DESIGN AND IMPLEMENTATION OF A VERB LEXICON AND VERB SENSE DISAMBIGUATOR FOR TURKISH

A THESIS

SUBMITTED TO THE DEPARTMENT OF COMPUTER

ENGINEERING AND INFORMATION SCIENCE

AND THE INSTITUTE OF ENGINEERING AND SCIENCE

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

by

Okan Yılmaz

September, 1994

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Asst. Prof. Kemal Oflazer (Advisor)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Asst. Prof. Halil Altay Güvenir

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Asst. Prof. Cem Bozşahin

Approved for the Institute of Engineering and Science:

Prof. Mehmet Baray
Director of the Institute

ABSTRACT

DESIGN AND IMPLEMENTATION OF A VERB LEXICON AND VERB SENSE DISAMBIGUATOR FOR TURKISH

Okan Yılmaz
M.S. in Computer Engineering and Information Science
Advisor: Asst. Prof. Kemal Oflazer
September, 1994

The lexicon has a crucial role in all natural language processing systems and has special importance in machine translation systems. This thesis presents the design and implementation of a verb lexicon and a verb sense disambiguator for Turkish. The lexicon contains only verbs because verbs encode events in sentences and play the most important role in natural language processing systems, especially in parsing (syntactic analyzing) and machine translation. The verb sense disambiguator uses the information stored in the verb lexicon that we developed. The main purpose of this tool is to disambiguate senses of verbs having several meanings, some of which are idiomatic. We also present a tool implemented in Lucid Common Lisp under X-Windows for adding, accessing, modifying, and removing entries of the lexicon, and a semantic concept ontology containing semantic features of commonly used Turkish nouns.

Keywords: Natural Language Processing, Machine Translation, Lexicon, Lexical Ambiguity, Ontology.

ÖZET

TÜRKÇE İÇİN
EYLEM SÖZLÜĞÜ
VE
EYLEM ANLAM ÇÖZÜMLEYİCİSİNİN
TASARIM VE GERÇEKLEŞTİRİLMESİ

Okan Yılmaz
Bilgisayar ve Enformatik Mühendisliği, Yüksek Lisans
Danışman: Yard. Doç. Dr. Kemal Oflazer
Eylül 1994

Bilgisayar sözlüğü özellikle bilgisayarlı çeviri gibi doğal dil işleme sistemlerinde önemli bir göreve sahiptir. Bu tezde biz türkçe için bir eylem belirleme sözlüğü ve eylem anlam çözümleyicisini tasarlayıp gerçekleştirdik. Eylemler olayları tümce içinde simgeleyip, özellikle sözdizimsel ayrıştırma ve bilgisayarlı çeviri gibi doğal dil işleme sistemlerinde en önemli göreve sahip olduklarından, sözlüğümüzü yanlızca eylemlerden oluşturduk. Eylem anlam çözümleyicimiz oluşturduğumuz eylem sözlüğündeki bilgileri kullanır. Bu uygulamanın temel amacı çok anlamlı ya da deyimsel anlamlar içeren eylemlerin anlam çözümlemesini yapmaktır. Bununla birlikte sözlüğe kayıt ekleme, kayıtlara erişme, kayıtları güncelleme ve silme görevini yapan *Lucid Common Lisp*'te *X-Windows* altında geliştirilmiş bir yazılım ve Türkçede çok kullanılan adların özelliklerini içeren bir bilgi yapısını da sunacağız.

Anahtar Sözcükler: Doğal dil işleme, bilgisayarlı çeviri, sözlük, sözcüksel çokanlamlılık, anlambilimsel bilgi yapısı.

ACKNOWLEDGMENTS

I would like to thank my advisor Asst. Prof. Kemal Oflazer who has provided a stimulating research environment and motivating support during my M.S. study.

I would also like to thank Asst. Prof. Halil Altay Güvenir and Asst. Prof. Cem Bozşahin for their valuable comments on this thesis.

Finally, I would like to specially thank to Ş. Türkol for her technical support, motivation and hope-giving participations during long sleepless nights. I would also like to thank my colleagues S. Hüsrevoğlu, M. Surav, S. Çil, B. Yalçıngediz, Y. Okur, and Y. Kara for their intellectual support.

I dedicate this thesis to my family.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Natural Language Processing (NLP) is a science and engineering discipline that aims to build systems for processing natural human languages for a variety of applications such as machine translation, spelling correction, etc. Most common components of NLP applications are:

- syntactic analysis,
- semantic analysis,
- generation,
- transfer component of machine translation.

Verbs play the most important role in all of these processes. In syntactic analysis, argument structures of the sentences depend on the sense of the verb.

(1) a. Birini geçirmek
to say goodbye to someone

b. Birseyi bir yerden bir yere geçirmek
to pass something from somewhere to somewhere

For example, in (1a), *geçirmek* is used in the sense *to say goodbye*, and in this sense the object must be in accusative or nominative case. Furthermore, in

VERB	# SENSES	VERB	# SENSES	VERB	# SENSES
çık	40	bak	13	an	10
at	32	bağla	12	bul	10
geç	27	in	12	dayan	9
al	21	gir	11	kır	8
gel	20	gör	11	kaldır	8
bırak	18	git	11	dokun	8
kaç	13	kur	10	dağıt	8

Table 1.1. Verbs with greatest number of senses in the Lexicon

(1b), the verb *geçirmek* is used in a totally different sense and argument structure. Although, in (1a) *geçirmek* subcategorizes objects in dative and ablative cases, this is not grammatical in (1b). In Turkish, just like any other language, verbs often have several meanings and most of them become idiomatic when they are used with special objects or subjects. Table 1.1 lists the verbs in Turkish having relatively more senses as given in *Türk Dil Kurumu Dictionary*. Since quite common verbs have a large number of meanings, the verb sense disambiguation process becomes an important step in machine translation between Turkish and other languages. The variations in the senses of verbs assign a crucial role to sense disambiguation process. For example, the two senses of *yemek* are totally different in (2).

(2) a. yemek
to eat

b. Parayı yemek
to spend money

In the analysis of a natural language text, we deal with ambiguous interpretations of words and sentences. Disambiguation is the process of resolving the lexical and syntactic (structural) ambiguities. In lexical ambiguities one word can be interpreted in more than one way. In NLP, there are three types of lexical ambiguity: *polysemy*, *homonymy*, and *categorical ambiguity*.

- *Polysemous words* have several meanings that are related or close to each other. For example, the Turkish verb *almak* has many senses concerning *taking with*, *getting*, *buying* and so on.

- *Homonymous words* have several meanings that have no obvious relationship one to another. For instance, in (1a) and (1b) the Turkish verb *geçirmek* has senses concerning *to say goodbye* and *to pass* with no obvious relationship.

- *Categorically ambiguous words* are those who have multiple syntactic categories. For example, the Turkish word *at* can be a noun meaning *horse* or a verb meaning *to throw*. Clearly, categorical ambiguity is orthogonal to the other types and is mainly a problem in parsing. Note that in the case of Turkish the morphotactical and syntactic restrictions help resolve such ambiguities in many cases.

