories. In Miriam Butt, Mary Dalrympe, and Tracy Holloway King, editors, *Intelligent Linguistic Architectures: Variations on themes by Ronald M. Kaplan*. CSLI Publications, 2006.

Joan Bresnan and Annie Zaenen. Deep unaccusativity in LFG. In Katarzyna Dziwirek, Patrick Farrell, and Errapel Mejías-Bikandi, editors, *Grammatical Relations: A Cross-Theoretical Perspective*, pages 45–57. CSLI Publications, Stanford, 1990.

Michael Burke, Olivia Lam, Rowena Chan, Aoife Cahill, Ruth ODonovan, Adams Bodomo, Josef van Genabith, and Andy Way. Treebank-based acquisition of a Chinese Lexical-Functional grammar. In *The 18th Pacific Asia Conference on Language, Information and Computation*, Tokyo, Japan, 2004.

Miriam Butt. From the lexicon to the syntax and what is a complex predicate, 2003. in ESSLLI 2006 Lecture Notes.

Miriam Butt. *Theories of Case*. Cambridge University Press, Cambridge, 2006.

Miriam Butt. *The Structure of Complex Predicates in Urdu*. CSLI Publications, Stanford, California, 1995.

Miriam Butt. Constraining argument merger through aspect. In Erhard Hinrichs, Andreas Kathol, and Tsuneko Nakazawa, editors, *Complex Predicates in Nonderivational Syntax*, pages 73–113. Academic Press, New York, 1998.

Miriam Butt and Tracy Holloway King. Restriction for morphological valency alternations: The Urdu causative. In Miriam Butt, Mary Dalrympe, and Tracy Holloway King, editors, *Intelligent Linguistic Architectures: Variations on Themes by Ronald M. Kaplan*, pages 235–258. CSLI Publications, Stanford, 2006.

Miriam Butt, María-Eugenia Niño, and Frédérique Segond. *A Grammar Writer's Cookbook*. CSLI Publications, Stanford, 1999.

Miriam Butt, Tracy Holloway King, and Gillian Ramchand. Complex predication: Who made the child pinch the elephant? In Linda Uyechi and Lian Hee Wee, editors, *Reality Exploration and Discovery Pattern: Interaction in Language and Life*. CSLI Publications, Stanford, 2008.

Aoife Cahill, Martin Forst, Michael Burke, Madeira McCarthy, Ruth O'Donovan, Christian Rohrer, Josef van Genabith, and Andy Way. Treebank-based acquisition of multilingual unification-based grammar resources. *Journal of Research on Language and Computation; Special Issue on Shared Representations in Multilingual Grammar Engineering*, 3(2):247–279, 2005.

Aoife Cahill, Michael Burke, Ruth O'Donovan, Stefan Riezler, Josef van Genabith, and Andy Way. Wide-coverage deep statistical parsing using automatic dependency structure annotation. *Computational Linguistics*, 34(1):81–124, 2008.

Ruken Çakıcı. Automatic induction of a CCG grammar for Turkish. In *Proceedings of the ACL Student Research Workshop*, pages 73–78, Ann Arbor, Michigan, June 2005. Association for Computational Linguistics.

Özlem Çetinoğlu and Miriam Butt. Turkish non-canonical objects. In *Proceedings of LFG08 Conference*, Sydney, Australia, 2008. CSLI Publications.

Özlem Çetinoğlu, Miriam Butt, and Kemal Oflazer. Mono/bi-clausality of Turkish causatives. In *Proceedings of the 14th International Conference on Turkish Linguistics*, Antalya, Turkey, 2008. Harrassowitz Verlag.

Özlem Çetinoğlu and Kemal Oflazer. Integrating derivational morphology into syntax. In Nicolas Nicolov, Galia Angelova, Kalina Bontcheva, and Ruslan Mitkov, editors, *Proceedings of the International Conference of Recent Advances in Natural Language Processing 2008 (RANLP 2008)*, Amsterdam, 2009. John Benjamins. to appear.

Özlem Çetinoğlu and Kemal Oflazer. Morphology-syntax interface for Turkish LFG. In *Proceedings of COLING/ACL 2006*, Sydney, 2006.

Seongsook Choi. Complex aspectual structures in Korean, 2002. Talk given in Workshop on Complex Predicates, Particles and Subevents.

Seongsook Choi. *Multiple Verb Constructions in Korean*. PhD thesis, University of Sussex, 2005.

Grzegorz Chrupała and van Genabith. Improving treebank-based automatic LFG induction for Spanish. In *Proceedings of the Tenth International Conference on Lexical Functional Grammar*, Bergen, Norway, 2006.

Mary Dalrymple. *Lexical Functional Grammar*, volume 34 of *Syntax and Semantics*. Academic Press, New York, 2001.

