

University of Warsaw
Faculty of Psychology

Berke Sensekerçi
Record book number: 400466

Coordination of Unlikes in Turkish

Master's thesis
in the field of Cognitive Science

The thesis was written under the supervision of
prof. dr hab. Adam Przepiórkowski
Faculty of Philosophy, University of Warsaw
Institute of Computer Science, Polish Academy of Sciences

Warsaw, August 2023

Summary

Coordination is a grammatical structure where two or more linguistic elements are joined together, typically via elements called conjunctions. Despite its prevalence in natural languages, coordination remains poorly understood. One widely accepted view posits that the grammaticality of a coordinate structure depends on the coordinated elements (conjuncts) having the same grammatical properties, such as case and syntactic category. Although this view has been repeatedly questioned in the past, the debate still centers on limited data mainly from Polish and English. To further challenge the assumption that conjuncts must be alike, the present study introduces two types of novel empirical data from Turkish to the current debate: corpus examples and judgments of native speakers.

The study hypothesizes that the grammaticality of coordinate structures in Turkish is contingent upon the grammatical function of the conjuncts, rather than their syntactic categories or cases. To investigate this hypothesis, a large web corpus of 3.3 billion words (Turkish Web 2012) was searched using tailored Corpus Query Language queries. This search yielded 137 instances of unlike category coordination and 51 instances of unlike case coordination. Notably, almost all extracted examples (with five exceptions) feature conjuncts with matching grammatical functions.

To assess the acceptability of different configurations of unlike category/case coordination, a formal acceptability judgment survey was conducted with 48 native speakers. Distinct sentence stimuli were rated on a 7-point Likert scale (from -3 to 3). An analysis of the results indicates that mismatching functions significantly affect the acceptability ratings, supporting the significance of grammatical function in coordination.

Finally, the observed empirical facts are formalized and analyzed within the framework of Lexical-Functional Grammar, adhering to the approach proposed by Przepiórkowski & Patejuk (2021). To validate the analysis, a computational implementation was carried out using Xerox Linguistic Environment, a specialized platform for implementing grammars based on Lexical-Functional Grammar.

In conclusion, this thesis adds to the growing body of research challenging the widely assumed notion that conjuncts must be alike in their cases and categories. Moreover, the implications of this research extend to formal linguistics and grammar engineering.

Key words

Coordination of unlike categories, coordination of unlike cases, corpus linguistics, experimental syntax, Lexical-Functional Grammar, Turkish syntax, grammar engineering.

The title of the thesis in Polish

Koordinacja różnych kategorii i przypadków gramatycznych w języku tureckim

Streszczenie

Koordinacja to struktura gramatyczna, w której połączone są dwa lub więcej składników, zwykle za pomocą elementów zwanych spójnikami. Pomimo jej powszechności w językach naturalnych koordynacja pozostaje słabo zrozumiana. Jedną powszechnie akceptowaną teorią zakłada, że gramatyczność koordynacji zależy od tego, czy skoordynowane elementy (koniunkty) mają takie same cechy gramatyczne, przede wszystkim takie same przypadki i klasy morfosyntaktyczne. Choć ta teoria wielokrotnie była w przeszłości kwestionowana, debata wciąż koncentruje się na ograniczonych danych głównie z języka polskiego i angielskiego. W celu pełniejszego podważenia założenia, że koniunkty muszą być składniowo podobne, niniejsze badanie wprowadza do debaty dwa rodzaje nowych danych empirycznych z języka tureckiego: autentyczne przykłady korpusowe i osądy rodzimych użytkowników języka.

Badanie zakłada, że gramatyczność koordynacji w języku tureckim zależy od funkcji gramatycznej koniunktów, a nie od ich klas czy przypadków gramatycznych. Aby zbadać tę hipotezę, przeszukano duży korpus internetowy liczący 3,3 miliarda słów (Turkish Web 2012), używając odpowiednich zapytań w języku Corpus Query Language. W wyniku tego przeszukania uzyskano 137 przypadków koordynacji różnych klas oraz 51 przypadków koordynacji różnych przypadków. Warto zauważyć, że prawie wszystkie wydobyte przykłady (za wyjątkiem pięciu) zawierały koniunkty o zgodnych funkcjach gramatycznych.

Aby ocenić akceptowalność różnych konfiguracji koordynacji różnych klas i przypadków, przeprowadzono formalne badanie akceptowalności z udziałem 48 rodzimych użytkowników języka tureckiego. Różne bodźce zdaniowe były oceniane na 7-stopniowej skali Likerta (od -3 do 3). Analiza wyników wskazuje, że niezgodność funkcji gramatycznych znacząco wpływa negatywnie na oceny akceptowalności, wspierając istotne znaczenie tożsamości funkcji gramatycznych – ale nie klas czy przypadków – w koordynacji.

Zaobserwowane fakty empiryczne zostały sformalizowane i zanalizowane w ramach teorii Lexical-Functional Grammar (LFG), zgodnie z podejściem zaproponowanym przez Przepiórkowskiego i Patejuk (2021). Aby zweryfikować analizę, przeprowadzono implementację obliczeniową przy użyciu środowiska Xerox Linguistic Environment, specjalistycznej platformy do implementacji gramatyk LFG.

Podsumowując, niniejsza praca dyplomowa przyczynia się do rosnącego zbioru badań kwestionujących powszechnie przyjmowaną tezę, że spójniki muszą być podobne pod względem przypadków i klas gramatycznych.

Słowa kluczowe

Koordinacja różnych kategorii, koordynacja różnych przypadków, lingwistyka korpusowa, składnia eksperymentalna, Lexical-Functional Grammar, składnia języka tureckiego, inżynieria lingwistyczna.

The title of the thesis in English

Coordination of Unlikes in Turkish

Contents

1	Introduction	6
1.1	Background	6
1.1.1	The law of the coordination of likes	7
1.1.2	Alternative accounts of unlike coordination	8
1.2	Aim and scope of the thesis	12
1.2.1	Problems and deficits	12
1.2.2	Hypothesis statement and the overview of the thesis	12
1.3	Methodological approach	14
2	Corpus study	15
2.1	Overview of the chapter	15
2.2	Methodology	15
2.2.1	The web corpus: Turkish Web 2012	15
2.2.2	Treebanks and valency dictionary	17
2.3	Turkish cases and suspended affixation	18
2.3.1	Turkish case system	18
2.3.2	Suspended affixation	19
2.4	Results	20
2.4.1	Coordination of unlike categories	20
2.4.2	Coordination of unlike cases	26
2.5	Discussion and conclusion	30
3	Acceptability judgment task	32
3.1	Overview of the chapter	32
3.2	Methodology	32
3.2.1	Experimental design	32
3.2.2	Materials	33
3.2.3	Survey structure and presentation	36
3.2.4	Participants	37
3.2.5	Participant elimination	38
3.2.6	Statistical tests	39
3.3	Results	40
3.3.1	Fillers	40
3.3.2	Coordination of unlike categories	40
3.3.3	Coordination of unlike cases	47
3.4	Conclusion	49

4	Formal analysis	51
4.1	Overview of the chapter	51
4.2	Frameworks of syntax	51
4.3	Lexical-Functional Grammar	53
4.3.1	Overview of LFG architecture	53
4.3.2	LFG grammars	54
4.3.3	Coordination in LFG	59
4.4	Analysis of unlike coordination in Turkish	61
4.4.1	Evaluating possible solutions	61
4.4.2	Analysis of distinct unlike coordination configurations	65
4.5	Example analysis	69
4.5.1	Coordination of unlike categories	69
4.5.2	Coordination of unlike cases	71
4.6	Conclusion	72
5	Implementation	73
5.1	Introduction	73
5.2	C-structure rules	73
5.3	Coordination of unlike adjuncts	75
5.3.1	Verbal adjuncts	75
5.3.2	Nominal adjuncts	77
5.4	Implementing the conservative solution	79
5.4.1	Off-path constraints	79
5.5	Coordination of unlike arguments	81
5.5.1	Predicative arguments	81
5.5.2	Object arguments	82
5.5.3	Oblique arguments	83
5.6	Conclusion	84
6	Conclusion	85
A	Corpus study	93
A.1	CQL queries	93
A.1.1	Interpreting CQL queries	93
A.1.2	Coordination of unlike categories	94
A.1.3	Coordination of unlike cases	96
A.2	Valency dictionary	97
A.2.1	Treebanks	97
A.2.2	Procedure	97
B	Acceptability judgment task	99
B.1	Sub-surveys	99
B.1.1	Category block	99
B.1.2	Case block	100
B.2	Instructions	100
B.2.1	Turkish original	100
B.2.2	English translation	101
B.3	Glossed materials	103
B.3.1	Coordination of unlike categories	103

B.3.2	Coordination of unlike cases	110
C	Implementation	118
C.1	Grammar	118
C.1.1	Preamble	118
C.1.2	C-structure rules	118
C.1.3	Templates	120
C.1.4	Lexicon	120
C.2	Test sentences	125

Chapter 1

Introduction

1.1 Background

Coordination refers to a grammatical structure wherein two or more linguistic elements are conjoined, typically through coordinating elements known as *conjunctions*. These conjoined elements, referred to as *conjuncts*, can be single words, phrases, or even complete sentences. Despite the ubiquity of coordination in natural languages, the syntactic and semantic characteristics of coordinate structures remain a topic of extensive debate within theoretical linguistics. Within the context of the present thesis, the central aim is to undertake an empirical investigation and formal analysis of an atypical form of coordination, termed “unlike coordination,” in Turkish data.

Unlike coordination occurs when the coordinated elements do not align either in terms of their syntactic categories (i.e., “coordination of unlike categories”) or in their grammatical cases (i.e., “coordination of unlike cases”). Illustrative examples of unlike coordination can be seen in the examples in (1). In the English sentence in (1a), the elements conjoined by *and* differ in their syntactic category since the first conjunct, *a Republican*, is a noun phrase (NP) while the second conjunct, *proud of it*, is an adjectival phrase (AdjP). The Polish example in (1b), on the other hand, exemplifies unlike case coordination. In this instance, both conjuncts are NPs, but they differ with respect to the grammatical cases that they bear.

- (1) a. Pat is [a Republican]_{NP} and [proud of it]_{AdjP}.
(Sag et al., 1985, p. 117, ex. (2b))

- b. Dajcie wina i całą świnie!
give wine.GEN and whole.ACC pig.ACC

‘Serve (some) wine and a whole pig!’

(Przepiórkowski, 1999, p. 175, ex. (5.269))

It should be noted that unlike coordination is not exempt from constraints. Not all configurations of unlike coordination are valid, as exemplified by the ungrammatical sentence in (2c). While each of the conjuncts in (2c) retains grammaticality when considered individually in (2a) and (2b), respectively; their coordination, which is an unlike category coordination (prepositional phrase (PP)

- (4) a. A plumber and making a fortune though Bill may be, he's not going to be invited to my party. (Peterson, 2004, p. 648, ex. (10a))
- b. Bill is a plumber and making a fortune, and so is John. (Peterson, 2004, p. 648, ex. (11a))
- c. Bill is, and John soon will be, a master plumber and making a fortune. (Peterson, 2004, p. 649, ex. (12a))

1.1.2 Alternative accounts of unlike coordination

In non-transformational frameworks, such as GPSG², HPSG³ and LFG⁴, unlike coordination has been acknowledged as a genuine case of coordination, and various analyses to account for its behavior have been offered.

Sag et al. (1985)

Sag et al.'s (1985) GPSG analysis of unlike coordination constituted one of the first systematic attempts at formally modeling the phenomenon in question. In their analysis, the sentence (5a) is grammatical because the individual conjuncts meet the only requirement set by the verb BECOME: they are predicative arguments of the verb. Since BECOME does not impose any constraint regarding the syntactic category of its predicative argument, unlike category coordination becomes possible as long as the individual conjuncts fulfill the external requirements imposed on them. In this line of reasoning, the ungrammaticality of (5b) can also be explained since the individual conjuncts fail to satisfy the more specific constraints imposed by BECOME. Namely, the conjuncts in the argument position should not only be predicative elements but also either nominals or adjectivals. This grammatical relation between the verb and its coordinated arguments is established through the mother node immediately dominating the conjuncts, which bears only the features that are common to all conjuncts.⁵

- (5) a. Pat became a Republican and quite conservative. (Sag et al., 1985, p. 142, ex. (67a))
- b. *Tracy has become a Republican and of the opinion that we must place nuclear weapons in Europe. (Sag et al., 1985, p. 142, ex. (67b))

Bayer (1996)

Sag et al.'s (1985) work demonstrated that there is no intrinsic categorial constraint imposed upon the conjuncts by the coordination itself. The constraints on coordinate structures can instead be explained and modeled in terms of more fundamental syntactic relations between the coordination and its external governor. Sag et al.'s (1985) analysis, however, was driven primarily by English data

²Generalized Phrase Structure Grammar; Gazdar et al., 1985

³Head-Driven Phrase Structure Grammar; Pollard and Sag, 1994; Müller et al., 2021

⁴Lexical-Functional Grammar; Kaplan and Bresnan, 1982; Dalrymple et al., 2019

⁵The formal details behind Sag et al.'s GPSG analysis are grossly simplified here, as a detailed treatment of their analysis is beyond the scope of the present chapter.

and, as a consequence, effectively only dealt with unlike category coordination in complement position.

The analysis proposed by Bayer (1996) frames the phenomenon solely as a relationship between individual conjuncts and their external governor. Under his analysis, syntactic constraints are encoded strictly in terms of disjunctive or conjunctive statements that are, in effect, evaluated for each conjunct. For instance, BECOME specifies that its predicative complement must satisfy the disjunctive statement $NP \vee AP$, which, in the case of coordination, is evaluated for the category of each conjunct. This simple mechanism predicts that among the two sentences represented in (5), (5a) is the valid one and not the (5b).

Peterson (2004)

One of the first detailed LFG treatments of unlike coordination can be found in Peterson (2004), where (un-)grammaticality of unlike coordination is shown to be a natural consequence of the built-in formal features of LFG for modeling coordination phenomena. According to Peterson, the parameters of unlike coordination do not hinge upon syntactic category considerations but instead upon the grammatical functions of individual conjuncts. In his LFG analysis, the sentence (6a) is grammatical because the particular model of coordination he offers only allows conjuncts with matching functions to connect to their external dependents. In this example, since both conjuncts, *awake* and *asking for you*, have predicative functions, they can be associated with their subject, *the children*. If the conjuncts have different grammatical functions, as in (6b), where we have a coordination of a direct object, *a unicorn*, and a predicate, *happy*, *happy* cannot be associated with its intended subject, *John*.

- (6) a. The children are awake and asking for you. (Peterson, 2004, p. 664, ex. (45))
 b. *John saw a unicorn and happy. (Peterson, 2004, p. 664, ex. (46b))

Patejuk (2015) and Dalrymple (2017)

Peterson's strong claim regarding the functional equivalence requirement between the conjuncts has been questioned in the LFG literature. First, Patejuk (2015, ch. 5) points out that cross-linguistic data, primarily from Russian (Chaves and Paperno, 2007) and Polish (Patejuk and Przepiórkowski, 2012, pp. 2200–2202), show that different grammatical functions can be coordinated in certain syntactic contexts.⁶ Second, Dalrymple (2017) draws attention to the fact that Peterson's phrase-structure rule for coordination (see (7)), where the category of the coordination is equated with the category of the first conjunct, allows unlike coordination so long as the category of the first conjunct is compatible with the external categorial requirements imposed on the coordination.

⁶See Patejuk (2015, ch.5) for a detailed catalog of papers dealing with the coordination of unlike grammatical functions in various European languages.

- (7) $X \rightarrow X \text{ Conj } Y$ (Peterson, 2004, p. 652, ex. (20))

As a consequence, Peterson's coordination rule represented in (7) classifies the ungrammatical coordinate structure in (6b) as a licit one because the category of the coordination (NP), derived from the category of the first conjunct, is compatible with the category requirement imposed by *SEE*. In order to deal with these issues, Patejuk (2015) introduces a new syntactic category label, termed UP, for unlike category coordination.

Dalrymple's (2017) account regards Patejuk's UP proposal as insufficient and instead replaces atomic syntactic categories with feature matrices where, for instance, the atomic categories of N and NP are uniformly represented as $[N +, V -, P -, \text{ADJ} -, \text{ADV} -]$. As for coordination, the feature matrix associated with the coordination node is understood as a "collection hub" where categorial information present in individual conjuncts is aggregated. For example, the feature matrix of an unlike coordination conjoining a nominal and an adjectival element would be encoded as $[N +, \text{ADJ} +]$.

Complex categorial constraints imposed by governors are also encoded in terms of these feature matrices. So, when a governor constrains the category of its arguments, it states what categories are *not* allowed and leaves the co-occurrence of allowed categories as a viable possibility. For instance, the sentence (8a) is grammatical because the category of the coordination, $[N +, \text{ADJ} +]$, does not violate the categorial constraint imposed by *BECOME*, which is $[V -, P -, \text{ADV} -]$. The same mechanism predicts the ungrammaticality of the sentence (8b) because the category of coordination, $[N +, P +]$, clashes with the constraints imposed by *BECOME*.⁷

- (8) a. Fred became a professor and proud of his work.
(Dalrymple, 2017, p. 35, ex. (6a))
- b. *Fred became a professor and in line for a promotion.
(Dalrymple, 2017, p. 35, ex. (6c))

Przepiórkowski and Patejuk (2021)

Nevertheless, based on cross-linguistic data found in corpora, Przepiórkowski and Patejuk (2021) question Dalrymple's analysis by pointing out that the constraints imposed by governors are usually not limited to basic syntactic categories but extend to specific references to complementizers, prepositions, and grammatical cases as well.

Consider the English verb *wait*. Its argument position does not indiscriminately accommodate any PP or CP; rather, it specifically necessitates a PP headed by *for* or a CP headed by *until*. This is evident from the observation that substituting other prepositions or complementizers for *for* or *until* in a sentence containing *wait* results in ungrammaticality. Moreover, this constraint also applies to coordinated arguments of *wait*. In (9), we observe that the argument position of *wait* features a coordination of unlike categories (PP & CP). Nonetheless, this

⁷Like other analyses presented in this chapter, this is a highly simplified account of Dalrymple's (2017) analysis.

construction is grammatical since each conjunct individually satisfies the complex morphosyntactic requirements imposed by *wait* on its arguments.

- (9) Next time you're thinking about seeing that upcoming sequel, consider waiting [for the DVD]_{PP[for]} or [until it's out on Netflix]_{CP[until]}.
(Przepiórkowski and Patejuk, 2021, p. 211, , ex. (14))

Consequently, Przepiórkowski and Patejuk (pp. 216–220) propose an LFG-based solution, wherein both categorical and other functional constraints, such as case and preposition form, are handled at the level of individual conjuncts through constraining statements that do not refer to the features carried by the mother node. Their analysis effectively characterizes unlike category coordination as a structure devoid of any specific syntactic category and models syntactic category as a functional rather than a strictly phrase-structural term. Their approach, which is akin to Bayer's, serves as the foundation for the formal analysis of Turkish unlike coordination in the context of the present thesis.

Przepiórkowski (2022)

Apart from syntactic category considerations, in morphosyntactically more complex languages, the possibility of conjoining nominals bearing different grammatical cases enters into the equation as well (i.e., unlike case coordination), which can be observed in the Polish examples in (10).

- (10) a. Przyjedzie albo późnym wieczorem, albo
come.FUT.3SG or late.INS.SG.M evening.INS.SG.M or
następnej zimy.
next.GEN.SG.F winter.GEN.SG.F
'(S)he will come either late in the evening, or next winter'
(Przepiórkowski, 1999, p. 175, ex. (5.270))

- b. Ja i trzech innych nosimy
I.NOM.SG and three.ACC.PL.M others.GEN.PL.M carry.1PL
ją w lektyce ...
she.ACC in litter
'Me and three others are carrying her in a litter . . .'
(Przepiórkowski, 2022, p. 607, ex. (36))

In the literature on unlike coordination, there has been a greater focus on unlike category coordination compared to unlike case coordination. This discrepancy could be attributed to the extensive discussion of unlike coordination being based primarily on English, a language that practically lacks a system of grammatical case. However, a recent paper by Weisser (2020) proposes a cross-linguistic generalization called *Symmetry of Case in Conjunction* (p. 43). According to Weisser, the grammatical cases of the conjuncts always match at an underlying level, even in apparent mismatches. Weisser suggests that these apparent mismatches can be explained through ellipsis or other superficial morphological processes.

Przepiórkowski (2022) argues against Weisser’s generalization and presents genuine instances of unlike case coordination primarily from Polish. Przepiórkowski asserts that these instances cannot be explained through ellipsis or other superficial morphological operations, leading to the conclusion that there is no cross-linguistic requirement for conjuncts to bear the same grammatical case. While Przepiórkowski’s data indeed call into question the *Symmetry of Case in Conjunction* generalization, further cross-linguistic data is needed to draw more robust conclusions about the parameters governing the grammatical cases of conjuncts.

1.2 Aim and scope of the thesis

1.2.1 Problems and deficits

Despite numerous analyses of unlike coordination proposed over the past four decades, the study of coordination of unlikes remains active and the subject continues to be a matter of contention (Przepiórkowski, 2022, p. 593). Nevertheless, the progress of the subject seems to be hindered by two distinct empirical challenges: 1) a lack of diverse linguistic data; 2) a lack of proper experimental approaches to data validation.

The importance of the first challenge can be easily appreciated when we take into account the fact that the emergence of new data has consistently played a pivotal role in driving further insights and refinements within the literature. Yet, the current discussion on unlike coordination is still exclusively driven by data from English and Polish. The second challenge, however, is more controversial, as there does not seem to be a clear consensus among researchers regarding the adoption of formal experimental methodologies in syntax research.⁸ Since the present thesis is categorically situated within the field of cognitive science and regards linguistics as one of the cognitive sciences, it employs formal experimentation methods to study the phenomena that constitute the research object.

1.2.2 Hypothesis statement and the overview of the thesis

In light of these issues, the present thesis attempts to further refute the assumed coordination of likes constraint based on novel empirical and experimental data from Turkish. Consequently, the primary hypothesis put forward in this thesis is that conjuncts may differ both in their syntactic categories and cases in Turkish as long as they match in their grammatical functions.

While there is no universally agreed-upon set of grammatical functions, the present work acknowledges a subset of grammatical functions commonly assumed in nontransformational and traditional frameworks of grammar, such as dependency grammar (Tesnière, 1959) and LFG.⁹ The recognized set of grammatical functions in this work includes five elements: subject, object, oblique,

⁸For arguments in favor of formal experimentation methods, see Schütze (1996), Cowart (1997), Wasow and Arnold (2005), Gibson and Fedorenko (2013), and Hitz and Francis (2016). For arguments that cast doubt upon their necessity, see Phillips (2010) and Sprouse et al. (2013).

⁹A motivation as to why these frameworks were chosen as reference points is provided in §4.2.

predicate, and adjunct. Notably, the first four functions are categorized as arguments, assuming the argument-adjunct dichotomy. This dichotomy, however, remains a controversial in the syntax literature (see Vater 1978; Haspelmath 2014; Przepiórkowski 2016), primarily due to the absence of reliable tests for accurately classifying arguments and adjuncts. However, this thesis abstracts away from these debates, as the argument-adjunct dichotomy does not play a central role in the investigation. The hypothesis primarily operates on the functional labels of conjuncts and their (in)compatibility – i.e., the hypothesis does not assert that argumenthood or adjuncthood plays a key role in Turkish coordination.

When defining and classifying grammatical functions, there is also no universally agreed-upon set of principles or tests, as languages employ diverse strategies to encode these functions. The present work, however, primarily relies on morphosyntactic tests for classifying grammatical functions.

In the present study, the identification of the subject function relies primarily on case-marking¹⁰ (nominative in matrix clauses or genitive in subordinate clauses) and verbal agreement involving number and person. Distinguishing objects involves the passivization test as well as the identification of case-marking (typically accusative, but also nominative with non-specific objects). The oblique function in Turkish is realized by NPs that can assume different cases, typically the dative case. However, distinct from adjuncts, which can also be realized by NPs with different cases, the case of Turkish obliques is either determined by the governing verb or semantically assigned based on their thematic roles. When it comes to predicative arguments, they can be easily distinguished within the context of copular constructions. In Turkish, this copular item can be a suffix, such as *-y-* marked directly on the predicative arguments, a separate verbal item like *ol-* ‘be/become’, or can be simply omitted, similar to Russian and Hebrew. The predicates themselves can be realized by AdjPs, PPs, and NPs in the nominative case. Consequently, the combined use of these morphosyntactic tests determines the classification of grammatical functions throughout the investigation into the coordination of unlikes.

The hypothesis itself cannot be considered a novel one, as Peterson (2004) makes a similar claim for English coordination. However, to the best of this author’s knowledge, neither the coordination of unlikes nor the specific hypothesis of the thesis have previously been tested against Turkish data. In this regard, the present thesis constitutes the first systematic attempt at investigating and analyzing unlike coordination within the domain of Turkish, a language natively spoken by around 82 million people (Eberhard et al., 2022).

The hypothesis is handled in four separate, yet inter-related dimensions in the body of the thesis: 1) Data exploration in annotated Turkish corpora; 2) Validation of the collected corpus data through a formal acceptability judgment survey filled out by native Turkish speakers; 3) Lexical-Functional Grammar analysis of Turkish coordination accounting for the validated empirical phenomena and native speaker intuitions; 4) Computational implementation of the proposed analysis in XLE (Xerox Linguistic Environment; Crouch et al., 2017). The first two dimensions not only form the empirical basis of the present research but also introduce novel coordination data to the ongoing debate. The remaining dimen-

¹⁰A brief overview of Turkish cases and their relation to grammatical functions is provided in §2.3.1. For a more detailed overview, see Göksel and Kerslake (2010, pp. 45–58).

sions provide a formal foundation that can be used for further experimentation through parsing test suites and improving the existing large-scale, computational grammars of Turkish, such as the one developed by Çetinoğlu (2009) in XLE.

1.3 Methodological approach

The present thesis aligns itself with the empirically-oriented part of theoretical linguistics, where the sole judgment of the linguist is no longer accepted as a valid source of information. Accordingly, the hypothesis asserted in this thesis is tested against two types of empirical evidence obtained under two distinct methodological approaches: a corpus analysis and an acceptability judgment experiment.

While both types of methodologies are commonly employed in empirically-oriented works, corpus-based methodologies are much more prevalent due to understandable reasons. Corpus-based methodologies are not only easier to deploy, but also better suited for quantitative analyses. Despite these advantages, corpora should not be regarded as the ultimate sources of linguistic evidence (for an alternative view, see Sampson, 2002, pp. 1–13). As Schütze and Sprouse (2013, p. 29) note, the existence (or non-existence) of a specific syntactic construction in a corpus does not constitute ultimate evidence proving/disproving a hypothesis on that construction. After all, if language is acknowledged as inherently a generative and compositional system, then no corpus will ever be able to store all the conceivable sentences of a given language. In the same vein, there is no guarantee that a corpus will contain only the valid sentences of a language, as automatically compiled and spoken-language corpora tend to contain ill-formed sentences.

This necessitates supplementing the corpus-based methodology with acceptability judgment data, where hypotheses are additionally validated by native speaker judgments. This integration does not only enable the researcher to validate examples discovered in corpora but also test rare constructions that fail to appear in them. Just like corpus-based methods, acceptability judgment data, regardless of the formality of the setting, also suffers from various problems due to judgment data being a type of self-report data. In this respect, the variability between speaker responses (Cowart, 1997, pp. 3–7) seems to be a major concern. Nevertheless, the present thesis regards both corpus data and acceptability judgment data as complementary sources of evidence that should be jointly employed to help balance out their respective shortcomings. In this connection, the thesis also rejects the claim that a single data collection method can reliably and accurately capture the notoriously complex domain of natural languages.

Chapter 2

Corpus study

2.1 Overview of the chapter

The corpus study primarily plays an exploratory role in the thesis, contributing to both the formulation and evaluation of the hypothesis.

The methodology employed in the corpus study is elucidated in §2.2. §2.3 offers a concise introduction to the Turkish case system and Turkish suspended affixation. The study's findings are presented in §2.4. Lastly, §2.5 provides a brief discussion, evaluating the results within the context of the hypothesis put forward in the thesis.

2.2 Methodology

In the investigation of syntactic phenomena, using raw text corpora is essentially futile, as the focus is not on the surface forms of individual tokens but rather in the abstract structures underlying them. In this regard, to carry out corpus-oriented investigations of syntactic phenomena, linguists often turn to two types of annotated corpora: part-of-speech (POS) tagged corpora and treebanks.

Part-of-speech (POS) tagged corpora are annotated with the morphosyntactic features of each lexical item, either through automated or manual methods. This allows linguists to move beyond surface forms and focus on the syntactic relations between lexical items. On the other hand, treebanks contain syntactically parsed sentences, typically following either phrase-structure or dependency structure. Treebanks provide a more detailed and precise analysis, but are often smaller in size compared to POS-tagged corpora. As a result, when investigating rare phenomena, it may be necessary to supplement a treebank analysis with POS-tagged corpora. Overall, the combination of POS-tagged corpora and treebanks provides linguists with powerful tools for conducting corpus-oriented investigations of syntactic phenomena. For that reason, the corpus study utilizes both types of syntactically annotated corpora.

2.2.1 The web corpus: Turkish Web 2012

The POS-tagged corpus utilized in the study, Turkish Web 2012 (trTenTen; Baisa and Suchomel, 2012), is a large web corpus consisting of 3.3 billion automatically extracted and annotated tokens (Çöltekin, 2010). Automatically constructed and

annotated corpora often suffer from low reliability and present several challenges to linguists. The process of automatic extraction can lead to the inclusion of texts from questionable sources, potentially even machine-generated content. Moreover, the automatic annotator might produce numerous inaccurate annotations, further complicating the utility of the corpus. Nevertheless, Turkish Web 2012 remains the only POS-tagged corpus used throughout the investigation as the best alternatives, such as Turkish National Corpus (TNC; Aksan et al., 2012) and TS Corpus (Sezer, 2017), are either drastically smaller or not compatible with complex syntactic queries.

The query tool: Corpus Query Language

Queries formulated in Corpus Query Language (CQL) were the main tools in the exploration of the corpus. CQL queries enable the linguist to combine lexico-syntactic information with logical operators in a relatively straightforward fashion. This renders CQL queries powerful tools for searching specific syntactic constructions in large annotated corpora.

Within the context of the present investigation, various CQL queries targeting unlike category and case coordination have been constructed. These rather general CQL statements, however, had to be constantly revised due to the systematic annotation mistakes found in the corpus. As a consequence, some of the constructed CQL queries attempted to target coordinate structures controlled by specific verbs, which were speculated to allow arguments of different categories/cases. Such speculations were based on the cross-linguistic data from previous works on unlike coordination and the constructed valency dictionary, which is the topic of the subsequent section. Ultimately, the instances of unlike coordination were investigated solely based on 15 distinct CQL queries, which were determined, through experimentation, to be useful and sufficiently comprehensive. While 9 queries were designed to investigate distinct unlike category coordination configurations, the remaining 6 queries explored different unlike case coordination possibilities.¹

The example in Figure 2.1 illustrates one of the shortest queries among the 15 CQL queries. When we look at the first line of the query, the section after [tag != '<V>'] specifies in three separate query blocks that we are looking for a sequence where a noun is optionally (specified by the “?” sign after the blocks) preceded by an indefinite determiner² or an adjective, or both. This part collectively encodes the anatomy of a simple and most common type of Turkish noun phrase. The [tag != '<V>'] block preceding the Turkish NP section indicates that the token immediately preceding an NP must not bear a verbal tag. This block ensures that we do not have a sentential first conjunct where the NP is the argument of the preceding verb. The second line specifies that the NP must be followed by a coordinating conjunction, while the third line forces the second conjunct to be an adjective phrase immediately followed by a verb whose stem is *ol* ‘be/become’. In conclusion, the query in Figure 2.1 specifically targets an NP & AdjP coordination that is governed by the verb *ol* ‘be/become’.

¹ Refer to Appendix A.1 for the queries.

² The query does not leave the determiner requirement underspecified on purpose. The only determiner that can be used in that position is the indefinite article as Turkish lacks a definite

```
[tag != '<V>'] [tag = '<Adj>']? [tag = '<Det:indef>']? [tag = '<N>']
[tag = '<Cnj:coo>']
[tag = '<Adv>']? [tag = '<Adj>'] [stem = 'ol' & tag = '<V>']
```

Figure 2.1: 'NP & AdjP BE/BECOME' Query

Sampling

While the number of sentences returned by a few queries was smaller than 300, most of the queries returned thousands of sentences. Due to the widespread annotation mistakes observed in the corpus, formulating highly specific queries was not preferred, as such queries were observed during the query design process to miss occurrences of unlike coordination. Instead, the study sampled a specified number of sentences from the queries that returned more than 500 sentences. Subsequently, each sample was manually checked to see whether the sampled sentence contained genuine unlike coordination.

Verification

During the manual verification process, the sentences that were irrelevant due to annotation mistakes were not taken into account. The validity of promising occurrences, however, was subject to the following criteria: 1) unlike coordination instances that were deemed to be outright ungrammatical (as determined by the author of the thesis, who is a native speaker of Turkish) were considered invalid; 2) unlike coordination instances that resolve to like coordination due to the process of suspended affixation were considered invalid; 3) unlike coordination instances that incorporated fixed and/or idiomatic expressions were considered invalid. If an evaluated instance passed all these criteria, then it was recorded as a genuine case of unlike coordination.

2.2.2 Treebanks and valency dictionary

When formulating verb-oriented queries, valency dictionaries might offer great assistance, especially in the case of unlike case coordination. This is because valency dictionaries provide a highly comprehensive list of lexical items and their combinatorial possibilities with other items, expressed in terms of morphosyntactic and syntactico-semantic relations. For example, a valency dictionary not only indicates how many arguments a verb requires but also lists the specific morphosyntactic and syntactico-semantic properties of those arguments. In the context of the present study, if a valency dictionary indicates, for instance, that the direct object of a verb can be realized by different syntactic categories or nominals in different cases, we can expect to observe instances of unlike direct objects being coordinated within the context of this verb.

Unfortunately, at the time of writing, there is no Turkish valency dictionary that can be employed for such purposes. As a solution, a small-scale “valency” dictionary of Turkish verbs was constructed by automatically parsing 5 prominent Turkish dependency treebanks.³ The resulting dictionary was used in com-

article and the use of demonstratives complicates the anatomy of an NP.

³See Appendix A.2 for more details about the construction procedure.

ination with previous Polish and English findings from the literature to guide the query design process. It is important to note here, however, that the constructed valency dictionary is neither fully accurate nor sufficiently comprehensive simply due to the small size of the treebanks. Thus, the valency dictionary played solely an informative role in the present study, providing clues as to where to look for unlike coordination in the POS-tagged corpus.

2.3 Turkish cases and suspended affixation

To gain a better understanding of the investigation's findings, it is essential to provide a brief overview of two aspects of the Turkish language: its case system and the phenomenon of suspended affixation. The Turkish case system is a crucial element of its grammar, and understanding it is vital for analyzing the coordination of nominal constituents with different grammatical cases. Additionally, suspended affixation is a prevalent phenomenon in Turkish that impinges on coordination.

2.3.1 Turkish case system

Currently, there does not seem to be a clear consensus in the literature as to how many cases there are in Turkish. This disagreement specifically stems from the dubious morphosyntactic status of the bound morpheme *-(y)lA*, which is either classified as the cliticized version of the postposition *ile* (Lewis, 1967; Kornfilt, 1997), or an instrumental/comitative case marker (Göksel and Kerslake, 2010, p. 54). Regrettably, no empirical work has yet been done to determine the exact morphosyntactic status of this form and its relation to the Turkish case system. For this reason, the present work conforms to the most recent treatment of this problem by Göksel and Kerslake (2010) and acknowledges an instrumental/-comitative case realized by *-(y)lA*. Additionally, the automatic POS-tagger employed in Turkish Web 2012 (Çöltekin, 2010) also tags this bound morpheme as a marker of instrumental/comitative case suffix. Nevertheless, since there is no empirical work confirming the existence of an instrumental/comitative case, it is possible that all NPs labeled as instrumental/comitative in this study could alternatively be classified as PPs headed by *ile*. In the latter scenario, the examples of unlike case coordination incorporating instrumental NPs would be reinterpreted as instances of unlike category coordination.

Hence, the present work recognizes 7 grammatical cases in Turkish. With the exception of the nominative/absolutive case, each case is realized by an overt case suffix, as illustrated in Figure 2.2. It is important to note here that the suffixes undergo morphophonological transformation (Turkish vowel harmony) as governed by the base they attach to.

The syntactic functions of noun phrases in Turkish are almost exclusively determined by the case suffix that they bear (Kornfilt, 1997, p. 212), rather than their phrase-structural positions. Each Turkish case is commonly associated with a particular syntactic function, i.e., a given syntactic function is predominantly realized by one particular grammatical case.