In this thesis, we deal with the resolution of the senses of polysemy and homonymy verbs in Turkish. The categorical ambiguity of the words are assumed to be resolved in syntactic and morphological processing steps (although this may not always be possible). We present the design and implementation of a verb lexicon for sense resolution of verbs in Turkish using morphological, syntactic, and semantic information available in the context of the verb. In the lexicon, all senses of verbs are stored in the same entry and a two-level semantic network is used for disambiguation. Verb senses are determined by testing semantic, syntactic and morphological constraints defined for arguments of the verbs. A tool has been implemented using *Lucid Common Lisp* (LCL) under *X-Windows*. The system has been developed in object-oriented programming style and for this purpose *Common Lisp Object System (CLOS)* is used. A semantic concept hierarchy has also been developed using the facilities of *LOOM* [1]. A noun lexicon containing semantic features of commonly used nouns is developed and inserted into *LOOM* as instances.

The outline of the thesis is as follows: A general overview of the concept of a lexicon, and related work is covered in Chapter 2. The semantic structure of Turkish language and the lexicon that has been developed for Turkish are described in Chapter 3. In Chapter 4, the sense disambiguation process and the structure of our ontological database are described. Chapter 5 contains the description of the verb entry and sense disambiguation tool, and sample runs.

We then conclude this work and give suggestions for future directions in the last chapter. The appendices present concept ontology and the list of Turkish verbs covered in the lexicon.

Chapter 2

The Lexicon

A lexicon is a collection of lexical units of a language with information about their morphological, syntactic, and semantic properties relevant to the processing involved. The lexicon has a very important role in all natural language processing systems, and most importantly, in machine translation (MT) systems.

In this chapter, we discuss the concept and the role of the lexicon in natural language processing, mainly in parsing and machine translation. We will first go over the concept of lexicon, explain the function of lexicon in the parsing process. We will then present a brief overview of machine translation systems, and then discuss the role of the lexicon in MT. Finally, the lexicon of the DIANA (a DIstributed ANAlysis System) semantic analysis system [8] will be illustrated as an example.

2.1 Lexicon

A lexicon of a natural language lists the lexical items occuring in the language. In a typical traditional dictionary, entries are identified by a base ('canonical') form of the word. This sometimes (though not always) corresponds to the uninflected root (as in English). In French dictionaries, for example, verbs are listed under one of their inflected form (usually the infinitive, e.g., *manger* [6]). In Latin dictionaries, nouns are given in the nominative singular (e.g., *equus*),

and verbs in the 1^{st} person singular present tense active voice (e.g., *habeo*). Traditional dictionary entries indicate pronunciations, give grammatical categories, provide definitions, and supply etymological and stylistic information.

The lexicon in a NLP system is substantially different from the lexicon in typical daily or linguistic usage. For some languages, an NLP system has *full-form* lexicons which lists the words as they actually occur, with corresponding grammatical information. Thus, for example, the lexicon might separately lists the words *play, plays, playing.* However, this is not at all attractive for agglutinative languages like Turkish, since these languages have very productive morphology and each lexical root may give rise to hundreds or thousands of forms. As an example from Turkish, *gel* (*to come*) has many forms: *gel* (*come* (*imperative*)), *geliyorum* (*I am coming*), *geliyorsun* (*you are coming*), *gelir* (*he/she/it comes*), *gelecekler* (*they will come*), *geliyorken* (*while they are coming*), etc.

2.2 The Function of Lexicon in Syntactic Analysis

A major component of any NLP system is the *parsing* or syntactic analysis component, which takes a *grammar* (a set of rules which describe the acceptable combinations and sequences of words that are acceptable) and a lexicon as data, and a text (e.g. sentence) as input, produces an analysis of the structure of the text as output. Grammars for natural language usually express structures of well-formed strings by derivation rules annotated with feature constraints. The role of lexicon in a parser is to maintain the information about the features associated with individual lexical items. In fact, most systems have a great number of lexical entries and very few general rules, relying extensively on the lexicon.

Here we give a very simple example of the usage of a lexicon in parsing from the Lexical Functional Grammar (LFG) parser developed for Turkish by Güngördü [4]. Although the lexical entries used in this system were very simple, they nevertheless illustrate the role of a verb lexicon in a parser. In the verb lexicon, argument structures of each of senses of verbs are stored. Along with the objects, an entry which contains one or more senses of the verb are kept for each verb. An explanation of the meaning and the objects to be taken are

```
(``al''
  (SENS ((``to take''
            (ARGS (((*CASE* (NOM ACC))
                    (*TYPE* DIRECT)
                    (*OCC* OBLIGATORY)
                    (*ROLE* THEME))
                   ((*CASE* ABL)
                    (*TYPE* INDIRECT)
                    (*OCC* OPTIONAL)
                    (*ROLE* SOURCE)))))))
```

Figure 2.1. The argument structure of *al* in the verb lexicon of LFG parser for Turkish.

indicated for each sense. An object is specified by its case (e.g. NOMinative, ACCusative, etc.), type (i.e. direct, indirect or oblique), thematic role (deep case relation) (see Section 3.1.1), and a flag which indicates whether the verb optionally or obligatorily subcategorizes for the object.

The argument structure of the verb *almak* (*take*) is illustrated in Figure 2.1. It obligatorily subcategorizes for a nominative or accusative marked direct object, and optionally subcategorizes for an ablative marked indirect object. The thematic roles of a direct object is theme and that of the indirect object is source. For example, in (3) where *kitap* (*book*) is the direct object and *masa* (*table*) is the indirect object.

(3) Ben kitabı masadan aldım.
I book+ACC table+ABL take+PAST+1SG.
I took the book from the table.

By using the output of morphological analyzer and argument structures kept in the lexicon for verbs, the analysis process determines whether a sentence is grammatical or not. For example, (4a) is determined as grammatical, although (4b) is not. The lexicon can be used to resolve ambiguous outputs of the parser. For instance, the predicate of (5) may be *kalın* or *kal*. This ambiguity can be resolved by comparing the argument structures of these predicates against the lexicon.

(4)

a. Kalemi aldım.
pencil+ACC take+PAST+1SG
I took the pencil.

b. ? Kalemde aldım.
pencil+LOC take+PAST+1SG
? I took at the pencil.

(5) O gece evde kalındı.
(they) stayed at the home at that night.
that night home+LOC stay+PASS+PAST
or
? that night home+LOC thick+PAST

Note that the second interpretation of (5) is semantically nonsense.

2.3 The Role of Lexicon in Machine Translation

Machine Translation (MT) is the traditional and standard name for computer systems responsible for the production of translations from one natural language into another, with or without human assistance. There are three basic MT strategies, namely *direct method, transfer method*, and *interlingua method.* The oldest one is the direct approach adopted by most MT systems that have come to be known as the *first generation* MT systems. The inadequate results of this strategy have led to the development of the transfer-based and interlingua-based approaches. This kind of systems are sometimes referred to as *second generation* systems. The basic differences of these strategies lie under their approaches to the three components of the translation process: *analysis, transfer* and *generation.* Figure 2.2 illustrates the differences among these approaches.