Mary Dalrymple, Ronald M. Kaplan, John T. Maxwell, and Annie Zaenen, editors. *Formal Issues in Lexical-Functional Grammar*. CSLI Publications, Stanford, CA, 1995.

Müşerref Dede. Grammatical relations and surface cases in Turkish. In *Proceedings of BLS*, volume 7, Berkeley, 1981. Berkeley Linguistic Society.

Müşerref Dede. Causatives in Turkish. In Ayhan Aksu Koç and Eser Erguvanlı Taylan, editors, *Türk Dil Bilim Konferansı Bildirileri*. Boğazişi University Publications, 1984.

Faruk Duman. *Seslerde Başka Sesler Var (There are Other Sounds in Sounds)*. Can Sanat Yayınları, Istanbul, 1997.

Mürvet Enç. The semantics of specificity. *Linguistic Inquiry*, 22(1):1–25, 1991.

Gülşen Eryiğit and Kemal Oflazer. Statistical dependency parsing for Turkish. In *Proceedings of EACL 2006 - The 11th Conference of the European Chapter of the Association for Computational Linguistics*, Trento, Italy, 2006. Association for Computational Linguistics.

Gülşen Eryiğit, Joakim Nivre, and Kemal Oflazer. Dependency parsing of Turkish. *Computational Linguistics*, 34(3):357–389, 2008.

Yehuda N. Falk. *Lexical-Functional Grammar: An Introduction to Parallel Constraint-Based Syntax*. CSLI Publications, Stanford, CA, 2001.

Anette Frank, Tracy Holloway King, Jonas Kuhn, and John T. Maxwell III. Optimality Theory style constraint ranking in large-scale LFG grammars. In Peter Sells, editor, *Formal and Empirical Issues in Optimality Theoretic Syntax*. CSLI Publications, Stanford, CA, 2001.

Jeanne Gibson and İnci Özkaragöz. The syntactic nature of the Turkish causative construction. In *Proceedings of CLS*, volume 17, Chicago, 1981. Chicago Linguistic Society.

Aslı Göksel. *Levels of Representation and Argument Structure in Turkish*. PhD thesis, SOAS, University of London, 1993.

Aslı Göksel and Celia Kerslake. *Turkish: A Comprehensive Grammar*. Routledge Publishers, London, New York, 2005.

Tuba Gümüş. LFG for Turkish point-in-time expressions. Master's thesis, ITU Department of Computer Engineering, 2007.

Zelal Güngördü and Elisabeth Engdahl. A relational approach to relativization in Turkish. In *Joint Conf. Formal Grammar, HPSG and Categorial Grammar*, Saarbrücken, Germany, August 1998.

Zelal Güngördü and Kemal Oflazer. Parsing Turkish using the Lexical Functional Grammar formalism. *Machine Translation*, 10(4):515–544, 1995.

Yuqing Guo, Josef van Genabith, and Haifeng Wang. Acquisition of wide-coverage, robust, probabilistic Lexical-Functional grammar resources for Chinese. In *Proceedings of the 12th International Lexical Functional Grammar Conference*, Palo Alto, CA, 2007.

Şebnem İşigüzel. *Öykümü Kim Anlatacak (Who will Tell My Story).* Can Yayınları, Istanbul, 1994.

Barış Kabak. Turkish suspended affixation. *Linguistics*, 45:311–347, 2007.

Ron Kaplan and Paula Newman. Lexical resource conciliation in the xerox linguistic environment. In *Proceedings of the ACL Workshop on Computational Environments for Grammar Development and Engineering*, Madrid, Spain, 1997. Association for Computational Linguistics.

Ron Kaplan, Stefan Riezler, Tracy King, John Maxwell, Alexander Vasserman, and Richard Crouch. Speed and accuracy in shallow and deep stochastic parsing. In *Proceedings of the Human Language Technology Conference and the 4th Annual Meeting of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL'04)*, Boston, MA, 2004a.

Ronald M. Kaplan and Joan Bresnan. Lexical-Functional Grammar: A formal system for grammatical representation. In Joan Bresnan, editor, *The Mental Representation of Grammatical Relations*, pages 173–281. The MIT Press, Cambridge, MA, 1982.

Reprinted in Mary Dalrymple, Ronald M. Kaplan, John Maxwell, and Annie Zaenen, eds., *Formal Issues in Lexical-Functional Grammar*, 29–130. Stanford: CSLI Publications. 1995.

Ronald M. Kaplan and John T. Maxwell. Constituent coordination in Lexical-Functional Grammar. In *Proceedings of the 12th International Conference on Computational Linguistics (COLING '88)*, pages 303–305, Budapest, 1988. Reprinted in Mary Dalrymple, Ronald M. Kaplan, John Maxwell, and Annie Zaenen, eds., *Formal Issues in Lexical-Functional Grammar*, 199–210. Stanford: CSLI Publications. 1995.