Nominative/absolutive case is primarily used for the subjects of finite, ma-

Nominative/Absolutive	-Ø
Accusative	-(y)I
Dative	-(y)A
Genitive	-(n)In
Instrumental	-(y)lA
Locative	-DA
Ablative	-DAn

Figure 2.2: Turkish case suffixes

trix clauses and for non-specific objects (i.e., differential object marking).⁴ Accusative case is strictly used for direct objects, including the derived direct objects of causative constructions. Dative case is used for indirect objects and oblique arguments of certain verbs. In addition to being used to mark the possessor of possessive noun phrase, genitive case is used to mark the subjects of nominalized subordinate clauses. The remaining cases, instrumental, locative, and ablative, are typically used to mark various types of adjuncts, such as adjuncts of instrument, manner, location, source, and goal. However, certain verbs, such as *vazgeç*- ‘renounce/to give up’ and *karar kıI*- ‘decide on’, exclusively take ablative and locative arguments, respectively.

2.3.2 Suspended affixation

A pervasive morphosyntactic phenomenon in Turkish, suspended affixation (SA) occurs when an affix (or affixes) marked on one of the edgemost conjuncts takes phrasal scope over the coordinate structure. In Turkish, SA is strictly limited to the affixes marked on the rightmost conjunct. Exactly what types of suffixes can take part in SA, however, is still a debated topic in the literature. Although Kabak (2007) and Kornfilt (2012) claim that only inflectional suffixes are available for suspension, recent work by Akkuş (2016) and Şenşekerci (2022) questions this claim on the basis of attested corpus examples of suspended affixation incorporating derivational suffixes. Based on available research, it is reasonable to infer that Turkish SA allows for all nominal inflections (i.e., number, possession, and case) as well as certain derivational suffixes to be suspended. A classical example of suspended affixation can be seen in (11) where multiple nominal inflections are subject to suspension at the same time.

- (11) Tebrik ve teşekkür-**ler-im-i** sun-uyor-um.
 Congratulation and thank-PL-1SG.POSS-ACC offer-PRES.PROG-1SG
 ‘(I) am offering my congratulations and thanks.’ (Lewis, 1967, p. 40)

SA is crucial to consider when investigating instances of unlike case coordination in Turkish, as failure to take SA into account may lead to erroneous conclusions. For instance, the coordination in (11) looks like an instance of unlike case coordination on the surface. The first conjunct, *tebrik* ‘congratulation’, bears

⁴It is worth noting that a non-specific object in the nominative case should immediately precede its verb. If the non-specific object is separated from its verb, the accusative marking on the object is obligatory.

no case suffix (i.e., it seems to bear the nominative/absolute case) while the second conjunct, *teşekkürlerimi* ‘my thanks’, is marked with the accusative case suffix. This particular coordinate structure, however, can only be classified as a coordination of like cased conjuncts since the first conjunct, being a bare noun that carries no inflections, automatically assumes all the nominal features of the rightmost conjunct due to the mechanism of Turkish SA. Here, however, one may argue that this example may be ambiguous between like and unlike case coordination depending on the activation of SA. This argument, however, can be invalidated when we explicitly disable SA by switching the order of the conjuncts. The resulting structure would have the nominative conjunct as its second and the accusative conjunct as its first conjunct. If the nominative and accusative cased conjuncts can be coordinated without any problem, this altered structure should be grammatical as well since the order of the conjuncts should not matter. However, this altered construction, presented in (12), is ungrammatical, thus confirming that an underlying unlike case coordination is not possible.

- (12) *Teşekkür-ler-im-i ve tebrik sun-uyor-um.
 thank-PL-1SG.POSS-ACC and congratulation offer-PRES.PROG-1SG

Intended reading: ‘(I) am offering my thanks and congratulation.’

For this reason, the present study does not classify coordinate structures that incorporate SA as unlike case coordination. Specifically, the coordinate structures where the rightmost conjunct has a case suffix and the preceding conjunct(s) are case-unmarked are treated as like case coordination. Likewise, coordinate structures where the final conjunct is case-unmarked and the preceding conjunct(s) are explicitly case marked are not considered a valid configuration to investigate in the corpus study as they are uncontroversially ungrammatical.⁵

2.4 Results

2.4.1 Coordination of unlike categories

As can be seen in Table 2.1, 9 queries in total targeted distinct unlike category coordination configurations in the corpus. Among the 227,177 sentences (hits) collectively returned by the queries, 1687 were sampled, and 137 instances out of 1687 were determined (based on the criteria explained in the previous section) to be valid instances of unlike category coordination.

The primary factor contributing to the imbalance in the number of samples across different configurations is the occurrence of queries that generated an excessive number of duplicate or poorly annotated sentences. This necessitated a flexible approach to sampling rates for various queries. However, in the case of the PP & NP configuration, the sheer number of queries naturally resulted in a higher sample count. The limitation to the specified search configurations is driven by the potential for alternative configurations to result in inaccurate outcomes and false positives. For instance, configurations like NP & PP (the reverse of PP & NP) tend to produce examples where the NP serves as the complement

⁵Moreover, a quick investigation of this configuration in corpus revealed no valid results.

of the postpositional head. Similarly, the reverse of NP & AdjP and NP & AdvP configurations can trigger erroneous instances due to nominalizing suffixes that extend their scope over the entire conjunct, transforming the initial AdjP or AdvP conjuncts into NPs derived from adjectival or adverbial roots.

Categories	Queries	Hits	Sampled	Verified
PP & NP	4	173,774	527	33
PP & AdvP	1	1,523	240	20
PP & AdjP	1	25,272	240	29
NP & AdjP	1	26,318	390	29
NP & AdvP	2	290	290	26
Total	9	227,177	1687	137

Table 2.1: General results of unlike category investigation in the corpus

PP & NP

The unlike category coordination where the first conjunct is a postpositional phrase and the second conjunct is a noun phrase was searched by 4 distinct queries (see Appendix A.1 for the queries). Out of 4 queries, 3 targeted PP & NP coordinate structures that occupy the typical complement position of verbs with the following stems: *ol-* ‘be/become’, *düşün-* ‘think/consider’, *konuş-* ‘talk/converse/speak’. Out of 527 sampled sentences, 33 occurrences were verified.

As can be seen in Table 2.2, all the validated occurrences of PP & NP coordination incorporate conjuncts that match in their functions.

Functions	Number of examples
Predicate & Predicate	14
Object & Object	3
Adjunct & Adjunct	16
Total	33

Table 2.2: Functional roles of PP & NP conjuncts

There are in total 14 PP & NP coordination instances where both conjuncts are predicative arguments. They were found exclusively through a query based on the verbal stem *ol-* ‘be/become’. An example of PP & NP coordination where both conjuncts are predicative complements can be seen in (13).

- (13) Bu iş [sevgi ile]_{PP} ve [gönül-den]_{NP} ol-malı.
 This job love with and soul-ABL be-NECESS

‘This work should be done with love and from heart.’ (trTenTen)

In 3 examples, PP & NP conjuncts function as objects. These examples were discovered within the context of a query targeting the typical complement position of verbs whose stem are *konuş-* ‘talk/converse’. An example of this configuration can be seen in (14).

- (14) Kolektör-ler sık sık [antik enstrüman-lar
 collector-PL frequent frequent antique instrument-PL
 hakkında]_{PP} veya [ticari bilgi]_{NP} konuş-ur-lar.
 about or commercial information talk-AOR-3PL
 ‘Collectors frequently talk about antique instruments or about commercial information.’ (trTenTen)

The PP conjunct in the glossed sentence in (14) can undergo passivization, which is an indication that it is an object. It is important to note here, however, that the postpositional head of the first conjunct, *hakkında* ‘about’, has been systematically classified as a locative noun by the automatic POS-tagger employed in the corpus. The present study classifies *hakkında* as a postposition since it is classified as a type of postposition by virtually all recent descriptive grammars of Turkish.⁶ The second conjunct is an object, as it not only bears the accusative case, which is the only case used for marking specific direct objects in Turkish, but can undergo passivization as well.

In 16 out of 33 examples, PP & NP conjuncts are identified as adjuncts. Such examples were primarily obtained within the context of a general query that did not contain any specifications external to the coordination. A (simplified & modified) example of PP & NP coordination where both conjuncts are adjuncts is represented in (15).

- (15) Suy-u [çamaşır yıka-mak için]_{PP} ve [banyo-da]_{NP}
 water-ACC clothes wash-INF for and bathroom-LOC
 kullan-ıyor-um.
 use-PRES.PROG-1SG
 ‘I use the water for doing laundries and in the bathroom.’ (trTenTen)

PP & AdvP

The unlike category coordination where the first conjunct is a PP and the second conjunct is an AdvP was explored via one query only. This query, however, was a rather general one and did not specify any external governors. From the results of this general query, 240 sentences were sampled, and 20 genuine cases of PP & AdvP coordination were obtained.

Among the verified sentences, all the PP & AdvP conjuncts are adjuncts in their respective sentences. Examples of this configuration can be seen in the two attested sentences presented below. In sentence (16), while the first conjunct is a PP headed by the postposition *birlikte* ‘together’, the second conjunct is an adverb derived from a verbal stem, *açıkla-* ‘explain’ through an adverbializing suffix *-arak*.

⁶See Kornfilt (1997, pp. 100–102), Göksel and Kerslake (2005, pp. 225–226), and Göksel and Kerslake (2010, p. 107).

- (16) Kural-lar-ı [öğrenci-ler-le birlikte] _{PP} ve [açıkla-(y)arak] _{AdvP}
 rule-PL-ACC student-PL-INS together and explain-ADVZ
 koy-malı.
 establish-NECESS

‘Rules must be established with students and by providing explanations.’
 (trTenTen)

The sentence (17) incorporates a PP headed by *gibi* ‘like’ and an AdvP, *profesyonelce* ‘professionally’, derived from an adjectival base through *-ce*, which is an adverbializing suffix.

- (17) [Gerçek bir iş insanı gibi] _{PP} ve
 genuine INDE.DET business person like and
 [profesyonel-ce] _{AdvP} tokalaş-ın.
 professional-ADVZ shake hands-2P.IMP

‘Shake hands like a proper business person and professionally.’
 (trTenTen)

PP & AdjP

There was only one general query investigating the PP & AdjP coordination. Among the 240 sampled sentences, 29 occurrences were determined to be genuine instances of PP & AdjP coordination. As demonstrated in Table 2.3, all the verified instances incorporate conjuncts with matching functions.

Functions	Number of examples
Adjunct & Adjunct	13
Predicate & Predicate	16
Total	29

Table 2.3: Functional roles of PP & AdjP conjuncts

In 13 unlike category coordination instances, the conjuncts are analyzed as adjuncts. A (simplified & modified) example of this configuration is demonstrated in (18). Here, the first conjunct is a PP headed by *göre* ‘according to’, and the second is an AdjP whose head is a derived adjective, *ölçülü* ‘with measure’.

- (18) Gözlüğ-ü [bir gaye-ye göre] _{PP} ve
 eyeglasses-ACC INDE.DET purpose-DAT according to and
 [ölçü-lü] _{AdjP} yap-tı.
 measure-ADJZ make-PST

‘(He/she) made the eyeglasses with a purpose and with measure.’
 (trTenTen)

The remaining 16 instances, however, contain conjuncts that are determined to be predicative arguments governed by *ol-* ‘be/become’ words. In example (19),

the first conjunct is a PP whose head is the postposition *yönelik* ‘for/towards’, while the second is a simple AdjP headed by a simple adjective *net* ‘plain/clear’.

- (19) Konuşma-lar-ınız [hedef-e yönelik]_{PP} ve [net]_{AdjP}
 speech-PL-2PL.POSS goal-DAT towards and plain
 ol-malı.
 be-NECESS
 ‘Your speeches should be to the point and plain.’ (trTenTen)

NP & AdjP

NP & AdjP coordination was investigated with only one query, which targeted unlike category coordination occupying the typical complement position of *ol* ‘be/become’. The query was looking for constructions comparable to the NP & AdjP coordination typically discussed within the context of English data in the literature, such as the ones illustrated in (20).

- (20) a. Pat is [a Republican]_{NP} and [proud of it]_{AdjP}.
 (Sag et al., 1985, p. 117, ex. (2b))
 b. Fred became [a professor]_{NP} and [proud of his work]_{AdjP}.
 (Dalrymple, 2017, p. 35, ex. (6a))

Among the sampled 390 sentences, 29 occurrences of NP & AdjP coordination were determined to be valid. As expected, in all examples, the conjuncts are predicative complements. In (21), two attested instances of Turkish NP & AdjP coordination can be observed.

- (21) a. Bu [çok büyük bir proje]_{NP} ve [çok
 this very big INDE.DET project and very
 masraf-lı]_{AdjP} ol-acak.
 cost-ADJZ be-FUT
 ‘This will be a very big project and very costly.’ (trTenTen)
 b. Fırça [doğal kıl-dan]_{NP} ve [yumuşak]_{AdjP} ol-malı.
 brush natural bristle-ABL and soft be-NECESS
 ‘(A) brush must be of natural bristle and soft.’ (trTenTen)

NP & AdvP

The NP & AdvP coordination was located via two distinct verb-specific queries. One of the queries targeted the typical complement positions of verbs with the stem *sür-*, which can either mean ‘last/continue’ or ‘spread/apply’. The query, however, attempted to constrain the search to the former use, ‘last/continue’, as this use had previously been observed to take oblique arguments realized either

as NPs or AdvPs. The other query searched for NP & AdvP coordination within the context of verbs with the stem *hallet-* ‘handle’.

As shown in Table 2.4, 26 genuine instances of NP & AdvP coordination were extracted from the sample. However, not all verified examples contain conjuncts with matching functional roles. Although the majority of examples are classified as having matching functions, 4 instances have conjuncts with discrepant functions.

Functions	Number of examples
Adjunct & Adjunct	19
Oblique & Adjunct	4
Oblique & Oblique	3
Total	26

Table 2.4: Functional roles of NP & AdvP conjuncts

There were in total 3 NP & AdvP examples where both conjuncts in each instance are classified as the oblique arguments of *sür-* ‘last/continue’. A (simplified & modified) example can be seen in (22), where the first conjunct is an NP headed by *hafta* ‘week’, while the second conjunct is an AdvP headed by the adverb *yıllarca* ‘for years’, which is derived through an adverbializing suffix *-ca*.

- (22) Bu program [her hafta]_{NP} ve [yıl-lar-ca]_{AdvP}
 this programme every week and year-PL-ADVZ
 sür-ecek.
 last-FUT

‘This programme will run every week and for years.’ (trTenTen)

The majority of the verified NP & AdvP instances incorporate conjuncts that function as adjuncts. While oblique & oblique instances were extracted from the *sür-* query only, adjunct & adjunct instances were supplied by both *sür-* and *hallet-* queries. An example of NP & AdvP coordination where both conjuncts are adjuncts can be seen in (23).

- (23) Her şey [eğitim-le]_{NP} ve [konuş-arak]_{AdvP}
 every thing education-INS and talk-ADVZ
 halled-il-ebil-ir.
 handle-PASS-ABIL-AOR

‘Everything can be handled with education and communication.’ (trTenTen)

In 4 examples, the conjuncts differ not only in their categories but also in their functions. This configuration, as exemplified in (24), is rather unexpected, as all the verified examples we have seen so far have conjuncts with matching functions.

- (24) İtalyan-lar-a karşı yerli savaşı [çetin biçim-de]_{NP}
 Italian-PL-DAT against domestic war bitter shape-LOC
 ve [yıl-lar-ca]_{AdvP} sür-dü.
 and year-PL-ADVZ last-PST

‘The domestic struggle against the Italians lasted bitterly and for years.’
 (trTenTen)

To account for such examples, it is possible to posit that, for some native speakers, the verb *sür-* ‘last/continue’ can function as a one-argument verb that only requires a subject. This interpretation would classify both conjuncts as adjuncts. This hypothesis gains some support from the fact that all four instances were found within the context of the verb *sür-* ‘last/continue,’ and they all exhibited an adjunct & oblique configuration similar to the example presented in (24). While the one-argument interpretation of *sür-* does not align with the author’s native speaker intuitions, this possibility should not be entirely disregarded.

2.4.2 Coordination of unlike cases

As Table 2.5 shows, the coordination of unlike cases was investigated in six different conditions where the combinatorial possibilities between different Turkish cases⁷ are modeled in terms of unlike case coordination. The encoding scheme of the conditions in Table 2.5 is explained as follows: Let us assume that ϕ is a variable that can assume any Turkish case; then $\neg\text{NP}_\phi$ & NP_ϕ signifies a coordinate structure where the second conjunct is an NP in the ϕ case while the first conjunct is an NP that bears a case that is not ϕ . For example, coordinate structures whose first conjunct is an NP headed by a non-accusative noun and whose second conjunct is an NP headed by an accusative noun are represented by $\neg\text{NP}_{\text{acc}}$ & NP_{acc} .

6 queries corresponding to six conditions returned 75,753 hits, out of which 840 were sampled for manual verification. From this sample, 51 examples are classified as genuine instances of unlike case coordination. There were, however, three conditions for which no examples could be found. This result, as discussed in the next section, is expected due to the nature of Turkish cases.

Cases Configurations	Queries	Hits	Sampled	Verified
$\neg\text{NP}_{\text{acc}}$ & NP_{acc}	1	3,969	140	0
$\neg\text{NP}_{\text{dat}}$ & NP_{dat}	1	15,659	140	0
$\neg\text{NP}_{\text{gen}}$ & NP_{gen}	1	22,455	140	0
$\neg\text{NP}_{\text{ins}}$ & NP_{ins}	1	9,524	140	30
$\neg\text{NP}_{\text{abl}}$ & NP_{abl}	1	8,437	140	6
$\neg\text{NP}_{\text{loc}}$ & NP_{loc}	1	15,709	140	15
Total	6	75,753	840	51

Table 2.5: General results of unlike case coordination investigation in the corpus

⁷The nominative case is excluded from the queries as it is susceptible to suspended affixation (see §2.3).

Unfruitful conditions: [\neg NP_{acc} & NP_{acc}], [\neg NP_{dat} & NP_{dat}], [\neg NP_{gen} & NP_{gen}]

The fact that certain conditions did not produce any results can be explained based on the morphosyntactic nature of Turkish cases. As previously discussed, there is a rather strict association between grammatical cases and functions in Turkish, where one grammatical case is predominantly used for one grammatical function.

Based on these facts, the unfruitful conditions denote a coordinate structure where the functions of conjuncts are forced to conflict with each other in most cases. To illustrate, let us look at \neg NP_{acc} & NP_{acc}. In this condition, the second conjunct is expected to be a specific direct object, as this grammatical function is exclusively marked with the accusative case. Since this condition constrains the first conjunct to bear a non-accusative case, the first conjunct will never be a specific direct object whichever case it may bear and, consequently, be in functional conflict with the second conjunct. Although the object function can also be realized by the nominative case to mark non-specificity, in a configuration where the first conjunct is in nominative and the second is in accusative can only be interpreted as a coordination of like cases due to suspended affixation. The same principle applies to other unfruitful conditions as well, since the functions that are mapped to dative and genitive cases are almost exclusively realized by them.

[\neg NP_{ins} & NP_{ins}]

[\neg NP_{ins} & NP_{ins}] was a relatively fruitful condition.⁸ A total of 30 examples are classified as genuine cases of unlike coordination, which is a considerable amount considering the small sample size of 140 sentences.

As can be seen in Table 2.6, all the unlike coordination examples in this condition incorporate conjuncts that match in their function, which is adjunct across examples. As far as the grammatical cases of the individual conjuncts are considered, the overwhelming majority of the coordinate structures appear to conjoin a locative and an instrumental noun, while a few examples incorporate an ablative and an instrumental conjunct.

Functions	Conjunct cases	Number of examples
Adjunct & Adjunct	Locative & Instrumental	26
	Ablative & Instrumental	4
Total		30

Table 2.6: [\neg NP_{ins} & NP_{ins}] results

The (simplified) sentence in (25) contains an unlike case coordination of a locative and an instrumental noun, where both conjuncts act as adjuncts.

⁸As discussed earlier, NP_{ins} could be reinterpreted as PP depending on one's specific theoretical perspective on the Turkish case system. However, if we were to reinterpret NP_{ins} as PP, all the examples in this section (including other examples containing an instrumental NP conjunct) would be rendered as instances of coordination of unlike categories (NP & PP).

- (25) [Doğru yer-de]_{NP_{loc}} ve [doğru antrenör-le]_{NP_{ins}}
 right place-LOC and right trainer-INS
 çalış-(1)yor-uz.
 work-PRES.PROG-1PL
 ‘(We) work in the right place and with the right trainer.’ (trTenTen)

The (simplified) sentence in (26) is an unlike case coordination where an ablative and an instrumental noun are conjoined. Likewise, the conjuncts are classified as adjuncts.

- (26) Proje-ler-iniz-i [el-den]_{NP_{abl}} veya [kargo-(y)la]_{NP_{ins}}
 project-PL-2PL.POSS-ACC hand-ABL or cargo-INS
 teslim ed-ebil-ir-siniz.
 delivery do-ABIL-AOR-2PL
 ‘(You) can deliver your projects by hand or by courier.’ (trTenTen)

[¬NP_{abl} & NP_{abl}]

Among the 140 sentences sampled for the condition [¬NP_{abl} & NP_{abl}], only 6 genuine cases of unlike coordination were found. As can be seen in Table 2.7, although 5 of the verified examples incorporate conjuncts with matching functions, the conjuncts of one instance have conflicting functions.

Functions	Conjunct cases	Number of examples
Adjunct & Adjunct	Locative & Ablative	3
	Instrumental & Ablative	2
Oblique & Adjunct	Dative & Ablative	1
Total		6

Table 2.7: [¬NP_{abl} & NP_{abl}] results

The (simplified) sentence in (27) contains an unlike case coordination where the first conjunct is a locative noun and the second conjunct is an ablative noun. They both act as adjuncts in the sentence.

- (27) Proje-ye [doğru zaman-da]_{NP_{loc}} ve [doğru fiyat-tan]_{NP_{abl}}
 project-DAT right time-LOC and right price-ABL
 gir-di-m.
 enter-PST-1SG
 ‘I entered the project at the right time and at the right price.’ (trTenTen)

The following is a (simplified) instance of unlike case coordination where the first conjunct is an instrumental noun and the second is an ablative noun. Again, both conjuncts are adjuncts.

- (28) O-na [saf bir zihin-le]_{NP_{ins}} ve [gönül-den]_{NP_{abl}}
 s/he-DAT pure INDEF.DET mind-INS and heart-ABL
 yönel-di-k.
 turn-PST-1PL

‘(We) turned to her/him with a pure mind and from heart.’ (trTenTen)

Finally, the (simplified) sentence in (29) is the only example of absolute unlike case coordination where conjuncts also differ in terms of their functions. The first conjunct, *aynı yöne* ‘to the same direction’, is the obligatory oblique argument of the matrix verb while the second conjunct, *aynı yerden* ‘from the same place’, is an adjunct in the ablative case.

- (29) Öğretmen-im-le [aynı yön-e]_{NP_{dat}} ve [aynı
 teacher-POSS.1SG-with same direction-DAT and same
 yer-den]_{NP_{abl}} bak-mı-yor-uz.
 place-ABL look-NEG-PRES.PROG-1PL

‘My teacher (and I) do not look in the same direction and from the same place.’ (trTenTen)

Although the acceptability of this example is more controversial compared to other examples presented so far, it neither is a downright ungrammatical sentence nor leads to considerable parsing difficulty. A parallel explanation, akin to the one provided for counter-examples in the context of unlike category coordination (involving the verb *sür-*), can be extended to this example as well. Namely, it is possible that, for certain native speakers, one-argument interpretation of the verb *bak-* ‘look’ exists, leading to the interpretation of conjuncts as adjuncts.

[¬NP_{loc} & NP_{loc}]

Out of 140 sampled sentences, 15 unlike case coordination instances were classified as genuine cases of the condition [¬NP_{loc} & NP_{loc}]. As summarized in Table 2.8, all the verified instances incorporate conjuncts that are adjuncts. Across all examples, the first conjuncts are observed to be either in the instrumental or in the ablative case. In this respect, instrumental NP seems to be predominantly coordinated with a locative NP, which is also supported by the findings from the condition [¬NP_{ins} & NP_{ins}] (see Table 2.6).

Functions	Conjunct cases	Number of examples
Adjunct & Adjunct	Instrumental & Locative	13
	Ablative & Locative	2
Total		15

Table 2.8: [¬NP_{loc} & NP_{loc}] results

Sentence (30) exemplifies the configuration where the first conjunct is an instrumental NP and the second is a locative NP. The conjuncts are both adjuncts, like the other examples categorized under this configuration.

- (30) Pamuk-lu çarşaf-lar-ı [yumuşak deterjan-la]_{NP_{ins}} ve
cotton-ADJZ sheet-PL-ACC soft detergent-INS and
[soğuk su-da]_{NP_{loc}} yıka-yın.
cold water-LOC wash-2P.IMP
‘Wash the cotton sheets with mild detergent and in cold water.’
(trTenTen)

In the (simplified & modified) sentence in (31), we can observe the configuration where the first conjunct is an ablative NP and the second is a locative NP. Likewise, both conjuncts are adjuncts.

- (31) Tüm özellik-ler [aynı arayüz-den]_{NP_{abl}} ve [basit
all feature-PL same interface-ABL and simple
bir şekil-de]_{NP_{loc}} yönet-il-ir.
INDEF.DET form-LOC manage-PASS-AOR
‘All the features are managed from the same interface and in a simple manner.’
(trTenTen)

2.5 Discussion and conclusion

The corpus investigation resulted in 188 verified instances of unlike coordination. Out of 188 verified instances, 137 are examples of unlike category coordination, while there are only 51 examples of unlike case coordination. This substantial difference, however, can be attributed to the smaller sample size and more general queries adopted in unlike case coordination part of the corpus investigation. With more specific queries and a larger sample size, we can expect this difference to become insignificant. Although the total absence of examples in three unlike case conditions is a significant finding, it is not unexpected.

It is visible in Table 2.9 and Table 2.10 that the overwhelming majority of both unlike category and case coordination instances incorporate conjuncts that match with respect to their grammatical functions. Against 133 instances of unlike category coordination where conjuncts match in their functions, there are only 4 instances of absolute unlike category coordination where conjuncts, in addition to their categories, differ with respect to their functions. When it comes to unlike case coordination, all the verified examples have conjuncts with matching functions, except for one example.

Categories of conjuncts	Like Functions	Unlike Functions
PP & NP	33	0
PP & AdvP	20	0
PP & AdjP	29	0
NP & AdjP	29	0
NP & AdvP	22	4
Total	133	4

Table 2.9: Summary of the results of unlike category coordination

Cases of conjuncts	Like Functions	Unlike Functions
Ablative & Instrumental	6	0
Dative & Ablative	0	1
Locative & Ablative	5	0
Locative & Instrumental	39	0
Total	50	1

Table 2.10: Summary of the results of unlike case coordination

The results obtained from the corpus corroborate the hypothesis that Turkish coordination is driven by the underlying grammatical functions of conjuncts rather than their surface-level morphosyntactic properties, such as case or syntactic category. The strength of the empirical support, however, differs between the two predictions underlying the hypothesis. First, the hypothesis predicts finding genuine instances of unlike coordination in Turkish corpora as some grammatical functions in Turkish can surface in different syntactic categories and grammatical cases. This prediction seems to be borne out by the sheer number of genuine instances of unlike coordination extracted from the corpus. Second, the hypothesis additionally predicts that the verified occurrences of unlike coordination will have conjuncts with matching functions. The empirical support for this claim, however, is not as strong as the support for the first prediction. Even though the overwhelming majority of the extracted instances conform to the second prediction, the existence of a few counter-examples is sufficient to question the validity of the hypothesis. Nevertheless, the explanation positing a 1-argument interpretation of the verbs involved remains plausible, given the scope and scale of the corpus.

As discussed previously, neither the absence nor the existence of a syntactic construction in corpora can be regarded as conclusive evidence. This point is especially relevant in the case of complex and atypical syntactic phenomena, which do not easily lend themselves to straightforward corpus analyses. Furthermore, the validity of the examples extracted from corpora cannot be taken for granted, especially if the corpus in question is made up of textual materials automatically collected from the Internet. Overall, although the findings of the corpus study lend considerable support to the hypothesis, they must be supplemented by native speaker judgments before any conclusions are drawn.

Chapter 3

Acceptability judgment task

3.1 Overview of the chapter

The empirical backbone of the thesis is built upon the acceptability judgment task. The experimental setup not only enables the validation of the examples found in the corpus but also makes it possible for testing structures that were absent in the corpus.

Details about the specific methodology employed for the acceptability judgment task are elucidated in §3.2. Subsequently, §3.3 presents and discusses the findings obtained from the experiment. Finally, §3.4 provides a summary of the results within the context of the hypothesis proposed in the thesis.

3.2 Methodology

3.2.1 Experimental design

The acceptability judgment task implemented in the thesis followed a specific form of repeated measures factorial design. The hypothesis was broken down into two conceptual blocks to facilitate the factorial design: 1) the category block where the coordination of unlike categories is investigated; 2) the case block where the coordination of unlike cases is investigated. The category block followed a 2×2 factorial design where the interactions between the independent variables of syntactic category and syntactic function are modeled. While the 2×2 factorial design was pursued in the case block as well, this could not be achieved due to the nature of Turkish case and function mapping. As a result, the case block was conceived as a single factor with three levels (distinct sentence types), which will be discussed later. The measured (dependent) variable was the judgment of native speakers and it was a quantitative one as the participants expressed their judgments on a 7-point Likert scale (from -3 to $+3$).

The category block encompassed four sentence types (as per the factorial design) in total, which are illustrated in Table 3.1 along with their respective codes. In the table, the sentence type encoded as LCAT-LF denotes a coordination where conjuncts match both in their syntactic category and function. The conjuncts found in sentence type UCAT-UF, however, neither match in their syntactic category nor their syntactic function. Within the context of the hypothesis tested in the experiment, the sentence types listed under the “Like Functions” column

(LCAT-LF and UCAT-LF) should have similar scores and collectively receive significantly higher scores than the ones under the column “Unlike Functions”. This stems from the fact that it is the discrepant functions that are hypothesized to cause unacceptability rather than the individual categories of conjuncts.

	Like Functions	Unlike Functions
Like Categories	LCAT-LF	LCAT-UF
Unlike Categories	UCAT-LF	UCAT-UF

Table 3.1: The category block

The same factorial design was also pursued but could not be enforced in the case block, as can be seen in Table 3.2. LCASE-UF type of coordination, where conjuncts match in their cases but not in their functions, seems to be impossible in Turkish as each Turkish case is commonly associated with a particular grammatical function.

	Like Functions	Unlike Functions
Like Cases	LCASE-LF	LCASE-UF
Unlike Cases	UCASE-LF	UCASE-UF

Table 3.2: The case block

The nominative case, however, is a significant exception to this strict association between cases and functions in Turkish. As discussed earlier, in addition to its typical mapping to the subject function, the nominative is also used to mark non-specific objects in Turkish. Nevertheless, coordination of a nominative object with a nominative subject can only be postulated on a theoretical level as this coordination results either in both being interpreted as non-specific objects (see (32)) or subjects if one can ignore the missing obligatory object of the verb *gör* ‘see’.

- (32) Kedi ve köpek gör-dü.
 cat.NOM and dog.NOM see-PST-3P

‘Intended meaning: The cat saw some dog.’

‘Available meanings: (He/she/it) saw some cat and some dog.’

For that reason, the case block contains only three sentence types that are conceptually identical to the ones observed in the category block. In the case block, however, the category variable is replaced with the case variable, as this block models the interactions between the cases and functions of conjuncts. Likewise, for the hypothesis to be supported, the sentence types listed under the “Like Functions” should receive significantly higher scores by native speakers compared to the sentence type UCASE-UF.

3.2.2 Materials

The material preparation was heavily based on the token-set methodology proposed by Wayne Cowart (1997) in his seminal work, *Experimental Syntax*. According to this methodology, the experiment material should be devised as a

(33) a. [Saat-im-i]_{NPobject} watch-1SG.POSS-ACC ve and [telefon-um-u]_{NPobject} nazikçe gently
çekmece-m-e koy-du-m.
drawer-1SG.POSS-DAT put-PST-1SG
‘I gently put my watch and my cellphone in my drawer.’ **LCAT-LF**

b. Saat-im-i [nazik-çe]_{AdvPadjunct} ve [dikkat-le]_{NPadjunct}
watch-1SG.POSS-ACC gentle-ADVZ and care-INS
çekmece-m-e koy-du-m.
drawer-1SG.POSS-DAT put-PST-1SG
‘I put my watch in my drawer gently and with care.’ **UCAT-LF**

c. [Saat-im-i]_{NPobject} ve [çekmece-m-e]_{NPoblique} nazikçe
watch-1SG.POSS-ACC and drawer-1SG.POSS-DAT gently
koy-du-m.
put-PST-1SG
‘I gently put my watch and in my drawer.’ **LCAT-UF**

d. [Saat-im-i]_{NPobject} ve [nazik-çe]_{AdvPadjunct} çekmece-m-e
watch-1SG.POSS-ACC and gentle-ADVZ drawer-1SG.POSS-DAT
koy-du-m.
put-PST-1SG
‘I put my watch and gently in my drawer.’ **UCAT-UF**

(34) a. Oğlum-a internet-ten [şapka]_{NOMobject} ve [ayakkabı]_{NOMobject}
son-DAT internet-ABL hat and shoe
al-dı-m.
buy-PST-1SG
‘I bought my son a pair of shoes and a hat through the internet.’ **LCASE-LF**

- b. Oğlum-a [internet-ten]_{ABLadjunct} ve [kredi kartı-yla]_{INSadjunct}
 son-DAT internet-ABL and credit card-INS
 ayakkabı al-dı-m.
 shoe buy-PST-1SG

‘I bought my son a pair of shoes through the internet and by credit card.’
UCASE-LF

- c. [Oğlum-a]_{DAToblique} ve [iste-diğ-i]
 son-DAT and want-PTCP-3SG.POSS
 ayakkabı-ylı_{ACCobject} internetten al-dı-m.
 shoe-ACC internet-ABL buy-PST-1SG

‘I bought my son and the pair of shoes he wanted through the internet.’
UCASE-UF

However, the material for the judgment task must not be limited to only one token set. Ultimately, the rationale behind constructing token sets is to test the participant’s ‘reaction’ to the sentence stimuli across different sentences in a controlled and standardized manner. Therefore, for each block, 12 distinct token sets were constructed, which amounted to 24 token sets and 84 sentences in total.

The research designs based on repeated measures are highly susceptible to sequence effects and often need carefully designed counterbalancing schemes (Graziano and Raulin, 2014, pp. 305–306). In the present study, the counterbalancing scheme involved the use of (an equal number of) filler sentences interspersed (randomized) with the sentence stimuli. Their function was to eliminate priming effects as well as to gauge the participant’s understanding of the task and attention throughout the task. To control for learning effects, all the fillers contained coordination as well, but they were not constructed based on any particular sentence type, incorporating different types of coordination. Moreover, the fillers represented the extremities of acceptable/unacceptable dichotomy. While half of them were outright unacceptable (to the point of being incomprehensible), the other half were perfectly acceptable. Additionally, some of the fully grammatical filler sentences incorporated semantic oddities and region-specific uses, which were used to evaluate the participant’s sensitivity to semantic content and regional variance. In conclusion, the typical issues inherent in the research design were tackled on the level of individual participant’s individual responses to filler sentences.

In addition to stimuli and filler sentences, a set of eight demographic questions was developed. These questions aimed to gather information about the participants’ demographics, including their sex, age, native language, potential bilingualism, the region where their native language(s) were acquired, education level, and educational background/major. The inclusion of these questions served two purposes: obtaining essential demographic information and controlling for potential confounding variables, such as bilingualism and education level.

3.2.3 Survey structure and presentation

The way the material is presented to the participant is equally important in the implementation of token-set methodology. First, the participant must judge exactly one sentence type from a given token set. Therefore, multiple sentence types from the same token set cannot be presented to a single participant. Second, it must be ensured that the participant judges all the sentence types tested in the study. Namely, the participant must encounter all the sentence types demonstrated in Tables 3.1 and 3.2 throughout the survey. Last but not least, the participant must judge the same number of sentence tokens for each type. The last requirement imposes a “democratic” representation scheme where each sentence type must be represented by an equal number of representatives in the survey. For example, if the LCAT-LF type is represented by 4 distinct sentence tokens, the other types must be represented by 4 representatives as well.

To conform to these requirements, 4 different surveys (sub-experiments) were constructed following a Latin square procedure.¹ Each survey contained distinct 21 sentence stimuli. While 12 of them were sentences taken from the collection of category block token sets, 9 were from the collection of the case block. To comply with the democratic representation scheme, each sentence type (from both blocks) is represented by 3 distinct sentences extracted from distinct token sets.²

The survey was presented such that each sentence stimulus was followed by a filler sentence. This ensured that no sentence stimulus was followed by another sentence stimulus. While this order of presentation was fixed beforehand, exactly which sentence stimulus or filler appears in an appropriate slot was randomly chosen from the pool of 21 sentence stimuli and the pool of 22 filler sentences, respectively. Apart from fillers and sentence stimuli, 3 practice sentences were also included at the beginning of the survey for the participant to acclimatize to the survey platform. In conclusion, each survey presented 46 distinct sentences to be judged, including the practice sentences.