- The *direct approach* has no intermediate stage in translation process. In systems that use this approach, the input text is directly translated to

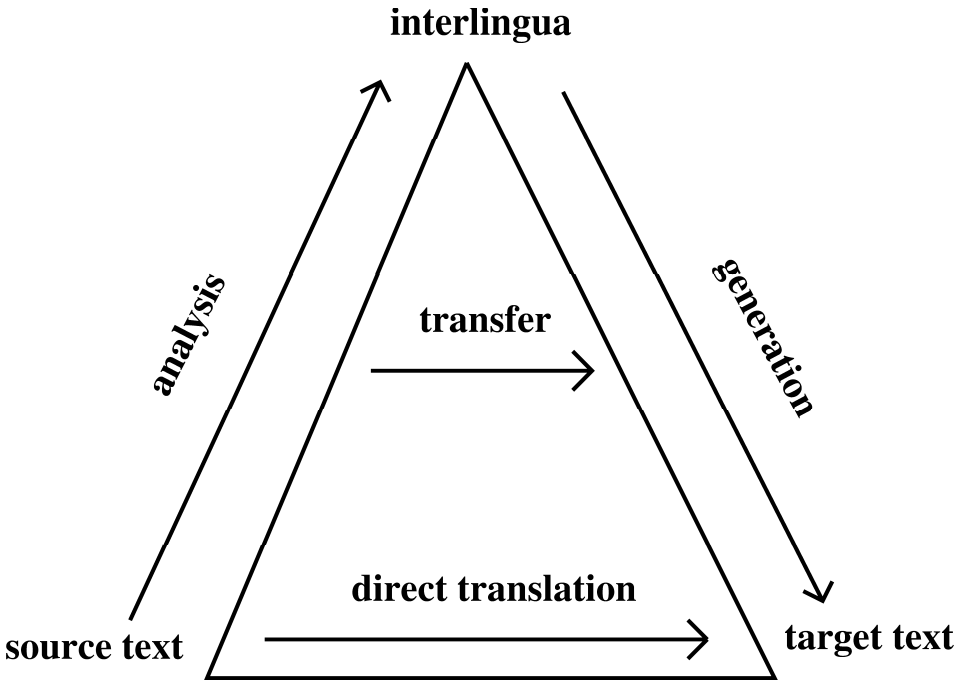

Figure 2.2. Transfer and interlingua 'Pyramid' diagram

the desired target language output text almost word by word with certain structural change.

- The *interlingua-based approach* consists of two steps. In the first step, the source text is analyzed and translated into an intermediate representation. In the second step, the target text is generated from the intermediate representation without referring to the original text. The strict separation of the analysis and generation is a disadvantage due to two reasons: i) The analysis process can not be oriented towards a particular target language. ii) It is not desirable to orient the generation process by looking back at the original source language text. The interlingua representation must include all the information necessary in the course of the generation of any target language text. In effect, this high degree of language-independence and neutrality means that interlingua must be striven towards universality in lexicon and structure.

- In the *transfer method*, the source text is analyzed and an abstract representation of the source text is generated. This intermediate representation is converted into abstract representation of the target language by transfer modules. Finally, the target text is generated from the abstract representation of the target language.

The analysis and generation processes rely heavily on lexicons. Transfer-based MT systems use bilingual transfer lexicons, in which the translation components from lexical units of the source language into lexical units of the target language are listed. In some MT systems using the interlingua approach, two monolingual lexicons can be used: one for analysis and the other for generation. All the lexicons for analysis contain morphological, syntactic, semantic and pragmatic information about the lexical entry. On the other hand, generation lexicons support text planning, including lexical selection and realization in generation.

Lexicons are also used in sense disambiguation process of MT systems. Sense disambiguators resolve ambiguities by using the information stored in lexicon. In the verb sense disambiguation process the syntactic, semantic, and morphological features of the arguments of a verb are used as constraints. The correct sense is determined when the constraints of the arguments of verbs are satisfied.

2.4 An Example Lexicon

In NLP systems various type of lexicons are used. In many systems, more than one lexicons are used for analysis and generation. For example, ULTRA [2] uses three lexicons, the intermediate representation lexicon, the Spanish lexicon for analysis, and the English lexicon used by the generator, Penman [7]. In the Spanish lexicon, nouns and pronouns are stored in an entry having five components. These components are the lexical item, person and gender information, case information and corresponding interlingua token. Verbs as well as adjectives are represented in ten tuples. These ten fields indicate the lexical token, whether the verb is stative or dynamic, agreement information, information on tense, aspect, mood, and voice as well as the corresponding interlingual token. The intermediate representation contains nouns and verbs. The fields of a noun entry encode a semantic category, whether the noun is proper or common, whether it is mass or count. The fields of a verb and adjective mark the sense token, whether the sense is dynamic or stative, a semantic classification for the verb, the semantic roles of its arguments, and the semantic classification of the entities filling those roles. The English generation lexicon contains entries for Penman.

An other dictionary example is IPAL [9] developed for verbs in Japanese. In this dictionary, case frames for 861 typical Japanese verbs are stored. For each Japanese verb, surface cases, some semantic markers and several typical example sentences are given in each case slot.

In this section, we will illustrate the structure of an analysis lexicon developed for DIANA natural language analysis system. This lexicon has been developed at Carnegie Mellon University [8] and designed for analysis of English texts. In this process, both semantic and pragmatic concerns have been taken into account. As a result of this analysis, an interlingua text (ILT) is generated in a specially designed text meaning language TAMERLAN [11]. Even though the former lexicon is developed for analysis purposes, the knowledge about language and meaning represented are considerably independent of processing considerations. This methodology allows the use of the lexicon for both analysis and generation.

The lexicon is a set of `SUPERENTRIES` which are the basic units. Each

```
LEXICON ENTRY: SMELL (SUPERENTRY INDEX)
; shown here is the index to the superentry ``smell'' followed
; by entry for smell-v1.

; INDEX TO SUPERENTRY ``SMELL''

; v1 DEF use olfactory sense voluntarily
;     EX Here... smell this liquid
; v2 DEF use olfactory sense involuntarily
;     EX I smell garlic
; v3 DEF emit gases that one can smell-v1/v2
;     EX The flower smells sweet
; v4 DEF smell-v3 in an unpleasant way
;     EX UGH!! Fred smells!
; v5 DEF to perceive something negative intuitively
;     EX I could smell trouble brewing
; v6 DEF to give a negative impression
;     EX The whole thing smells fishy to me


; n1 DEF physiological sense of perceiving with the nose
;     EX Because of this cold, my sense of smell is gone today
; n2 DEF voluntary use of olfactory sense
;     EX Have a smell of this wine
; n3 DEF attribute perceived by one smell-n1/n2
;     EX Delicious smell of fresh-brewed coffee
; n4 DEF impression, aura, feel, quality...
;     EX Everything he does has the smell of success about it
```

Figure 2.3. Index to the superentry *SMELL*

```
LEXICON ENTRY: SMELL-v1
(smell
    (make-frame-old
        +smell-v1
        (CAT (value v))
        (STUFF
            (DEFN ``use olfactory sense voluntarily'')
            (EXAMPLES ``smell this liquid...what do you
                        think it is?'')
            (TIME-STAMP ``ingrid feb 12 90'')
        )
        (MORPH
            (IRREG (*v+past* smelt optional)
                   (*v+past-part* smelt optional)
            )
        )
        (SYN)
        (SYN-STRUCT
            (LOCAL
                ((root $var0)
                 (subj ((root $var1) (CAT n))
                 (obj  ((root $var2 optional) (CAT n)))
                )
            )
        )
        (SEM
            (LEX-MAP
                (%voluntary-olfactory-event
                    (AGENT (value ^$var1)
                        (SEM (*OR* *mammal *bird
                                   *reptile *amphibian))
                    ; only classes of animals that have
                    ; an olfactory organ
                    ; (e.g. not ?*fish, ?*protozoan)
                    (THEME (value ^$var2)
                       (SEM *physical-object)
                    )
                    (INSTRUMENT (SEM *olfactory-organ))))))))
```

Figure 2.4. Entry for *SMELL*

```
(LEXICON
      (SUPERENTRY 1                        ; headword 1
            (make-frame +ENTRY-x1 ...)     ; (cat x, sense 1)
            (make-frame +ENTRY-x2 ...)     ; (cat x, sense 2)
            (make-frame +ENTRY-y1 ...)     ; (cat y, sense 1)
            (make-frame +ENTRY-y2 ...)     ; (cat y, sense 2)
      (SUPERENTRY 2                        ; headword 2
             etc ... ))
```

Figure 2.5. The structure of the lexicon

SUPERENTRY has a HEADWORD and a list of ENTRIES. This list comprises one or more ENTRIES, each having a unique identifier called LEXEME and denoting different grammatical categories or senses of the lexeme. For the superentries, having more than one entry, a superentry index, e.g., a list of the various lexemes, each with an abbreviated definition is given along with a short example. Index to the superentry "smell" and entry for smell-v1 (the first verb sense of *smell*) are illustrated in Figures 2.3 and 2.4, respectively.