Ronald M. Kaplan and Jürgen Wedekind. Restriction and correspondence-based translation. In *Proceedings of the 6th Conference of the Association for Computational Linguistics European Chapter (EACL)*, pages 193–202, Utrecht University, 1993.

Ronald M. Kaplan and Annie Zaenen. Long-distance dependencies, constituent structure, and functional uncertainty. In Mark Baltin and Anthony Kroch, editors, *Alternative Conceptions of Phrase Structure*, pages 17–42. Chicago University Press, Chicago, 1989. Reprinted in Mary Dalrymple, Ronald M. Kaplan, John Maxwell, and Annie Zaenen, eds., *Formal Issues in Lexical-Functional Grammar*, 137–165. Stanford: CSLI Publications. 1995.

Ronald M. Kaplan, John T. Maxwell III, Tracey Holloway King, and Richard Crouch. Integrating finite-state technology with deep LFG grammars. In *Proceedings of the Workshop on Combining Shallow and Deep Processing for NLP*, Nancy, France, 2004b. ESSLLI.

Paul Kiparsky. Partitive case and aspect. In Miriam Butt and Wilhelm Geuder, editors, *The Projection of Arguments: Lexical and Compositional Factors*, pages 265–307. CSLI Publications, Stanford, 1998.

Laura Knecht. Lexical causatives in Turkish. In Dan I. Slobin and Karl Zimmer, editors, *Studies in Turkish Linguistics*. John Benjamins Publishing Company, 1986.

Jaklin Kornfilt. Functional projections and their subjects in Turkish clauses. In Eser E.

Taylan, editor, *The Verb in Turkish*. John Benjamins Publishing Company, Amsterdam/Philadelphia, 2002.

Jaklin Kornfilt. *Turkish (Descriptive Grammar)*. Routledge, New York, 1997.

Jaklin Kornfilt. A note on subject raising in Turkish. *Linguistic Inquiry*, 8(4):736–742, 1977.

Jaklin Kornfilt. Remarks on headless partitives and case in Turkish. In Joan Mascaro and Marina Nespor, editors, *Grammar in Progress — GLOW Essays for Henk van Riemsdijk*, pages 285–303. GLOW, 1990. Syntax and Semantics Volume 30.

Kimmo Koskenniemi. *Two-level Morphology: A general computational model for word form recognition and production*. PhD thesis, Department of General Linguistics, University of Helsinki, 1983.

Rochelle Lieber. Phrasal compounds in English and the morphology-syntax interface. In Diane Brentari, Gary N. Larson, and Lynne A. MacLeod, editors, *Papers from the Parasession on Agreement in Grammatical Theory (CLS 24)*, pages 193–202, Chicago, 1988.

Rochelle Lieber. *Deconstructing Morphology: Word Formation In Syntactic Theory*. The University of Chicago Press, Chicago, 1992.

Mitchell Marcus, Grace Kim, Mary Ann Marcinkiewicz, Robert MacIntyre, Ann Bies, Mark Ferguson, Karen Katz, and Britta Schasberger. The Penn treebank: Annotating predicate argument structure. In *Proceedings of the Human Language Technology Workshop*, San Francisco, CA, 1994.

Yo Matsumoto. A reexamination of the cross-linguistic parameterization of causative predicates: Japanese perspectives. In *Proceedings of LFG98 Conference*, Brisbane, Australia, 1998. CSLI Publications.

John T. Maxwell and Ronald M. Kaplan. A method for disjunctive constraint satisfaction. In Masaru Tomita, editor, *Current Issues in Parsing Technology*, pages 173–190. Kluwer Academic Publishers, Dordrecht, 1991. Revised version of 'An overview

of disjunctive constraint satisfaction', Maxwell and Kaplan, 1989. Proceedings of the International Workshop on Parsing Technologies. Reprinted in Mary Dalrymple, Ronald M. Kaplan, John Maxwell, and Annie Zaenen, eds., *Formal Issues in Lexical-Functional Grammar*, 381–401. Stanford: CSLI Publications. 1995.

John T. Maxwell and Ronald M. Kaplan. The interface between phrasal and functional constraints. *Computational Linguistics*, 19(4):571–590, 1993. Reprinted in Mary Dalrymple, Ronald M. Kaplan, John Maxwell, and Annie Zaenen, eds., *Formal Issues in Lexical-Functional Grammar*, 403–429. Stanford: CSLI Publications. 1995.

John T. Maxwell and Ronald M. Kaplan. An efficient parser for LFG. In Miriam Butt and Tracy Holloway King, editors, *The Proceedings of the LFG '96 Conference*, Rank Xerox, Grenoble, 1996.

Tara Mohanan. *Argument Structure in Hindi*. CSLI Publications, Stanford, 1994.

John Moore. Turkish copy-raising and a-chain locality. *Natural Language and Linguistic Theory*, 16:149–189, 1998.