The described survey design was implemented on LimeSurvey Community Edition (*ver.* 3.28) hosted by *kognilab*³ and constructed surveys were distributed through the same platform. The distribution process ensured that each participant was assigned to only one survey and all surveys were assigned the same number of participants. All surveys commenced with the presentation of demographic questions.

In the immediately following section, the instructions supplemented by example evaluations were shown.⁴ The chief aim of the instructions was to ex-

¹2-survey configuration was also considered, since it is also compatible with the specific design and requirements of the token-set methodology. 2-survey design, however, was deemed to be too exhausting for the participant due to the sheer number of sentences in the material pool. Furthermore, the schematic representation of how the collection of token sets from the material pool is equally distributed among the 4 surveys in a Latin square procedure can be seen in the Appendix B.1.

²The category block tested 4 different sentence types. 12/4 gives us 3 representatives for each type. In contrast, the case block tested 3 sentence types, which gives us 3 representatives for each type again due to 9/3.

³Kognilab is an experimental philosophy lab run by researchers from University of Warsaw. For more information, the following link can be visited: www.kognilab.pl

⁴See Appendix B.2 for the translated version of the instructions.

plain that the task was neither about applying the (typically prescriptive) grammar/orthography rules learned at school nor about evaluating the propositions expressed in the sentences. To this end, instructions prompted participants to imagine a situation in their daily lives where they encounter the presented sentence. Through this embodied context, the meaning of the 7-point Likert scale (from -3 to +3) was explained. The highest score, 3, was to be used for natural-sounding sentences that are likely to be constructed by a native speaker of Turkish. The lowest score, -3, however, was to be used for highly-problematic (borderline incomprehensible) sentences that can only be formulated by a foreigner who just started learning Turkish. To illustrate that the task was not about evaluating the proposition expressed by the sentence, a fully grammatical sentence expressing a controversial political opinion was used as an example. Here the instructions explicitly informed that such a sentence should always receive the highest score despite its semantic content may not truthfully reflect the actual states of affairs or be conflicting with the participant's world-view.

Right after the instructions section, the sentences were presented (starting with the 3 practice sentences) to the participant one by one. Although the participant was allowed to save their partial answers and later return to the survey for completion, they were not allowed to go back and change their scores.

3.2.4 Participants

During the recruitment process, the convenience sampling method was used and the voluntary nature of the participation was emphasized throughout the advertisement process. Bursa Uludağ University, which is one of the biggest universities in Turkey, constituted the primary source of recruitment because of the author's personal connections with the faculty staff. The other sources of recruitment were friends, relatives and their acquaintances.

83 participants attempted the online survey, and 67 of them completed it. The participants who gave partial answers were automatically eliminated from the study. Of the remaining 67 participants, 19 more participants were eliminated. Of those 19, two were eliminated as they were not native speakers of Turkish and 17 were determined to have provided invalid answers on the basis of the scores that they gave to filler sentences (the details of this elimination procedure is explained the following selection). As a result, the answers of 48 participants were considered as valid and were subject to the statistical tests.

The sample consisted of relatively well-educated participants as 32 were undergraduate students/graduates and 12 were post-graduate students/graduates. Consequently, this led to a relatively young sample where the mean age was 30.25 years ($SD = 11.22$). The majority of the participants were female. Out of 48 participants, 31 were females, compared to 15 males and 2 participants who marked the 'other' option. The demographic information regarding bilingualism showed that most of the participants seem to have acquired Turkish in a strictly monolingual environment ($n=40$). Even though the recruitment process took place only in one city (Bursa) in the Marmara region of Turkey, 18 out 48 participants reported to have acquired Turkish in other regions of Turkey, which contributed to the representativeness rate of the sample. The global rise in the number of bilinguals (at least in the educated population) was observed in the present study as

well. 34 participants reported that they speak (mostly at an intermediate level) either one or two foreign languages. When it comes to the academic majors of the participants, social sciences was the most common academic major observed among the participants ($n=31$), which was followed by fine arts with 9 participants. The least common academic majors were physical sciences and humanities as each was represented by only 2 participants.

3.2.5 Participant elimination

Before conducting statistical tests, it is crucial to detect and eliminate participants who did not fill out the surveys correctly. In the case of acceptability judgment tasks, this procedure is mainly done on the basis of the scores that a participant gives to grammatical/ungrammatical filler sentences.

One straightforward procedure is to assume arbitrary threshold values for grammaticality and ungrammaticality and eliminate participants whose mean scores of grammatical/ungrammatical fillers violate these thresholds. However, this method has limitations, as the chosen threshold values lack statistical justification and different participants tend to interpret the rating scale differently. Alternatively, outliers can be detected based on basic statistical methods of outlier detection, such as z -score and Interquartile Range (IQR). However, blindly relying on these methods can be problematic within the context of acceptability judgment tasks as such methods tend to disregard the unique distribution of data points and overlook the conceptual significance of scale points.

To illustrate, consider the commonly employed IQR outlier detection method. First, the dataset is sorted in ascending order. Then, the first quartile ($Q1$) and the third quartile ($Q3$) are calculated by finding the medians of the lower and upper halves of the dataset, respectively. The IQR is then obtained by subtracting $Q1$ from $Q3$. Subsequently, a threshold multiplier is defined to determine the range for outliers, which is typically assumed to be 1.5 or 3. The lower threshold (LT) and upper threshold (UT) are obtained by subtracting and adding the threshold multiplier multiplied by the IQR to $Q1$ and $Q3$, respectively. Finally, outliers are identified as any data points falling below the lower bound or above the upper bound. The steps of the IQR outlier detection method for a given distribution are outlined below:

$$\begin{aligned}
 (35) \quad Q1 &= \text{Mdn}(\text{lower half of dataset}) \\
 Q3 &= \text{Mdn}(\text{upper half of dataset}) \\
 IQR &= Q3 - Q1 \\
 LT &= Q1 - (\text{threshold multiplier} \times IQR) \\
 UT &= Q3 + (\text{threshold multiplier} \times IQR)
 \end{aligned}$$

In the given equations, several variables are defined: $Q1$ represents the first quartile, $Q3$ represents the third quartile, IQR represents the interquartile range, LT represents the lower threshold, UT represents the upper threshold, and the threshold multiplier is the value chosen to determine the range for outliers.

By adjusting the threshold multiplier, the stringency of the outlier detection procedure can be controlled. A higher threshold multiplier results in a more lenient detection approach, while a lower threshold multiplier leads to a more conservative approach. For instance, a threshold multiplier of 1.5 is relatively conservative, as it classifies more data points as outliers. Conversely, selecting a threshold multiplier of 3 allows for a broader range of values to be considered normal, resulting in a more lenient detection approach, where fewer data points may be identified as outliers.

In the context of this study, implementing the IQR method with a threshold multiplier of 3 yields an upper threshold of 1.63 for ungrammatical fillers and a lower threshold of -0.27 for grammatical fillers. In other words, this implementation suggests keeping participants that scored, on average, fully grammatical sentences above -0.27 and participants who gave positive scores on average (i.e., 1.63) to outright ungrammatical sentences. Hence, it is clear that the threshold multiplier of 3 fails to capture the meaning of scale points. The commonly used threshold multiplier of 1.5 results in more meaningful threshold values, with an upper threshold of -0.40 for ungrammatical fillers and a lower threshold of 1.09 for grammatical fillers. Nonetheless, this standard approach proved overly conservative and led to the exclusion of 20 participants.

As a result, for this study, the threshold multiplier was adjusted to 1.6 to achieve a more balanced outlier detection procedure. This modification aimed to align with the meaning of the scale points while maintaining a reasonable balance between strictness and leniency in identifying outliers. This adjustment led to the exclusion of participants with mean ungrammatical filler scores exceeding -0.27 or mean grammatical filler scores falling below 1. Breaching either threshold resulted in participant elimination.

$$(36) \quad \begin{aligned} LT_{\text{gram_fillers}} &= Q3_{\text{gram_fillers}} - (1.6 \times IQR_{\text{gram_fillers}}) \\ UT_{\text{ungram_fillers}} &= Q1_{\text{ungram_fillers}} + (1.6 \times IQR_{\text{ungram_fillers}}) \end{aligned}$$

3.2.6 Statistical tests

The numerical judgment data collected via Likert scale violates some of the mathematical assumptions of parametric statistical tests. Crucially, parametric tests assume that the responses are normally distributed and are on a scale that qualifies as interval. However, the judgment data obtained via Likert scale rarely satisfies the first assumption⁵ and never fulfills the second assumption as Likert scale is ordinal.

Nevertheless, parametric tests have proven themselves to be fairly resistant to these violations (Norman, 2010) and produce results congruent with their non-parametric counterparts (Schütze and Sprouse, 2013). For this reason, *linear mixed-effects models*,⁶ which is a parametric linear model allowing to account

⁵The data obtained within the context of the present study also violates the normality assumptions as validated by Q-Q plots and Shapiro-Wilk tests conducted.

⁶The utilization of linear-mixed effects models is well-justified in experimental syntax, as well as in other experimental approaches to linguistic investigation (Gibson et al., 2011; Cummings, 2012).

for random effects, constituted the primary statistical tool for hypothesis testing in the present study. In the model, participant and sentence ID are encoded as random effects while the case/category and grammatical function as fixed effects. Additionally, Wilcoxon signed-rank test (Wilcoxon, 1945), which is the non-parametric alternative of dependent samples *t*-test, was employed in various comparison configurations involving two conditions.

While the data exploration, preparation and preprocessing were done in Python (*version* 3.8.12) with the help of *pandas* (Wes McKinney, 2010), *numpy* (Harris et al., 2020) and *scipy* (Virtanen et al., 2020) libraries, statistical analyses and plotting were achieved in R (*version* 4.2.3) with *lme4* (Bates et al., 2015) and *ggplot2* (Wickham, 2016) packages, respectively.

3.3 Results

3.3.1 Fillers

The scores given to the filler sentences help establish baseline values for the extreme points of acceptability. The grammatical fillers received a mean score of 2.08 ($SD = 1.66$) while the ungrammatical fillers received a mean score of -1.99 ($SD = 1.41$). However, as a measure of central tendency, median is more suitable for finding the actual centre of ordinal data. In this respect, the medians of both the grammatical and ungrammatical fillers give us the extreme points of the scale, which are 3 and -3 , respectively. Further descriptive statistics of the fillers can be seen in Table 3.3.

Filler Type	N	Mean	SD	Min	Max	Median
Grammatical	528	2.08	1.66	-3	3	3
Ungrammatical	528	-1.99	1.41	-3	3	-3

Table 3.3: Descriptive statistics of the filler sentences

3.3.2 Coordination of unlike categories

Descriptive statistics

Consistent with expectations, the LCAT-LF condition, representing the standard coordination type in Turkish, achieved the highest average score of 2.45 ($SD = 1.10$, $Mdn = 3$). Its unlike counterpart, UCAT-LF, where conjuncts differed in their categories but matched in their functions, was rated slightly less acceptable but still positive with a mean score of 1.88 ($SD = 1.53$, $Mdn = 2.5$). However, the participants gave poor ratings to the two conditions where the conjuncts had mismatching functions. While LCAT-UF received a mean score of -0.80 ($SD = 1.91$, $Mdn = -1$), UCAT-UF was rated as the least acceptable coordination type with an average score of -0.84 ($SD = 1.94$, $Mdn = -1$). Nevertheless, we cannot regard these two conditions as outright unacceptable when we compare their mean scores with the mean score of fully ungrammatical filler sentences, which

should be considered as the baseline value for extreme unacceptability. The descriptive statistics of unlike category coordination part of the experiment can be observed in the following table.

Coordination type	N	Mean	SD	Min	Max	Median
LCAT-LF	144	2.45	1.10	-3	3	3.00
UCAT-LF	144	1.88	1.53	-3	3	2.50
LCAT-UF	144	-0.80	1.91	-3	3	-1.00
UCAT-UF	144	-0.84	1.94	-3	3	-1.00

Table 3.4: Descriptive statistics of the unlike category coordination experiment

As can be seen in the boxplot in Figure 3.1, the participants seem to be in strong agreement when it comes to the high acceptability of LCAT-LF as the IQR box is encapsulated between the scores of 2 and 3 with a relatively short lower whisker length. Likewise, the distribution of scores given to UCAT-LF sentences also follows a clear left-skewed distribution towards the positive end of the scale, although now the IQR range is slightly extended. The participants seem to be in relative disagreement about the acceptability of LCAT-UF and UCAT-UF conditions as scores given to these conditions explore the full range of the Likert scale. Ultimately, however, both LCAT-UF and UCAT-UF received negative mean and median scores that were significantly lower than their counterparts.

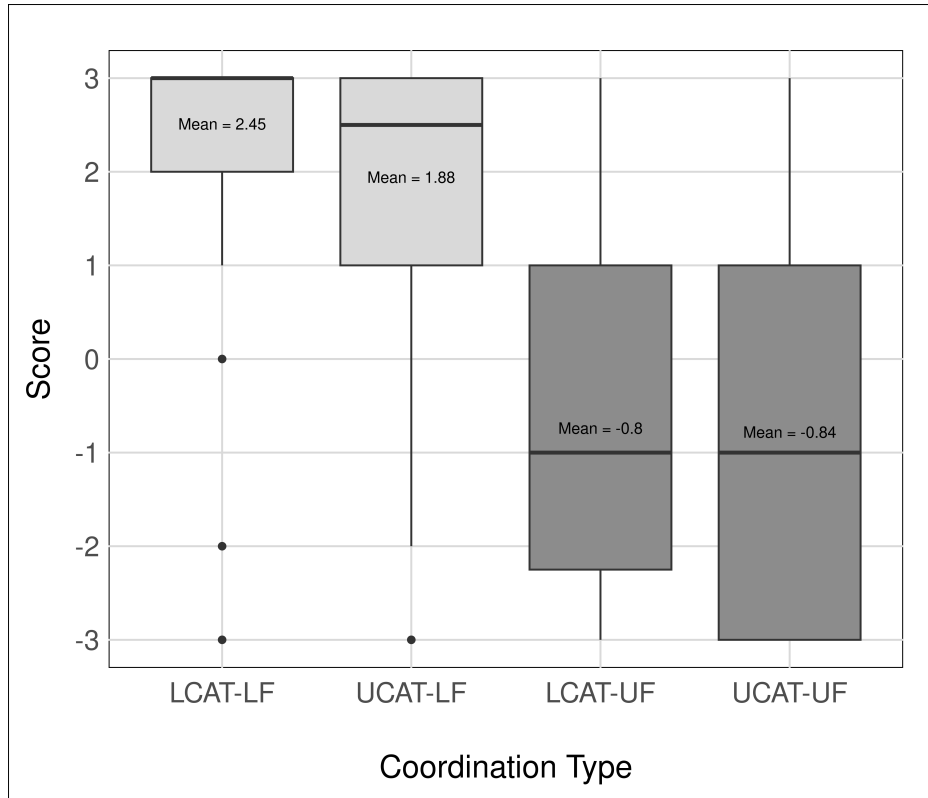


Figure 3.1: Boxplot for unlike category coordination experiment

Inferential statistics

Before fitting the linear mixed-effects models, several relevant comparisons between different pairs of conditions were carried out. A Wilcoxon signed-rank test was conducted to compare the scores for the LCAT-LF and UCAT-LF conditions. The test revealed a significant difference between the two conditions, with a test statistic (V) of 2563 ($p < 0.001$). This finding is in conflict with the part of the hypothesis that there should be no significant difference in acceptability so long as the conjuncts match in their functions. However, the same test demonstrated that UCAT-LF was also significantly different from both LCAT-UF ($V = 7735$, $p < 0.001$) and UCAT-UF ($V = 7394$, $p < 0.001$).

Three distinct linear models were fitted to unlike category coordination data. In the first model, the distinct coordination configurations (i.e., LCAT-LF, UCAT-LF, LCAT-UF, UCAT-UF) were encoded as the levels of one independent variable called “conditions.” While this independent variable was specified as the fixed effect, IDs of the participants and token set indices were encoded as random effects. The exact R code used to fit the model can be seen in Figure 3.2. Since the default optimiser of the ‘lmer()’ function, the “nloptwrap” algorithm, failed to produce a fully convergent model, the “bobyqa” algorithm was used.

```
control <- lmerControl(optimizer = "bobyqa")

model_cat_1factor <- lmer(rating ~ condition +
  (1 + condition|participant_ID) +
  (1 + condition|tokenset),
  data = lm_cat_1factor,
  control = control)
```

Figure 3.2: R code for Model I of unlike category coordination experiment data

The relevant specifications of the first model (including some random effect statistics) are summarized in Figure 3.5.

<i>Predictors</i>	acceptability		
	<i>Estimates</i>	<i>CI</i>	<i>p</i>
(Intercept)	2.45	2.20 - 2.71	< 0.001
condition [UCAT-LF]	-0.58	-1.04 - -0.13	0.012
condition [LCAT-UF]	-3.25	-3.75 - -2.75	< 0.001
condition [UCAT-UF]	-3.29	-3.82 - -2.76	< 0.001
Random Effects			
σ^2	1.59		
τ_{00} participant_ID	0.04		
τ_{00} token_set	0.06		
N participant_ID	48		
N token_set	12		
Observations	576		
Marginal R^2 / Conditional R^2	0.587 / NA		

Table 3.5: One independent variable model specifications

In the context of a linear mixed effects model, the fixed effects refer to the parameters that are estimated to have a consistent effect on the outcome variable across all levels of the random effects. The respective coefficients of each level of the fixed effect are interpreted as follows:

- **LCAT-LF:** The intercept coefficient of the model, which is 2.45 ($SE = 0.13$, 95% $CI [2.20, 2.71]$, $p < 0.001$), represents the average rating for the LCAT-LF condition. This means that, on average, participants rated the LCAT-LF stimuli 2.45 on the scale.
- **UCAT-LF:** The coefficient for the UCAT-LF condition is -0.58 ($SE = 0.17$, 95% $CI [-1.04, -0.13]$, $p = 0.012$), indicating that participants rated the UCAT-LF stimuli slightly lower by 0.58 points, on average, compared to the LCAT-LF stimuli.
- **LCAT-UF:** The coefficient for the LCAT-UF condition is -3.25 ($SE = 0.25$, 95% $CI [-3.75, -2.75]$, $p < 0.001$), which represents the difference in average scores between LCAT-UF and LCAT-LF when all other predictors are controlled for. The negative coefficient suggests that participants rated the LCAT-UF stimuli 3.25 points lower than the LCAT-LF stimuli, on average.
- **UCAT-UF:** Finally, the coefficient for the UCAT-UF condition is -3.29 ($SE = 0.25$, 95% $CI [-3.82, -2.76]$, $p < 0.001$), indicating that participants rated the UCAT-UF stimuli 3.29 points lower than the LCAT-LF stimuli, on average.

Table 3.6 below provides pairwise comparisons of conditions in the first model. Each row represents a specific comparison between two conditions. The “Estimate” column displays the estimated difference between the conditions. For example, in the LCAT-LF ~ LCAT-UF comparison, the estimate of 3.2517 suggests that, on average, LCAT-LF condition is estimated to receive a higher score than the LCAT-UF condition. The “SE” column shows the standard error of the estimate while “t-ratio” column displays the calculated t-ratio for the comparison. The “p-value” column indicates the probability of obtaining the observed t-ratio or a more extreme value if the null hypothesis (no difference between the conditions) is true.

From the table, we can observe that the LCAT-LF ~ LCAT-UF and LCAT-LF ~ UCAT-UF comparisons have significant differences, as evinced by the very small p-values (both < 0.0001). However, the LCAT-LF ~ UCAT-LF comparison does not attain statistical significance ($p = 0.1114$), suggesting that the difference between these conditions is not statistically significant. Similarly, the LCAT-UF ~ UCAT-UF comparison shows that their negative impact on scores is virtually the same ($p = 0.9983$). Moreover, the LCAT-UF ~ UCAT-LF and UCAT-LF ~ UCAT-UF comparisons both demonstrate significant differences, with p-values less than 0.0001.

The second model had two independent variables as fixed effects. These were the binary categorical variables of “function” and “category” where 1 indicated that conjuncts matched in their grammatical functions or categories, respectively. This model makes it possible to better isolate and estimate the individual impacts of functional and categorical matching between the conjuncts on the acceptability scores of the participants. While Figure 3.3 shows the exact R code to fit the model, Figure 3.7 lists the exact specifications of the second model. As

Comparison	Estimate	SE	t-ratio	p-value
LCAT-LF ~ UCAT-LF	0.5810	0.232	2.508	0.1114
LCAT-LF ~ LCAT-UF	3.2517	0.256	12.682	<0.0001
LCAT-LF ~ UCAT-UF	3.2941	0.270	12.186	<0.0001
UCAT-LF ~ UCAT-UF	2.7132	0.372	7.298	<0.0001
LCAT-UF ~ UCAT-LF	-2.6708	0.299	-8.934	<0.0001
LCAT-UF ~ UCAT-UF	0.0424	0.258	0.164	0.9983

Table 3.6: Pairwise comparisons of conditions in Model I

this model could reach full convergence with the default optimizer algorithm, no alterations were introduced.

```
model_cat <- lmer(rating ~
  function_match + category_match +
  (function_match + category_match | participant_ID) +
  (function_match + category_match | tokenset),
  data = lm_cat)
```

Figure 3.3: R code for Model II of unlike category coordination experiment data

<i>Predictors</i>	acceptability		
	<i>Estimates</i>	<i>CI</i>	<i>p</i>
(Intercept)	2.32	2.00 - 2.65	<0.001
function mismatch	-2.99	-3.46 - -2.51	<0.001
category mismatch	-0.31	-0.59 - -0.03	0.029
Random Effects			
σ^2	1.79		
τ_{00} participant_ID	0.16		
τ_{00} token_set	0.18		
N participant_ID	48		
N token_set	12		
Observations	576		
Marginal R ² / Conditional R ²	0.558 / NA		

Table 3.7: Two independent variables model specifications

The intercept of the second model, which represents the average score when both functions and categories of the conjuncts match, was found to be 2.32 ($SE = 0.16$, 95% $CI [2.00, 2.65]$, $p < 0.001$). The estimated coefficient for function mismatch was -2.99 ($SE = 0.23$, 95% $CI [-3.46, -2.51]$, $p < 0.001$), indicating that when the functions of the conjuncts do not match, the average rating is expected to decrease by 2.99. The coefficient for category mismatch was -0.31 ($SE = 0.14$, 95% $CI [-0.59, -0.03]$, $p = 0.029$), suggesting that the average rating is expected to decrease by 0.31 when the conjuncts have different syntactic categories.

The results of this model suggest that both function mismatch and category mismatch have a significant impact on the acceptability ratings. When the functions of the conjuncts do not match, however, the acceptability scores dramat-

ically decrease, while a mismatch in syntactic categories leads to only a slight decrease in scores.

Note that the intercept of the second model refers to the score of the condition LCAT-LF. It is, however, different from the intercept of the first model, which corresponds to the same variable. This stems from the fact that the second model does not take into account the interaction between the two fixed effects. For this reason, a third model was also fit to the data to account for the interaction between the fixed effects of function mismatch and category mismatch as well. The Figure 3.4 demonstrates the exact R code used to fit the model. The Table 3.8, on the other hand, provides the estimates, confidence intervals, and p-values for each predictor in this model.

```
model_cat_interaction <- lmer(rating ~
  function_match * category_match +
  (function_match + category_match | participant_ID) +
  (function_match + category_match | tokenset),
  data = lm_cat)
```

Figure 3.4: R code for Model III of unlike category coordination experiment data

<i>Predictors</i>	acceptability		
	<i>Estimates</i>	<i>CI</i>	<i>p</i>
(Intercept)	2.45	2.11 – 2.80	<0.001
function mismatch	-3.26	-3.78 – -2.73	<0.001
category mismatch	-0.58	-0.94 – -0.22	0.002
function mismatch × category mismatch	0.54	0.10 – 0.98	0.015
Random Effects			
σ^2	1.77		
τ_{00} participant_ID	0.16		
τ_{00} token_set	0.18		
N participant_ID	48		
N token_set	12		
Observations	576		
Marginal R2 / Conditional R2	0.563 / NA		

Table 3.8: Two independent variables + interaction model specifications

The intercept of the third model (see Table 3.8) indicates that the expected acceptability rating is 2.45 when conjuncts match both in their functions and categories ($SE = 0.17$, 95% CI [2.11, 2.80], $p < 0.001$). The estimate for the function mismatch variable is -3.26 ($SE = 0.26$, 95% CI [-3.78 , -2.73], $p < 0.001$), indicating that a mismatch of functions is associated with an (average) decrease of 3.26 in the acceptability rating when other predictors are held constant. The estimate for the category mismatch is -0.58 ($SE = 0.18$, 95% CI [-0.94 , -0.22], $p = 0.002$). That is, mismatching categories of conjuncts is associated with a decrease of 0.58 in the acceptability rating.

The interaction predictor, function mismatch × category mismatch, has an estimate of 0.54 ($SE = 0.22$, 95% CI [0.10, 0.98], $p = 0.015$). This means that,

on average, when both the functions and the categories of conjuncts mismatch, their cumulative negative effect on the acceptability rating becomes less severe. To calculate the combined effect, we would sum the estimates of the individual main effects (-3.26 for function mismatch and -0.58 for category mismatch) along with the estimate for the interaction term, which is 0.54 .

Discussion

The results obtained from the three models demonstrate that a functional match between the conjuncts is a crucial factor in the acceptability rate of a coordinate structure. In the first model, the conditions where the conjuncts have different functions (i.e., LCAT-UF and UCAT-UF) have large negative impacts on the acceptability ratings. The predictors of the second model indicates that the functional match between conjuncts has the biggest impact on the acceptability ratings with a coefficient estimate of -2.99 . In the third model, the negative impact of mismatching functions becomes more pronounced at an estimated coefficient of -3.26 when the model accounted for the interaction effects between the predictors.

With the exception of the pairwise comparisons in Model I (see Table 3.6), all the analyses show that there is a significant difference between LCAT-LF and UCAT-LF in their acceptability rates. Thus, a categorical incongruity between the conjuncts seems to have a statistically significant impact on the acceptability rate. This impact, however, is not substantial as can be seen from the relevant coefficients of categorical predictors in the models. Furthermore, UCAT-LF is by no means an utterly unacceptable construction as it received a mean score of 1.88 and a median of 2.5 compared to the negative mean scores of LCAT-UF and UCAT-UF conditions.

Based on the results, two distinct conclusions can be drawn. First, we can challenge the original hypothesis and propose a more flexible version of it, implying that the mismatch between syntactic categories or cases could potentially exert a significant influence on acceptability. Alternatively, we can seek to retain the original hypothesis by considering the impact of frequency factors on acceptability. Within the context of the present findings, LCAT-LF, which is the standard type of coordination in Turkish, is undeniably much more frequent in Turkish corpora than UCAT-LF. In experimental syntax literature, there is a substantial body of evidence indicating that structures with low frequencies tend to receive lower acceptability ratings when compared to high-frequency structures, assuming both structures are grammatical (Bresnan, 2007; Divjak, 2016; Francis, 2021). Therefore, a minor disparity in favor of the LCAT-LF condition is expected.

As a result, the present thesis leans towards the latter conclusion, thereby upholding the original hypothesis. This stance essentially asserts that the slight variation in acceptability ratings between the conditions could be attributed to the frequency difference between the LCAT-LF and UCAT-LF conditions.

3.3.3 Coordination of unlike cases

Descriptive statistics

As demonstrated in Table 3.9, the mean rating for LCASE-LF coordination type is 2.32, with a standard deviation of 1.38. The median rating is 3, indicating a generally positive evaluation of this coordination type, which is not surprising as this is the standard type of coordination in Turkish. On the other hand, the mean rating for UCASE-LF coordination type is 1.58, with a higher standard deviation of 1.71. The median rating is 2.00, suggesting a positive evaluation of this coordination type, but with more variability than LCASE-LF. Finally, the mean rating for UCASE-UF coordination type is negative at -0.37 , with the highest standard deviation of 2.04. The median rating is -1.00 , indicating an overall negative evaluation of this coordination type.

Coordination type	N	Mean	SD	Min	Max	Median
LCASE-LF	144	2.32	1.38	-3	3	3.00
UCASE-LF	144	1.58	1.71	-3	3	2.00
UCASE-UF	144	-0.37	2.04	-3	3	-1.00

Table 3.9: Descriptive statistics of unlike case coordination experiment

Participants strongly agree that LCASE-LF is highly acceptable, as can be seen in Figure 3.5. The scores of UCASE-LF, however, show more variability as the IQR box is relatively more spread out between the scores of 1 and 3. UCASE-LF, however, still seems to be an acceptable type as the distribution is again considerably left-skewed towards the positive end of the scale. UCASE-UF, on the other hand, seems to be slightly right-skewed towards the negative end of the scale. The variability of UCASE-UF scores, however, is quite high as IQR box is extended between the scores of -2 and 1.

Inferential Statistics

Similar to the analyses of unlike category coordination experiment, in addition to fitting a linear mixed-effects model, several Wilcoxon signed-rank tests were conducted to compare relevant pairs of conditions. The difference between LCASE-LF and UCASE-LF was significant ($V = 2655$, $p < .001$), confirming that mismatching cases of conjuncts had a significant negative impact on acceptability. However, the difference between UCASE-LF and UCASE-UF was also significant ($V = 5859$, $p < .001$), suggesting functional incongruity between the conjuncts also had a considerable negative impact on acceptability.

As opposed to the unlike category coordination experiment, only one linear mixed-effects model was fitted to the data as unlike case coordination experiment did not have a proper 2×2 factorial design (see §3.2.1). This model is practically the same as the first model used to analyze the unlike category coordination experiment.

This model had one independent variable called “condition” encoded as the fixed effect. The distinct types of coordination (i.e., LCASE-LF, UCASE-LF, UCASE-UF) were analyzed as the levels of this variable. The random effects were the same as the previous models: IDs of the participants and the token set indices.

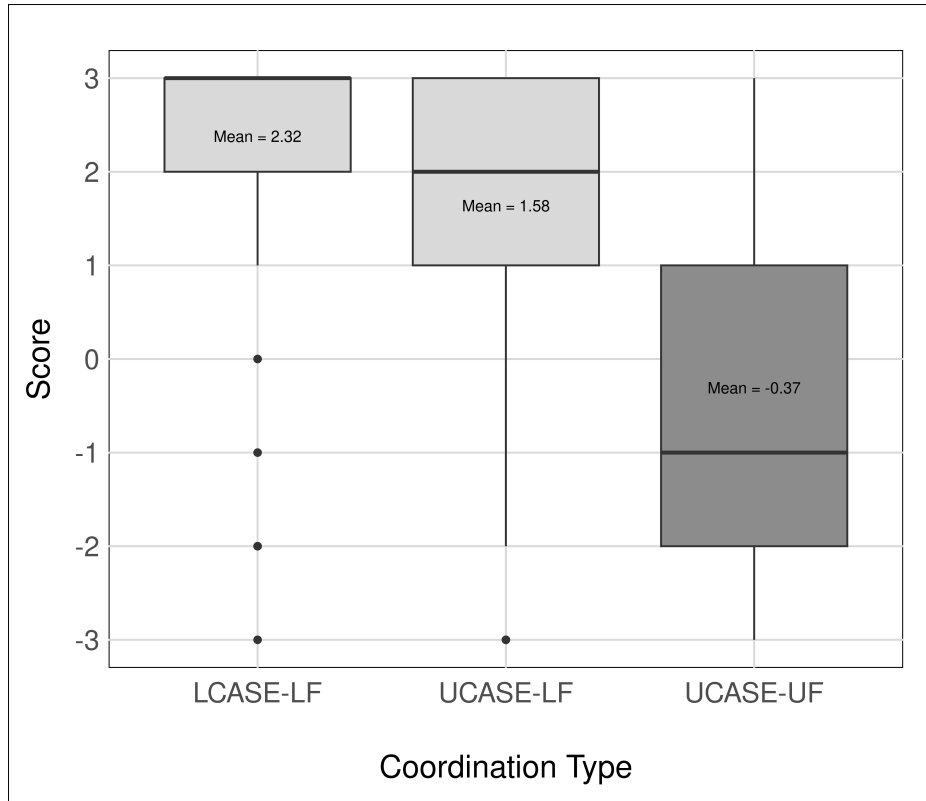


Figure 3.5: Boxplot for unlike case coordination experiment

Since this model also did not fully converge with the default optimizer algorithm, “bobyqa” algorithm had to be manually introduced. The exact code used to fit the model can be seen in Figure 3.6. The specifications of the fitted model are summarized in Table 3.10.

```
control <- lmerControl(optimizer = "bobyqa")

model_case = lmer(rating ~ condition +
  (1 + condition|participant_ID) +
  (1 + condition|tokenset),
  data = lm_case,
  control = control)
```

Figure 3.6: R code for the model of unlike case coordination experiment data

As revealed by the model, participant scores were significantly affected by the experimental conditions UCASE-LF and UCASE-UF. The intercept, representing the expected score for LCASE-LF, was estimated to be 2.32 ($SE = 0.15$, 95% CI [2.02, 2.62], $p < 0.001$). The scores given to UCASE-LF stimuli were 0.75 points lower ($SE = 0.20$, 95% CI [-1.15, -0.34], $p < 0.001$) than LCASE-LF condition, on average. UCASE-UF stimuli received 2.70 points lower scores ($SE = 0.31$, 95% CI [-3.31, -2.09], $p < 0.001$) than LCASE-LF condition, on average.

These coefficients indicate that both experimental conditions have significant negative effects on participant scores, compared to LCASE-LF, which is the standard type of coordination in Turkish. However, the negative effect is consid-

<i>Predictors</i>	acceptability		
	<i>Estimates</i>	<i>CI</i>	<i>p</i>
(Intercept)	2.32	2.02 – 2.62	<0.001
condition [UCASE-LF]	-0.75	-1.15 – -0.34	<0.001
condition [UCASE-UF]	-2.70	-3.31 – -2.09	<0.001
Random Effects			
σ^2	1.93		
τ_{00} participant_ID	0.22		
τ_{00} token_set	0.06		
N token_set	12		
Observations	432		
Marginal R2 / Conditional R2	0.300 / 0.554		

Table 3.10: One independent variable case model specifications

erably larger for the UCASE-UF condition than for the UCASE-LF condition.

Discussion

The combined results from the linear mixed-effects model and Wilcoxon signed-rank test comparisons draw a picture similar to the unlike category coordination experiment. Functional matching is indeed the most important factor when it comes to the acceptability of Turkish coordinate structures. Mismatching case configuration alone was not able to engender a highly negative response from the participants despite being a significant and slightly negative predictor of acceptability ratings. Furthermore, the UCASE-UF condition cannot also be considered as an outright unacceptable condition, as the mean score of UCASE-UF is still significantly higher than the mean score of ungrammatical fillers.

Once again, these findings lend support to the hypothesis that regards functional matching between the conjuncts as the essential factor behind the acceptability of Turkish coordinate structures.

3.4 Conclusion

This chapter discussed the methodology and results of the formal acceptability judgment experiment conducted within the context of the present thesis. The experiment provided data that is valuable not only within the scope of the present thesis but also in terms of the general theoretical research on coordination.

The results strongly support the claim that the functional congruence between the conjuncts is the main driving factor behind the acceptability of Turkish coordinate structures. However, the claim that functional congruence is the only significant factor behind acceptability could be problematic if one does not acknowledge the impact of frequency factors on acceptability ratings. Furthermore, both UCAT-LF and UCASE-LF received considerably positive ratings and should ultimately be considered as highly acceptable constructions.

The acceptability experiment further evaluated the acceptability rates of coordination configurations featuring mismatching functions, denoted by conditions ending with UF. The tested configurations also included different types of

absolute unlike coordination that were not encountered within the corpus. The results of the experiment suggest that all such configurations have a significantly negative effect on acceptability. Notably, the rarity of these constructions in the corpus appropriately mirrors their ungrammatical status. Overall, the results of the acceptability experiment align with the results of the corpus investigation.

Chapter 4

Formal analysis

4.1 Overview of the chapter

This chapter presents a formal analysis of Turkish unlike coordination phenomena within the framework of Lexical-Functional Grammar (LFG). In §4.2, the rationale for adopting this particular theoretical framework is justified. Next, in §4.3, the fundamental properties of LFG are introduced, providing essential background information for the analysis. §4.4 presents a detailed formal analysis of Turkish unlike coordination within the context of distinct syntactic configurations. Finally, §4.5 demonstrates how the analysis works within the context of actual examples of unlike coordination.

4.2 Frameworks of syntax

Modern approaches to natural language syntax can be broadly classified into two overarching frameworks (Pullum and Scholz, 2001): *generative-enumerative syntax* and *model-theoretic syntax*. Although these two frameworks share the common goal of capturing the syntactic parameters of natural languages, they differ significantly in their approaches to achieving this objective.