2.4.1 The Structure of an Entry

In the lexicon of the DIANA system, each entry is a frame identified by a lexeme which is a headword symbol preceded by '+', plus an indicator of grammatical category, plus a numerical index, e.g., +smell-v1, +smell-n1. The structure of the lexicon is summarized in Figure 2.5.

Each entry has at most ten zones, corresponding to a slot in the entry frame. These zones and corresponding slots are:

1. *the grammatical category zone*, represented as the CAT slot, denotes grammatical category of the lexeme.

2. *the user information zone*, represented as the STUFF slot, contains information for the human user. The information consists of one or more definitions for the verb sense, examples, and some administrative data.

3. *the orthography zone*, represented as the `ORTH` slot, stores acceptable orthographic variants and accepted abbreviations of the lexeme.

4. *the phonology zone*, represented as the `PHON` slot, is used when the phonology of a word form is not entirely predictable from the orthography.

5. *the morphology zone*, represented as the `MORPH` slot, contains irregular forms, stem variants, and formation paradigms of the lexeme. This zone is needed for languages where each word has a very small number of morphologically inflected form.

6. *the syntactic feature zone*, represented as the `SYN` slot, contains the syntactic features of the lexeme. For example, the information which shows the lexeme in category noun is countable is stored in this zone.

7. *the syntactic structure zone*, represented as the `SYN-STRUCT` slot, contains a Lexical-Functional Grammar like argument structure of associated lexeme.

8. *the semantic zone*, represented as the SEM slot, containing a declarative specifications of meaning through a mapping to the ontology or a mapping directly into interlingua structures or a combination of both.

9. *the lexical relations zone*, represented as the `LEXICAL-RELATIONS` slot, is designed to show various kinds of relations between word senses.

10. *the pragmatics zone*, represented as the `PRAGM` slot, contains pragmatic information about the lexeme.

Figure 2.4 illustrates the structure of an entry. `SMELL-v1` denotes that this is the first entry of *smell* in the grammatical category of verb. This is also stated in `CAT` slot. In the `STUFF` zone the meaning of *smell* is defined as *use olfactory sense voluntarily.* An example and the entry date are given in this slot, too. Since *smell* is an irregular verb, its past and past-participle forms are stored as morphological features. No syntactic feature is stated. The argument structure of *smell* is specified in the `SYN-STRUCT` zone. The arguments of *smell* are a subject and an `optional` object. The category of both the subject and the object is noun. The `LEX-MAP` slot of the `SEM` zone contains the detailed semantic information to reference the ontology used. The above lexical

mapping says that the given sense of *smell* is mapped in TAMERLAN as an instance of the %voluntary-olfactory-event ontological concept. Moreover, the semantic interpretation of whatever occupied the `subj` position in f-structure should be assigned as the value of the `AGENT` thematic role. The `SEM` zone of the `AGENT` slot denotes that this argument should be a `mammal`, a `bird`, a `reptile`, or an `amphibian`. The `THEME` slot states that the meaning of whatever occupied the `obj` position in the f-structure should be assigned as the value of the theme thematic role. In the `SEM` slot the `THEME` of the sentence is specified as `physical-object`. The `INSTRUMENT` slot specifies the `INSTRUMENT` of the sentence as an `olfactory-organ`.

In DIANA, an entry is kept for each sense of the verb. This causes data repetition for homonymous words and verbs having idiomatic senses. Another storage problem arises while storing words having so many senses, because a different entry is generated for each of them. Moreover, morphological constraints are not considered in this design. Since Turkish verbs have so many senses and some of those meanings are idiomatic and since morphological constraints have an important role in NLP systems for agglutinative languages like Turkish, the structure of this lexicon is not suitable for Turkish.

Chapter 3

A Verb Lexicon for Turkish

In the syntactic and semantic analysis of a sentence, verbs play the most important role. Almost all Turkish verbs have several meanings some of which are idiomatic. For instance, the verb *gelmek* has 20 different senses (see Figures 3.1 [1] and 3.2). This assigns an important role to verb sense disambiguation step in the analysis process. In Turkish language, semantic roles of subject and objects of a sentence must be well understood in order to determine the semantic information that is to be included in a verb lexicon. In this chapter, we will present the structure of the verb lexicon developed for Turkish. First, we will study *thematic roles* (also called *deep case relations, semantic cases, semantic roles, thematic relations*, and *theta roles*) which are semantic relations connecting entries to events/processes/states denoted by verbs. We will then study semantic categories of Turkish verbs and relationship between grammatical relations and thematic roles. Later, the structure of the lexicon will be illustrated. Finally, we will present the usage of the lexicon in parsing and machine translation.

3.1 Semantic Analysis of Thematic Roles in Turkish

Not only the grammatical relations but also the thematic roles and surface case marking play an important role in the analysis process of natural languages. There have been many studies about the thematic roles (e.g., for English [3]).

[1] Idiomatic senses of *gelmek* are also given

- **Sense:** to feel
  Example: Uykum geldi.
  (I feel sleepy - lit. My sleep came.)
- **Sense:** to be bored
  Example: Gına geldi artık.
  (I got bored.)
- **Sense:** to weigh
  Example: Adam 80 kilo geliyormuş.
  (The man weighs 80 kilo.)
- **Sense:** to affect in a negative way
  Example: Kurşun koluma geldi.
  (The bullet hurt my arm.)
- **Sense:** to survive
  Example: Günümüze birçok anıt geldi.
  (So many monuments survive today.)
- **Sense:** to be
  Example: Saat sabahın 8'ine geldi.
  (It is 8 in the morning.)
- **Sense:** to come to
  Example: Adam ana konuya gelemedi.
  (The man couldn't come to the main topic.)
- **Sense:** to stand
  Example: Çocuk soğuğa gelemez.
  (The child can not stand the cold.)
- **Sense:** to accept
  Example: Bu adam hiç şakaya gelmez.
  (That man never takes joke.)

Figure 3.1. Senses of the verb *gelmek*.