Jean Mulder. Raising in Turkish. In *Proceedings of the Second Annual Meeting of the Berkeley Linguistics Society*, pages 298–307, Berkeley, 1976.

Ruth O'Donovan, Aoife Cahill, Josef van Genabith, and Andy Way. Automatic acquisition of Spanish LFG resources from the CAST3LB treebank. In *Proceedings of the Tenth International Conference on Lexical Functional Grammar*, Bergen, Norway, 2005.

Kemal Oflazer. Dependency parsing with an extended finite-state approach. *Computational Linguistics*, 29(4):515–544, 2003.

Kemal Oflazer. Two-level description of Turkish morphology. *Literary and Linguistic Computing*, 9(2):137–148, 1994.

Kemal Oflazer and Sharon Inkelas. The architecture and the implementation of a finite state pronunciation lexicon for Turkish. *Computer Speech and Language*, 20(1), 2006.

Kemal Oflazer, Bilge Say, Dilek Zeynep Hakkani-Tür, and Gökhan Tür. Building a Turkish treebank. In Anne Abeille, editor, *Building and Exploiting Syntactically-annotated Corpora.* Kluwer Academic Publishers, Dordrecht, 2003.

Masanori Oya and Josef van Genabith. Automatic acquisition of Lexical-Functional grammar resources from a Japanese dependency corpus. In *PACLIC 2007: Proceedings of the 21st Pacific Asia Conference on Language, Information and Computation*, 2007.

İnci Özkaragöz. Monoclausal double passives in Turkish. In Dan I. Slobin and Karl Zimmer, editors, *Studies in Turkish Linguistics*. John Benjamins Publishing Company, 1986.

İnci Özkaragöz. Evidence for 2-3 retreat in Turkish, 1979. unpublished paper.

Balkız Öztürk. *Case, Referentiality and Phrase Structure*. John Benjamins, Amsterdam, 2005.

Zuhal Peksoy. *Parfüm Reyonu (Perfume Section)*. Bilgi Yayınevi, Istanbul, 2000.

Paul Postal. Anaphoric islands. In Georgia M. Green Robert I. Binnick, Alice Davison and Jerry L. Morgan, editors, *Proceedings of CLS*, volume 5. Chicago Linguistic Society, Chicago, 1969.

Alan Prince and Paul Smolensky. *Optimality Theory: Constraint Interaction in Generative Grammar*. Blackwell Publishers, 2004.

Stefan Riezler and Alexander Vasserman. Incremental feature selection and l1 regularization for maximum-entropy modeling. In *Proceedings of EMNLP 2004*, Barcelona, Spain, 2004.

Stefan Riezler, Tracy H. King, Ronald M. Kaplan, Richard Crouch, John T. Maxwell, and Mark Johnson. Parsing the Wall Street Journal using a Lexical-Functional Grammar and discriminative estimation techniques. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02)*, Philadephia, 2002.

Sara Rosen. *Argument Structures and Complex Predicates*. PhD thesis, Brandeis University, 1989.

Bilge Say, Deniz Zeyrek, Kemal Oflazer, and Umut Özge. Development of a corpus and a treebank for present-day written Turkish. In Kamile İmer and Gürkan Doğan, editors, *Proceedings of the 11th International Conference on Turkish Linguistics*, pages 399–412, Northers Cyprus, 2002. Eastern Mediterranean University.

Natalie Schluter and Josef van Genabith. Treebank-based acquisition of LFG parsing resources for French. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Morocco, 2008.

Peter Sells. *Lectures on Contemporary Syntactic Theories*. CSLI Publications, Stanford, CA, 1985. CSLI Lecture Notes, number 3.

Jane Simpson. *Aspects of Warlpiri Morphology and Syntax*. PhD thesis, MIT, 1983.

Jane Simpson. *Warlpiri Morpho-Syntax: A Lexicalist Approach*. Kluwer, Dordrecht, 1991.

Andrew Spencer. Bracketing paradoxes and the English lexicon. *Language*, 64:663–682, 1988.

Andrew Spencer. *Morphological Theory*. Basil Blackwell, Oxford, 1991.

Lamia Tounsi, Muhammed Attia, and Josef van Genabith. Automatic treebank-based acquisition of Arabic LFG dependency structures. In *Proceedings of EACL 2009 Workshop on Computational Approaches to Semitic Languages*, Athens, Greece, 2009.

Kenji Yokota. Complex-predicate formation and some consequences in Japanese. In *Proceedings of LFG01 Conference*, Hong Kong, 2001. CSLI Publications.

Karl E. Zimmer. Some constraints on Turkish causativization. In Masayoshi Shibatani, editor, *The Grammar of Causative Constructions*, pages 399–412. Academic Press, 1976. Syntax and Semantics Volume 6.