Generative-enumerative syntax (henceforth abbreviated as GES) has its origins in the syntactic component of mathematical logic and formal language theory, with foundational contributions made by Chomsky (1959). Within the GES framework, language is viewed as a (potentially infinite) collection of strings generated by a formal “device.” This formal device consists of a finite set of rewrite rules, often referred to as phrase-structure rules, and lexical items. Consequently, the responsibility of this device is to generate (and enumerate) all and only the grammatical strings of a given language. Accordingly, linguists working within the GES framework endeavor to formulate a precise set of rules for this device that can adequately account for grammatical strings while avoiding the generation of ungrammatical ones. Prominent theoretical approaches following GES include Minimalism (Chomsky, 1993) and its precursor, Government and Binding Theory (Chomsky, 1981), as well as Tree-Adjoining Grammars (Joshi, 1987) and various forms of categorial grammars.

In contrast to GES, *model-theoretic syntax* (henceforth abbreviated MTS) primarily originates from the semantic domain of mathematical logic. Unlike GES,

MTS does not rely on rewrite rules but rather employs a set of constraints to define syntax. These constraints do not aim to exhaustively enumerate all and only the grammatical strings of a given language. Instead, they collectively define syntax by specifying the necessary conditions for the desired syntactic structure of a valid natural language string. Consequently, a natural language string is deemed grammatical if it satisfies the constraints formulated by a linguist, while it is considered ungrammatical if it violates them. It is important to note that these constraints do not generate syntactically well-formed strings, neither conceptually nor formally. Rather, they model possible syntactic structures that can characterize a well-formed string. Notable MTS-based approaches include Head-Driven Phrase Structure Grammar (HPSG; Pollard and Sag, 1994; Müller et al., 2021) and Lexical-Functional Grammar (LFG; Kaplan and Bresnan, 1982; Dalrymple et al., 2019).

MTS-based approaches offer several advantages over their GES-based counterparts both in formal and empirical domains (for an overview, see Pullum and Scholz, 2001; Sag and Wasow, 2011). One notable advantage of MTS-based approaches lies in their ability to account for gradience in ungrammaticality, a feature lacking in GES-based approaches. In GES-based approaches, grammars generate all and only the grammatical strings that form a set. Thus, a given sentence either belongs to this set (grammatical) or not (ungrammatical). However, this binary classification fails to capture the varying degrees to which a sentence violates syntactic rules. To exemplify this issue, consider the following sentences in (37).

- (37) a. A wrong decision is better than indecision.
 b. *A wrong decision are better than indecision.
 c.** Decision a better wrong is indecision than.

In a GES-based grammar of English, both (37b) and (37c) would be classified as ungrammatical since they do not belong to the generated set of grammatical strings. While both sentences are indeed ungrammatical, the extent to which they violate the syntactic rules of English differs. Sentence (37b) still retains some ability to express a proposition, albeit with a subject-verb agreement error. In contrast, sentence (37c) fails to convey any coherent meaning at all. However, within the GES framework, there is no inherent mechanism to account for gradience in ungrammaticality, as grammaticality is treated as a binary property tied to set membership.

On the other hand, an MTS-based grammar provides the necessary tools to model degrees of ungrammaticality. A given string does not need to violate all syntactic constraints simultaneously. For instance, sentence (37b) violates only the constraint that governs the number and person agreement between subject and verb in English, while satisfying all other syntactic constraints. In contrast, sentence (37c) violates numerous constraints of English syntax. Consequently, an MTS-based grammar can effectively explain and measure the degrees of grammaticality by assessing the number of constraint violations, rather than relying solely on set membership.

Another advantage of MTS stems from its ability to effectively model the incremental processing of language, a crucial aspect of real-life language compre-

hension. When individuals encounter a sentence, they begin constructing its meaning and syntactic structure as soon as they start perceiving the linguistic input, without waiting for the full utterance to conclude. This incremental processing involves constant revision and updating of the partial structure in their minds as new linguistic items unfold.¹

To effectively model this dynamic and incremental process, a formal mechanism is required to construct and evaluate partial structures. GES-based approaches, however, fall short in this regard since their rewrite rules solely enumerate complete expressions and exclude the fragments that constitute them. For instance, the string *a wrong decision is* does not exist within the set of grammatical strings in GES-based grammars. Consequently, it is treated no differently than a randomly ordered sequence of symbols.

In contrast, MTS-based approaches, such as HPSG and LFG, are equipped to handle partial structures. In these approaches, the string *a wrong decision is* is analyzable and already satisfies numerous constraints pertaining to the structure of English NPs and subject-verb agreement, among others. Moreover, lexical items in MTS-based approaches inherently carry diverse morphosyntactic information regardless of whether they are part of a larger syntactic construction. This means that even as isolated expressions, lexical items contribute to (partial) meaning and morphosyntactic structure. This feature-rich nature of lexical items enables MTS-based approaches to account for partial expressions giving rise to both meaning and morphosyntactic structure. Consequently, it can be claimed that MTS-based approaches enable linguists to more accurately model the intricacies of linguistic knowledge in humans.

Building upon the advantages offered by MTS foundations, this thesis proposes a formal analysis of Turkish coordination within an MTS-based framework, namely Lexical-Functional Grammar (LFG). While LFG also incorporates elements from GES, the analysis presented here integrates mechanisms proposed in previous works to align LFG more closely with the core principles of MTS. Thus, the analysis not only provides a formalized account of the empirical findings regarding Turkish coordination but also exemplifies a form of LFG that more closely adheres to the core principles of MTS.

4.3 Lexical-Functional Grammar

4.3.1 Overview of LFG architecture

LFG emerged as a compelling alternative to the prevailing transformative framework in 1970s. It was proposed by Joan Bresnan, a linguist, and Ronald Kaplan, a psychologist. The motivation behind their proposal primarily arose from the transformative framework's shortcomings in terms of psychological plausibility and its limited ability to model crosslinguistic generalizations effectively (Dalrymple et al., 2019, pp. 1–2). Unlike transformative frameworks, as well as other constraint-based frameworks, LFG posits a modular architecture of grammar, wherein distinct forms of linguistic knowledge are represented at different formal

¹This phenomenon of incremental processing has been extensively supported by a wealth of psycholinguistic research. See Levelt (1993), Tanenhaus et al. (1995), and von der Malsburg and Vasishth (2011).

levels. Crucially, these levels are interconnected through well-defined mappings (*projections*), enabling the exchange of information across levels.

In LFG, syntax is handled through two key levels: c-structure (constituent structure) and f-structure (functional structure). To elucidate these distinct structures, let us consider the sentence (38) and its corresponding LFG analysis, as portrayed in Figure 4.1.

(38) Tony likes the book.

The syntax of the sentence is captured by two distinct data structures that are systematically connected. C-structure (depicted on the left) is a phrase-structure tree, generated using a specific type of phrase-structure rules encoded in the grammar. At the c-structure level, the information is primarily concerned with the linear order of words and the underlying constituent structure of the sentence. By contrast, the f-structure (presented on the right) represents more abstract syntactic relations and notions in the form of an unordered collection of pairs of attributes and values. Importantly, the f-structure is constructed (denoted by arrows in the diagram) based on the information present on specific c-structure nodes. Let us now elucidate how LFG grammars can construct such analyses.

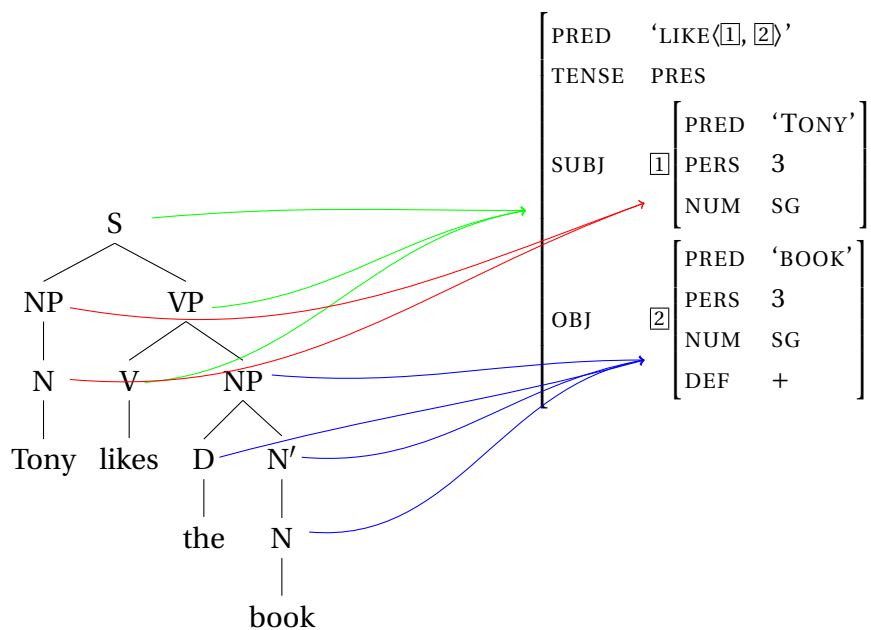


Figure 4.1: C-structure and f-structure of sentence (38)

4.3.2 LFG grammars

LFG grammars consist of two fundamental components: a set of c-structure rules and a set of lexical entries. While this setup may initially resemble traditional phrase-structure grammars, LFG introduces several key distinctions. To gain a better understanding of LFG grammars and how they function, let us construct a toy grammar for English in LFG step-by-step. The goal of our toy grammar will be to parse the sentence *Tony likes the book*, as shown in Figure 4.1.

C-structure rules

First, let us consider the conventional phrase-structure rules that would allow our grammar to parse the sentence *Tony likes the book*. The rules provided in (39) can produce the simple phrase-structure tree depicted in Figure 4.1.

- (39) $S \longrightarrow NP \quad VP$
 $VP \longrightarrow V \quad NP$
 $NP \longrightarrow \{N \mid D N'\}$
 $N' \longrightarrow N$

In LFG, however, phrase-structure rules have to be somehow connected to f-structure since syntax in LFG is decomposed into two distinct levels of representation. To establish this connection between c-structure and f-structure, LFG introduces specific annotations called functional annotations on phrase-structure rules. Consequently, the traditional phrase-structure rules in (39) would be represented in LFG grammars as follows:

- (40) $S \longrightarrow \begin{array}{cc} NP & VP \\ (\uparrow \text{ SUBJ}) = \downarrow & \uparrow = \downarrow \end{array}$
 $VP \longrightarrow \begin{array}{cc} V & NP \\ \uparrow = \downarrow & (\uparrow \text{ OBJ}) = \downarrow \end{array}$
 $NP \longrightarrow \begin{array}{ccc} \{ N & \mid & D \quad N' \} \\ \uparrow = \downarrow & & \uparrow = \downarrow \quad \uparrow = \downarrow \end{array}$
 $N' \longrightarrow \begin{array}{c} N \\ \uparrow = \downarrow \end{array}$

In the above representation, c-structure nodes are annotated with two distinct symbols: \downarrow and \uparrow . The symbol \downarrow refers to the f-structure associated with the c-structure node itself, while \uparrow refers to the f-structure of the mother node, i.e., the node immediately dominating the c-structure node in question.

In the first rule, the annotation $\uparrow = \downarrow$, directly below the VP node, indicates that the f-structure of VP is to be equated with the f-structure of its mother node, S. This means that the information associated with VP will be integrated into the overall f-structure of the S node. Intuitively, the annotation $\uparrow = \downarrow$ serves the function of propagating information up the tree through unification of f-structures.

The annotation $(\uparrow \text{ SUBJ}) = \downarrow$ below the NP node is slightly more complex. The first part of the annotation, $(\uparrow \text{ SUBJ})$, indicates that there is a SUBJ attribute within the f-structure of the mother node. The remaining part of the annotation, $= \downarrow$, specifies that the value of that SUBJ attribute is the f-structure of the NP node. This rule states that the first NP is the subject of the sentence. Similarly, the second rule specifies that the f-structure of the V node is directly equated with the

f-structure of its mother node, which is VP. Additionally, the f-structure associated with the immediately following NP node is assigned as the value of the OBJ attribute within the mother node's f-structure. Consequently, this rule states that the NP immediately following V is the object of the sentence. As for the NP and N' rules, the $\uparrow = \downarrow$ annotations on the daughters simply serve to “carry” and “unify” the f-structures of the daughters at the level of their respective mother nodes.

Lexical entries

The f-structure information on the nodes is mainly supplied by the lexical entries of the words appearing in a sentence. These lexical entries play a crucial role in providing the necessary linguistic information to build the f-structures. These lexical entries typically include both *defining equations* (signified by the use of the plain “=” sign), which create attribute-value pairs in the f-structure, and *constraining equations* (signified by the use of the “=_c” sign), which do not create but only verify the presence of specific attribute-value pairs in the f-structure. Let us examine the lexical entry of the word *likes* in (41) to better understand how the equations in lexical entries work.

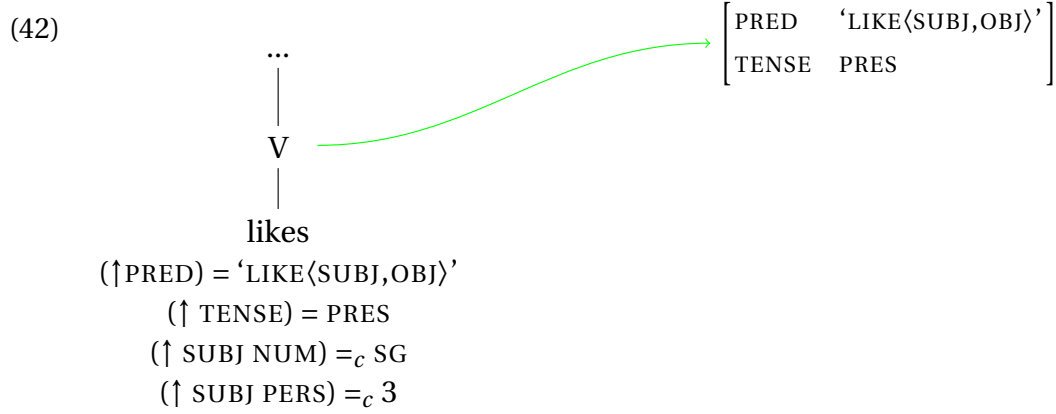
- (41) likes V (↑ PRED) = 'LIKE(SUBJ, OBJ)'
 (↑ TENSE) = PRES
 (↑ SUBJ NUM) =_c SG
 (↑ SUBJ PERS) =_c 3

In the lexical entry above, the first is a defining equation that creates an attribute-value pair $\langle \text{PRED}, \text{'LIKE'}(\text{SUBJ}, \text{OBJ}) \rangle$. Intuitively, this equation indicates that the word *likes* represents the (abstract) predicate LIKE and requires both a subject and an object. Similarly, the following defining equation generates the attribute-value pair $\langle \text{TENSE}, \text{PRES} \rangle$, signifying that the word is in the present tense.

However, the last two equations are constraining equations, responsible for verifying the presence (or absence) of specific features once the entire f-structure of the sentence is constructed. The equation $(\uparrow \text{SUBJ NUM}) =_c \text{SG}$ ensures that the subject of the verb *likes* must be in the singular form. The presence of this equation means that the f-structure associated with the subject should contain an attribute-value pair $\langle \text{NUM}, \text{SG} \rangle$. If the sentence’s f-structure does not satisfy this constraint, the analysis will be disallowed.

Similarly, the equation $(\uparrow \text{SUBJ PERS}) =_c 3$ ensures that the subject of the verb *likes* must be in the third person. It requires the f-structure value of the SUBJ attribute of the sentence to have an attribute-value pair $\langle \text{PERS}, 3 \rangle$. If a sentence violates this constraint, the analysis is ruled out.

Importantly, all the equations employ the \uparrow symbol, allowing the word *likes* to instantiate its f-structure on its mother node, which in this case can only be a V node. The visualization in (42) demonstrates how a word instantiates its f-structure on its preterminal node. As depicted, the instantiated f-structure on the V node does not have a SUBJ attribute, as constraining equations cannot generate attribute-value pairs.



Let us now expand the lexicon of our toy grammar by adding other lexical entries so that our grammar can recognize the sentence *Tony likes the book*. All the lexical entries in our toy grammar are listed in (43).

- (43)
- | | | |
|-------|---|---|
| Tony | N | $\begin{aligned} (\uparrow \text{PRED}) &= \text{'TONY'} \\ (\uparrow \text{PERS}) &= 3 \\ (\uparrow \text{NUM}) &= \text{SG} \end{aligned}$ |
| | | |
| likes | V | $\begin{aligned} (\uparrow \text{PRED}) &= \text{'LIKE<SUBJ, OBJ>'} \\ (\uparrow \text{SUBJ NUM}) &=_{\text{c}} \text{SG} \\ (\uparrow \text{SUBJ PERS}) &=_{\text{c}} 3 \\ (\uparrow \text{TENSE}) &= \text{PRES} \end{aligned}$ |
| | | |
| the | D | $(\uparrow \text{DEF}) = +$ |
| | | |
| book | N | $\begin{aligned} (\uparrow \text{PRED}) &= \text{'BOOK'} \\ (\uparrow \text{PERS}) &= 3 \\ (\uparrow \text{NUM}) &= \text{SG} \end{aligned}$ |

Full analysis

Now that we have expanded the lexicon of our toy grammar with additional lexical entries, we can analyze the sentence *Tony likes the book* using the formulated c-structure rules and lexical entries. The fully annotated c-structure, along with the corresponding f-structure, is presented in Figure 4.2. The f-structure on the right makes explicit the mapping between the c-structure nodes and the elements in the f-structure through indices on nodes.²

²The indices on c-structure nodes are not part of the grammar. They are inserted on the nodes to facilitate understanding the mapping between the c-structure and f-structure.

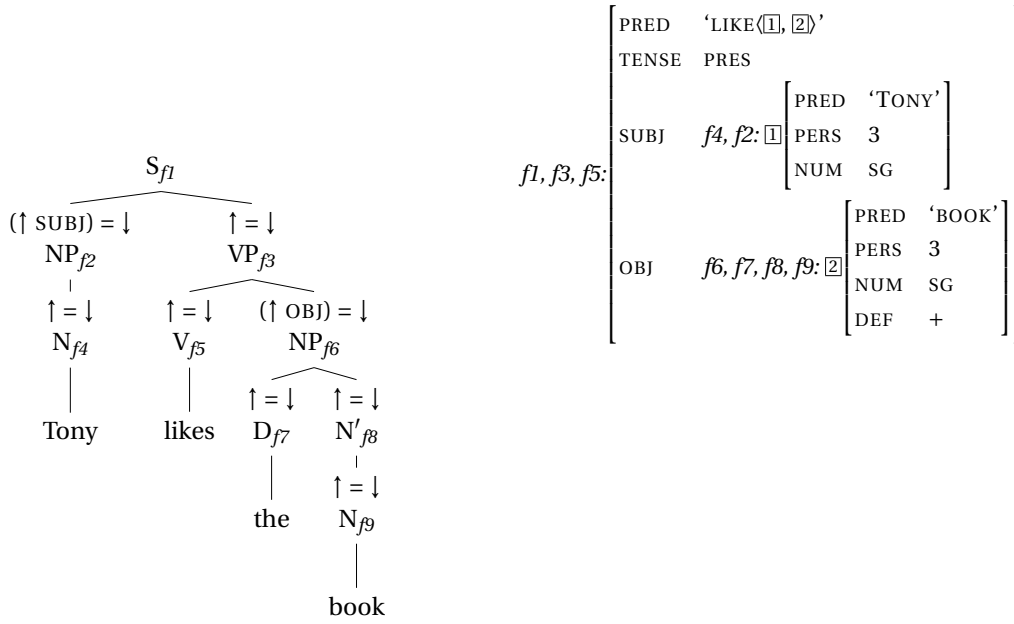


Figure 4.2: Annotated c-structure and f-structure of sentence (38)

The c-structure represents the hierarchical syntactic structure of the sentence. At the top, we have the S node, representing the sentence as a whole. The S node has two daughters: NP_{f2} and VP_{f3} . The NP_{f2} node represents the subject *Tony*, and its f-structure is supplied by the lexical entry for *Tony*, which is relayed through the N_{f4} node. The VP_{f3} node represents the verb phrase *likes the book*, and its f-structure is built from the lexical entry for *likes* and the f-structure of the NP_{f6} node, which is assigned as the value of the OBJ attribute. The f-structure associated with the NP_{f6} node is constructed by the integration of the information provided by the lexical entries for *the* and *book*.

Well-formedness conditions

In LFG, the syntactic well-formedness of a sentence depends on several factors. If a sentence cannot be parsed validly according to our c-structure rules, it is not recognized by our grammar. This means our grammar fails to construct both the c-structure and f-structure for the sentence. However, being parsed by the grammar is not sufficient for a string to be considered syntactically well-formed; the constructed f-structure must also satisfy certain conditions.

An f-structure is deemed valid if it fulfills three overarching conditions simultaneously: Uniqueness, Completeness, and Coherence. The Uniqueness condition stipulates that each attribute in an f-structure should be associated with only one value. Consequently, while different attributes may map to the same value, an attribute cannot be mapped to two distinct values. The Completeness condition ensures that the obligatory arguments of a predicate are present in the f-structure. For instance, if the predicate *like* requires both a subject and an object, the relevant attribute-value pairs incorporating SUBJ and OBJ should be present in the f-structure. Consequently, the sentence *Tony likes* violates the Completeness condition as the OBJ requirement of *likes* is not fulfilled. The Coherence condition mandates that no additional governable function (i.e., SUBJ, OBJ, OBL

etc.) should be present in the f-structure. This condition rules out sentences like *my mother cried sadness*, as the verb *cry* does not take an object.

Additionally, f-structures can be rendered invalid if they violate constraining equations imposed on them. For example, if we introduce the word *children* to our lexicon as in (44), our toy grammar would still parse the sentence *Children likes the books* given our c-structure rules and lexical entries and produce a corresponding f-structure representation. However, the sentence would still be considered ungrammatical by our grammar since the plural subject *children* is incompatible with the constraining equation, $(\uparrow \text{ SUBJ NUM}) =_c \text{ SG}$, found in the lexical entry of the verb *likes*.

- (44) children N $(\uparrow \text{ PRED}) = \text{'CHILD'}$
 $(\uparrow \text{ PERS}) = 3$
 $(\uparrow \text{ NUM}) = \text{PL}$

- (45) Children likes the books.

PRED	'LIKE(<u>1</u> , <u>2</u>)'								
TENSE	PRES								
SUBJ	<u>1</u> <table> <tr> <td>PRED</td><td>'CHILD'</td></tr> <tr> <td>PERS</td><td>3</td></tr> <tr> <td>NUM</td><td>PL</td></tr> </table>	PRED	'CHILD'	PERS	3	NUM	PL		
PRED	'CHILD'								
PERS	3								
NUM	PL								
OBJ	<u>2</u> <table> <tr> <td>PRED</td><td>'BOOK'</td></tr> <tr> <td>PERS</td><td>3</td></tr> <tr> <td>NUM</td><td>SG</td></tr> <tr> <td>DEF</td><td>+</td></tr> </table>	PRED	'BOOK'	PERS	3	NUM	SG	DEF	+
PRED	'BOOK'								
PERS	3								
NUM	SG								
DEF	+								

4.3.3 Coordination in LFG

Sets and distributivity

In LFG, coordinate structures are represented as sets in the f-structure. This representation decision is based on two fundamental observations: 1) There is no limit to the number of elements that can be coordinated in a given coordinate structure (abstracting away from working memory limitations); 2) The conjuncts in a coordinate structure are equal, meaning there is typically no hierarchical relationship between them.

LFG sets are subject to fundamental set-theoretic operations such as union and intersection. However, sets representing coordinate structures in LFG are unique in that they can have their own attribute-value pairs alongside their elements. This distinction arises from the observation that coordinate structures may have properties independent from their conjuncts. Hence, the sets representing coordinate structures are called *hybrid objects* (Dalrymple et al., 2019, pp. 49–50).

Consider the f-structure (46) for the sentence *Tony and Paulie run*. The SUBJ attribute maps to an f-structure that encompasses both a set (of conjuncts) and its own features $\langle \text{NUM}, \text{PL} \rangle$ and $\langle \text{CONJ}, \text{AND} \rangle$. This hybrid structure allows us to

explain why the verb *run* is grammatical instead of *runs* despite both conjuncts being singular nouns. The constraining equation associated with *runs* demands that its subject must be singular, represented by the necessary presence of $\langle \text{NUM}, \text{SG} \rangle$ in the subject f-structure. However, in the hybrid structure associated with the SUBJ attribute, we find the presence of $\langle \text{NUM}, \text{PL} \rangle$, which clashes with the constraint imposed by *runs*.

$$(46) \left[\begin{array}{l} \text{PRED} \quad \text{'RUN'} \\ \text{TENSE} \quad \text{PRES} \\ \\ \text{SUBJ} \quad \boxed{\text{I}} \end{array} \left[\begin{array}{l} \text{CONJ} \quad \text{AND} \\ \text{NUM} \quad \text{PL} \\ \left\{ \begin{array}{l} \left[\begin{array}{l} \text{PRED} \quad \text{'TONY'} \\ \text{PERS} \quad 3 \\ \text{NUM} \quad \text{SG} \end{array} \right] \\ \left[\begin{array}{l} \text{PRED} \quad \text{'PAULIE'} \\ \text{PERS} \quad 3 \\ \text{NUM} \quad \text{SG} \end{array} \right] \end{array} \right\} \end{array} \right] \right]$$

These hybrid objects are constructed with set-theoretic annotations on c-structure rules. Consider the basic NP coordination rule in (47). The NP nodes corresponding to conjuncts are adorned with the special notation $\uparrow \in \downarrow$, indicating that the associated f-structure of the NP node belongs to a set.

$$(47) \quad \text{NP} \longrightarrow \text{NP}^+ \quad \text{Cnj} \quad \text{NP} \\ \uparrow \in \downarrow \quad \uparrow = \downarrow \quad \uparrow \in \downarrow$$

The information introduced by the lexical entry of the conjunction is directly mapped to the f-structure that encompasses the set, as denoted by “ $\uparrow = \downarrow$ ” beneath the Cnj node. As demonstrated in (46), the information associated with the lexical entry of *and* directly maps to the hybrid object and is not encapsulated as an additional member of the set.

$$(48) \quad \text{and} \quad \text{Cnj} \quad \begin{array}{l} (\uparrow \text{ CONJ}) = \text{AND} \\ (\uparrow \text{ NUM}) = \text{PL} \end{array}$$

Furthermore, the properties associated with the coordination are classified into two classes: 1) distributive and 2) non-distributive properties. They are formally defined in Dalrymple and Kaplan (2000, p. 779) as follows:

(49) *Distributive* and *nondistributive* features:

For any *distributive* property P and set s , $P(s)$ iff $\forall f \in s. P(f)$.

For any *non-distributive* property P and set s , $P(s)$ iff P holds of s itself.

Thus, if a feature is encoded as distributive in an LFG grammar, it applies to all conjuncts encapsulated in the set. For example, if CASE is encoded as a distributive property and it is assigned to coordinate structure, then all conjuncts must

bear the assigned case value if there is no overriding constraint. By contrast, non-distributive properties do not distribute to the elements of the set (f-structure of individual conjuncts) and only act as independent properties of the set. In this regard, when a property is encoded as non-distributive, conjuncts are allowed to have discrepant features pertaining to that property.

CAT predicate

CAT predicate is a formal device implicated in the analysis of unlike coordination in LFG (see Dalrymple and Kaplan, 2000; Dalrymple, 2017). The device provides a means for specifying the simple categorical requirements of a given predicate on its arguments.

Formally, CAT predicate is a function that takes two arguments: an f-structure f and a set of syntactic categories C . Its purpose is to determine whether there exists a syntactic category in the input set C that also maps to the given f-structure f . For example, the fact that the verb *become* constrains its predicative argument (which is encoded as the value of PREDLINK attribute in multiple LFG analyses) to either be an NP or an AdjP can be captured with the statement $\text{CAT}((\uparrow \text{PREDLINK}), \{\text{NP}, \text{AdjP}\})$, which evaluates to true if the f-structure associated with PREDLINK function is mapped to either an NP or an AP.

4.4 Analysis of unlike coordination in Turkish

Before proposing a formal analysis that captures empirical facts obtained from corpus and acceptability judgment studies, let us outline some desiderata for the analysis, which are crucial to ensure both theoretical parsimony and broad applicability.

Firstly, to ensure theoretical simplicity and broad applicability, the analysis should avoid introducing coordination mechanisms or overarching grammar parameters specific to Turkish. Instead, basic coordination facts observed in Turkish should naturally fall out from more fundamental constraints pertaining to the morphosyntactic relation between Turkish predicates and their arguments. In other words, the focus should be on developing accurate predicate constraints that can be evaluated for each conjunct in a coordination.

Secondly, the proposed analysis should avoid introducing significant extensions to the existing LFG architecture. In this way, it can be ensured that the analysis remains compatible with XLE (Xerox Linguistic Environment; Crouch et al., 2017), which is a computational platform for implementing LFG grammars.

4.4.1 Evaluating possible solutions

CAT predicate

Modeling unlike category coordination through the mechanism of CAT predicate was pursued in Dalrymple (2017). Under her analysis, syntactic category labels are replaced with feature matrices. For example, instead of using simple categories, like N or NP, Dalrymple's approach represents them as complex feature matrices, such as $[N +, V -, P -, \text{ADJ} -, \text{ADV} -]$.

In the context of coordination, the feature matrix associated with the coordination node aggregates via a specific coordination rule the categorical information of each conjunct. For instance, the coordinate structure incorporating a noun and an adjective would bear the feature matrix [N +, ADJ +].

Categorical constraints imposed by the predicates are also expressed using these feature matrices. When a predicate constrains the category of its arguments, it indicates via the CAT predicate the categories that are not allowed, leaving room for the co-occurrence of permitted categories. For instance, the predicative argument of the Turkish verb *ol-* ‘be/become’, can be realized by an NP, AdjP, PP or their coordination. This categorical constraint would be encoded under Dalrymple’s analysis as in (50).³

- (50) CAT((↑ PREDLINK), %C)
 (%C V) = –
 (%C ADV) = –

Dalrymple’s CAT predicate analysis coupled with complex feature matrices offers an elegant analysis of unlike category coordination. However, its coverage is far from comprehensive, as pointed out by Przepiórkowski and Patejuk (2021), who argue that predicates not only impose general categorical constraints on their arguments but also intricate morphosyntactic requirements.

To illustrate this point, consider the Turkish sentences in (51) and (52). In (51), both examples involve a coordinate structure consisting of a PP and an NP that serves as the object argument of the verb *konuş-* ‘talk/converse’. However, not all NPs or PPs can occupy the object argument position of this verb. Specifically, the head of the PP must be either the postposition *hakkında* ‘about’, as seen in (51a–b), or *üzerine* ‘upon/over’, as demonstrated in (52). Additionally, the object NP argument must be in either the nominative case, as seen in (51a) (indicating non-specificity), or the accusative case, as observed in (51b) (indicating specificity).

- (51) a. Kolektör-ler sık sık [antik enstrüman-lar
 collector-PL frequent frequent antique instrument-PL
 hakkında]_{pp} veya [ticari bilgi]_{NP_{nom}} konuş-ur-lar.
 about or commercial information talk-AOR-3PL
 ‘Collectors frequently talk about antique instruments or about commercial information.’
 (trTenTen)
- b. Kendi-si ile [Sofya baş müftü yardımcısı Necati
 self-3P with Sofia chief mufti deputy Necati
 Ali hakkında]_{pp} ve [yap-tık-ları hizmet-ler-i]_{NP_{acc}}
 Ali about and do-PTCP-3PL.POSS service-PL-ACC
 konuş-tu-k.
 talk-PST-1PL
 ‘We talked about Necati Ali, the deputy chief mufti of Sofia, and the services they provide.’
 (trTenTen)

³%C acts as a placeholder name (as indicated by the % sign) for the f-structure that is the value of the PREDLINK attribute.

- (52) Sayın Babacan ile Ortadoğu-da-ki son
 honourable Babacan with Middle East-LOC-ADJZ last
 gelişme-ler üzerine konuş-tu-k.
 development-PL over talk-PST-1PL

‘With honourable Babacan, we talked about the latest developments in the Middle East.’ (trTenTen)

Przepiórkowski and Patejuk (2021) (henceforth abbreviated as P&P) provide further examples from Polish and English that illustrate similarly complex morphosyntactic constraints and subsequently consider a potential amendment to Dalrymple’s account in terms of complex categories. This solution involves representing various morphosyntactic properties, such as case and the form of the postposition, in the syntactic category labels. For instance, this solution would necessitate encoding the category of *hakkında* not as a simple P but as a complex category P[*hakkında*]. Similarly, we would need to encode the grammatical case of an NP in terms of complex categories. Since practically all Turkish grammatical cases are found in unlike coordination configurations, this would require 7 distinct nominal categories in the form of NP[*x*], where *x* stands for a Turkish case.

However, as P&P observe, this solution presents challenges. First, it contradicts one of the fundamental assumptions of LFG, which advocates representing universal and abstract grammatical features, like CASE, at the level of f-structure. Introducing numerous complex c-structure categories where such information is represented in category labels renders the f-structure redundant. Second, even if complex categories were adopted, they would not guarantee comprehensive coverage. P&P support this argument with Polish examples illustrating how the assignment of case to a nominal can change depending on the presence of negation, further complicating the complex category and CAT predicate analysis.

Solution proposed by Przepiórkowski and Patejuk (2012; 2021)

To address the complications arising from the CAT predicate solution, P&P offer a robust solution.⁴ They propose representing syntactic categories exclusively as values of a distributive CAT attribute in the f-structure.

In their work, P&P present two conceptually equivalent solutions. The first solution, known as the ‘liberal’ solution, makes use of local names and constraining equations. While the ‘liberal’ solution slightly extends the default assumptions of LFG, it succinctly captures the underlying mechanism involved. The second solution, referred to as the ‘conservative’ solution, involves using a specific LFG device called off-path constraints. The ‘conservative’ solution remains fully compatible with the default assumptions of LFG, making it a desirable option for computational implementation. For the purposes of this work, the ‘liberal’ solution is utilized in the current chapter due to its conceptual clarity and intuitiveness. However, in the subsequent chapter on implementation, the ‘conservative’ solution is adopted as the implementation method.

⁴The foundations of this solution were laid in their earlier work: Przepiórkowski and Patejuk (2012). The present work, however, refers to their more recent paper as the present analysis draws upon that version.

To illustrate their analysis within the context of the ‘liberal’ solution, let us consider the examples in (53).

- (53) a. We all believe [in positive energy]_{PP[in]} and [that what you give comes back]_{CP[that]}.
 b. We also believe [that you learn from your mistakes]_{CP[that]} and [in second chances]_{PP[in]}.
 (Przepiórkowski and Patejuk, 2021, p. 210, ex. (11) and (12))

These examples demonstrate that the OBL (oblique) argument of the verb *believe* can manifest as a CP headed by the complementizer *that*, or a PP with the preposition *in*. This conclusion is supported by the fact that replacing the complementizer *that* with *if*, or substituting the preposition *in* with *at*, leads to ungrammaticality. In light of these observations, P&P propose the following statement that captures the constraints imposed by *believe* on its OBL argument(s):

- (54) $(\uparrow \text{ OBL}) = \%C \wedge$
 $[[(\%C \text{ CAT}) =_c V \wedge (\%C \text{ COMP-FORM}) =_c \text{ THAT}] \vee$
 $[(\%C \text{ CAT}) =_c P \wedge (\%C \text{ PFORM}) =_c \text{ IN}]]$

The constraining statement starts by assigning a local name, %C, to the f-structure value of the OBL attribute. This allows us to refer to this specific f-structure using the name %C throughout the statement. To be considered a valid f-structure, %C must satisfy one of the conditions presented in the second and third lines, as indicated by the disjunction symbol, \vee .

The first condition, encoded in the second line, states that %C must contain two specific attribute-value pairs simultaneously, as indicated by the conjunction symbol, \wedge . These pairs are CAT attribute with the value “V” (indicating verbal category) and COMP-FORM attribute with the value “that” (indicating the presence of the complementizer *that*). The second condition, on the other hand, mandates that %C must have two particular attribute-value pairs together. These are CAT attribute with the value “P” (indicating prepositional category) and PFORM with the value “in” (signifying the preposition *in*).

In simpler terms, the constraining statement states that the OBL argument of the verb *believe* can take one of two forms: either a CP projected by *that* or a PP projected by *in*. Below are the visual representations of the two possible f-structures.

- (55) First condition:

$$\left[\begin{array}{l} \text{PRED} \quad \text{'BELIEVE<SUBJ, OBL>'} \\ \text{OBL} \quad \left[\begin{array}{ll} \text{CAT} & V \\ \text{COMP-FORM} & \text{THAT} \\ \dots & \end{array} \right] \\ \dots \end{array} \right]$$

- (56) Second condition:

$$\left[\begin{array}{l} \text{PRED} \quad \text{'BELIEVE<SUBJ, OBL>'} \\ \text{OBL} \quad \left[\begin{array}{ll} \text{CAT} & P \\ \text{PFORM} & \text{IN} \\ \dots & \end{array} \right] \\ \dots \end{array} \right]$$

However, the intended functionality of the statement encounters a challenge when dealing with coordinated oblique arguments (i.e., when %C corresponds to a hybrid object). In standard LFG assumptions, the statement is evaluated once

for the entire coordinate structure, not for each individual conjunct as intended. Consequently, the coordinated oblique arguments are required to be either all CP[*that*] or all PP[*in*], rather than allowing a mix of the two categories.