- **Sense:** to understand
  Example: Sonunda dediğime geldiniz.
  (Finally, you understood what I said.)
- **Sense:** to fit
  Example: Ayakkabı ayağıma geldi.
  (The shoe fit my foot.)
- **Sense:** to seem
  Example: Yalan gibi geliyor.
  (It seems to be a lie.)
- **Sense:** to cost
  Example: Bardakların tanesi 10000 liraya geliyor.
  (Each of the glasses costs 10000 liras.)
- **Sense:** to occur
  Example: Bu evde bir patlama meydana gelmiş.
  (An explosion has been occured at this house.)
- **Sense:** to be remembered
  Example: Hatırıma gelmedi.
  (I did not remember - lit. It did not came to my memory.)
- **Sense:** to be deceived
  Example: Oyuna geldiler.
  (They were deceived - lit. They came to a trick.)
- **Sense:** to result from
  Example: Bütün güzelliği topraktan geliyor.
  (All its beauty comes from the soil.)
- **Sense:** to act as if
  Example: Görmezlikten geldiler.
  (They acted as if they did not see.)
- **Sense:** to be the first, to come first
  Example: Adam bu yarışta da başta geldi.
  (He was the first in this race, too.)
- **Sense:** to come from/to
  Example: Babam okuldan eve gelmiş.
  (My father has come home from school.)

Figure 3.2. Senses of the verb *gelmek* continued.

In some of these studies, the number of thematic roles have been quoted as from 18 to 25 for English and 33 for Japanese [10]. Yalçın [14] specifies seven basic deep case relations for Turkish but certainly these can be extended with a finer resolution of the roles. According to Yalçın, the thematic roles used as the obligatory ones are *agent*, *patient*, *experiencer*, *beneficiary*, *complement*, *location*, and the optional one is *instrument*. In this study, we extended these roles by adding *value-designator*, and subdividing *patient*, and *location* in three groups. The subcategories of *patient* are *patient*, *theme*, *recipient*, and *location* are *location*, *source*, and *goal*.

In the following sections we will present twelve thematic roles:

1. agent,
2. patient,
3. theme,
4. experiencer,
5. beneficiary,
6. recipient,
7. source,
8. goal,
9. location,
10. instrument,
11. complement, and
12. value-designator.

We will categorize Turkish verbs in sixteen groups:

1. state verbs,
2. process verbs,
3. action verbs,
4. process-action verbs,
5. state-experiential verbs,
6. process-experiential verbs,
7. action-experiential verbs,
8. state-benefactive verbs,
9. process-benefactive verbs,

10. process-action-benefactive verbs,
11. state-completable verbs,
12. action-completable verbs,
13. state-locative verbs,
14. process-locative verbs,
15. action-locative verbs, and
16. action-process-locative verbs.

Finally, we will study the relationship between grammatical objects and thematic roles.

3.1.1 Thematic Roles in Turkish

In Turkish, the noun phrases (NPs) and sometimes post-positional phrases (PPs) function as thematic role fillers: For example, sometimes the subject (*babası* (*his father*)) is an agent, the direct object (*o* (*he*)) is a patient, and the action is performed by using an instrument (*sopa* (*stick*)) as in (6).

(6) Babası onu sopayla dövmüştü.
His father had beaten him with a stick.

Thematic roles in Turkish are as follows:

- **Agent**

  According to Frawley, the agent is the *deliberate*, *potent*, *active* instigator of the predicate: the primary, involved doer [3]. The verb categories which involve an action require the occurrence of agent along with the other deep case relations. Agents are typically animate and agency is often connected with *volition*, *will*, *intentionality*, and *responsibility*. The following sentences illustrate the agency:

  (7) a. Hakan kitabı dört günde okudu.
  Hakan read the book in four days.

b. Kedimiz sonunda eve döndü.
Our cat finally returned home.

c. Adam o akşam polis tarafından yakalandı.
The man was caught by the police at that night.

In (7a) and (7b), *Hakan* and *kedimiz* (*our cat*) stand for the agent because they take the action willingly and intentionally. In (7c), even though *polis* is not the subject, they take the action, and hence stand for the agent. In general, agents are in nominal case when they are subjects and argument to a specific PP (in (7c) post-positional form = *taraf* + POSS + ABL) in passive sentences.

- **Theme**

Let us consider the following sentences:

(8) a. Buz eridi.
The ice melted.

b. Oyun bitti.
The game is over.

In (8a) and (8b), *buz* (*ice*) and *oyun* (*game*) stand for the themes, because they do not perform any action or are not directly affected by the agent of any action. Also in (7a), *kitap* (*book*) is not directly affected by the action of *Hakan* and there is no change of shape or state as the result of the action. Therefore, *kitap* (*book*) in (7a) also stands for the theme.

- **Patient**

In some cases, an argument which can be a direct object or a subject is changed by or directly affected by a predicate. That argument is called as the patient. The patient suffers from the situation or comes out changed as a result of the action of the predicate. In examples (9) *araba* (*car*), *karlar* (*snow*), and *kuş* (*bird*) stand for the patient.

(9) a. Babam arabasını yıkadı.
My father cleaned his car.

b. Güneş karları eritti.
The sun melted the snow.

c. Kuş çocuklar tarafından vuruldu.
The bird was shot by the children.

- **Experiencer**

Let us consider the following sentences:

(10) a. O tuhaf kokuyu ben de duydum.
I smelled that strange odor, too.

b. Kötü haber beni üzdü.
Bad news upset me.

In (10a), *ben* (*I*) is mentally disposed by a mental experience and *ben*'s mental process is effected by 'bad news' in (10b). When someone is disposed in some way just like *ben* (*I*) in (10a) and (10b), it is called as the experiencer of the predicate.

- **Beneficiary**

In (11), *ben* (appearing in dative form *bana*) benefits from others' help. The person benefiting from a state or an action is the beneficiary of the predicate.

(11) Lütfen bana yardım edin!
Please help me!

- **Recipient**

Generally recipients have an animate nature and actually are receivers of physical objects; for example, in (12) *ben* is the receiver of *kitap* (*book*) and named as the recipient of *vermek* (*to give*).

(12) Kitabı bana verir misin?
Could you give me the book?

- **Source**

Let us consider the following sentences:

(13) a. Ben kediyi kasaptan evime getirdim.
I brought the cat home from the butcher.

b. Ben bu kitabı Yavuz'dan aldım.
I took this book from Yavuz.

(13a) and (13b) represent a displacement of *kedi* (*cat*) and *kitap* (*book*), respectively, and *kasap* (*butcher*) and *Yavuz* indicate the points of the origin of the displacements. The arguments such as *kasap* in (13a) and *Yavuz* in (13b) state the source of the predicate.

- **Goal**

  Goal represents the destination of the displacement. In (13a), *ev* (*home*) is the destination of the indicated displacement and the goal of the predicate *getirmek* (*to bring*). However, we classify *ben* (*I*) in (13b) as the recipient instead of a goal.

- **Location**

  Let us consider the following sentence.

  (14) Kedi şimdi evde uyuyor.
  The cat is sleeping at home now.

  The thematic role of arguments which denote spatial position of the predicate is location. Since in (14) *ev* (*home*) is the spatial position of *uyumak* (*to sleep*), it is the location.

- **Instrument**

  (15) Sinan saçlarını saç kurutma makinasıyla kuruttu.
  Sinan dried his hair with the hair dryer.

  According to Frawley, if an argument describes the means by which a predicate is carried out, it has the thematic role of instrument [3], i.e. the action is taken by using an instrument. In (15) *Sinan* takes the action, *saç kurutmak* (*hair drying*), by using a device *saç kurutma makinası* (*hair dryer*), so that *saç kurutma makinası* has the thematic role of instrument. These arguments are sometimes marked with the instrumental postclitic *-(y)le/ile* (*with*). They may also be followed by a noun *vasıtasıyla* (*by means of*) or *sayesinde* (*due to*).

- **Value-Designator**

  Most verbs can be used with a value marker. A special thematic role value-designator is used when an action is taken for some money, or the action costs that much money. In (16a) and (16b) *8.000.000 lira* (*8,000,000 liras*) and *10 dolar* (*10 dollars*) are value-designators.