To address this issue, P&P expand the notion of distributivity to encompass statements. This revised definition implies that not only features like CAT or CASE can distribute to conjuncts, but also statements such as the one shown in (54).

Furthermore, P&P's solution also effectively eliminates the troublesome issue of assigning a syntactic category label to the node dominating the conjuncts. In the analysis proposed by Dalrymple, the mother coordination node collects in the form of a feature matrix the category of all its daughters. Alternatively, Peterson's (2004) account assigns the category of the first conjunct to the entire coordination. P&P's proposal, however, adopts a drastically different approach. Under their analysis, the hybrid object corresponding to coordination does not possess a syntactic category label itself; instead, individual conjuncts are assigned their respective syntactic categories.

This solution is attractive in that it allows both categorical and other morphosyntactic constraints to be unified within a single statement, representing a conceptually superior approach compared to the use of the CAT predicate. Moreover, within vanilla LFG, syntactic categories and bar levels of nodes are actually represented in a separate level called l-structure, which projects from c-structure nodes, much like f-structure (Lowe and Lovestrand, 2020, p. 16). Therefore, P&P's proposal essentially transfers the CAT attribute from l-structure to f-structure.⁵ Thus, the present thesis adopts the solution proposed by P&P to analyze Turkish unlike coordination facts.

4.4.2 Analysis of distinct unlike coordination configurations

Predicative arguments

In the case of coordination of unlike arguments, the most common type observed in the corpus was predicative argument coordination. These predicative arguments specifically occur in relation to the verb *ol-* 'be/become' and can occur as PPs, NPs, and AdjPs.

When it comes to PP predicates, virtually any Turkish postposition can be used to project the PP in the predicative position. Some verified examples of these postpositional heads include *yönelik* 'towards', *ile* 'with', *kadar* 'until', *birlikte* 'together', *karşı* 'against', *göre* 'according to', *için* 'for', *gibi* 'like', *dolayı* 'due to'. Accordingly, we can assume that the verb, *ol-*, leaves PFORM attribute of its PREDLINK argument (predicative argument) underspecified. Similarly, there appears to be no morphosyntactic constraint on predicative AdjPs.

In contrast, when it comes to NPs, the verb *ol-* specifies that its NP argument can bear any case except accusative or dative. This is because nominals marked with accusative and dative cases are consistently assigned to the functions of OBJ and OBL, respectively, and not the predicative function, PREDLINK. The absence of corpus examples featuring accusative or dative nominals in the predicative position further supports this constraint.

⁵The question as to whether bar levels should also be relocated to the f-structure goes beyond the scope of this thesis. However, such a modification would have significant implications for the overall architecture of LFG.

The distinct morphosyntactic constraints imposed by *ol-* on its predicative arguments can be effectively encoded using the disjunctive statement presented in (57).

- (57) $(\uparrow \text{ PREDLINK}) = \%C \wedge$
 $[[(\%C \text{ CAT}) =_c N \wedge \neg[(\%C \text{ CASE}) =_c \text{ ACC} \vee (\%C \text{ CASE}) =_c \text{ DAT}]] \vee$
 $(\%C \text{ CAT}) =_c P \vee$
 $(\%C \text{ CAT}) =_c \text{ Adj}]$

Importantly, since the definition of distributivity is extended to statements under the present analysis, these disjuncts are evaluated for individual conjuncts if the value of the oblique is a hybrid object. As a result, the coordination of predicative arguments is well-formed if each conjunct individually satisfies exactly one of the disjuncts in the statement.

Object arguments

The coordination of unlike objects was observed specifically within the context of the verb *konuş-* ‘talk/converse’, which subcategorizes for PP and NP objects. As mentioned earlier, the constraints imposed by *konuş-* on its OBJ are not limited to syntactic categories alone. The PP argument must be introduced by the postpositions *hakkında* ‘about’ or *üzerine* ‘over/upon’. Likewise, the case of the nominal argument is restricted to accusative and nominative cases. These requirements imposed by *konuş-* can be represented in the following disjunctive statement:

- (58) $(\uparrow \text{ OBJ}) = \%C \wedge$
 $[[(\%C \text{ CAT}) =_c P \wedge [(\%C \text{ PFORM}) =_c \text{ HAKKINDA} \vee (\%C \text{ PFORM}) =_c \text{ ÜZERINE}]] \vee$
 $[(\%C \text{ CAT}) =_c N \wedge [(\%C \text{ CASE}) =_c \text{ ACC} \vee (\%C \text{ CASE}) =_c \text{ NOM}]]]$

Oblique arguments

In the acceptability judgment study, unlike oblique arguments were tested within the context of two verbs: *sür-* ‘last/continue’ and *saydır-* ‘curse (someone)’. However, there was a notable difference in the participants’ responses to the sentences. The sentence testing *saydır-* received somewhat unfavorable scores ($M = -0.18$, $SD = 2.31$), whereas the sentence testing *sür-* was more positively received ($M = 1.72$, $SD = 1.55$). Despite this, the average score for the *saydır-* sentence is not low as the average score for ungrammatical fillers ($M = -1.99$, $SD = 1.41$), suggesting that *saydır-* might be a verb capable of taking objects with distinct categories. Thus, an analysis for this verb can also be proposed. The sentence that tested *saydır-* is provided in (59).

- (59) Kürsü-de [yerel minibüs şoför-ler-i hakkında]_{PP} ve
 podium-LOC local minibüs driver-PL-3P about and
 [belediye-ye]_{NP} saydır-dı.
 municipality-DAT curse-PST.3SG

‘On the podium, he/she cursed at local minibüs drivers and municipality.’

In this example, the first conjunct, *yerel minibüs şoförleri hakkında*, is a PP introduced by the postposition *hakkında*. The second conjunct, *belediyeye*, is a dative NP, which is the typical argument of the verb *saydır-*. Tampering with the postposition or the case of the NP negatively affects the grammaticality of the sentence. Therefore, we can conclude that the argument of *saydır-* must be either a PP projected by *hakkında* or a dative NP. To capture these facts, the following disjunctive constraint is proposed:

- (60) $(\uparrow \text{OBL}) = \%C \wedge$
 $[[(\%C \text{ CAT}) =_c N \wedge (\%C \text{ CASE}) =_c \text{DAT}] \vee$
 $[(\%C \text{ CAT}) =_c P \wedge (\%C \text{ PFORM}) =_c \text{HAKKINDA}]]$

In contrast, the verb *sür-* has different requirements for its OBL argument. It requires that its OBL argument be filled only by an AdvP, an NP in the nominative case, and a PP projected by *boyunca* ‘throughout/during’. Furthermore, the arguments must express duration or frequency. However, according to LFG’s modular architecture, these semantic constraints should not be encoded at the f-structure level but rather at the s(ematic)-structure or a(rgument)-structure. Thus, the constraining statement capturing these facts would be encoded as follows:

- (61) $(\uparrow \text{OBL}) = \%C \wedge$
 $[[(\%C \text{ CAT}) =_c N \wedge (\%C \text{ CASE}) =_c \text{NOM}] \vee$
 $[(\%C \text{ CAT}) =_c P \wedge (\%C \text{ PFORM}) =_c \text{BOYUNCA}] \vee$
 $(\%C \text{ CAT}) =_c \text{Adv}]$

Adjuncts

Within the scope of the study, unlike adjuncts were observed to function as either verbal or nominal modifiers. However, they are subject to different constraints.

In the case of verbal modifiers, the conjuncts can be realized as PPs, NPs, or AdvPs. Consider sentence (62a), where the PP conjunct *teknelele* is coordinated with the AdvP *yürüyerek* to modify the main predicate *ulaşılır*. Similarly, in sentence (62b), the PP conjunct, *kan ürünleri ile*, is coordinated with the ablative NP, *frengili anneden*, collectively modifying the main verb *bulaşabilir*.

- (62) a. ... şehir merkez-i-ne tekne-ler ile veya yürü-yerek
 ... city center-3SG-DAT boat-PL with or walk-ADVZ
 ulaş-ıl-ır.
 reach-PASS-AOR

‘One can reach the city center with boats or on foot.’ (trTenTen)

- b. Hastalık kan ürün-ler-i ile ve frengi-li
 illness blood product-PL-3P with and syphilis-ADJZ
 anne-den çocuğ-a bulaş-abil-ir.
 mother-ABL child-DAT spread-ABIL-AOR

‘The disease can be transmitted to the child by blood products and from the mother with syphilis.’ (trTenTen)

By contrast, nominal modifiers can be realized either as PPs or AdjPs. Take sentence (63a) as an example, where the PP projected by *boyunca* is coordinated with the AdjP projected by *sınırsız*. Together, they modify the noun *gezme*. Likewise, sentence (63b) demonstrates a coordinate structure incorporating PP and AdjP conjuncts, headed by *ait* and *olmamış*, respectively.

- (63) a. ... bir yıl boyunca ve sınır-sız gezme ...
 ... one year throughout and limit-less sightseeing
 ‘... limitless sightseeing for a year ...’ (trTenTen)
- b. Daha sonraki dönem-e ait veya tahrip
 more after era-DAT belong to or wreck
 ol-ma-mış plak-lar ...
 be-NEG-ADJZ record-PL
 ‘Records that belong to the later era or are undamaged ...’ (trTenTen)

Overall, the constraints imposed on adjuncts appear to be primarily related to syntactic categories, with other morphosyntactic properties such as PFORM and CASE being underspecified. Moreover, these categorical constraints seem to depend solely on the phrase-structure positions of the adjuncts. Hence, the constraints on adjuncts can be succinctly expressed through functional annotations on c-structure rules rather than at the level of individual predicates. Accordingly, the relevant c-structure rule for verbal modifiers could be encoded as follows:

$$(64) \quad V' \longrightarrow \{PP \mid NP \mid AdvP\} \quad V'$$

$$\downarrow \in (\uparrow ADJ) \quad \uparrow = \downarrow$$

This formulation, however, is not compatible with the approach proposed by Przepiórkowski and Patejuk since the rule represents syntactic categories (e.g., PP or NP) at the level of node labels as in vanilla LFG. To align with their approach, we should transfer the syntactic category information to functional annotations, retaining only the bar-level information at the node labels. Consequently, the appropriate rule is presented in (65).

$$(65) \quad X' \longrightarrow \begin{array}{ccc} XP & & X' \\ \downarrow \in (\uparrow ADJ) & & \uparrow = \downarrow \\ (\downarrow CAT) \in_c \{P, N, Adv\} & & (\downarrow CAT) =_c V \end{array}$$

In the modified rule, the first daughter node, XP, is a maximal node. Directly below, the first functional annotation, $\downarrow \in (\uparrow ADJ)$, indicates that the structure associated with this node is a member of the set of adjuncts, as specified by the ADJ attribute. This node can map to a singleton in the case of a single adjunct or a hybrid object representing the coordination of multiple adjuncts. Crucially, the second functional annotation restricts XP to have the syntactic categories P, N, or Adv. In the case of coordination, this categorical constraint is distributed to all conjuncts, as CAT is encoded as a distributive attribute. The X' node, which immediately follows the XP node, is crucially constrained to be a verbal node.

This constraint, on the other hand, ensures that the adjunct node can only attach to a verbal node.

The rule for nominal modifiers, which is presented in (66) shares similarities with the rule for verbal modifiers. However, the nominal modifier rule differs in that the XP node must either be a postpositional or an adjectival phrase. Additionally, the subsequent node is constrained to have the category N.

$$(66) \quad X' \longrightarrow \begin{array}{ccc} & XP & X' \\ & \downarrow \in (\uparrow \text{ADJ}) & \uparrow = \downarrow \\ (\downarrow \text{CAT}) \in_c \{P, \text{Adj}\} & & (\downarrow \text{CAT}) =_c N \end{array}$$

4.5 Example analysis

4.5.1 Coordination of unlike categories

In sentence (67), both conjuncts serve as predicative arguments of the verb *ol-malı*. However, they differ in terms of their syntactic categories. The first conjunct, *doğal kıldan*, is an NP headed by an ablative noun, while the second conjunct, *yumuşak*, is an AdjP.

- (67) Fırça [doğal kıl-dan]_{NP} ve [yumuşak]_{AdjP} ol-malı.
brush natural bristle-ABL and soft be-NECESS
‘(A) brush must be of natural bristle and soft.’ (trTenTen)

Let us start analyzing this sentence by providing lexical entries for all the words present in the sentence except for the verb.

- (68)
- | | | | | | |
|--------|---|---|---------|---|---|
| fırça | X | (↑ PRED) = ‘BRUSH’
(↑ PERS) = 3
(↑ CAT) = N
(↑ CASE) = NOM
(↑ NUM) = SG | yumuşak | X | (↑ PRED) = ‘SOFT’
(↑ CAT) = Adj |
| kıldan | X | (↑ PRED) = ‘BRISTLE’
(↑ PERS) = 3
(↑ CAT) = N
(↑ CASE) = ABL
(↑ NUM) = SG | doğal | X | (↑ PRED) = ‘NATURAL’
(↑ CAT) = Adj |
| | | | ve | X | (↑ CONJ) = AND
(↑ CAT) = CNJ
(↑ NUM) = PL |

As can be observed in the lexical entries, the syntactic categories of words are represented through defining equations with the CAT attribute. As a result, the conventional LFG designation for indicating the syntactic category of words, which is directly next to the word itself, is now occupied by the placeholder symbol X.

Regarding the verb *olmalı*, the constraint presented in (57) within the context of *ol-* words is appended to the lexical entry of the verb. This incorporation ensures that the specified constraint is consistently applied to any sentence containing the verb *olmalı*.

- (69) *olmalı* X (\uparrow PRED) = 'BE(\uparrow SUBJ, PREDLINK)'
 (\uparrow CAT) = V
 (\uparrow TENSE) = PRES
 (\uparrow NECESS) = +
 (\uparrow PREDLINK) = %C \wedge
 [[(%C CAT) =_c N \wedge \neg [(%C CASE) =_c ACC \vee (%C CASE) =_c DAT]] \vee
 (%C CAT) =_c P \vee
 (%C CAT) =_c Adj]]

Given the relative complexity of c-structure rules required for parsing this sentence, a detailed exploration of these rules falls outside the scope of this chapter.⁶ Therefore, our focus here is directed towards the relationship between the f-structure of the sentence and the constraining statement embedded within the lexical entry of *olmalı*. Let us proceed by assuming that our c-structure rules have the capability to construct the following f-structure for this sentence:

- (70)

PRED	'BE(\uparrow 1, \uparrow 2)'																				
CAT	V																				
NECESS	+																				
TENSE	PRES																				
SUBJ	<table border="0" style="border-collapse: collapse;"> <tr> <td style="vertical-align: top; padding-right: 10px;">PRED</td> <td style="vertical-align: top;">'BRUSH'</td> </tr> <tr> <td style="vertical-align: top; padding-right: 10px;">CASE</td> <td style="vertical-align: top;">NOM</td> </tr> <tr> <td style="vertical-align: top; padding-right: 10px;">CAT</td> <td style="vertical-align: top;">N</td> </tr> <tr> <td style="vertical-align: top; padding-right: 10px;">NUM</td> <td style="vertical-align: top;">SG</td> </tr> <tr> <td style="vertical-align: top; padding-right: 10px;">PERS</td> <td style="vertical-align: top;">3</td> </tr> </table>	PRED	'BRUSH'	CASE	NOM	CAT	N	NUM	SG	PERS	3										
PRED	'BRUSH'																				
CASE	NOM																				
CAT	N																				
NUM	SG																				
PERS	3																				
CONJ	AND																				
NUM	PL																				
PREDLINK	<table border="0" style="border-collapse: collapse;"> <tr> <td style="vertical-align: top; padding-right: 10px;">PRED</td> <td style="vertical-align: top;">'BRISTLE'</td> </tr> <tr> <td style="vertical-align: top; padding-right: 10px;">CASE</td> <td style="vertical-align: top;">ABL</td> </tr> <tr> <td style="vertical-align: top; padding-right: 10px;">CAT</td> <td style="vertical-align: top;">N</td> </tr> <tr> <td style="vertical-align: top; padding-right: 10px;">NUM</td> <td style="vertical-align: top;">SG</td> </tr> <tr> <td style="vertical-align: top; padding-right: 10px;">PERS</td> <td style="vertical-align: top;">3</td> </tr> <tr> <td style="vertical-align: top; padding-right: 10px;">ADJUNCT</td> <td style="vertical-align: top;"> <table border="0" style="border-collapse: collapse;"> <tr> <td style="vertical-align: top; padding-right: 10px;">PRED</td> <td style="vertical-align: top;">'NATURAL'</td> </tr> <tr> <td style="vertical-align: top; padding-right: 10px;">CAT</td> <td style="vertical-align: top;">Adj</td> </tr> </table> </td> </tr> <tr> <td style="vertical-align: top; padding-right: 10px;">PRED</td> <td style="vertical-align: top;">'SOFT'</td> </tr> <tr> <td style="vertical-align: top; padding-right: 10px;">CAT</td> <td style="vertical-align: top;">Adj</td> </tr> </table>	PRED	'BRISTLE'	CASE	ABL	CAT	N	NUM	SG	PERS	3	ADJUNCT	<table border="0" style="border-collapse: collapse;"> <tr> <td style="vertical-align: top; padding-right: 10px;">PRED</td> <td style="vertical-align: top;">'NATURAL'</td> </tr> <tr> <td style="vertical-align: top; padding-right: 10px;">CAT</td> <td style="vertical-align: top;">Adj</td> </tr> </table>	PRED	'NATURAL'	CAT	Adj	PRED	'SOFT'	CAT	Adj
PRED	'BRISTLE'																				
CASE	ABL																				
CAT	N																				
NUM	SG																				
PERS	3																				
ADJUNCT	<table border="0" style="border-collapse: collapse;"> <tr> <td style="vertical-align: top; padding-right: 10px;">PRED</td> <td style="vertical-align: top;">'NATURAL'</td> </tr> <tr> <td style="vertical-align: top; padding-right: 10px;">CAT</td> <td style="vertical-align: top;">Adj</td> </tr> </table>	PRED	'NATURAL'	CAT	Adj																
PRED	'NATURAL'																				
CAT	Adj																				
PRED	'SOFT'																				
CAT	Adj																				

The validity of this f-structure is two-fold. First, it adheres to all well-formedness conditions. Second, it satisfies the morphosyntactic constraints set by *olmalı* on its PREDLINK argument. The f-structure of the initial conjunct fulfills the first condition stipulated in the constraining statement: it possesses the attribute-value pair \langle CAT, N \rangle , and its case attribute value is not ACC or DAT. Meanwhile, the f-structure of the second conjunct meets the third condition from the constraining statement by containing the attribute-value pair \langle CAT, A \rangle .

Now let us alter the morphosyntactic properties of conjuncts and make the first conjunct an accusative NP and the second an AdvP. This sentence, presented below in (71), is outright ungrammatical.

⁶The relevant c-structure rules are presented later in §5.2.

- (71) *Fırça [doğal kıl-ı]_{NP} ve [yumuşak-ça]_{AdvP} ol-malı.
brush natural bristle-ACC and soft-ly be-NECESS

Intended meaning: ‘(A) brush must be of natural bristle and soft.’

By considering the f-structures of the modified conjuncts, as illustrated in (72), we can understand the source of this ungrammaticality. The f-structure of the first conjunct is illicit as it bears one of the disallowed attribute-value pairs, namely $\langle \text{CASE}, \text{ACC} \rangle$. Similarly, the f-structure of the second conjunct violates the constraints imposed by the verb due to the presence of the attribute-value pair $\langle \text{CAT}, \text{Adv} \rangle$.

- (72) a. *doğal kılı*

b.

PRED	‘BRISTLE’
CASE	ACC
CAT	N
NUM	SG
PERS	3
ADJUNCT	$\left\{ \begin{bmatrix} \text{PRED} & \text{‘NATURAL’} \\ \text{CAT} & \text{Adj} \end{bmatrix} \right\}$

- (73) a. *yumuşakça*

b.

PRED	‘SOFTLY’
CAT	Adv

4.5.2 Coordination of unlike cases

In the context of coordination of unlike cases, when the mismatching conjuncts are situated within the c-structure as verbal modifiers, as per the rule outlined in (65), the cases of the conjuncts can freely mismatch without encountering a constraint. In this respect, the following sentence is deemed grammatical as the mismatching conjuncts occupy the position of verbal adjuncts, wherein no case-related constraint is operational.

- (74) Sonra [robot-ta]_{NP_{loc}} ve [çatal-la]_{NP_{ins}} ez-meli-yiz.
then robot-LOC and fork-INS crush-NECESS-1PL

‘Then we should crush (it) in the (kitchen) appliance and with a fork.’

(trTenTen)

However, the proposed analysis does exhibit a specific limitation that deserves attention. The analysis notably permits coordination between nominative and accusative objects, particularly evident in the case of the verb *konuş-* ‘talk/-converse’. This prediction is problematic, as evidenced by the ungrammaticality of the example provided in (75).

- (75) *Kolektör-ler sık sık [antik enstrüman-lar-ı]_{NP_{acc}}
collector-PL frequent frequent antique instrument-PL-ACC
veya [ticari bilgi]_{NP_{nom}} konuş-ur-lar.
or commercial information talk-AOR-3PL

‘Collectors frequently talk about antique instruments or about commercial information.’

One way of tackling this problem would involve enforcing congruence in the specificity of the conjuncts, thereby preventing the coordination of nominative and accusative objects. This solution entails encoding the specificity attribute, SPEC, as a distributive property and incorporating the subsequent constraint into the lexical entries of verbs that subcategorize for objects.

$$(76) \quad (\uparrow \text{OBJ SPEC}) = + \vee (\uparrow \text{OBJ SPEC}) = -$$

It is important to highlight that this constraint is distinct from the preceding distributive constraints that have been presented. Its evaluation occurs only once for the whole coordination, which enforces that either all object conjuncts must be specific or non-specific.

Notably, the introduction of this constraint carries an implicit assertion pertaining to Turkish grammar: Turkish conjuncts occupying the object position must share identical specificity values. The validation of this assertion, however, requires careful empirical scrutiny, which falls beyond the scope of the present thesis and, therefore, remains a subject for future research.

4.6 Conclusion

In this chapter, we have attempted to formally model empirical facts about Turkish coordination within the framework of LFG, focusing on theoretical parsimony and broad applicability.

The proposed analysis exemplifies theoretical simplicity by deriving basic coordination facts from fundamental morphosyntactic constraints. It avoids assuming distinct overarching parameters for unlike and like coordination, recognizing that if a specific syntactic position is constrained by disjunctive morphosyntactic rules, then coordination occupying that position can incorporate unlike conjuncts. The central notion behind the analysis lies in evaluating the relevant constraints not on the entire coordinate structure, but on individual conjuncts, which aligns with the empirical observations.

Chapter 5

Implementation

5.1 Introduction

Prior to this work, a comprehensive LFG grammar for Turkish had been developed in XLE by Çetinoğlu (2009). However, the unavailability of the source files for that implementation necessitated the creation of a small-scale Turkish grammar from scratch for the current implementation.

Building grammars in XLE involves two essential steps: formulating c-structure rules and constructing a lexicon with lexical entries. In the present implementation, the Turkish grammar naturally consisted of a small set of c-structure rules and a limited lexicon. Despite its basic nature, the implemented grammar is still of significant size, making it impractical to explore in its entirety within a single chapter. Hence, this chapter will primarily focus on the relevant implementation features related to the analysis presented in the previous chapter. For a comprehensive view of the implementation, readers can refer to Appendix C.1.

Furthermore, the verification of the grammar involved extensive testing on sentences representing diverse configurations. The testing process included both well-formed and ill-formed sentences to assess the selectivity of the implementation. A total of 53 sentences were tested, including 22 ill-formed ones. Although providing a detailed analysis of each test sentence is beyond the scope of this chapter, interested readers can find them in Appendix C.2.

5.2 C-structure rules

The implementation faced a significant challenge in encoding c-structure rules, particularly because the analysis adopted in the previous chapter required the exclusion of syntactic categories represented at node labels, which is a standard XLE implementation feature. To illustrate this, let us consider the rule in (77), which is intended to parse simple Turkish sentences.

$$(77) \quad S \longrightarrow \begin{array}{cc} (NP) & VP \\ (\uparrow \text{ SUBJ}) = \downarrow & \uparrow = \downarrow \\ (\downarrow \text{ CASE}) =_c \text{ NOM} & \end{array}$$

This rule posits that a sentence consists of an NP and VP. The NP is identified as the subject of the sentence, and a constraining equation ensures that this position can only be filled by a nominative NP. Crucially, the NP can be omitted (as indicated by parentheses) since the inclusion of the subject is optional in Turkish.

However, according to the analysis proposed by Przepiórkowski and Patejuk (2021), this rule has to be modified, as shown in (78).

$$(78) \quad S \longrightarrow \begin{array}{cc} (XP) & XP \\ (\uparrow \text{ SUBJ}) = \downarrow & \uparrow = \downarrow \\ (\downarrow \text{ CASE}) =_c \text{ NOM} & (\downarrow \text{ CAT}) =_c \text{ V} \\ (\downarrow \text{ CAT}) =_c \text{ N} & \end{array}$$

The XLE version of the sentence rule, as presented in (79), is not vastly different from the original LFG rule. However, some elucidation is still needed to explain the parallelisms between LFG and XLE notations.

$$(79) \quad S \dashrightarrow \begin{array}{l} (X_MAX : (\wedge \text{ SUBJ}) = ! \\ (\wedge \text{ CASE}) =_c \text{ NOM} \\ (\wedge \text{ CAT}) =_c \text{ N} \\ \\ X_MAX : \wedge = ! \\ (\wedge \text{ CAT}) =_c \text{ V} . \end{array}$$

In XLE notation, functional annotations are not located below their respective nodes, as is typical in standard LFG rules. Instead, functional annotations are attached to their nodes using the semicolon placed next to the node labels. The metavariables \uparrow and \downarrow in LFG notation are represented in XLE by the caret symbol (\wedge) and exclamation mark (!), respectively. It is worth mentioning that in XLE, there is no need to explicitly signify conjunction (\wedge) through symbols such as &. Instead, the conjunction between constraints (or functional equations) is implied when they are placed adjacent to each other. In this instance, despite there being no conjunction symbols between functional equations, XLE interprets them as logical conjuncts. Additionally, the presented XLE rule encodes XP as X_MAX since the implemented grammar followed a notation that used X_n to represent bar-levels, with X_1 representing X' and X_0 representing X.

When it comes to coordination rule, the most straightforward way of encoding it would be as follows:

$$(80) \quad XP \longrightarrow \begin{array}{ccc} XP^+ & X & XP \\ \uparrow \in \downarrow & \uparrow = \downarrow & \uparrow \in \downarrow \\ (\downarrow \text{ CAT}) =_c \text{ CNJ} & & \end{array}$$

In this representation, the rule does not impose constraints on the syntactic categories of the conjuncts, as such constraints depend on the specific syntactic position they occupy. However, the head of the phrase (coordination) is constrained to have the category CNJ. When transferred to XLE, this rule would be expressed as in (81).

(81) X_MAX --> X_MAX+: ! \$ ^;
 X: ^ = !
 (! CAT) =_c CNJ
 X_MAX: ! \$ ^.

However, this rule clashed with other X_MAX rules (see Appendix C.1.2 for other X_MAX rules) for different phrase types, for reasons that remain unclear. Consequently, the grammar failed to produce parses for sentences containing coordination. To overcome this issue, the rule had to be adapted as shown in (82). Importantly, this modification does not imply that the current thesis asserts coordination to have the category COORD. Rather, the rule should be interpreted solely as a technical workaround.

(82) COORD --> X_MAX+: ! \$ ^;
 CNJ: ^ = !;
 X_MAX: ! \$ ^.

5.3 Coordination of unlike adjuncts

5.3.1 Verbal adjuncts

In the previous section, it was put forward that coordination of unlike verbal adjuncts can be best accounted for through c-structure rules rather than at the level of lexical entries. The XLE representation of the rule (83) is provided in (84).

(83) X' → XP⁺ X'
 ↓ ∈ (↑ ADJ) ↑ = ↓
 (↓ CAT) ∈_c {P, N, Adv} (↓ CAT) =_c V

(84) X_MAX --> { "VP"
 {
 "Verbal adjuncts"
 (COORD: ! \$ (^ ADJUNCT)
 (! CAT) \$ {P N ADV})
 |
 (X_MAX: ! \$ (^ ADJUNCT)
 (! CAT) \$ {P N ADV})
 }
 "Verbal head"
 X_1: ^=!
 (! CAT) =_c V

It is worth noting that there are some differences between the original formulation and the XLE representation. First, X_MAX is used as the mother node instead of X' (X_1) and the verbal adjuncts are marked as optional elements instead of being adorned with Kleene plus. This decision was made to keep c-structure

rules and testing as simple as possible. Second, the verbal adjunct position preceding the verbal head is allowed to rewrite either to COORD or a single X_MAX (indicated by the use of " | "). This is due to coordination being encoded as a separate rule rather than being part of the X_MAX rules.

Now, let us put our grammar to the test and examine how it handles different instances of verbal adjuncts. Consider the sentences provided in (85), which allow us to test different configurations for the implemented analysis.

- (85) a. *Kural-lar-a göre ve zor çalış-ıyor-uz.
 rule-PL-DAT according to and difficult work-PRES.PROG-1 PL
 ‘We work according to the rules and hard.’
- b. Yürü-yerek veya tekne-ler ile merkez-e ulaş-ır-sın.
 walk-ADVZ or boat-PL with center-DAT reach-AOR-2SG
 ‘(You) reach the center on foot or by boats.’
- c. Yürü-yerek ve yorul-madan merkez-e ulaş-ır-sın.
 walk-ADVZ and be tired-ADVZ center-DAT reach-AOR-2SG
 ‘(You) reach the center on foot and without getting tired.’

The sentence (85a) is ungrammatical as coordinating a PP with an AdjP violates the rule in (84). Thus, XLE should not produce a parse for this sentence. Similarly, (85b) features a coordination of unlike verbal adjuncts, where the first conjunct, *belediyeye birlikte*, is a PP, and the second conjunct, *yorulmadan*, is an AdvP. The implemented grammar, however, should correctly parse this sentence, as PPs and AdvPs can be coordinated in the verbal adjunct position. Finally, the sentence (85c) contains a coordination of two adverbs that modify the predicate *ulaşırsın* ‘you reach’. This sentence should be recognized as a licit structure by our grammar, as it represents a standard coordination of adverbs.

As expected, XLE fails to produce a parsing solution for the first sentence, as the category of the second conjunct does not satisfy the functional annotation (! CAT) \$ {P N ADV}. However, for the second sentence (85b), XLE successfully generates both the correct c-structure (Figure 5.1) and f-structure (Figure 5.2). The coordinated elements are identified as verbal adjuncts and they differ in their CAT value. For the third sentence, a similar parse is obtained, with the exception that both adjunct conjuncts have the same CAT value.

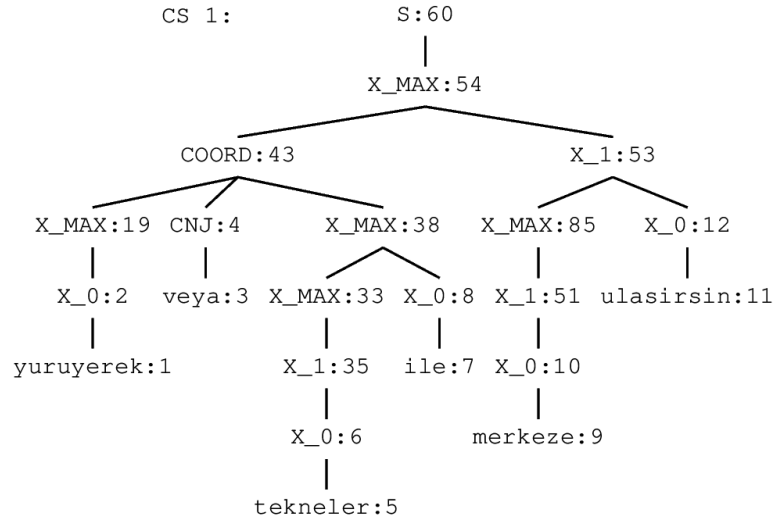


Figure 5.1: XLE screenshot of c-structure of sentence (85b)

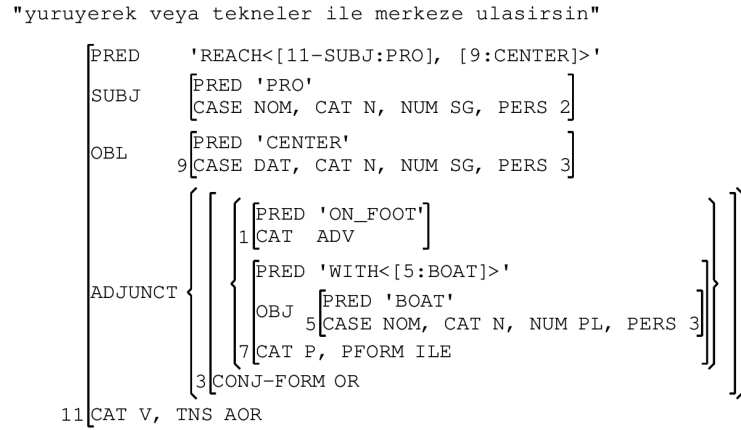


Figure 5.2: XLE screenshot of f-structure of sentence (85b)

5.3.2 Nominal adjuncts

The implementation of the rule governing nominal adjuncts is structurally quite similar to the one for verbal adjuncts, as both types of adjuncts are constrained through their c-structure positions. The rule proposed in the analysis chapter (repeated here in (86)) is transferred to XLE as shown in (87). In this rule, the head of the phrase is constrained to be a nominal, as indicated by $(! \text{ CAT}) =_c \text{ N}$, and the adjuncts can only be either a PP or an AdjP, as indicated by $(! \text{ CAT}) \$ \{P \text{ A}\}$.

$$\begin{array}{lcl}
 (86) & X' & \longrightarrow \quad \begin{array}{c} XP^+ \\ \downarrow \in (\uparrow \text{ ADJ}) \\ (\downarrow \text{ CAT}) \in_c \{P, \text{ Adj}\} \end{array} \quad \begin{array}{c} X' \\ \uparrow = \downarrow \\ (\downarrow \text{ CAT}) =_c \text{ N} \end{array}
 \end{array}$$

```

(87)      X_MAX  --> {      "NP "
                           {
                           "Nominal adjuncts"
                           (COORD: ! $ (~ ADJUNCT)
                           (! CAT) $ {P A})
                           |
                           (X_MAX: ! $ (~ ADJUNCT)
                           (! CAT) $ {P A})
                           }
                           "Nominal head"
                           X_1: ^=!
                           (! CAT) =c N }

```

Let us now test this rule on different configurations provided in (88). The first sentence (88a) is ungrammatical because the nominal adjunct position is occupied by a coordination of AdjP and AdvP, which is not permitted according to the rule. However, the sentence (88b) incorporates a coordination of a PP and an AdjP, which should be correctly parsed by the grammar. Finally, the last sentence involves a coordination of AdjPs, which should also not pose a problem for the grammar.

- (88) a. *Güzel ve profesyonel-ce eser-ler-i konuş-tu-k.
 beautiful and professional-advz work-PL-ACC talk-PST-1 PL

‘(We) talked about beautiful and professionally works of art.’

- b. Antik dönem-e ait ve eskimiş
 antiquated era-DAT belong to and obsolescent
 eser-ler-i konuş-tu-k.
 work-PL-ACC talk-PST-1 PL

‘(We) talked about works of art that belong to ancient history and are obsolescent’

- c. Güzel ve antik enstrüman-lar-ı konuş-tu-k.
 beautiful and antiquated instrument-PL-ACC talk-PST-1 PL

‘(We) talked about beautiful and antiquated instruments.’

The first sentence (88a) does not receive a parsing solution, as the nominal adjunct position is occupied by a coordination of an AdjP and an AdvP, which is not permitted. The second sentence (88b) receives the desired parsing solution, as indicated by its c-structure (Figure 5.3) and f-structure (Figure 5.4).¹ The last sentence (88c) is also correctly parsed.

¹The grammar produces another solution for (88b) where *antik* ‘antiquated’ is parsed as one of the conjuncts rather than the adjectival modifier of *döneme* ‘era’. This alternative reading is indeed easily recoverable in spoken language if a prosodic break (i.e., a pause) is inserted between *antik* and *döneme*.