  (16) a. O evde 8.000.000 liraya oturuyorlarmış.
  They live in that apartment for 8,000,000 liras.

  b. Oralarda 10 dolar için adam öldürürler.
  They kill people for 10 dollars there.

In Turkish, the argument structures of a verb depends on its senses. For example, in (17a) *götür* (*to take from somewhere to somewhere*) is used with all arguments it subcategorizes for, but in (17b), it is used in the sense *to take away*. In (17a), *otobüs* (*bus*) is the instrument and *10 lira* (*10 liras*) is the value-designator of *götür*. Almost all the Turkish verbs can be accompanied by a *value-designator* and an *instrument*.

(17) a. Ben seni evden okula otobüsle 10 liraya götürdüm.
I took you from home to school by bus for 10 liras.

b. Adam arabayı götürdü.
The man took the car away.

3.1.2 Verb Categories in Turkish

When we semantically analyze Turkish verbs, we see that their semantic structures are very different. For example, in (18a), there is an *action* taken by someone. However, when we analyze (18b) and (18c), we see no *action* is taken, because *adam* (*the man*) is not really doing anything. In (18b) and (18c), a state and a process are denoted by the predicate of the sentences.

(18) a. Adam öldürüldü.
The man was killed.

b. Adam ölü.
The man is dead.

c. Adam ölüyor.
The man is dying.

Turkish verbs can be categorized in three basic groups: *state, process*, and *action*; also in sixteen subgroups according to their accompanying subject and objects [14]. These groups are:

- **State**

  Consider the following examples:

  (19) a. Demet çok akıllı.
  Demet is very smart.

  b. Su 15 dakikada kaynadı.
  Water boiled in 15 minutes.

  c. Hakan çok okur.
  Hakan reads a lot.

  d. Yıldız hanım bulaşıkları yıkadı.
  Mrs. Yıldız washed the dishes.

  In (19a) the noun *Demet* is in a certain state or condition which is *akıllı* (*smart*). Here the verb is indicated as state and the subject as its theme, i.e. *the theme specifies what/who is in that state.* Such state predicates have mostly simple adjectives like *iyi* (*good*), *kötü* (*bad*), *sıcak* (*hot*), *çok* (*many*), fazla (*excessive*), etc. The verbs in the remaining sentences, (19b), (19c) and (19d) are not specified as states. Non-states can be distinguished from states by asking the questions "What happened?", "What is happening?". There is another test, called the progressive form test. In many cases, a non-state can occur in the progressive form which is unavailable to a state. In (20b), (20c), and (20d), the non-states in (19b), (19c), and (19d) occur in progressive form. Since the predicate of (19a) denotes a state, its progressive form in (20a) is not grammatical.

(20) a. * Demet çok akıllıyor.

b. Su kaynıyor.
Water is boiling.

c. Hakan çok okuyor.
Hakan reads a lot.

d. Yıldız hanım bulaşıkları yıkıyor.
Mrs. Yıldız is washing the dishes.

- **Process**

  In (20b) the subject *su* (*water*) changes its state from *not boiled* to *boiled.* The verbs such as *kaynamak* (*to boil*), *donmak* (*to freeze*), *pişmek* (*to cook* (*of food*)), *solmak* (*to discolor*), *erimek* (*to melt*), etc. are categorized as process verbs. This kind of verbs express the change in the state of the accompanying subject. Since a process involves a relation between the noun, which is the subject of the sentence, and a state, the subject is still the theme of the verb.

- **Action**

  The role of the verb in (19c) is different from those of (19a) and (19b). In (19c), there is no state, or change of state, instead, an activity or an action taken by someone is expressed, i.e., *Hakan* does the activity reading. Examples of this kind of verbs are *koşmak* (*to run*), *ötmek* (*to chirp*), *okumak* (*to read*), *yatmak* (*to lie*), etc.

  In order to distinguish an action from a process or a state, the question "What did X do?", where X is the subject of the sentence, can be asked. This question can be answered in action sentences, but not in process or state sentences. For example, the following questions can be asked for (19a), (19b), and (19c), respectively.

What did Hakan do?	he read.

However, the questions below can not be answered.

What did water do?	no answer

What did Demet do?	no answer

On the contrary, process (but not the state or action) sentences answer the question "What happened to X?". In the following sentences, these questions are asked to the sentences of Example (19). The action sentence (19c) and the state sentence (19a) do not answer this question, though the process sentence (19b) does.

What happened to Hakan?	no answer
What happened to the water?	It boiled.
What happened to Demet?	no answer

Since the subject of an action sentence specifies something which is neither in some state nor changing its state, it is no longer the theme. Thus, states and processes are accompanied by themes while actions accompanied by agents.

- **Process-Action**

  Some sentences are both process and action sentences. In (19d) (*Yıldız hanım bulaşıkları yıkadı*), *Yıldız hanım*, the subject, does an action of washing (*yıkamak*) and the state of the direct object, *bulaşıklar* changes from dirty to clean. This kind of sentences are classified as process-action sentences. *Bozmak* (*to damage*), *dikmek* (*to set up*), *yıkamak* (*to wash*) are examples of such verbs. The subject is specified as the agent; the direct objects of them sometimes have the patient (e.g., *Kadir bardağı kırdı.* (*Kadir broke the glass*)) or the theme role (e.g. *Ali topu tuttu.* (*Ali caught the ball*)) [2]. Both of these sentences answer the questions "What did X do?", where X is the subject of the sentence, and "What happened to Y?", where Y is the direct object of the sentence questions.

[2] Note that this is the case only when the sentence is in active voice.

What did Kadir do?	He broke the glass.
What happened to the glass?	It was broken.
What did Ali do?	He caught the ball.
What happened to the ball?	It was caught.

- **State-Experiential**

  Let us consider the sentences:

  (21) a. Adam kıza aşık.
  The man is in love with the girl.

  b. Gün geçtikçe seni daha çok seviyorum.
  I love you more and more everyday.

  c. Beni çok üzdün.
  You made me very upset.

  The subject *adam* in (21a) is not an agent, a patient or a theme. He is someone who is mentally disposed in some way. The arguments *adam* (*man*) and *kız* (*girl*) are the experiencer and the theme, respectively. The predicates like *aşık* (*in love*), *memnun* (*pleased*), *razı* (*content*), *sevdalı* (*in love*), etc, are classified as state-experiential predicates, because they express both the state of the object and the emotional experience of the subject simultaneously.

- **Process-Experiential**

  In (21b), *sen* is the theme and *ben*, the hidden subject, is the experiencer of the sentence. The experiential verb in (21b) is also a process verb and is categorized as process-experiential.

- **Action-Experiential**

  An example of an emotional experience, caused by an action, speech, or attitude, is given in (21c). The hidden subject is the agent and *ben* is the experiencer of the sentence. Some other Turkish verbs in this category are *kırmak* (*to break*), *sıkmak* (*to bore*), *üzmek* (*to make sad*), etc.

- **State-Benefactive**

  Let us consider the following sentences:

  (22) a. Çocuğun kırmızı bir balonu var.
  The child has a red balloon.

  b. Kasap dün 50,000 TL kazanmış.
  The butcher earned 50.000 TL yesterday.

  c. Kardeşime hediyesini gönderdim.
  I sent my sister her present.