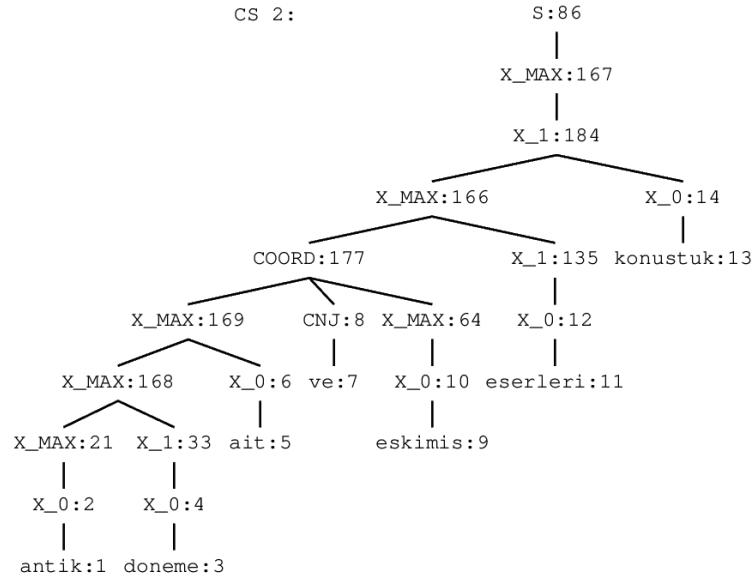


Figure 5.3: XLE screenshot of c-structure of sentence (88b)

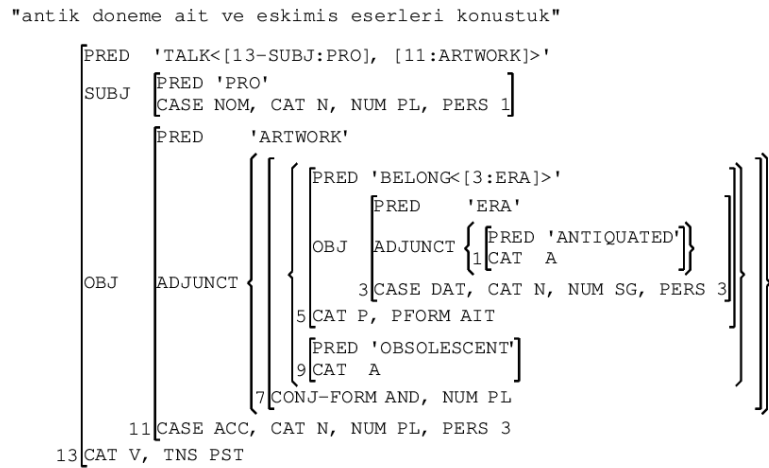


Figure 5.4: XLE screenshot of f-structure of sentence (88b)

5.4 Implementing the conservative solution

5.4.1 Off-path constraints

In the preceding chapter, among the analyses offered by Przepiórkowski and Patejuk, the choice was made to embrace the liberal solution when formally analyzing Turkish unlike coordination, due to its conceptual advantages. In contrast to the liberal solution, the conservative solution offers the capacity to enforce morphosyntactic constraints on the f-structure of individual conjuncts without extending the formalism of LFG. While the conservative solution involves more complicated formal machinery, it is fully compatible with XLE, making it the ideal option for implementation.

The conservative solution relies on using a specific constraint type called off-path constraints. Before demonstrating their application, consider the examples

(89) and (90) to understand the basic mechanism of off-path constraints on f-structures.

$$\begin{array}{ll}
 (89) \text{ a. } (\uparrow \text{ A B C}) =_c + & (90) \text{ a. } (\uparrow \text{ A B C}) =_c + \\
 (\leftarrow \text{ D}) =_c \text{ E} & (\rightarrow \text{ D}) =_c \text{ E} \\
 \text{b. } \left[\begin{array}{c} \text{A} \left[\begin{array}{c} \text{B} \left[\begin{array}{c} \text{C} \end{array} \right] + \end{array} \right] \\ \text{D E} \end{array} \right] & \text{b. } \left[\begin{array}{c} \text{A} \left[\begin{array}{c} \text{B} \left[\begin{array}{c} \text{C} \end{array} \right] + \end{array} \right] \\ \text{D E} \end{array} \right]
 \end{array}$$

The constraints themselves can be observed in (89a) and (90a). The minimal f-structures that satisfy them are given in (89b) and (90b), respectively. The first lines of the constraints are simply constraining equations. The second lines of the constraints, however, are where the off-path constraints are found, which are placed directly below the attribute that they apply.

Crucially, off-path constraints use distinct metavariables in the form of horizontal arrows. The left arrow (\leftarrow) denotes the f-structure that contains the attribute that the off-path constraint applies. The right arrow (\rightarrow), on the other hand, refers to the value of the attribute that the off-path constraint applies.

In both examples, although the off-path constraint is assigned to the attribute A, they use different horizontal arrows. In (89a), the use of \leftarrow states that there must be an attribute-value pair $\langle \text{D}, \text{E} \rangle$ in the f-structure that contains the attribute A, which is satisfied by the f-structure in (89b). In (90a), \rightarrow that the attribute-value pair $\langle \text{D}, \text{E} \rangle$ must instead be found in the f-structure that corresponds to the value of the attribute A.

From distributive statements to off-path constraints

As previously discussed, the liberal solution proposed by P&P involves extending the definition of distributivity to constraining statements. The original statement that they proposed for the constraints imposed by *believe* on its oblique argument is shown again in (91).

$$\begin{array}{l}
 (91) \quad (\uparrow \text{ OBL}) = \%C \wedge \\
 \quad \quad \quad [[(\%C \text{ CAT}) =_c \text{ V} \wedge (\%C \text{ COMP-FORM}) =_c \text{ THAT}] \vee \\
 \quad \quad \quad [(\%C \text{ CAT}) =_c \text{ P} \wedge (\%C \text{ PFORM}) =_c \text{ IN}]]
 \end{array}$$

This statement can be “translated” to an off-path constraint to achieve the same effect, as shown below.

$$\begin{array}{l}
 (92) \quad (\uparrow \text{ OBL} \quad \quad \quad \text{PRED} \quad \quad \quad) \\
 \quad \quad \quad \{(\leftarrow \text{ CAT}) =_c \text{ V} \wedge [(\leftarrow \text{ COMP-FORM}) =_c \text{ THAT}] \\
 \quad \quad \quad \vee [(\leftarrow \text{ CAT}) =_c \text{ P} \wedge (\leftarrow \text{ PFORM}) =_c \text{ IN}]
 \end{array}$$

Within the context of this off-path constraint, \leftarrow refers to the f-structure of the oblique argument of *believe*. Once this off-path constraint is added to lexical entry for *believe*, the OBL argument of *believe* can be either a CP projected by *that*, as indicated by the COMP-FORM attribute, or a PP projected by *in*, as indicated by the PFORM attribute. The constraint is anchored to the PRED attribute, as any lexical item mapped to the OBL attribute inherently carries its respective PRED

attribute. This anchoring ensures that the off-path constraint is evaluated for each f-structure containing PRED if the OBL attribute is mapped to a hybrid object corresponding to coordination. In other words, this constraint is recycled for each conjunct, as each conjunct has the PRED attribute.

5.5 Coordination of unlike arguments

5.5.1 Predicative arguments

In the previous chapter, the morphosyntactic restrictions imposed by the Turkish verb *ol-* 'be/become' upon its predicative arguments were formulated as a distributive constraint (see (93)). This distributive formulation can be converted into an off-path constraint, as demonstrated in (94).

$$(93) \quad (\uparrow \text{ PREDLINK}) = \%C \wedge \\ \quad [((\%C \text{ CAT}) =_c N \wedge \neg[(\%C \text{ CASE}) =_c \text{ACC} \vee (\%C \text{ CASE}) =_c \text{DAT}])] \vee \\ \quad (\%C \text{ CAT}) =_c P \vee \\ \quad (\%C \text{ CAT}) =_c \text{Adj}]$$

$$(94) \quad (\uparrow \text{ PREDLINK} \quad \text{PRED}) \\ \quad \{(\leftarrow \text{CAT}) =_c N \wedge \neg[(\leftarrow \text{CASE}) =_c \text{ACC} \vee (\leftarrow \text{CASE}) =_c \text{DAT}]\} \\ \quad \vee [(\leftarrow \text{CAT}) =_c P \vee (\leftarrow \text{CAT}) =_c \text{Adj}]$$

This off-path constraint states that for each PRED containing f-structure that is the value of the PREDLINK attribute, exactly one of the following conditions must be satisfied:

- It must contain a CAT attribute with the value N (representing nominal category), and it must **not** possess the attribute-value pairs $\langle \text{CASE}, \text{ACC} \rangle$ or $\langle \text{CASE}, \text{DAT} \rangle$.
- It must contain a CAT attribute with the value P (representing postpositional category).
- It must contain a CAT attribute with the value Adj (representing adjectival category).

The XLE implementation of the constraint, as shown in (95), utilizes specific notations to represent various linguistic features. Notably, XLE employs the notation \leftarrow to signify the leftward pointing horizontal arrow (\leftarrow), establishing the connection between the f-structure(s) and the PREDLINK argument. Additionally, the negation symbol “ \neg ” is represented in XLE using the tilde symbol “ \sim ”. As previously mentioned, XLE interprets the adjacency (as indicated by the space) between statements as an indication of logical conjunction, which renders “ $(\leftarrow \text{CAT}) =_c N$ ” and “ $\neg[(\leftarrow \text{CASE}) =_c \text{ACC} \vee (\leftarrow \text{CASE}) =_c \text{DAT}]$ ” as logical conjuncts. Furthermore, the entire disjunctive constraint is anchored to the PRED attribute through the use of a semicolon adjacent to the PRED attribute.

- (95) (\sim PREDLINK PRED :
 $\{(\leftarrow \text{CAT}) =_c \text{N} \sim \{(\leftarrow \text{CASE}) =_c \text{ACC} \mid (\leftarrow \text{CASE}) =_c \text{DAT}\} \mid$
 $(\leftarrow \text{CAT}) =_c \text{P} \mid$
 $(\leftarrow \text{CAT}) =_c \text{A}\} \}.$

This constraint has demonstrated its effectiveness in handling distinct configurations represented by a total of specifically designed 9 test sentences. For well-formed sentences, the grammar correctly provides valid parses, and, in the case of ill-formed strings, the grammar behaves as expected by not providing any solutions.

5.5.2 Object arguments

The verb *konuş-* ‘talk/converse’ imposes intricate morphosyntactic constraints on its object(s). These constraints were captured in the preceding chapter using the subsequent distributive constraining statement:

- (96) (\uparrow OBJ) = %C \wedge
 $[[(\leftarrow \text{CAT}) =_c \text{P} \wedge [(\leftarrow \text{PFORM}) =_c \text{HAKKINDA} \vee (\leftarrow \text{PFORM}) =_c \text{ÜZERINE}]] \vee$
 $[(\leftarrow \text{CAT}) =_c \text{N} \wedge [(\leftarrow \text{CASE}) =_c \text{ACC} \vee (\leftarrow \text{CASE}) =_c \text{NOM}]]]$

This constraining statement can be equivalently transformed into the following off-path constraint:

- (97) (\uparrow OBJ PRED)
 $[(\leftarrow \text{CAT}) =_c \text{P} \wedge [(\leftarrow \text{PFORM}) =_c \text{HAKKINDA} \vee (\leftarrow \text{PFORM}) =_c \text{ÜZERINE}]]$
 $\vee \{(\leftarrow \text{CAT}) =_c \text{N} \wedge [(\leftarrow \text{CASE}) =_c \text{ACC} \vee (\leftarrow \text{CASE}) =_c \text{NOM}]\}$

This off-path constraint stipulates that for each PRED containing f-structure that is the value of the OBJ attribute, exactly one of the following constraints must hold:

- It must contain a CAT attribute with the value P, and it must either possess the attribute-value pair $\langle \text{PFORM}, \text{HAKKINDA} \rangle$ or $\langle \text{PFORM}, \text{ÜZERINE} \rangle$.
- It must contain a CAT attribute with the value N, and it must possess either the attribute-value pair $\langle \text{CASE}, \text{ACC} \rangle$ or $\langle \text{CASE}, \text{NOM} \rangle$.

The corresponding XLE implementation of this constraint is presented below:

- (98) (\sim OBJ PRED :
 $\{(\leftarrow \text{CAT}) =_c \text{N} \{(\leftarrow \text{CASE}) =_c \text{ACC} \mid (\leftarrow \text{CASE}) =_c \text{NOM}\} \mid$
 $(\leftarrow \text{CAT}) =_c \text{P} \{(\leftarrow \text{PFORM}) =_c \text{HAKKINDA}$
 $\mid (\leftarrow \text{PFORM}) =_c \text{ÜZERINE}\} \} \}.$

The effectiveness of this constraint was verified through testing on 8 different configurations (sentences), where it demonstrated its capability to handle them effectively.

5.5.3 Oblique arguments

Coordination of unlike oblique arguments was examined in the previous chapter within the context of two verbs: *saydır-* ‘curse/swear’ and *sür-* ‘last/continue’.

The morphosyntactic constraints imposed by the verb *saydır-* on its oblique argument was formulated through the distributive statement in (99), which is translated into the off-path constraint in (100).

$$(99) \quad (\uparrow \text{OBL}) = \%C \wedge \\ [[(\%C \text{ CAT}) =_c N \wedge (\%C \text{ CASE}) =_c \text{DAT}] \vee \\ [(\%C \text{ CAT}) =_c P \wedge (\%C \text{ PFORM}) =_c \text{HAKKINDA}]]$$

$$(100) \quad (\uparrow \text{OBL} \quad \text{PRED}) \\ [(\leftarrow \text{CAT}) =_c N \wedge (\leftarrow \text{CASE}) =_c \text{DAT}] \\ \vee [(\leftarrow \text{CAT}) =_c P \wedge (\leftarrow \text{PFORM}) =_c \text{HAKKINDA}]$$

This off-path constraint states that for each PRED containing f-structure that is the value of the attribute OBL, exactly one of the following constraints must hold:

- It must contain a CAT attribute with the value N, and it must possess the attribute-value pair $\langle \text{CASE}, \text{DAT} \rangle$.
- It must contain a CAT attribute with the value P, and it must possess the attribute-value pair $\langle \text{PFORM}, \text{HAKKINDA} \rangle$.

The XLE version of this constraint is provided in (101). This constraint was tested on 8 distinct configurations and effectively handled both well-formed and ill-formed strings.

$$(101) \quad (\sim \text{OBL} \text{ PRED} : \\ \{ (\leftarrow \text{CAT}) =_c N \mid (\leftarrow \text{CASE}) =_c \text{DAT} \mid \\ (\leftarrow \text{CAT}) =_c P \mid (\leftarrow \text{PFORM}) =_c \text{HAKKINDA} \}) .$$

Turning our attention to the verb *sür-*, its constraints pertaining to its OBL argument(s) were formulated in the form of a distributive statement in the previous chapter as presented below:

$$(102) \quad (\uparrow \text{OBL}) = \%C \wedge \\ [[(\%C \text{ CAT}) =_c N \wedge (\%C \text{ CASE}) =_c \text{NOM}] \vee \\ [(\%C \text{ CAT}) =_c P \wedge (\%C \text{ PFORM}) =_c \text{BOYUNCA}] \vee \\ (\%C \text{ CAT}) =_c \text{Adv}]$$

The off-path constraint equivalent of this distributive formulation is provided below:

$$\begin{aligned}
 (103) \quad & (\uparrow \text{OBL} \quad \text{PRED}) \\
 & [(\leftarrow \text{CAT}) =_c \text{N} \quad \wedge \quad (\leftarrow \text{CASE}) =_c \text{NOM}] \\
 & \vee [(\leftarrow \text{CAT}) =_c \text{P} \quad \wedge \quad (\leftarrow \text{PFORM}) =_c \text{BOYUNCA}] \\
 & \vee (\leftarrow \text{CAT}) =_c \text{Adv}
 \end{aligned}$$

This off-path constraint states that for each PRED containing f-structure that is the value of the attribute OBL, exactly one of the following constraints must hold:

- It must contain a CAT attribute with the value N, and it must possess the attribute-value pair $\langle \text{CASE}, \text{NOM} \rangle$.
- It must contain a CAT attribute with the value P, and it must possess the attribute-value pair $\langle \text{PFORM}, \text{BOYUNCA} \rangle$.
- It must contain a CAT attribute with the value Adv.

The corresponding XLE implementation of this constraint, demonstrated in (101), was tested on 12 distinct configurations, successfully handling both well-formed and ill-formed strings.

$$\begin{aligned}
 (104) \quad & (\wedge \text{OBL PRED} : \\
 & \{ (\leftarrow \text{CAT}) =_c \text{N} \quad (\leftarrow \text{CASE}) =_c \text{NOM} \mid \\
 & (\leftarrow \text{CAT}) =_c \text{P} \quad (\leftarrow \text{PFORM}) =_c \text{BOYUNCA} \mid \\
 & (\leftarrow \text{CAT}) =_c \text{ADV} \}).
 \end{aligned}$$

5.6 Conclusion

This chapter provided an overview of the XLE implementation of the proposed analysis from the preceding chapter. The success of the implementation is ensured through rigorous testing involving 53 sentences. The ‘conservative solution’ proposed by Przepiórkowski and Patejuk enables the representation of intricate morphosyntactic constraints associated with specific verbs while maintaining full XLE compatibility.

An especially noteworthy accomplishment of this implementation is the effective transfer of syntactic category information from node labels to the f-structure. This accomplishment not only validates the viability of the analysis proposed by Przepiórkowski and Patejuk, but also underscores the versatility of the LFG and XLE framework.

Chapter 6

Conclusion

To the best of this author's knowledge, this thesis constitutes the first attempt at investigating unlike coordination phenomena in Turkish. To maximize scientific validity, two widely employed methodologies in linguistics were utilized: corpus study and acceptability judgment study. Furthermore, these methodologies were employed in a way that complemented each other.

The results from both the corpus study and the acceptability judgment task provide compelling evidence that Turkish indeed allows for coordination of unlike elements, both in terms of unlike category and unlike case coordination. However, unlike coordination in Turkish is subject to certain restrictions; not all elements can be coordinated with one another. The grammatical function of the conjuncts appears to play a significant role in determining the acceptability of such structures, as hypothesized at the outset of the thesis. This hypothesis finds strong support both in the corpus study and acceptability judgment task. In the corpus study, the vast majority of examples feature conjuncts with matching functions. Additionally, the linear-mixed effects models applied to the acceptability judgment task data further confirm this observation, as mismatching functions were found to be a negative predictor of acceptability ratings.

In conclusion, the empirical investigations confirm the hypothesis put forward in this research, also providing support for the increasingly prevalent view that coordination does not universally require conjuncts to match in their morphosyntactic properties. However, further empirical data from other languages would indeed be valuable in strengthening this position.

To analyze and formalize the observed empirical facts, the thesis adopts the framework of Lexical-Functional Grammar and follows the approach proposed by Przepiórkowski and Patejuk (2021). The analysis refrains from encoding an overarching principle that only conjuncts with like functions can be coordinated, as was suggested by Peterson (2004). Instead, it proposes that each conjunct must meet certain constraints within the syntactic position they occupy, which is a position aligned with Przepiórkowski (2022). Consequently, this approach elegantly accounts for the empirical observations on Turkish coordination without postulating Turkish- or coordination-specific parameters.

Furthermore, the implementation chapter demonstrates that the proposed analysis of unlike coordination can be implemented. The logic behind the implementation was determined to be straightforward and fully compatible with XLE. However, a noteworthy challenge that requires further investigation involves in-

tegrating an approach that transfers syntactic categories from c-structure to f-structure within existing XLE implementations and grammar engineering frameworks.

Bibliography

- Akkuş, F. (2016). Suspended affixation with derivational suffixes and lexical integrity. *Mediterranean Morphology Meetings*, 10(0):1–15.
- Aksan, Y., Aksan, M., Koltuksuz, A., Sezer, T., Mersinli, Ü., Demirhan, U. U., Yilmazer, H., Atasoy, G., Öz, S., Yıldız, İ., and Kurtoğlu, Ö. (2012). Construction of the Turkish National Corpus (TNC). In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 3223–3227. European Language Resources Association (ELRA).
- Baisa, V. and Suchomel, V. (2012). Large corpora for Turkic languages and unsupervised morphological analysis. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, pages 28–32. European Language Resources Association (ELRA).
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.
- Bayer, S. (1996). The coordination of unlike categories. *Language*, 72(3):579–616.
- Beavers, J. and Sag, I. A. (2004). Coordinate ellipsis and apparent non-constituent coordination. In Müller, S., editor, *Proceedings of the 11th International Conference on Head-Driven Phrase Structure Grammar, Center for Computational Linguistics, Katholieke Universiteit Leuven*, pages 48–69, Stanford, CA. CSLI Publications.
- Bresnan, J. (2007). Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In *Roots*, pages 75–96. Mouton de Gruyter.
- Bruening, B. and Khalaf, E. A. (2020). Category mismatches in coordination revisited. *Linguistic Inquiry*, 51(1):1–36.
- Çetinoğlu, Ö. (2009). *Building a Lexical Functional Grammar for Turkish*. Ph.D. dissertation, Sabancı University.
- Chaves, R. P. and Paperno, D. (2007). On the Russian hybrid coordination construction. In Müller, S., editor, *Proceedings of the 14th International Conference on Head-Driven Phrase Structure Grammar, Stanford Department of Linguistics and CSLI's LinGO Lab*, pages 46–64, Stanford, CA. CSLI Publications.
- Chomsky, N. (1957). *Syntactic Structures*. Mouton & Co., The Hague.
- Chomsky, N. (1959). On certain formal properties of grammars. *Information and Control*, 2(2):137–167.

- Chomsky, N. (1981). *Lectures on Government and Binding: The Pisa Lectures*. De Gruyter Mouton.
- Chomsky, N. (1993). A minimalist program for linguistic theory. In Hale, K. and Keyser, S. J., editors, *The View From Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*. MIT Press.
- Çöltekin, Ç. (2010). A freely available morphological analyzer for Turkish. In Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M., and Tapias, D., editors, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).
- Çöltekin, Ç. (2015). A grammar-book treebank of Turkish. In *Proceedings of the 14th workshop on Treebanks and Linguistic Theories (TLT 14)*, pages 35–49. IPI PAN.
- Cowart, W. (1997). *Experimental Syntax: Applying Objective Methods to Sentence Judgments*. SAGE Publications.
- Crouch, D., Dalrymple, M., Kaplan, R., King, T. H., Maxwell III, J., and Newman, P. (2017). *XLE documentation*. Palo Alto Research Center.
- Cummings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, 28(3):369–382.
- Dalrymple, M. (2017). Unlike phrase structure category coordination. *Bergen Language and Linguistics Studies*, 8(1).
- Dalrymple, M. and Kaplan, R. M. (2000). Feature indeterminacy and feature resolution. *Language*, 76(4):759.
- Dalrymple, M., Lowe, J. J., and Mycock, L. (2019). *The Oxford Reference Guide to Lexical-Functional Grammar*. Oxford University Press.
- Dik, S. (1968). *Coordination: Its Implications for a Theory of General Linguistics*. North-Holland.
- Divjak, D. (2016). The role of lexical frequency in the acceptability of syntactic variants: Evidence from that-clauses in Polish. *Cognitive Science*, 41(2):354–382.
- Eberhard, D. M., Simons, G. F., and Fennig, C. D. (2022). *Turkish*. SIL International, Dallas, Texas, 25th edition.
- Francis, E. J. (2021). On the relationship between corpus frequency and acceptability. In *Gradient Acceptability and Linguistic Theory*, pages 103–125. Oxford University Press.
- Gazdar, G., Klein, E., Pullum, G. K., and Sag, I. A. (1985). *Generalized Phrase Structure Grammar*. Harvard University Press.
- Gibson, E. and Fedorenko, E. (2013). The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes*, 28(1–2):88–124.

- Gibson, E., Piantadosi, S., and Fedorenko, K. (2011). Using mechanical turk to obtain and analyze English acceptability judgments. *Language and Linguistics Compass*, 5(8):509–524.
- Göksel, A. and Kerslake, C. (2005). *Turkish: A Comprehensive Grammar*. Routledge.
- Göksel, A. and Kerslake, C. (2010). *Turkish: An Essential Grammar*. Routledge.
- Graziano, A. M. and Raulin, M. L. (2014). *Research Methods: A Process of Inquiry (8th ed.)*. Pearson Education Ltd.
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825):357–362.
- Haspelmath, M. (2014). Arguments and adjuncts as language-particular syntactic categories and as comparative concepts. *Linguistic Discovery*, 12(2).
- Hitz, J. and Francis, E. J. (2016). On the usefulness of formal judgment tasks in syntax and in second-language research: The case of resumptive pronouns in English, Turkish, and Mandarin Chinese. *Linguistics*, 54(6):1241–1280.
- Joshi, A. K. (1987). An introduction to Tree Adjoining Grammar. In *Mathematics of Language*, pages 87–115. John Benjamins Publishing Company.
- Kabak, B. (2007). Turkish suspended affixation. *Linguistics*, 45(2):311–347.
- Kaplan, R. M. and Bresnan, J. (1982). Lexical-Functional Grammar: A formal system for grammatical representation. In Bresnan, J., editor, *The Mental Representation of Grammatical Relations*. MIT Press.
- Kayadelen, T., Oztürel, A., and Bohnet, B. (2020). A gold standard dependency treebank for Turkish. In *Proceedings of The 12th Language Resources and Evaluation Conference (LREC’20)*, pages 5156–5163. European Language Resources Association (ELRA).
- Kornfilt, J. (1997). *Turkish*. Routledge.
- Kornfilt, J. (2012). Revisiting “suspended affixation” and other coordinate mysteries. In *Functional Heads, Volume 7: The Cartography of Syntactic Structures*. Oxford University Press.
- Levelt, W. J. M. (1993). *Speaking: From intention to articulation*. The MIT Press.
- Lewis, G. L. (1967). *Turkish Grammar*. Oxford University Press.
- Lowe, J. and Lovestrand, J. (2020). Minimal phrase structure: a new formalized theory of phrase structure. *Journal of Language Modelling*, 8(1):1–52.

- Müller, S., Abeillé, A., Borsley, R. D., and Koenig, J.-P., editors (2021). *Head-Driven Phrase Structure Grammar: The handbook*. Language Science Press, Berlin.
- Norman, G. (2010). Likert scales, levels of measurement and the “laws” of statistics. *Advances in Health Sciences Education*, 15(5):625–632.
- Patejuk, A. (2015). *Unlike Coordination in Polish: An LFG Account*. Ph.D. dissertation, Institute of Polish Language, Polish Academy of Sciences, Cracow.
- Patejuk, A. and Przepiórkowski, A. (2012). A comprehensive analysis of constituent coordination for grammar engineering. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012)*, pages 2191–2207, Mumbai, India.
- Peterson, P. (1981). Problems with constraints on coordination. *Linguistic Analysis*, 8:449–460.
- Peterson, P. (2004). Coordination: Consequences of a lexical-functional account. *Natural Language & Linguistic Theory*, 22(3):643–679.
- Phillips, C. (2010). Should we impeach armchair linguists. *Japanese/Korean Linguistics*, 17:49–64.
- Pollard, C. and Sag, I. A. (1994). *Head-Driven Phrase Structure Grammar*. The University of Chicago Press.
- Przepiórkowski, A. (1999). *Case Assignment and the Complement-Adjunct Dichotomy: A Non-Configurational Constraint-Based Approach*. PhD thesis, Universität Tübingen, Germany.
- Przepiórkowski, A. (2016). How *not* to distinguish arguments from adjuncts in LFG. In Arnold, D., Butt, M., Crysmann, B., King, T. H., and Müller, S., editors, *The Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar*, pages 560–580, Stanford, CA. CSLI Publications.
- Przepiórkowski, A. (2022). Coordination of unlike grammatical cases (and unlike categories). *Language*, 98(3):592–634.
- Przepiórkowski, A. and Patejuk, A. (2012). On case assignment and the coordination of unlikes: The limits of distributive features. In Butt, M. and King, T. H., editors, *The Proceedings of the LFG’12 Conference*, pages 479–489, Stanford, CA. CSLI Publications.
- Przepiórkowski, A. and Patejuk, A. (2021). Coordinate structures without syntactic categories. In Arka, I. W., Asudeh, A., and King, T. H., editors, *Modular Design of Grammar: Linguistics on the Edge*, pages 205–220. Oxford University Press, Oxford.
- Pullum, G. K. and Scholz, B. C. (2001). On the distinction between model-theoretic and generative-enumerative syntactic frameworks. In de Groote, P., Morrill, G., and Retoré, C., editors, *Logical Aspects of Computational Linguistics*, pages 17–43, Berlin, Heidelberg. Springer Berlin Heidelberg.

- Sag, I. A., Gazdar, G., Wasow, T., and Weisler, S. (1985). Coordination and how to distinguish categories. *Natural Language & Linguistic Theory*, 3(2):117–171.
- Sag, I. A. and Wasow, T. (2011). Performance-compatible competence grammar. In *Non-Transformational Syntax*, pages 359–377. Wiley-Blackwell.
- Sampson, G. (2002). *Empirical Linguistics*. Continuum.
- Schachter, P. (1977). Constraints on coordination. *Language*, 53:86–103.
- Schütze, C. T. (1996). *The empirical base of linguistics*. University of Chicago Press, Chicago.
- Schütze, C. T. and Sprouse, J. (2013). Judgment data. In Podesva, R. J. and Sharma, D., editors, *Research Methods in Linguistics*, pages 27–50. Cambridge University Press.
- Şenşekerci, B. (2022). L_RFG analysis of Turkish suspended affixation in nonverbal coordination. In Butt, M., Findlay, J. Y., and Toivonen, I., editors, *Proceedings of the LFG'22 Conference*. CSLI Publications.
- Sezer, T. (2017). TS corpus project: An online Turkish dictionary and TS DIY corpus. *European Journal of Language and Literature*, 9(1):18.
- Sprouse, J., Schütze, C. T., and Almeida, D. (2013). A comparison of informal and formal acceptability judgments using a random sample from linguistic inquiry 2001–2010. *Lingua*, 134:219–248.
- Sulubacak, U., Gokirmak, M., Tyers, E., Çöltekin, Ç., Nivre, J., and Eryiğit, G. (2016). Universal dependencies for Turkish. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 3444–3454, Osaka, Japan. The COLING 2016 Organizing Committee.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., and Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217):1632–1634.
- Tesnière, L. (1959). *Éléments de syntaxe structurale*. Klincksieck.
- Türk, U., Atmaca, F., Özateş, Ş. B., Berk, G., Bedir, S. T., Köksal, A., Başaran, B. Ö., Güngör, T., and Özgür, A. (2020). Resources for Turkish dependency parsing: Introducing the BOUN treebank and the BoAT annotation tool.
- Vater, H. (1978). On the possibility of distinguishing between complements and adjuncts. In *Studies in Language Companion Series*, page 21. John Benjamins Publishing Company.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R.,

- Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272.
- von der Malsburg, T. and Vasishth, S. (2011). What is the scanpath signature of syntactic reanalysis? *Journal of Memory and Language*, 65(2):109–127.
- Wasow, T. and Arnold, J. (2005). Intuitions in linguistic argumentation. *Lingua*, 115(11):1481–1496. Data in Theoretical Linguistics.
- Weisser, P. (2020). On the symmetry of case in conjunction. *Syntax*, 23(1):42–77.
- Wes McKinney (2010). Data structures for statistical computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56 – 61.
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.
- Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83.
- Williams, E. (1981). Transformationless grammar. *Linguistic Inquiry*, 12:645–653.

Appendix A

Corpus study

A.1 CQL queries

A.1.1 Interpreting CQL queries

CQL is a specialized language used for querying and searching linguistic data within a corpus. Queries in CQL are written in a code-like format and consist of several components, each serving a specific purpose:

Tags and tag patterns: Tags represent specific parts of speech (e.g., nouns, verbs, adjectives) and are enclosed in angle brackets (< >). Tag patterns can match multiple sub-types associated with a tag, which can be achieved by an asterisk symbol. For example, the block, `[tag='<Postp[:>].*']`, targets all types of tokens as long as they are tagged as postpositions. The block `[longtag='<Det:indef>']`, however, targets only indefinite determiners, which is a sub-type of determiners. The blocks that target sub-types use the keyword `longtag` and omit the asterisk symbol.

Optional elements: Optional elements in the query are denoted by placing a question mark (?) after the element. These elements may or may not be present in the pattern being searched. For instance, the block, `[tag='<Det:indef>']?` specifies that this position can be optionally filled by an indefinite determiner.

Exclusion: The “!=” symbol is used to exclude specific tags or patterns from the query, ensuring that certain elements are not present in the sentence. For example, the block, `[tag!='<V[:>].*']`, fails to match if the position is filled by any verbal element.

Logical operators: Logical operators such as “&” (AND) and “|” (OR) are used to combine different elements within the query and impose more sophisticated requirements on the patterns. For instance, the block `[tag='<N[:>].*' & tag!='<N><sg>']` targets non-singular nouns in the corpus.

Coordinating conjunction: The presence of a coordinating conjunction is indicated using the `Cnj : coo longtag`, representing conjunctions like *and* or *or*.

Stem: The stem attribute allows specific verb stems to be targeted, enabling the search for patterns incorporating particular verb forms. For instance, the block, [tag='<V[:>].*'] & stem='run'], targets all verbs with the stem *run*.

Word search: Specific words can be targeted using the word attribute, allowing searches for patterns containing specific lexical items. For example, the block, [word='alien'], matches the word “alien.”

The linear order of elements in the query directly corresponds to the linear order of elements searched in a sentence pattern. This means that the components in the query are interpreted in the same order they appear, allowing for precise control over the arrangement and occurrence of linguistic elements in the targeted patterns.

A.1.2 Coordination of unlike categories

PP & NP

```
[tag!="<V[:>].*"] [longtag="<Det:indef>"]?[tag="<Adj[:>].*"]?
[tag="<N[:>].*"] [tag="<Postp[:>].*"] [longtag="<Cnj:coo>"]
[tag="<Adj[:>].*"]? [longtag="<Det:indef>"]? [tag="<N[:>].*"]
[tag!="<Postp[:>].*"]
```

Figure A.1: General PP & NP Query

```
[tag!="<V[:>].*"] [tag="<Adj[:>].*"]? [longtag="<Det:indef>"]?
[tag="<N[:>].*"] [tag="<Postp[:>].*"] [longtag="<Cnj:coo>"]
[tag="<Adj[:>].*"]? [longtag="<Det:indef>"]? [tag="<N[:>].*"]
[stem="ol" & tag="<V[:>].*"] [tag!="<Postp[:>].*"]
```

Figure A.2: PP & NP TO BE Query

```
[tag="<Adj[:>].*"]? [longtag="<Det:indef>"]? [tag="<N[:>].*"]
[tag="<Postp[:>].*"] [longtag="<Cnj:coo>"] [tag="<Adj[:>].*"]?
[longtag="<Det:indef>"]? [tag="<N[:>].*"] [tag!="<Postp[:>].*"]
[tag="<V[:>].*" & stem="dusun"]
```

Figure A.3: PP & NP TO THINK/CONSIDER Query

The query in Figure A.4 additionally specifies that the postpositional head of the first conjunct may alternatively be occupied by the word *hakkında* ‘about’. This stems from the fact that the automatic POS-tagger employed in the corpus sometimes tags *hakkında* as a locative noun even though it was classified as a type of postposition by Kornfilt (1997, pp. 100–102).

```
[tag!="<V[:>].*"] [tag="<Adj[:>].*"]? [longtag="<Det:indef>"]?
[tag="<N[:>].*"] [tag="<Postp[:>].*" | word="hakkinda"]
[tag="<Cnj:coo>"] [tag="<Adj[:>].*"]? [longtag="<Det:indef>"]?
[tag="<N[:>].*"] [tag!="<Postp[:>].*"]
[tag="<V[:>].*" & stem="konu"]
```

Figure A.4: PP & NP TO TALK/CONVERSE/SPEAK Query

PP & AdvP

```
[tag="<Adj[:>].*"]? [longtag="<Det:indef>"]? [tag="<N[:>].*"]
[tag="<Postp[:>].*"] [longtag="<Cnj:coo>"] [tag="<Adv[:>].*"]
[tag="<V[:>].*"]
```

Figure A.5: General PP & AdvP Query

PP & AdjP

```
[tag="<Adj[:>].*"]? [tag="<Det:indef>"]? [tag="<N[:>].*"]
[tag="<Postp[:>].*"] [longtag="<Cnj:coo>"] [tag="<Adv[:>].*"]?
[tag="<Adj[:>].*" [tag="<V[:>].*"]
```

Figure A.6: General PP & AdjP Query

NP & AdjP

```
[tag != "<V[:>].*"] [tag = "<Adj[:>].*"]? [tag = "<Det:indef>"]?
[tag = "<N[:>].*"] [longtag = "<Cnj:coo>"] [tag = "<Adv[:>].*"]?
[tag = "<Adj[:>].*" [stem = "ol" & tag = "<V[:>].*"]
```

Figure A.7: NP & AdjP TO BE Query

NP & AdvP

```
[tag!="<V[:>].*"] [tag="<Num>"]? [tag="<N[:>].*"] [longtag="<Cnj:coo>"]
[tag="<Adv[:>].*"] [tag="<V[:>].*" & stem="sur"]
```

Figure A.8: NP (numerical) & AdvP TO LAST Query


```
[tag!="<V[:>].*"] [tag="<Adj[:>].*"]? [longtag='<Det:indef>']?
[tag="<N[:>].*"] [longtag="<Cnj:coo>"] [tag="<Adv[:>].*"]
[tag="<V[:>].*" & stem="hallet"]
```

Figure A.9: NP & AdvP TO HANDLE Query

A.1.3 Coordination of unlike cases

```
[tag != "<V[:>].*"] [tag = "<Adj[:>].*"]?
[longtag = "<Det:indef>"]? [(longtag = "<N><dat>") |
(longtag = "<N><gen>" | longtag = "<N><abl>") | longtag = "<N><ins>"
| longtag = "<N><loc>"] [longtag = "<Cnj:coo>"] [tag = "<Adj[:>].*"]?
[longtag = "<Det:indef>"]? [longtag = "<N><acc>"]
[tag != "<N[:>].*" & tag != "<Postp[:>].*"]
```

Figure A.10: NP(\neg acc) & NP(acc) Query

```
[tag != "<V[:>].*"] [tag = "<Adj[:>].*"]?
[longtag = "<Det:indef>"]? [(longtag = "<N><acc>") |
(longtag = "<N><gen>" | longtag = "<N><abl>") | longtag = "<N><ins>"
| longtag = "<N><loc>"] [longtag = "<Cnj:coo>"] [tag = "<Adj[:>].*"]?
[longtag = "<Det:indef>"]? [longtag = "<N><dat>"]
[tag != "<N[:>].*" & tag != "<Postp[:>].*"]
```

Figure A.11: NP(\neg dat) & NP(dat) Query

```
[tag != "<V[:>].*"] [tag = "<Adj[:>].*"]?
[longtag = "<Det:indef>"]? [(longtag = "<N><acc>") |
(longtag = "<N><dat>" | longtag = "<N><abl>") | longtag = "<N><ins>"
| longtag = "<N><loc>"] [longtag = "<Cnj:coo>"] [tag = "<Adj[:>].*"]?
[longtag = "<Det:indef>"]? [longtag = "<N><gen>"]
[tag != "<N[:>].*" & tag != "<Postp[:>].*"]
```

Figure A.12: NP(\neg gen) & NP(gen) Query

```
[tag != "<V[:>].*"] [tag = "<Adj[:>].*"]?
[longtag = "<Det:indef>"]? [(longtag = "<N><acc>") |
(longtag = "<N><dat>" | longtag = "<N><abl>") | longtag = "<N><gen>"
| longtag = "<N><loc>"] [longtag = "<Cnj:coo>"] [tag = "<Adj[:>].*"]?
[longtag = "<Det:indef>"]? [longtag = "<N><ins>"]
[tag != "<N[:>].*" & tag != "<Postp[:>].*"]
```

Figure A.13: NP(\neg ins) & NP(ins) Query

```
[tag != "<V[:>].*"] [tag = "<Adj[:>].*"]?
[longtag = "<Det:indef>"]? [(longtag = "<N><acc>") |
(longtag = "<N><dat>" | longtag = "<N><ins>") | longtag = "<N><gen>"
| longtag = "<N><loc>"] [longtag = "<Cnj:coo>"] [tag = "<Adj[:>].*"]?
[longtag = "<Det:indef>"]? [longtag = "<N><abl>"]
[tag != "<N[:>].*" & tag != "<Postp[:>].*"]
```

Figure A.14: NP(\neg abl) & NP(abl) Query

```
[tag != "<V[:>].*"] [tag = "<Adj[:>].*"]?
[longtag = "<Det:indef>"]? [(longtag = "<N><acc>") |
(longtag = "<N><dat>" | longtag = "<N><ins>") | longtag = "<N><gen>"
| longtag = "<N><abl>"] [longtag = "<Cnj:coo>"] [tag = "<Adj[:>].*"]?
[longtag = "<Det:indef>"]? [longtag = "<N><loc>"]
[tag != "<N[:>].*" & tag != "<Postp[:>].*"]
```

Figure A.15: NP(\neg loc) & NP(loc) Query

A.2 Valency dictionary

A.2.1 Treebanks

In order to construct the small-scale valency dictionary of Turkish verbs, 5 prominent Turkish dependency treebanks, which are listed in Table A.1, have been automatically parsed with a script written in Python (*version 3.8.12*).