  Some predicates, such as *sahip* (*owner, possessor*), *malik* (*owner, possessor*), *var* (*existent*), and *yok* (*lacking*), specify a state and express a benefactive situation. For example, in (22a) *çocuk* (*child*) has or owns a *kırmızı balon* (*red balloon*). Here *çocuk* is the beneficiary and *kırmızı balon* is the theme.

- **Process-Benefactive**

  In (22b), the verb *kazanmak* refers to a change in disposition of *50,000 TL*. The thematic role of *50,000 TL* is value-designator according to our thematic role specifications, and *kasap* is in a benefactive situation. Other examples of such verbs are *bulmak* (*to find*), *sahip olmak* (*to have, to own*), *elde etmek* (*to acquire*), etc.

- **Process-Action-Benefactive**

  This kind of verbs express a process, an action, and a benefactive situation at the same time. In (22c), *ben* is the agent, *kardeşim* (*my sister*) is the beneficiary, and *hediye* is the theme. Some other examples of this kind could be given as *almak* (*to take*), *göndermek* (*to send*), *satmak* (*to sell*), *vermek* (*to give*), etc.

- **State-Completable**

  Let us consider the following examples:

  (23) a. Karısının bilezikleri iyi para etti.
  His wife's bracelets were sold for a good sum of money.

  b. Dört kişi briç oynadılar.
  Four people played bridge.

These verbs declare a state which implies the coexistence of a certain concept. For example, in (23a), *etmek* specifies a state implying the coexistence of *para*. In this sentence, *karısının bilezikleri* (*his wife's bracelets*) is the theme and *para* (*money*) is the complement of *etmek* which is categorized as a state-completable verb. Examples of this kind are (*ağır*) *gelmek/çekmek* (*to be weighty*), (*zaman*) *sürmek* (*to last*), (*boyunda*) *olmak* (*to be tall as*), (*aklında*) *olmak* (*to remember*), etc.

- **Action-Completable**

  Some of the action verbs also imply the coexistence of a certain nominal concept by their nature. *Oynamak* (*to play*), for example, implies a game like *briç* (*bridge*), *satranç* (*chess*), or *futbol* (*football*). In (23b), *oynadılar* is an action-completable verb, *dört kişi* (*four people*) and *briç* are the agent and complement, respectively. Some examples of this kind of verbs are (*koşu* (*race*)) *koşmak* (*to run*), (*sayı* (*number*)) *say* (*to count*), (*eser* (*monument*)) *yapmak* (*to build*), and (*hayat* (*life*)) *yaşamak* (*to live*).

- **State-Locative**

  Let us consider the following examples:

  (24) a. Dolapta karpuz var.
  There is a watermelon in the fridge.

  b. Atatürk bu evde yaşamış.
  Atatürk has lived in this house.

  c. Çocuk aniden yolda durdu.
  The child suddenly stopped on the road.

  d. Yazar piposunu masaya koydu.
  The writer put his pipe on the table.

  Locative verbs are accompanied by objects which bear the relation location. In (24a), *dolap* (*fridge*) is the location, where the state takes place, *var olmak* (*to exist*) is categorized as state-locative predicates. *Yok olmak* (*to not exist*), can be categorized as state-locative according to their usage.

- **Process-Locative**

  In (24b), *yaşamak* (*to live*) express a change in state of *Atatürk* and the location of this process is *bu ev* (*this home*). The verbs, such as *durmak* (*to stop*), *çarpmak* (*to hit*), *düşmek* (*to fall down*), *oturmak* (*to sit*), and *yaşa* (*to live*) can be categorized in this type according to their usage.

- **Action-Locative**

  The verbs, categorized as action-locative verbs, state an action and give the concept of location of that action at the same time. In (24c), *durmak* (*to stop*) is an action verb having the agent *çocuk* (*child*) and *yol* (*road*) is the location where *çocuk* performs the action. According to their usage, *çıkmak* (*to come up*), *dönmek* (*to turn*), *durmak* (*to stop*), and *oturmak* (*to stay*) can be categorized as action-locative verbs.

- **Process-Action-Locative**

  These verbs indicate an action and a change in state implying the location of the event. *Koymak* (*to put*) in (24d), is an example of this kind of verb. *Yazar* (*writer*), *pipo* (*pipe*), and *masa* (*table*) are the agent, the theme, and the locative goal, respectively. Some other examples are *çarpmak* (*to hit*), *dayamak* (*to hold against*), *koymak* (*to put*), *sermek* (*to spread over*), etc.

3.1.3 Relationship between Grammatical Relations and Thematic Roles

Both thematic roles and grammatical relations are well-studied relations between things typically representing entities (noun phrases) and events or states (verbs). However, their domains are different. The grammatical roles are relations in *syntax* not in *semantics*, but thematic roles are semantic relations. Moreover, the grammatical roles and thematic roles are features of *sentences* and *predications*, respectively. For example, *Subject* is a relation between an NP and a verb. In this relation, the morphological form of the verb is *governed* or *controlled* by the NP. In (25a), *it* is the subject because it determines the singular form of the verb therefore (25b) is not grammatical. However, *it* has no thematic role in (25a) because it does not represent an argument.

Thus, thematic roles require predicates and arguments, not necessarily NPs and verbs; thematic roles can not be directly taken from grammatical roles:

(25) a. It rains ice in London.

b. * It rain ice in London.

The following examples (from [3]):

(26) a. I have the book.

b. U menya kniga.
me+DAT book

c. Mam książkę.
have-I book

illustrate semantically equivalent expressions of (26a) in Russian and Polish in (26b) and (26c) respectively. In (26a), both *I* and *the book* are in nominative case. However, in (26b) the word for *I*, *menya*, is coded morphologically in the *dative case*. In Polish which is a language very closely related to Russian, the expression equivalent to (26a) and (26b) surfaces as (26c) and the word for *book*, książkę, is in accusative case. In these sentences, we see that although the meanings of (26a), (26b), and (26c) are equivalent, the morphological cases of the arguments are not comparable. As a result, the thematic roles can not be derived directly from surface case markers (morphological cases).

The examples above illustrate that thematic roles, grammatical relations and surface case markings are different concepts. However, we can not say that surface case markers, grammatical relations, and thematic roles are completely unrelated. On the contrary, thematic roles follow grammatical constraints and hence there are relationships among thematic roles, grammatical relations, and surface case markings.

According to Frawley [3], thematic roles provide a way to think how the pieces of any situation go together in our mental models, beyond the machinery that languages have for putting forms together into *expressions about*

situations. However, neither grammatical roles nor morphological cases provide this. Thematic roles "configure" the protected world of reference, linking predicates to arguments in particular ways.

Let us consider the examples below:

(27) a. Adam çocuğu dövdü.
The man beat the child.

b. Çocuk adam tarafından dövüldü.
The child was beaten by the man.

c. Adam annesinin çocuğu dövmesine neden oldu.
The man caused the beating of the child by her mother.

In the passive causative sentences thematic roles of the entities are preserved, though the grammatical category of the entities are changed. For example, in (27a) *adam* (*man*), the subject, is the agent and *çocuk* (*child*), the direct object, is the patient. Since the meaning of the sentence is not changed, these entities play the same semantic roles in (27b), although their grammatical categories are changed to subject and object respectively. Sentence (27c) illustrates thematic roles in a causative. In this sentence, *adam*, the subject, is the agent and *annesinin çocuğu dövmesi* (*the beating of the child by her mother*), the direct object, is the theme of *neden olmak* (*to cause*). But in the gerund phrase *annesinin çocuğu dövmesi* (*the beating of the child by her mother*), *annesi* (*her mother*) and *çocuk* (*the child*) are the agent and patient of *dövmek* (*to beat*).