Treebank	Sentences	Tokens	Annotation	Document Type
<i>TWT</i> ^a	4,851	80,920	Manual	Web
<i>The Grammarbook</i> ^b	2,803	16,516	Semi-automatic	Textbook examples
<i>IMST-UD</i> ^c	5,635	56,423	Semi-automatic	News reports, novels
<i>BOUN</i> ^d	9,761	121,214	Manual	Various
<i>Kenet UD</i> ^e	18,700	178,700	Manual	Various

Table A.1: List of treebanks employed in the corpus study

^aKayadelen et al., 2020^bÇöltekin, 2015^cSulubacak et al., 2016^dTürk et al., 2020^eKenet-UD GitHub Repository

A.2.2 Procedure

A dedicated Python library named *conllu* (<https://pypi.org/project/conllu/>) was employed to parse the treebanks encoded in CoNLL-U format. The written script in Python automatically constructed the dictionary entries in the following steps: 1) it extracted the main verb/predicate of a sentence; 2) it identified the dependents of the main verb that serve the role of object, oblique and

open clausal complement (xcomp) in their relation to the verb; 3) it catalogued the morphosyntactic properties (case and category) of the identified dependents alongside their respective verbal governors.

Once the dictionary was constructed, each entry was semi-automatically evaluated. First, a Python script automatically eliminated verbal entries that imposed only one category and case requirement on its argument(s). Subsequently, the remaining entries were manually evaluated and CQL queries were designed for the promising verbs. Most of the designed queries resulted in no verified results. An example entry from the constructed dictionary can be seen in (105), which is the entry for the Turkish verb *bas* (to press/push).

```
(105) bas (V):
      {
        'ObjCat': {'NOUN'}, 'ObjCase': {'Dat', 'Acc'},
        'OblCat': {'NOUN'}, 'OblCase': {'Dat', 'Loc'},
        'XCompCat': {'NOUN'}, 'XCompCase': {'Dat'}
      }
```

According to (105), *bas* can take three kinds of arguments: OBJ(ect), OBL(ique), and XCOMP. While OBJ function is realised by a dative or an accusative noun, the OBL element should either be a dative or a locative noun. XCOMP argument, on the other hand, can only be a nominal element in the dative case.

Appendix B

Acceptability judgment task

B.1 Sub-surveys

In the following tables, each cell should be understood as a sentence token that belongs to a particular token set and type based on its row and column indices. For example, the cell whose row name is “TS-6” and whose column name “UCAT-LF” is a sentence token from the token set 6 and represents the type: unlike category like function coordination. A symbol inserted in this cell, which is α in this case, signifies that this sentence is assigned to the survey symbolised as α , which is Survey 1.

B.1.1 Category block

	LCAT-LF	UCAT-LF	LCAT-UF	UCAT-UF
TS-1	α	β	γ	δ
TS-2	δ	α	β	γ
TS-3	γ	δ	α	β
TS-4	β	γ	δ	α
TS-5	α	β	γ	δ
TS-6	δ	α	β	γ
TS-7	γ	δ	α	β
TS-8	β	γ	δ	α
TS-9	α	β	γ	δ
TS-10	δ	α	β	γ
TS-11	γ	δ	α	β
TS-12	β	γ	δ	α

Symbol	Survey
α	Survey 1
β	Survey 2
γ	Survey 3
δ	Survey 4

B.1.2 Case block

	LCASE-LF	UCASE-LF	UCASE-UF
TS-1	α	δ	γ
TS-2	β	α	δ
TS-3	γ	β	α
TS-4	δ	γ	β
TS-5	α	δ	γ
TS-6	β	α	δ
TS-7	γ	β	α
TS-8	δ	γ	β
TS-9	α	δ	γ
TS-10	β	α	δ
TS-11	γ	β	α
TS-12	δ	γ	β

Symbol	Survey
α	Survey 1
β	Survey 2
γ	Survey 3
δ	Survey 4

B.2 Instructions

B.2.1 Turkish original

Sayfa 1/2

Lütfen ankete başlamadan önce yönergeleri dikkatle okuyunuz ve anladığınıza emin olunuz.

Bu ankette size ardı ardına farklı Türkçe cümleler gösterilecek. Öncelikle gösterilen cümleyi dikkatle okuyunuz. Ardından bu cümleyi günlük hayatınızda duyduğunuz veya okuduğunuz bir durum hayal etmeye çalışın.

Örnek durumlardan bazıları:

- Sokaktaki birinin bu cümleyi kurmuş olması
- Afiş, mesaj, kitapçık, sosyal medya postu veya reklam panolarında okumuş olmanız.

Kendinizi bu duruma koyduktan sonra, bu cümleye olan dilsel tepkinizi (–3) ve (3) ölçeği üzerinden puanlandırınız. Ölçeğin uç noktaları şöyle yorumlanmalıdır:

(3): Anadili Türkçe olan biri tarafından kurulmuş gibi duyulan bir cümle. Tamamen doğal ve anlaşılabilir.

(–3): Anlaşılması oldukça güç (hatta anlam ifade etmeyen), garip duyulan ve Türkçeyi yeni öğrenen bir yabancı tarafından kurulmuş olması oldukça muhtemel.

Eğer gösterilen cümleye olan tepkiniz bu iki uç nokta arasında ise aradaki değerleri tepkinizin derecesine göre kullanınız. Örneğin, oldukça doğal duyulan fakat az da olsa bir gariplik barındıran cümle için (1) veya (2) kullanılabilir.

Anket sırasında verdiğiniz puanlar cümleden cümleye önemli değişiklikler gösterebilir.

Sayfa 2/2

ÖNEMLİ NOKTALAR:

- Bu ankette DOĞRU veya YANLIŞ cevap kavramı bulunmamaktadır. Cümleleri puanlarken okulda öğrenmiş olduğunuz çeşitli dilbilgisi ve yazım kurallarını göz önünde bulundurmayınız. Bu çalışma için önemli olan sizin anadili Türkçe olan biri olarak Türkçenin günlük hayattaki kullanımına dair içgüdüsel yaklaşımınızı tespit etmektir.
- Bazı cümlelerin ifade ettiği fikirlere kişisel olarak katılmıyor olabilirsiniz. Fakat bu çalışma için önemli olan cümlelerin ifade ettiklerine karşı olan tutumunuz **DEĞİL**, cümleleri ne kadar anlaşılabilir ve doğal bulduğunuzdur.

Bazı örnek cümlelerin değerlendirilmesi:

Gösterilen bazı cümlelerdeki yazım biçimleri anadili Türkçe olan kişilerin günlük hayattaki konuşmalarını yansıtacak şekilde olabilir. Eğer aşağıdaki cümle size anadili Türkçe olan biri tarafından kurulmuş gibi duyuluyorsa, yazım biçimini görmezden geliniz.

1. “Dersten sonra ananemlere uğrıycam.”

Aşağıdaki örnek cümleyi değerlendirirken bu cümlemin fikir düzeyinde ne ifade ettiğini dikkate almayınız. Eğer gösterilen cümle anlamını sorunsuz bir şekilde iletebiliyorsa, bu cümle kişisel görüşlerinizden bağımsız olarak yüksek bir puan almalıdır.

2. “Monarşilerin yıkılmasıyla beraber acınası ve zavallı bir dünyada yaşar olduk.”

B.2.2 English translation

Page 1/2

Please carefully read and make sure that you fully understand the instructions before starting the survey.

In this survey, you will be shown distinct sentences in Turkish one after the other. First of all, carefully read the sentence shown. Then try to imagine a situation where you hear or read this sentence in your daily life. Some example situations:

- Some stranger on the street utters this sentence
- You read it on a poster, message, booklet, social media post or billboard.

After you put yourself in a situation, rate your linguistic response to this sentence on a scale from (–3) to (3). The endpoints of the scale should be interpreted as follows:

(3): A sentence that sounds like it is constructed by a native Turkish speaker. It is completely natural and fully understandable.

(-3): It is quite likely that the sentence is uttered by a foreigner who just started learning Turkish and it is very difficult to understand (might even be outright incomprehensible) and sounds quite strange.

If your reaction to the sentence shown is somewhere between these two extremes, use the values in-between according to the degree of your reaction. For instance, (1) or (2) can be used for a sentence that sounds quite natural with a little bit of awkwardness. Feel free to use the intermediate values between the endpoints accordingly.

The scores you employ during the survey can vary significantly from sentence to sentence.

Page 2/2

Important points to consider:

- There is no concept of TRUE or FALSE answers in this questionnaire. When scoring sentences, do not take into account the various grammar and spelling rules you learned in your education. For the survey, what is important is your instinctive approach to the use of Turkish in daily life as a native speaker of Turkish.
- You may not personally agree with the ideas expressed by some sentences. However, what is important for this survey is **NOT** your attitude towards what the sentences express, but how understandable and natural you find the sentences.

Evaluation of some sample sentences:

The spellings in some of the sentences shown may reflect the daily speech of native Turkish speakers. If the following sentence sounds like it is constructed by a native Turkish speaker, the unorthodox spelling must be ignored when scoring.

1. "I'll stop by my grandma after class."¹

When evaluating the example sentence below, do not take into account what this sentence expresses at the idea level. If the sentence shown can convey its meaning without any problems and sounds natural, this sentence should receive a high rating regardless of your personal opinions/worldview.

2. "With the collapse of monarchies, we found ourselves living in a pathetic and poor world."

¹ Here the original Turkish sentence was: "Dersten sonra ananemlere uğrıycam." which utilises a type of spelling that better reflects the phonetic aspects of daily speech. However, this kind of spelling is frowned upon both in formal and semi-formal written contexts. Especially the "educated" stratum of the Turkish society tend to be extremely sensitive to various orthography (often confused with grammar) rules.

B.3 Glossed materials

B.3.1 Coordination of unlike categories

- (TS1) a. [Zengin bir memur]NP_{PRED} veya [çok saygın
rich INDE.DET officer or very respected
biri]NP_{PRED} ol-abil-ir.
person be-ABIL-AOR

‘He/she can be a rich officer or a very respected person.’

- b. [Zengin bir memur]NP_{PRED} veya [çok
rich INDE.DET officer or very
saygın]AdjP_{PRED} ol-abil-ir.
respected be-ABIL-AOR

‘He/she can be a rich officer or very respected.’

- c. [Saygın bir memur-u]NP_{OBJ} ve [kendi
respected INDE.DET officer-ACC and own
ofis-i-nde]NP_{ADJUNCT} tebrik et-ti-m.
office-3P-LOC congratulation do-PST-1SG

‘I congratulated a respected officer and in his/her own office.’

- d. [Saygın bir memur-u]NP_{OBJ} ve
respected INDE.DET officer-ACC and
[inanarak]AdvP_{ADJUNCT} tebrik et-ti-m.
believe-ADVZ congratulation do-PST-1SG

‘I congratulated a respected officer and faithfully.’

- (TS2) a. [[Araba-lar ve arkadaşlık] hakkında]PP_{OBJ} konuş-tu-k.
car-PL and friendship about talk-PST-1PL

‘We talked about cars and friendship.’

- b. Arkadaşlığ-ımız hakkında [saygı
friendship-1PL.POSS about respect
çerçeve-si-nde]NP_{ADJUNCT} ve [uzun-ca]AdvP_{ADJUNCT}
framework-3P-LOC and long-ADVZ
konuş-tu-k.
talk-PST-1PL

‘We talked respectfully and at length about our friendship.’

- c. [İki arkadaş gibi]PP_{ADJUNCT} ve [araba-lar
two friend like and car-PL
hakkında]PP_{OBJ} konuş-tu-k.
about talk-PST-1 PL

‘We talked about cars and like two friends.’

- d. [Uzun-ca]AdvP_{ADJUNCT} ve [arkadaşlığ-ımız-ı]NP_{OBJ}
long-ADVZ and friendship-1 PL.POSS-ACC
konuş-tu-k.
talk-PST-1 PL

‘We talked at length and our friendship.’

- (TS3) a. Bu-ndan sonra her şey-i [[kendi kariyer-im
this-ABL after every thing-ACC own career-1 SG.POSS
ve çıkar-lar-ım] için]PP_{ADJUNCT} düşün-eceğ-im.
and interest-PL-1 SG.POSS for think-FUT-1 SG

‘From now on, I will think of everything for my own career and interests.’

- b. Bu-ndan sonra her şey-i [kendi kariyer-im-e
this-ABL after every thing-ACC own career-1 SG.POSS-DAT
göre]PP_{ADJUNCT} ve [bencil-ce]AdvP_{ADJUNCT} düşün-eceğ-im.
according to and selfish-ADVZ think-FUT-1 SG

‘From now on, I will think of everything according to my own career and selfishly.’

- c. Bu-ndan sonra [her gün]NP_{ADJUNCT} ve [kendi
this-ABL after every day and own
kariyer-im-i]NP_{OBJ} düşün-eceğ-im.
career-1 SG.POSS-ACC think-FUT-1 SG

‘From now on, I will think every day and about my career.’

- d. Bu-ndan sonra [kendi kariyer-im-i]NP_{OBJ} ve
this-ABL after own career-1 SG.POSS-ACC and
[sakin-ce]AdvP_{ADJUNCT} düşün-eceğ-im.
calm-ADVZ think-FUT-1 SG

‘From now on, I will think about my own career and calmly.’

- (TS4) a. Bu [savaş-lar]NP_{SUBJ} ve [isyan-lar]NP_{SUBJ}
this war-PL and rebellion-PL
toprak-lar-ımız-da yıl-lar-ca sür-dü.
land-PL-1 PL.POSS-LOC year-PL-ADVZ continue-PST

‘These wars and rebellions continued for years in our lands.’

- b. Bu isyan-lar [yıl-lar boyunca]PP_{OBL} ve [her
this rebellion-PL year-PL throughout and every
gün]NP_{OBL} sür-dü.
day continue-PST

‘These rebellions lasted for years and every day.’

- c. [Bu savaş-lar]NP_{SUBJ} ve [toprak-lar-ımız-da]NP_{ADJUNCT}
this war-PL and land-PL-1PL.POSS-LOC
yıl-lar-ca sür-dü.
year-PL-ADVZ continue-PST

‘These wars and in our lands lasted for years.’

- d. [Bu savaş-lar]NP_{SUBJ} ve [yıl-lar-ca]AdvP_{ADJUNCT}
this war-PL and year-PL-ADVZ
sür-dü.
continue-PST

‘These wars and for years lasted.’

- (TS5) a. Baba-m [abi-m-e]NP_{OBL} ve [bana]NP_{OBL}
father-1SG.POSS brother-1SG.POSS-DAT and me.DAT
aylık harçlık-lar-ımız-ı banka transfer-i-(y)le
monthly allowance-PL-1PL.POSS-ACC bank transfer-3P-INS
yolla-dı.
send-PST

‘My father sent our monthly allowances to my brother and me by bank transfer.’

- b. Baba-m abi-m-e aylık harçlığı-ı-nı
father-1SG.POSS brother-1SG.POSS-DAT monthly allowance-3P-ACC
[banka transfer-i-(y)le]NP_{ADJUNCT} ve [aksatmadan]AdvP_{ADJUNCT}
bank transfer-3P-INS and without delay
yolla-(y)abil-miş.
send-ABIL-PST

‘My father was able to send his monthly allowance to my brother by bank transfer and without delay.’

- c. Baba-m [abi-m-e]NP_{OBL} ve [aylık
father-1SG.POSS brother-1SG.POSS-DAT and monthly
harçlığı-ı-nı]NP_{OBJ} banka transfer-i-(y)le yolla-(y)abil-miş.
allowance-3P-ACC bank transfer-3P-INS send-ABIL-PST.3SG

‘My father was able to send my brother and his monthly allowance by bank transfer.’

- d. Baba-m aylık harçlıđ-ı-nı [abi-m-e]NP_{OBL}
 father-1SG.POSS monthly allowance-3P-ACC brother-1SG.POSS-DAT
 ve [aksatmadan]AdvP_{ADJUNCT} yolla-mış.
 and without delay send-PST

‘My father sent his monthly allowance to my brother and without delay.’

- (TS6) a. Kardeş-im-in iş-ler-i-ni [hızlı-ca]AdvP_{ADJUNCT}
 brother-1SG.POSS-GEN business-PL-3P-ACC quick-ADVZ
 ve [yorulmadan]AdvP_{ADJUNCT} hallet-ti-k.
 and without getting tired take care-PST-1PL

‘We took care of my brother’s affairs quickly and without getting tired.’

- b. Kardeş-im-in iş-ler-i-ni [kendi-miz-i
 brother-1SG.POSS-GEN business-PL-3P-ACC self-1PL.POSS-ACC
 yormadan]AdvP_{ADJUNCT} ve [el-imiz-den gel-diğ-i
 without tiring and hand-1PL.POSS-ABL come-PTCP-3P
 kadar]PP_{ADJUNCT} hallet-ti-k.
 inasmuch as take care-PST-1PL

‘We took care of my brother’s affairs as best we could and without tiring ourselves.’

- c. [Kardeş-im-in iş-ler-i-ni]NP_{OBJ} ve [şehir
 brother-1SG.POSS-GEN business-PL-3P-ACC and city
 merkez-i-nde]NP_{ADJUNCT} hallet-ti-k.
 center-3P-LOC take care-PST-1PL

‘We took care of my brother’s affairs and in downtown.’

- d. [Kardeş-im-in iş-ler-i-ni]NP_{OBJ} ve
 brother-1SG.POSS-GEN business-PL-3P-ACC and
 [yorulmadan]AdvP_{ADJUNCT} hallet-ti-k.
 without getting tired take care-PST-1PL

‘We took care of my brother’s affairs and without getting tired.’

- (TS7) a. Hayat-ın gerçek-ler-i-ni [kendi felsefe-si-ne
 life-GEN reality-PL-3P-ACC own philosophy-3P-DAT
 göre]PP_{ADJUNCT} ve [ol-duğ-u gibi]PP_{ADJUNCT} kabul
 according to and be-PTCP-3P like accept
 et-miştı.
 do-PST.PERF

‘(He/she) had accepted the facts of life according to his/her philosophy and as they were.’

- b. Hayat-in gerçek-ler-i-ni [kendi felsefe-si-ne
 life-GEN reality-PL-3P-ACC own philosophy-3P-DAT
 göre]PP_{ADJUNCT} ve [yadırgamadan]AdvP_{ADJUNCT} kabul
 according to and without hesitation accept
 et-mişti.
 do-PST.PERF

‘He/she accepted the facts of life according to his/her own philosophy and without hesitation.’

- c. [Hayat-in gerçek-ler-i-ni]NP_{OBJ} ve [yaşa-dı-(ğ)ı en
 life-GEN reality-PL-3P-ACC and live-PST-3P most
 zor an-lar-da]NP_{ADJUNCT} kabul ed-ebil-mişti.
 difficult moment-PL-LOC accept do-ABIL-PST.PERF

‘(He/she) was able to accept the facts of life and in the most difficult moments of his/her life.’

- d. [Hayat-in gerçek-ler-i-ni]NP_{OBJ} ve [geçmiş-te-ki
 life-GEN reality-PL-3P-ACC and past-LOC-ADJZ
 gibi]PP_{ADJUNCT} kabul edi-yor-du.
 like accept do-PROG-PST

‘(He/she) accepted the facts of life and as they were in the past.’

- (TS8) a. [Saat-im-i]NP_{OBJ} ve [telefon-um-u]NP_{OBJ} nazikçe
 watch-1SG.POSS-ACC and phone-1SG.POSS-ACC gently
 çekmece-m-e koy-du-m.
 drawer-1SG.POSS-DAT put-PST-1SG

‘I gently put my watch and my cellphone in my drawer.’

- b. Saat-im-i [nazikçe]AdvP_{ADJUNCT} ve [dikkat-le]NP_{ADJUNCT}
 watch-1SG.POSS-ACC gently and care-INS
 çekmece-m-e koy-du-m.
 drawer-1SG.POSS-DAT put-PST-1SG

‘I put my watch in my drawer gently and with care.’

- c. [Saat-im-i]NP_{OBJ} ve [çekmece-m-e]NP_{OBL} nazikçe
 watch-1SG.POSS-ACC and drawer-1SG.POSS-DAT gently
 koy-du-m.
 put-PST-1SG

‘I gently put my watch and in my drawer.’

- d. [Saat-im-i]NP_{OBJ} ve [nazikçe]AdvP_{ADJUNCT}
 watch-1SG.POSS-ACC and gently
 çekmece-m-e koy-du-m.
 drawer-1SG.POSS-DAT put-PST-1SG

‘I put my watch and gently in my drawer.’

- (TS9) a. Günümüz salgın hastalık-lar-ı genelde ya
 modern-day contagious sickness-PL-3P generally either
 [hava]NP_{ADJUNCT} ya da [kan yol-u-yla]NP_{ADJUNCT}
 air or blood path-3P-INS
 bulaş-ır-lar.
 spread-AOR-3PL

‘Today’s epidemic diseases are generally transmitted either by air or blood.’

- b. Bu hastalık [kan yol-u ile]PP_{ADJUNCT} veya
 this sickness blood path-3P with or
 [mikrop bulaşmış gıda-lar-dan]NP_{ADJUNCT} yayıl-abil-ir.
 microbe contaminated food-PL-ABL spread-ABIL-AOR

‘This disease can be spread through blood or from contaminated food.’

- c. [Günümüz salgın hastalık-lar-ı]NP_{SUBJ} ve
 modern-day contagious sickness-PL-3P and
 [kalabalık hastane-ler-den]NP_{ADJUNCT} bulaş-abil-ir.
 crowded hospital-PL-ABL spread-ABIL-AOR

‘Today’s epidemic diseases and from crowded hospitals can be spread.’

- d. Günümüz [salgın hastalık-lar-ı]NP_{SUBJ} ve
 modern-day contagious sickness-PL-3P and
 [hızlı-ca]AdvP_{ADJUNCT} bulaş-abil-ir.
 quick-ly spread-ABIL-AOR

‘Today’s epidemic diseases and can be transmitted quickly.’

- (TS10) a. Köy sakin-ler-i tedarik ed-il-en su-yu
 village habitant-PL-3P supply do-PASS-ADJZ water-ACC
 [[çamaşır yıka-mak ve bahçe sula-mak]
 laundry wash-INF and garden irrigate-INF
 için]PP_{ADJUNCT} kullan-ır-lar-dı.
 for use-AOR-3PL-PST

‘Village residents used the supplied water for washing clothes and watering gardens.’

- b. Köy sakin-ler-i tedarik ed-il-en su-yu
village habitant-PL-3P supply do-PASS-ADJZ water-ACC
çoğunlukla [çamaşır yıka-mak için]PP_{ADJUNCT} veya
mostly laundry wash-INF for or
[bahçe-ye]NP_{ADJUNCT} kullan-ır-lar-dı.
garden-DAT use-AOR-3PL-PST

‘Village residents used the supplied water mostly for washing clothes or the garden.’

- c. Köy sakin-ler-i [tedarik ed-il-en su-yu]NP_{OBJ}
village habitant-PL-3P supply do-PASS-ADJZ water-ACC
ve [bahçe-ler-i-nde]NP_{ADJUNCT} kullan-ır-lar-dı.
and garden-PL-3.POSS-LOC use-AOR-3PL-PST

‘Village residents used the supplied water and in their gardens.’

- d. Köy sakin-ler-i [tedarik ed-il-en su-yu]NP_{OBJ}
village habitant-PL-3P supply do-PASS-ADJZ water-ACC
ve [tasarruf ed-erek]AdvP_{ADJUNCT} kullan-ır-lar-dı.
and austerity do-ADVZ use-AOR-3PL-PST

‘Village residents used the supplied water and sparingly.’

- (TS11) a. Ya [ders-ler-i-ni]NP_{OBJ} ya da [ev iş-ler-i-ni]NP_{OBJ}
either homework-PL-3P-ACC or house chore-PL-3P-ACC
yarım yamalak yap-ar.
half-baked do-AOR

‘(He/she) either does his/her homework or his/her chores half-baked.’

- b. Ders-ler-i-ni [acele-yle]NP_{ADJUNCT} ve
homework-PL-3P-ACC hurry-INS and
[yarım yamalak]AdvP_{ADJUNCT} yap-tı.
half-baked do-PST

‘(He/she) did his/her homework with haste and half-baked.’

- c. [Oda-sı-nda]NP_{ADJUNCT} ve [ders-ler-i-ni]NP_{OBJ} acele-yle
room-3P-LOC and homework-PL-3P-ACC hurry-INS
yap-tı.
do-PST

‘(He/she) did his/her homework and his/her room with haste.’

- d. [Ders-ler-i-ni]NP_{OBJ} ve [yarım yamalak]AdvP_{ADJUNCT}
 homework-PL-3P-ACC and half-baked
 yap-tı.
 do-PST

‘(He/she) did his/her homework and half-baked.’

- (TS12) a. Kürsü-de [yerel minibüs şoför-ler-i-ne]NP_{OBL} ve
 podium-LOC local minibus driver-PL-3P-DAT and
 [belediye-ye]NP_{OBL} saydır-dı.
 municipality-DAT curse-PST

‘(He/she) cursed the local minibus drivers and the municipality on the podium.’

- b. Kürsü-de [yerel minibüs şoför-ler-i hakkında]PP_{OBL}
 podium-LOC local minibus driver-PL-3P about
 ve [belediye-ye]NP_{OBL} saydır-dı.
 and municipality-DAT curse-PST

‘On the podium, he/she cursed at local minibus drivers and municipality.’

- c. Kürsü-de [yerel minibüs şoför-ler-i-ne]NP_{OBL} ve
 podium-LOC local minibus driver-PL-3P-DAT and
 [kızgın bir şekil-de]NP_{ADJUNCT} saydır-dı.
 angry INDEFDET manner-LOC curse-PST

‘On the podium, (he/she) cursed the local minibus drivers and angrily.’

- d. Kürsü-de [yerel minibüs şoför-ler-i-ne]NP_{OBL} ve
 podium-LOC local minibus driver-PL-3P-DAT and
 [öfke saç-arak]AdvP_{ADJUNCT} saydır-dı.
 anger spill-ADVZ curse-PST

‘On the podium, (he/she) cursed the local minibus drivers and indignantly.’

B.3.2 Coordination of unlike cases

- (TS1) a. Bugün [yeni araba müze-si](DAT)_{OBL} ve [kent
 today new car museum-3P and city
 orman-ı-na]DAT_{OBL} git-ti-k.
 forest-3P-DAT go-PST-1PL

‘Today we went to the new car museum and urban forest’

- b. Yeni araba müze-si-ne [metro hat-tı-ndan]ABL_{ADJUNCT}
 new car museum-3P-DAT subway line-3P-ABL
 ve [M1 tren-i-yle]INS_{ADJUNCT} git-ti-k.
 and M1 train-3P-INS go-PST-1PL

‘We went to the new car museum from the metro line and with the M1 train.’

- c. [Yeni araba müze-si-ne]DAT_{OBL} ve [biz]NOM_{SUBJ}
 new car museum-3P-DAT and we
 metro-yla git-ti-k.
 subway-INS go-PST-1PL

‘We went and to the new car museum by the subway.’

- (TS2) a. [Devlet-ler]NOM_{SUBJ} ve [iş adam-lar-ı]NOM_{SUBJ}
 government-PL and business man-PL-3P
 dünya-da-ki herkes-e ekonomik işkence
 world-LOC-ADJZ everyone-DAT economic torture
 uygulu-yor-lar.
 inflict-PRES.PROG-3PL

‘Governments and businessmen inflict economic torture on everyone in the world.’

- b. Devlet-ler dünya-da-ki herkes-e hem
 government-PL world-LOC-ADJZ everyone-DAT both
 [ekonomik boyut-ta]LOC_{ADJUNCT} hem de [sosyal
 economic dimension-LOC and social
 yol-lar-dan]ABL_{ADJUNCT} işkence uygulu-yor-lar.
 method-PL-ABL torture inflict-PRES.PROG-3PL

‘Governments torture everyone in the world, both economically and socially.’

- c. [Devlet-ler]NOM_{SUBJ} ve [politik neden-ler-le]INS_{ADJUNCT}
 government-PL and political reason-PL-INS
 dünya-da-ki herkes-e ekonomik işkence
 world-LOC-ADJZ everyone-DAT economic torture
 uygulu-yor-lar.
 inflict-PRES.PROG-3PL

‘Governments and with political reasons torture everyone in the world.’

- (TS3) a. Okul-dan arkadaş-lar-ım [ben-i]ACC_{OBJ} ve
 school-ABL friend-PL-1SG.POSS me-ACC and
 [ablam-ı]ACC_{OBJ} büyük okul parti-si-ne çağır-dı-lar.
 sister-ACC big school party-3P-DAT invite-PST-3PL

‘My friends from school invited me and my sister to the big school party.’

- b. Okul-dan arkadaş-lar-ım ben-i [heyecan
 school-ABL friend-PL-1SG.POSS me-ACC excitement
 içerisi-nde]LOC_{ADJUNCT} ve [sabırsızlık-la]INS_{ADJUNCT} büyük
 inside-LOC and impatience-INS big
 okul parti-si-ne çağır-dı-lar.
 school party-3P-DAT invite-PST-3PL

‘My friends from school excitedly and impatiently invited me to the big school party.’

- c. Okul-dan arkadaş-lar-ım [ben-i]ACC_{OBJ} ve [büyük
 school-ABL friend-PL-1SG.POSS me-ACC and big
 okul parti-si-ne]DAT_{OBL} çağır-dı-lar.
 school party-3P-DAT invite-PST-3PL

‘My friends from school invited me and to the big school party.’

- (TS4) a. Oğlum-a internet-ten [şapka]NOM_{OBJ} ve [ayakkabı]NOM_{OBJ}
 son-DAT internet-ABL hat and shoe
 al-dı-m
 buy-PST-1SG

‘I bought my son a pair of shoes and a hat through the internet.’

- b. Oğlum-a [internet-ten]ABL_{ADJUNCT} ve [kredi
 son-DAT internet-ABL and credit
 kart-ı-yla]INS_{ADJUNCT} ayakkabı al-dı-m
 card-3P-INS shoe buy-PST-1SG

‘I bought my son a pair of shoes through the internet and by credit card.’

- c. [Oğlum-a]DAT_{OBL} ve [iste-diğ-i ayakkabı-yı]ACC_{OBJ}
 son-DAT and want-PTCP-3P shoe-ACC
 internet-ten al-dı-m
 internet-ABL buy-PST-1SG

‘I bought my son and the pair of shoes he wanted through the internet.’

- (TS5) a. [Kuzen-im]_{NOM_{SUBJ}} ve [ben]_{NOM_{SUBJ}} komşu-nun
 cousin-1SG.POSS and I neighbor-GEN
 oğl-u-na eski kitap-lar-ımız-ı kapı-nın
 son-3P-DAT old book-PL-1PL.POSS-ACC door-GEN
 ağz-ı-ndan ver-di-k.
 mouth-3P-ABL give-PST-1PL
 ‘My cousin and I gave the neighbor’s son our old books on the doorstep.’

- b. Eski kitap-lar-ımız-ı komşu-nun oğl-u-na
 old book-PL-1PL.POSS-ACC neighbor-GEN son-3P-DAT
 [acele içi-nde]_{LOC_{ADJUNCT}} ve [kapı-nın
 haste inside-LOC and door-GEN
 ağz-ı-ndan]_{ABL_{ADJUNCT}} ver-di-k.
 mouth-3P-ABL give-PST-1PL
 ‘(We) gave our old books to the neighbor’s son in a hurry and on the doorstep.’

- c. Kuzen-im [komşu-nun oğl-u-na]_{DAT_{OBL}} ve
 cousin-1SG.POSS neighbor-GEN son-3P-DAT and
 [eski kitap-lar-ı-nı]_{ACC_{OBJ}} kapı-nın ağz-ı-nda
 old book-PL-3P-ACC door-GEN mouth-3P-LOC
 ver-di.
 give-PST
 ‘My cousin gave the neighbor’s son and his old books on the doorstep.’

- (TS6) a. Günümüz genç-ler-i ya [sahil
 modern-day youngster-PL-3P either coast
 kenarı-nda]_{LOC_{ADJUNCT}} ya da [alışveriş
 side-LOC or shopping
 merkez-ler-i-nde]_{LOC_{ADJUNCT}} yürü-r-ler.
 center-PL-LOC walk-AOR-3PL
 ‘Today’s young people either walk by the beach or in shopping malls.’

- b. Günümüz genç-ler-i [el-ler-i-nde
 modern-day youngster-PL-3P hand-PL-3P-LOC
 telefon-la]_{INS_{ADJUNCT}} ve [yol-un orta-sı-ndan]_{ABL_{ADJUNCT}}
 telephone-INS and road-GEN middle-3P-ABL
 yürüme-yi çok sev-er-ler.
 walking-ACC much like-AOR-PL
 ‘Today’s young people love to walk in the middle of the road and with their phones in their hands.’

- c. Günümüz genç-ler-i ya [sahil
modern-day youngster-PL-3P either coast
kenarı-nda]LOC_{ADJUNCT} ya da [alışveriş
side-LOC or shopping
merkez-ler-i-ne]DAT_{OBL} yürü-r-ler.
center-PL-DAT walk-AOR-3PL

‘Today’s young people either walk by the beach or to the malls.’

- (TS7) a. Yazar yeni kitab-ı-nı [bunalım-da-ki
author new book-3P-ACC depression-LOC-ADJZ
ergen-ler-e]DAT_{OBL} ve [genç yetişkin-ler-e]DAT_{OBL}
adolescent-PL-DAT and young adult-PL-DAT
ada-dı.
dedicate-PST

‘The author dedicated (his/her) new book to depressed adolescents and young adults.’

- b. Yazar yeni kitab-ı-nı bunalım-da-ki genç
author new book-3P-ACC depression-LOC-ADJZ young
yetişkin-ler-e [mütevazilik-le]INS_{ADJUNCT} ve [kalp-ten]ABL_{ADJUNCT}
adult-PL-DAT modesty-INS and heart-ABL
ada-dı.
dedicate-PST

‘The author humbly and wholeheartedly dedicated (his/her) new book to depressed young adults.’

- c. Yazar yeni kitab-ı-nı [bunalım-da-ki
author new book-3P-ACC depression-LOC-ADJZ
ergen-ler-e]DAT_{OBL} ve [kalp-ten]ABL_{ADJUNCT} ada-dı.
adolescent-PL-DAT and heart-ABL dedicate-PST

‘The author dedicated (his/her) new book to depressed adolescents and from the heart.’

- (TS8) a. [İhtiyar heyet-i]NOM_{SUBJ} ve [seçkin kimse-ler]NOM_{SUBJ}
elder council-3P and distinguished people-PL
el kadar çocuğ-u kral ilan
hand inasmuch as child-ACC king proclamation
et-ti-ler.
do-PST-3PL

‘The council of elders and distinguished people proclaimed the very young child as king.’

- b. İhtiyar heyet-i el kadar çocuğ-u [yüksek
elder council-3P hand inasmuch as child-ACC high
koltuk-lar-ı-ndan]ABL_{ADJUNCT} ve [gurur-lu bir
seat-PL-3P-ABL and proud-ADJZ INDE.DET
eda-yla]INS_{ADJUNCT} kral ilan et-ti-ler.
manner-INS king proclamation do-PST-3PL

'The council of elders proclaimed the very young child as king from their high seats and proudly.'

- c. İhtiyar heyet-i [el kadar çocuğ-u]ACC_{OBJ}
elder council-3P hand inasmuch as child-ACC
ve [oturum-da]LOC_{ADJUNCT} kral ilan
and session-LOC king proclamation
et-ti-ler.
do-PST-3PL

'The council of elders proclaimed the very young child and in the session.'

- (TS9) a. Beşiktaş defans-ı kale-si-ni [soğukkanlılık](INS)_{ADJUNCT}
Beşiktaş defense-3P goal-3P-ACC calmness
ve [disiplin-le]INS_{ADJUNCT} hayati bir gol-den
and discipline-INS vital INDE.DET goal-ABL
koru-du.
protect-PST

Beşiktaş defense protected its goal from an important score with composure and discipline.'

- b. Beşiktaş defans-ı kale-si-ni [soğukkanlılık-la]INS_{ADJUNCT}
Beşiktaş defense-3P goal-3P-ACC calmness-INS
ve [dayanışma içi-nde]LOC_{ADJUNCT} koruma-yı
and solidarity inside-LOC protection-ACC
başar-dı.
succeed-PST

Beşiktaş defense succeeded in protecting its goal with composure and in solidarity.'

- c. Beşiktaş defans-ı [kale-yi]ACC_{OBJ} ve [ceza
Beşiktaş defense-3P goal-ACC and penalty
saha-sı-nda]LOC_{ADJUNCT} hayati bir gol-den
area-3P-LOC vital INDE.DET goal-ABL
koru-yabil-di.
protect-ABIL-PST

'Beşiktaş defense was able to protect the goal and in the penalty area from a vital score.'

- (TS10) a. [Küçük çocuk]NOM_{SUBJ} ve [arkadaş-ı]NOM_{SUBJ} akşam
 little child and friend-3P evening
 yemeği-ni kendi masa-lar-ı-nda ye-di-ler.
 meal-ACC own table-PL-3P-LOC eat-PST-PL

‘The little child and their friend had dinner at their own table.’

- b. Küçük çocuk akşam yemeği-ni [kendi
 little child evening meal-ACC own
 masa-sı-nda]LOC_{ADJUNCT} ve [anne-si-nin
 table-3P-LOC and mother-3P-GEN
 el-i-nden]ABL_{ADJUNCT} ye-di.
 hand-3P-ABL eat-PST

‘The little child had dinner at his/her own table and from his/her mother’s hand.’

- c. Küçük çocuk [akşam yemeği-ni]ACC_{OBJ} ve
 little child evening meal-ACC and
 [anne-si-nin el-i-nden]ABL_{ADJUNCT} ye-di.
 mother-3P-GEN hand-3P-ABL eat-PST

‘The little child had dinner and from his/her mother’s hand.’

- (TS11) a. Dünya-da-ki [yenilik-ler-i]ACC_{OBJ} ve [gelişme-ler-i]ACC_{OBJ}
 world-LOC-ADJZ innovation-PL-ACC and development-PL-ACC
 internet-ten takip ed-iyor-um.
 internet-ABL follow do-PRES.PROG-1SG

‘I follow the innovations and developments in the world over the internet.’

- b. Dünya-da-ki gelişme-ler-i [küçük yurt
 world-LOC-ADJZ development-PL-ACC small dormitory
 oda-m-da]LOC_{ADJUNCT} ve [internet-ten]ABL_{ADJUNCT} takip
 room-1SG.POSS-LOC and internet-ABL follow
 ed-iyor-um.
 do-PRES.PROG-1SG

‘I follow the developments in the world in my small dormitory room and over the internet.’

- c. Dünya-da-ki [gelişme-ler-i]ACC_{OBJ} ve
 World-LOC-ADJZ development-PL-ACC and
 [internet-ten]ABL_{ADJUNCT} takip ed-iyor-um.
 internet-ABL follow do-PRES.PROG-1SG

‘I follow the developments in the world and over the internet.’

- (TS12) a. [Park görevli-si-ne]DAT_{OBL} ve [park-ın sahib-i-ne]DAT_{OBL}
 park attendant-3P-DAT and park-GEN owner-3P-DAT
 ödeme-yi mobil uygulama-dan
 payment-ACC mobile application-ABL
 yap-tığ-ım-ı söyle-di-m.
 make-PTCP-1SG.POSS-ACC say-PST-1SG

'I told the park attendant and the owner of the park that I made the payment through the mobile app.'

- b. Park görevli-si-ne ödeme-m-i [mobil
 park attendant-3P-DAT payment-1SG.POSS-ACC mobile
 uygulama-dan]ABL_{ADJUNCT} ve [hediye kart-ı-yla]INS_{ADJUNCT}
 application-ABL and gift card-3P-INS
 yap-tığ-ım-ı söyle-di-m.
 make-PTCP-1SG.POSS-ACC say-PST-1SG

'I told the parking attendant that I made my payment from the mobile app and with gift card.'

- c. Park görevli-si-ne [ödeme-m-i]ACC_{OBJ} ve
 park attendant-3P-DAT payment-1SG.POSS-ACC and
 [mobil uygulama-dan]ABL_{ADJUNCT} yap-tığ-ım-ı
 mobile application-ABL make-PTCP-1SG.POSS-ACC
 söyle-di-m.
 say-PST-1SG

'I told the parking attendant that I made my payment and from the mobile app.'

Appendix C

Implementation

C.1 Grammar

C.1.1 Preamble

```
UC    TURKISH    CONFIG (1.0)
ROOTCAT    S.
FILES      .
LEXENTRIES (UC TURKISH).
RULES      (UC TURKISH).
TEMPLATES  (UC TURKISH).
GOVERNABLERELATIONS SUBJ OBJ OBJ2 COMP XCOMP XCOMP-PRED OBL OBL-?+.
SEMANTICFUNCTIONS  ADJUNCT TOPIC FOCUS POSS STANDARD.
NONDISTRIBUTIVES   NUM PERS CONJ-FORM.
EPSILON    e.
OPTIMALITYORDER  NOGOOD.
```

C.1.2 C-structure rules

```
-----
UC    TURKISH    RULES (1.0)

"Sentence rule"
S --> (X_MAX: (^ SUBJ) = !
      (! CAT)=c N
      (! CASE)=c NOM)
      X_MAX: ^=!
      (! CAT) =c V.

"Coordination rule"
COORD --> X_MAX+: ! $ ^;
          CONJ
          X_MAX: ! $ ^.

"Bar level 2 (maximal) rules"
X_MAX --> { "VP"
            {
              "Verbal adjuncts"
              (COORD: ! $ (^ ADJUNCT)
                (! CAT) $ {P N ADV})
              |
              (X_MAX: ! $ (^ ADJUNCT)
                (! CAT) $ {P N ADV})
            }
            "Verbal head"
            X_1: ^=!
            (! CAT) =c V
```

```

|
    "NP"
    {
        "Nominal adjuncts"
        (COORD: ! $ (~ ADJUNCT)
            (! CAT) $ {P A})
        |
        (X_MAX: ! $ (~ ADJUNCT)
            (! CAT) $ {P A})
    }
    X_1: ^=!
        (! CAT) =c N

|

    "PP"
    {"Object of P head"
    COORD: (~ OBJ)=!
    |
    X_MAX: (~ OBJ)=!
    }
    X_0: ^=!
        (! CAT) =c P

|

    "Shallow AdvP"
    X_0: ^=!
        (! CAT) =c ADV

|

    "Shallow AdjP"
    X_0: ^=!
        (! CAT) =c A
}.

"Bar Level 1 Rules"
X_1 --> {
    "V-bar"
    {"Immediately preverbal position"
    {
        (COORD: (~ OBJ) = !)
        |
        (COORD: (~ OBL) = !)
        |
        (COORD: (~ XCOMP-PRED) = !)
    }
    |
    {
        (X_MAX: (~ OBJ) = !)
        |
        (X_MAX: (~ OBL) = !)
        |
        (X_MAX: (~ XCOMP-PRED) = !)
    }
    }
    "Verbal head"
    X_0: ^=!
        (! CAT) =c V

|

    "Optional determiner + nominal head"
    (X_0: ^\CAT = !\CAT
        (! CAT) =c D)
    X_0: ^=!
        (! CAT) =c N
}.

```


C.1.3 Templates

```

----
UC    TURKISH    TEMPLATES (1.0)

"Ordinary noun template"
NOUN(P CS NUMBER) = (^ PRED)='P',
                    (^ PERS) = 3
                    (^ CAT) = N
                    (^ CASE) = CS
                    (^ NUM) = NUMBER.

"Adjective template"
ADJ(P) = (^ PRED)='P',
         (^ CAT) = A.

ADV(P) = (^ PRED)='P',
         (^ CAT) = ADV.

"Pronoun template"
PRO(PERSON NUMBER CS) = (^ PRED) = 'PRO',
                        (^ CAT) = N
                        (^ CASE) = CS
                        (^ PERS) = PERSON
                        (^ NUM) = NUMBER.

"Postposition template"
POSTP(PR FORM CS) = (^ PRED)='PR<(^ OBJ)>',
                    (^ CAT) = P
                    (^ OBJ CAT) =c N
                    (^ OBJ CASE) =c CS
                    (^ PFORM) = FORM.

"Verb with 2 args"
TRANS(P TENSE ARG-FUNC) = (^ PRED)='P<(^ SUBJ)(^ ARG-FUNC)>',
                          (^ CAT) = V
                          (^ TNS) = TENSE.

"Verb with 1 arg"
INTRANS(P TENSE) = (^ PRED)='P<(^ SUBJ)>',
                  (^ CAT) = V
                  (^ TNS) = TENSE.

"Subj-verb agreement + subj pro-drop"
SUBJ_PRO(PERSON NUMBER) = {(^ SUBJ PRED) = 'PRO'}
                        (^ SUBJ NUM) = NUMBER
                        (^ SUBJ CASE) = NOM
                        (^ SUBJ PERS) = PERSON
                        (^ SUBJ CAT) = N.

```

C.1.4 Lexicon

```

----
UC    TURKISH    LEXICON (1.0)

"Pronouns"
ben      X_0      *      @(PRO 1 SG NOM).
biz      X_0      *      @(PRO 1 PL NOM).
sen      X_0      *      @(PRO 2 SG NOM).
siz      X_0      *      @(PRO 2 PL NOM).
o        X_0      *      @(PRO 3 SG NOM).
onlar    X_0      *      @(PRO 3 PL NOM).

beni     X_0      *      @(PRO 1 SG ACC).
bizi     X_0      *      @(PRO 1 PL ACC).
seni     X_0      *      @(PRO 2 SG ACC).
sizi     X_0      *      @(PRO 2 PL ACC).

```

onu	X_0	*	@(PRO 3 SG ACC).
onlari	X_0	*	@(PRO 3 PL ACC).
bana	X_0	*	@(PRO 1 SG DAT).
bize	X_0	*	@(PRO 1 PL DAT).
sana	X_0	*	@(PRO 2 SG DAT).
size	X_0	*	@(PRO 2 PL DAT).
ona	X_0	*	@(PRO 3 SG DAT).
onlara	X_0	*	@(PRO 3 PL DAT).
bende	X_0	*	@(PRO 1 SG LOC).
bizde	X_0	*	@(PRO 1 PL LOC).
sende	X_0	*	@(PRO 2 SG LOC).
sizde	X_0	*	@(PRO 2 PL LOC).
onda	X_0	*	@(PRO 3 SG LOC).
onlarda	X_0	*	@(PRO 3 PL LOC).
benle	X_0	*	@(PRO 1 SG INS).
bizle	X_0	*	@(PRO 1 PL INS).
senle	X_0	*	@(PRO 2 SG INS).
sizle	X_0	*	@(PRO 2 PL INS).
onla	X_0	*	@(PRO 3 SG INS).
onlarla	X_0	*	@(PRO 3 PL INS).
benden	X_0	*	@(PRO 1 SG ABL).
bizden	X_0	*	@(PRO 1 PL ABL).
senden	X_0	*	@(PRO 2 SG ABL).
sizden	X_0	*	@(PRO 2 PL ABL).
ondan	X_0	*	@(PRO 3 SG ABL).
onlardan	X_0	*	@(PRO 3 PL ABL).
"Verbs"			
"to talk"			
konustuk	X_0	*	@(TRANS TALK PST OBJ)
			@(SUBJ_PRO 1 PL)
			(~ OBJ PRED:
			{(<- CAT) =c N {(<- CASE) =c ACC (<- CASE) =c NOM}
			(<- CAT) =c P {(<- PFORM) =c HAKKINDA (<- PFORM) =c UZERINE}}).
konustular	X_0	*	@(TRANS TALK PST OBJ)
			@(SUBJ_PRO 3 PL)
			(~ OBJ PRED:
			{(<- CAT) =c N {(<- CASE) =c ACC (<- CASE) =c NOM}
			(<- CAT) =c P {(<- PFORM) =c HAKKINDA (<- PFORM) =c UZERINE}}).
"to last"			
surdu	X_0	*	@(TRANS LAST PST OBL)
			@(SUBJ_PRO 3 SG)
			(~ OBL PRED:
			{(<- CAT) =c N (<- CASE) =c NOM
			(<- CAT) =c P (<- PFORM) =c BOYUNCA
			(<- CAT) =c ADV}).
surduler	X_0	*	@(TRANS LAST PST OBL)
			@(SUBJ_PRO 3 PL)
			(~ OBL PRED:
			{(<- CAT) =c N (<- CASE) =c NOM
			(<- CAT) =c P (<- PFORM) =c BOYUNCA
			(<- CAT) =c ADV}).
"to swear"			
saydirdi	X_0	*	@(TRANS SWEAR PST OBL)
			@(SUBJ_PRO 3 SG)
			(~ OBL PRED:
			{(<- CAT) =c N (<- CASE) =c DAT
			(<- CAT) =c P (<- PFORM) =c HAKKINDA}}).
"to be"			
olacak	X_0	*	@(TRANS BE FUT XCOMP-PRED)
			@(SUBJ_PRO 3 SG)
			(~ PREDLINK PRED:
			{(<- CAT) =c N ~{(<- CASE) =c ACC (<- CASE) =c DAT}

```

      (<- CAT) =c P |
      (<- CAT) =c A}).

olmalı X_0 *      @(TRANS BE PRES XCOMP-PRED)
                  @(SUBJ_PRO 3 SG)
                  (^ NECESS) = +
                  (^ PREDLINK PRED:
                  {(<- CAT) =c N ~{(<- CASE) =c ACC | (<- CASE) =c DAT} |
                  (<- CAT) =c P |
                  (<- CAT) =c A}).

"to work/to train"
calisiyoruz X_0 *      @(INTRANS WORK PRES)
                        @(SUBJ_PRO 1 PL)
                        (^ ASPECT) = PROG.

"to reach"
ulasirsın X_0 *      @(TRANS REACH AOR OBL)
                      @(SUBJ_PRO 2 SG)
                      (^ OBL CASE) = DAT.

"Determiners"
bir X_0 *      (^ DEF) = -
               (^ NUM) = SG
               (^ CAT) = D.

her X_0 *      (^ SPEC PRED) = 'EVERY'
               (^ NUM) = SG
               (^ CAT) = D.

"Nouns"

"Car"
araba      X_0      *      @(NOUN CAR NOM SG).
arabalar   X_0      *      @(NOUN CAR NOM PL).
arabayı    X_0      *      @(NOUN CAR ACC SG).
arabaları  X_0      *      @(NOUN CAR ACC PL).
arabaya    X_0      *      @(NOUN CAR DAT SG).
arabalara  X_0      *      @(NOUN CAR DAT PL).
arabada    X_0      *      @(NOUN CAR LOC SG).
arabalarda X_0      *      @(NOUN CAR LOC PL).
arabayla   X_0      *      @(NOUN CAR INS SG).
arabalarla X_0      *      @(NOUN CAR INS PL).
arabadan   X_0      *      @(NOUN CAR ABL SG).
arabalardan X_0      *      @(NOUN CAR ABL PL).

"Driver"
sofor      X_0      *      @(NOUN DRIVER NOM SG).
soforler   X_0      *      @(NOUN DRIVER NOM PL).
soforu     X_0      *      @(NOUN DRIVER ACC SG).
soforleri  X_0      *      @(NOUN DRIVER ACC PL).
sofore     X_0      *      @(NOUN DRIVER DAT SG).
soforlere  X_0      *      @(NOUN DRIVER DAT PL).
soforde    X_0      *      @(NOUN DRIVER LOC SG).
soforlerde X_0      *      @(NOUN DRIVER LOC PL).
soforle    X_0      *      @(NOUN DRIVER INS SG).
soforlerle X_0      *      @(NOUN DRIVER INS PL).
soforden   X_0      *      @(NOUN DRIVER ABL SG).
soforlerden X_0      *      @(NOUN DRIVER ABL PL).

"Municipality"
belediye   X_0      *      @(NOUN MUNICIPALITY NOM SG).
belediyeler X_0      *      @(NOUN MUNICIPALITY NOM PL).
belediyeyi X_0      *      @(NOUN MUNICIPALITY ACC SG).
belediyeleri X_0      *      @(NOUN MUNICIPALITY ACC PL).
belediyeye X_0      *      @(NOUN MUNICIPALITY DAT SG).
belediyelere X_0      *      @(NOUN MUNICIPALITY DAT PL).
belediyede X_0      *      @(NOUN MUNICIPALITY LOC SG).
belediyelerde X_0      *      @(NOUN MUNICIPALITY LOC PL).
belediyeyle X_0      *      @(NOUN MUNICIPALITY INS SG).
belediyelerle X_0      *      @(NOUN MUNICIPALITY INS PL).

```

belediyeden	X_0	*	@(NOUN MUNICIPALITY ABL SG).
belediyelerden	X_0	*	@(NOUN MUNICIPALITY ABL PL).
"Era"			
donem	X_0	*	@(NOUN ERA NOM SG).
donemler	X_0	*	@(NOUN ERA NOM PL).
donemi	X_0	*	@(NOUN ERA ACC SG).
donemleri	X_0	*	@(NOUN ERA ACC PL).
doneme	X_0	*	@(NOUN ERA DAT SG).
donemlere	X_0	*	@(NOUN ERA DAT PL).
donemde	X_0	*	@(NOUN ERA LOC SG).
donemlerde	X_0	*	@(NOUN ERA LOC PL).
donemle	X_0	*	@(NOUN ERA INS SG).
donemlerle	X_0	*	@(NOUN ERA INS PL).
donemden	X_0	*	@(NOUN ERA ABL SG).
donemlerden	X_0	*	@(NOUN ERA ABL PL).
"Instrument"			
enstruman	X_0	*	@(NOUN INSTRUMENT NOM SG).
enstrumanlar	X_0	*	@(NOUN INSTRUMENT NOM PL).
enstrumani	X_0	*	@(NOUN INSTRUMENT ACC SG).
enstrumanlari	X_0	*	@(NOUN INSTRUMENT ACC PL).
enstrumana	X_0	*	@(NOUN INSTRUMENT DAT SG).
enstrumanlara	X_0	*	@(NOUN INSTRUMENT DAT PL).
enstrumanda	X_0	*	@(NOUN INSTRUMENT LOC SG).
enstrumanlarda	X_0	*	@(NOUN INSTRUMENT LOC PL).
enstrumanla	X_0	*	@(NOUN INSTRUMENT INS SG).
enstrumanlarla	X_0	*	@(NOUN INSTRUMENT INS PL).
enstrumandan	X_0	*	@(NOUN INSTRUMENT ABL SG).
enstrumanlardan	X_0	*	@(NOUN INSTRUMENT ABL PL).
"Laundry"			
camasir	X_0	*	@(NOUN LAUNDRY NOM SG).
camasirlar	X_0	*	@(NOUN LAUNDRY NOM PL).
camasiri	X_0	*	@(NOUN LAUNDRY ACC SG).
camasirlari	X_0	*	@(NOUN LAUNDRY ACC PL).
camasira	X_0	*	@(NOUN LAUNDRY DAT SG).
camasirlara	X_0	*	@(NOUN LAUNDRY DAT PL).
camasirda	X_0	*	@(NOUN LAUNDRY LOC SG).
camasirlarda	X_0	*	@(NOUN LAUNDRY LOC PL).
camasirla	X_0	*	@(NOUN LAUNDRY INS SG).
camasirlarla	X_0	*	@(NOUN LAUNDRY INS PL).
camasirdan	X_0	*	@(NOUN LAUNDRY ABL SG).
camasirlardan	X_0	*	@(NOUN LAUNDRY ABL PL).
"Love"			
sevgi	X_0	*	@(NOUN LOVE NOM SG).
sevgiler	X_0	*	@(NOUN LOVE NOM PL).
sevgiyi	X_0	*	@(NOUN LOVE ACC SG).
sevgileri	X_0	*	@(NOUN LOVE ACC PL).
sevgiye	X_0	*	@(NOUN LOVE DAT SG).
sevgilere	X_0	*	@(NOUN LOVE DAT PL).
sevgide	X_0	*	@(NOUN LOVE LOC SG).
sevgilerde	X_0	*	@(NOUN LOVE LOC PL).
sevgiyle	X_0	*	@(NOUN LOVE INS SG).
sevgilerle	X_0	*	@(NOUN LOVE INS PL).
sevgiden	X_0	*	@(NOUN LOVE ABL SG).
sevgilerden	X_0	*	@(NOUN LOVE ABL PL).
"Aim"			
hedef	X_0	*	@(NOUN AIM NOM SG).
hedefler	X_0	*	@(NOUN AIM NOM PL).
hedefi	X_0	*	@(NOUN AIM ACC SG).
hedefleri	X_0	*	@(NOUN AIM ACC PL).
hedefe	X_0	*	@(NOUN AIM DAT SG).
hedeflere	X_0	*	@(NOUN AIM DAT PL).
hedefte	X_0	*	@(NOUN AIM LOC SG).
hedeflerde	X_0	*	@(NOUN AIM LOC PL).
hedefle	X_0	*	@(NOUN AIM INS SG).
hedeflerle	X_0	*	@(NOUN AIM INS PL).
hedeften	X_0	*	@(NOUN AIM ABL SG).

hedeflerden	X_0	*	@(NOUN AIM ABL PL).
"Brush"			
firca	X_0	*	@(NOUN BRUSH NOM SG).
fircalar	X_0	*	@(NOUN BRUSH NOM PL).
fircayi	X_0	*	@(NOUN BRUSH ACC SG).
fircalari	X_0	*	@(NOUN BRUSH ACC PL).
fircaya	X_0	*	@(NOUN BRUSH DAT SG).
fircalara	X_0	*	@(NOUN BRUSH DAT PL).
fircada	X_0	*	@(NOUN BRUSH LOC SG).
fircalarda	X_0	*	@(NOUN BRUSH LOC PL).
fircayla	X_0	*	@(NOUN BRUSH INS SG).
fircalarla	X_0	*	@(NOUN BRUSH INS PL).
fircadan	X_0	*	@(NOUN BRUSH ABL SG).
fircalardan	X_0	*	@(NOUN BRUSH ABL PL).
"Bristle"			
kil	X_0	*	@(NOUN BRISTLE NOM SG).
killar	X_0	*	@(NOUN BRISTLE NOM PL).
kili	X_0	*	@(NOUN BRISTLE ACC SG).
killari	X_0	*	@(NOUN BRISTLE ACC PL).
kila	X_0	*	@(NOUN BRISTLE DAT SG).
killara	X_0	*	@(NOUN BRISTLE DAT PL).
kilda	X_0	*	@(NOUN BRISTLE LOC SG).
killarda	X_0	*	@(NOUN BRISTLE LOC PL).
killa	X_0	*	@(NOUN BRISTLE INS SG).
killarla	X_0	*	@(NOUN BRISTLE INS PL).
kildan	X_0	*	@(NOUN BRISTLE ABL SG).
killardan	X_0	*	@(NOUN BRISTLE ABL PL).
"Week"			
hafta	X_0	*	@(NOUN WEEK NOM SG).
haftalar	X_0	*	@(NOUN WEEK NOM PL).
haftayi	X_0	*	@(NOUN WEEK ACC SG).
haftalari	X_0	*	@(NOUN WEEK ACC PL).
haftaya	X_0	*	@(NOUN WEEK DAT SG).
haftalara	X_0	*	@(NOUN WEEK DAT PL).
haftada	X_0	*	@(NOUN WEEK LOC SG).
haftalarda	X_0	*	@(NOUN WEEK LOC PL).
haftayla	X_0	*	@(NOUN WEEK INS SG).
haftalarla	X_0	*	@(NOUN WEEK INS PL).
haftadan	X_0	*	@(NOUN WEEK ABL SG).
haftalardan	X_0	*	@(NOUN WEEK ABL PL).
"Misc nouns"			
bicimde	X_0	*	@(NOUN SHAPE LOC SG).
yerde	X_0	*	@(NOUN PLACE LOC SG).
antrenorle	X_0	*	@(NOUN TRAINER INS SG).
zamanda	X_0	*	@(NOUN TIME LOC SG).
fiyattan	X_0	*	@(NOUN PRICE ABL SG).
elden	X_0	*	@(NOUN HAND ABL SG).
kargoyla	X_0	*	@(NOUN CARGO INS SG).
konusma	X_0	*	@(NOUN SPEECH NOM SG).
muhendis	X_0	*	@(NOUN ENGINEER NOM SG).
memur	X_0	*	@(NOUN OFFICER NOM SG).
Bertan	X_0	*	@(NOUN BERTAN NOM SG).
film	X_0	*	@(NOUN MOVIE NOM SG).
savaslar	X_0	*	@(NOUN WAR NOM PL).
yil	X_0	*	@(NOUN YEAR NOM SG).
yillar	X_0	*	@(NOUN YEAR NOM PL).
gun	X_0	*	@(NOUN DAY NOM SG).
fakirlikten	X_0	*	@(NOUN POVERTY ABL SG).
aylar	X_0	*	@(NOUN MONTH NOM PL).
teknele	X_0	*	@(NOUN BOAT NOM PL).
cabayla	X_0	*	@(NOUN EFFORT INS SG).
yurekten	X_0	*	@(NOUN HEART ABL SG).
merkeze	X_0	*	@(NOUN CENTER DAT SG).
eserleri	X_0	*	@(NOUN ARTWORK ACC PL).
yerde	X_0	*	@(NOUN PLACE LOC SG).

kurallara	X_0	0	@(NOUN RULE DAT PL).
"Postpositions"			
hakkinda	X_0	*	@(POSTP ABOUT HAKKINDA NOM).
ile	X_0	*	@(POSTP WITH ILE NOM).
boyunca	X_0	*	@(POSTP THROUGHOUT BOYUNCA NOM).
yonelik	X_0	*	@(POSTP TOWARDS YONELIK DAT).
ait	X_0	*	@(POSTP BELONG AIT DAT).
kadar	X_0	*	@(POSTP UNTIL KADAR DAT).
karsi	X_0	*	@(POSTP AGAINST KARSI DAT).
gore	X_0	*	@(POSTP ACCORDING_TO GORE DAT).
alakali	X_0	*	@(POSTP RELATED_TO ALAKALI INS).
dolayi	X_0	*	@(POSTP DUE_TO DOLAYI ABL).
uzerine	X_0	*	@(POSTP UPON UZERINE NOM).
birlikte	X_0	*	@(POSTP TOGETHER BIRLIKTE INS).
"Conjunctions"			
ve	CONJ	*	(~ CONJ-FORM) = AND.
veya	CONJ	*	(~ CONJ-FORM) = OR.
"Adjectives"			
sonraki	X_0	*	@(ADJ LATER).
eskimis	X_0	*	@(ADJ OBSOLESCEMENT).
antik	X_0	*	@(ADJ ANTIQUATED).
eski	X_0	*	@(ADJ OLD).
net	X_0	*	@(ADJ PLAIN).
guzel	X_0	*	@(ADJ NICE).
buyuk	X_0	*	@(ADJ BIG).
masraflı	X_0	*	@(ADJ COSTLY).
dogal	X_0	*	@(ADJ NATURAL).
yumusak	X_0	*	@(ADJ SOFT).
cetin	X_0	*	@(ADJ BITTER).
dogru	X_0	*	@(ADJ RIGHT).
klasik	X_0	*	@(ADJ CLASSICAL).
saygin	X_0	*	@(ADJ RENOWNED).
zengin	X_0	*	@(ADJ RICH).
zor	X_0	*	@(ADJ DIFFICULT).
siki	X_0	*	@(ADJ HARD).
"Adverbs"			
yuruyerek	X_0	*	@(ADV ON_FOOT).
gunlerce	X_0	*	@(ADV FOR_DAYS).
haftalarca	X_0	*	@(ADV FOR_WEEKS).
aylarca	X_0	*	@(ADV FOR_MONTHS).
yillarca	X_0	*	@(ADV FOR_YEARS).
profesyonelce	X_0	*	@(ADV PROFESSIONALLY).
uzunca	X_0	*	@(ADV LENGTHILY).
yorulmadan	X_0	*	@(ADV TIRELESSLY).

C.2 Test sentences

KONUS- (TO TALK) (args: PP_hakkinda | PP_uzerine | NP_acc | NP_nom)

PARSE-YES:

1. antik donemler hakkında ve klasik arabalar hakkında konustuk.
(PP_hakkinda & PP_hakkinda)
2. antik donemleri ve klasik arabalari konustuk.
(NP_acc & NP_acc)
3. antik donemler uzerine ve klasik arabalar hakkında konustular.
(PP_uzerine & PP_hakkinda)
4. antik donemler hakkında ve klasik arabalari konustuk.
(PP_hakkinda & NP_acc)

PARSE-NO:

5. antik donemler hakkında ve klasik arabalardan dolayi konustuk.
(PP_hakkinda & PP_dolayi)
6. antik donemleri ve klasik arabalara konustuk.
(NP_acc & NP_dat)
7. antik donemlere gore ve klasik arabalari konustuk.
(PP_according & NP_acc)
8. antik donemler uzerine ve klasik arabalara kadar konustuk.
(PP_uzerine & PP_kadar)

OL- (TO BE/BECOME) (args: PP | AdjP | NP)

PARSE-YES:

9. Bertan zengin bir memur veya saygin bir muhendis olacak.
(NP_nom & NP_nom)
10. film eski bir doneme ait veya arabalar hakkında olacak.
(PP_ait & PP_hakkinda)
11. firca dogal ve yumusak olmalı.
(AdjP & AdjP)
12. firca dogal kildan ve yumusak olmalı.
(NP_abl & AdjP)
13. konusma hedefe yonelik ve net olmalı.
(PP & AdjP)
14. konusma hedefe yonelik ve dogru yerde olmalı.
(PP_yonelik & NP_ins)

PARSE-NO:

15. zengin bir memur veya profesyonelce olmalı.
(NP_nom & AdvP)
16. film arabalar hakkında ve profesyonelce olmalı.
(PP_hakkinda & AdvP)
17. profesyonelce ve klasik olmalı.
(AdvP & AdjP)

SAYDIR- (TO SWEAR) (args: PP_hakkinda | NP_dat)

PARSE-YES:

18. Bertan soforler ve belediyeler hakkında saydirdi
(PP_hakkinda & PP_hakkinda)
19. soforlere ve belediyelere saydirdi.
(NP_dat & NP_dat)
20. Bertan uzunca soforler hakkında ve belediyelere saydirdi.
(PP_hakkinda & NP_dat)
21. belediyelere ve soforler hakkında saydirdi.
(NP_dat & PP_hakkinda)

PARSE-NO:

22. soforlerle alakali ve belediyeler hakkında saydirdi
(PP_alakali & PP_hakkinda)
23. soforleri ve belediyelere saydirdi.
(NP_acc & NP_dat)
24. soforler hakkında ve belediyeleri saydirdi.
(PP_hakkinda & NP_acc)
25. soforlerden ve belediyeler hakkında saydirdi.
(NP_abl & PP_hakkinda)

SUR- (TO LAST) (args: PP_boyunca | AdvP | NP_nom)

PARSE-YES:

26. cetin savaslar aylarca ve yillarca surduler.
(AdvP & AdvP)
27. cetin savaslar her hafta ve her yil surduler.
(NP_nom & NP_nom)
28. cetin savaslar donemler boyunca ve aylar boyunca surduler.
(PP_boyunca & PP_boyunca)
29. cetin savaslar her gun ve yillarca surduler.
(NP_nom & AdvP)
30. cetin savaslar her hafta ve yillar boyunca surduler.
(NP_nom & PP_boyunca)

31. cetin savaslar gunlerce ve yillar boyunca surduler.
(AdvP & PP_boyunca)

PARSE-NO:

32. cetin savaslar aylarca ve fakirlikten dolayi surdu.
(AdvP & PP_dolayi)

33. cetin savaslar her yil ve her haftadan surdu.
(NP_nom & NP_abl)

34. cetin savaslar donemler boyunca ve fakirlikten dolayi surdu.
(PP_boyunca & PP_dolayi)

35. cetin savaslar haftalara ve yillarca surdu.
(AdvP & NP_dat)

36. cetin savaslar her yil ve fakirlikten dolayi surdu.
(NP_nom & PP_dolayi)

37. cetin savaslar gunlerce ve haftalara gore surdu.
(AdvP & PP_gore)

VERBAL ADJUNCTS (elems: AdvP | NP | PP)

PARSE-YES:

38. yuruyerek ve yorulmadan merkeze ulasirsin.
(AdvP & AdvP)

39. soforlerle birlikte ve arabalar ile merkeze ulasirsin.
(PP_birlikte & PP_ile)

40. sevgiyle ve yurekten calisiyoruz.
(NP_ins & NP_abl)

41. belediyeyle birlikte ve profesyonelce calisiyoruz.
(PP_birlikte and AdvP)

42. antrenorle ve yorulmadan calisiyoruz.
(NP_ins & AdvP)

43. soforlerle birlikte ve arabayla merkeze ulasirsin.
(PP_birlikte and NP_ins)

PARSE-NO:

44. masraflı ve zor merkeze ulasirsin.
(AdjP & AdjP -- should be parsed as nominal modifiers)

45. zor ve kurallara gore calisiyoruz.
(AdjP & PP_gore)

46. antrenorle ve siki calisiyoruz.
(NP_ins & AdjP)

NOMINAL ADJUNCTS (elems: AdjP | PP)

PARSE-YES:

47. guzel ve antik enstrumanlari konustuk.
(AdjP & AdjP)

48. eski doneme ait ve sevgiyle alakali eserleri konustuk.
(PP_ait & PP_alakali)

49. antik doneme ait ve eskimis eserleri konustuk.
(PP_ait & AdjP)

PARSE-NO:

50. guzel ve profesyonelce eserleri konustuk.
(AdjP & AdvP)

51. eski doneme ait ve yillarca eserleri konustuk.
(PP_ait & AdvP)

COMPLEX SENTENCES (all types of coordination at the same time)

52. yorulmadan ve uzunca antik ve cetin donemler hakkında ve klasik ve guzel arabalari konustuk.

(Like v_adjuncts + Like n_adjuncts + Like n_adjuncts)
(PP_hakkinda + NP_acc)

53. arabada ve uzunca klasik doneme ait ve guzel arabalari ve antik ve enstrumanlarla alakali donemler uzerine konustuk.

(Unlike v_adjuncts + unlike n_adjuncts + unlike n_adjuncts)
(NP_acc & PP_uzerine)