3.2 The Structure of the Lexicon

Our design for the lexicon has been inspired by the lexicon of DIANA system (see Section 2.4). In DIANA, each sense of the lexical entry is stored separately. This structure is not suitable for Turkish verbs because:

- Verbs have many senses (normal and idiomatic) in Turkish. If the lexicon of DIANA system were used so many entries would have been defined. This prevents spurious repetitive common features.

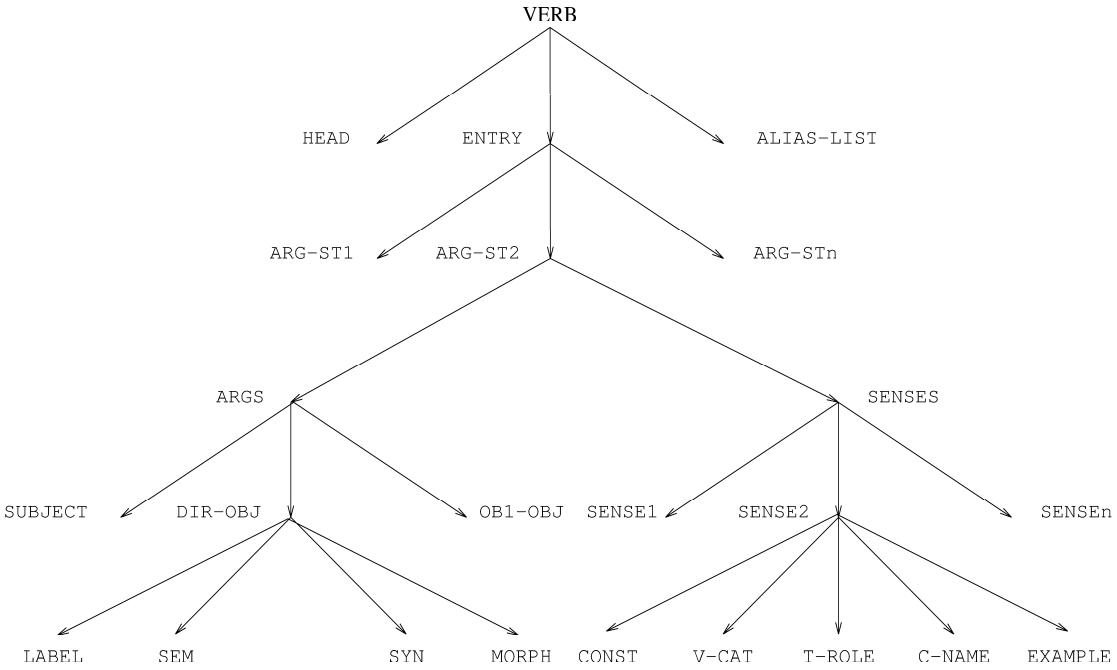

Figure 3.3. Tree structure of a lexical entry design.

- Morphological constraints on the arguments of verbs play an important role in the sense disambiguation process. For this reason, morphological constraints about the arguments of verbs should also be included in the verb lexicon.
- The senses of verbs can be classified according to argument structures, so that no redundant repetition in argument structures slot is made.

Figure 3.3 illustrates a tree structure for our lexical entry design. In this structure, in order to avoid redundant repetitions of similar argument structures, we define `ARG-ST` (argument structure) slots containing an `ARGS` (arguments) slot and a `SENSES` slot, and collect the senses having the same argument structure in the same `ARG-ST` slot.

The lexicon which consists of lexical items is structured as shown as a list in Figure 3.4. A lexical entry consists of:

1. head in the `HEAD` slot,
2. list of `ARG-ST`s (argument structures) in the `ENTRY` zone, and

```
(((HEAD "ak" )          ; lexeme of the verb ak
  (ENTRY ...            ; entry having all senses of the verb ak
         ... )
  (ALIAS-LIST ... ))    ; aliases
  (HEAD "at" )          ; lexeme of the verb at
  (ENTRY ...            ; entry having all senses of the verb at
         ... )
  (ALIAS-LIST ... ))    ; aliases
  ...)
```

Figure 3.4. The structure of the lexicon.

Argument St.	ARG-ST1	ARG-ST2	ARG-ST3
Subject (NOM)	yes	yes	yes
Dir. Obj. (NOM/ACC)	no	yes	yes
Oblique Obj. (DAT)	no	no	yes
Oblique Obj. (ABL)	no	no	no
Sense	to beat	to scream	to put
Example	kalbi atmak	nara atmak	yemeğe tuz atmak

Table 3.1. The first 3 argument structures of *atmak*.

3. list of aliases in the `ALIAS-LIST` zone.

The `HEAD` slot, which contains the lexeme of the entry, is stored as strings. The characters that are not valid in Latin alphabet are indicated as capital letters.

The `ARG-STx` slot where `x` denotes the index of the argument structure consists of:

1. list of arguments in the `ARGS` zone and
2. list of senses in the `SENSES` zone.

The argument structures are ordered from the more relaxed to specific. For example, *atmak* (*to throw*) has five argument structures. In the first one, the senses having no arguments other than a subject are stored. In the second argument structure, there is also a direct object in nominative or accusative

Argument St.	ARG-ST4	ARG-ST5
Subject (NOM)	yes	yes
Dir. Obj. (NOM/ACC)	yes	yes
Oblique Obj. (DAT)	no	yes
Oblique Obj. (ABL)	yes	yes
Sense	to throw out	to throw
Example	oyundan atmak	taşı oradan buraya atmak

Table 3.2. The fourth and the fifth argument structures of *atmak*.

case. The third one is more general and has an oblique object in dative case in addition to a subject and a direct object. The fourth one is more general than the second one as well, but in this case the oblique object has an ablative case marker. The most general one is always the last one. This structure has a subject, a direct object in nominative or accusative case and two oblique objects in dative and ablative case respectively. The argument structures of *atmak* (*to throw*) are illustrated in Tables 3.1 and 3.2.

The `ALIAS-LIST` slot is for storing alias definitions. These aliases are used for commonly used phrases.

The `ARGS` zone consists of:

1. the `LABEL` slot and
2. arguments of the verb.

The arguments can be a subject, a direct object, an indirect object, and one or two oblique objects, and represented as `SUBJECT` slot, `DIR-OBJ` slot, `IND-OBJ` slot, `OB1-OBJ` slot, and `OB2-OBJ slot`, respectively. The information about each arguments is stored in an association list (see Figure 3.5). In Turkish, most verbs are accompanied by NPs or PPs having thematic role of the value-designator or the instrument. The features of these arguments can be specified in `INST` and `VAL-DES` slots, respectively.

Figure 3.5 illustrates the structure of an argument zone which consists of:

- *The label zone*, represented as `LABEL` slot, contains the label of the argument. In each of the semantic, syntactic or morphological constraint

```
(<grammatical role>      ; can be SUBJECT, DIR-OBJ, IND-OBJ,
                         ; OB1-OBJ, OB2-OBJ, INST, VAL-DES.
    (LABEL ...)          ; label of the argument
    (SEM ...)            ; semantic features of the argument
    (SYN ...)            ; syntactic features of the argument
    (MORPH ...))         ; morphological features of the argument
```

Figure 3.5. The structure of each argument.

the label of an argument is specified. By using this label value of the argument that the constraint is applied is determined. For example, for a direct object labeled as `D`, the constraints are given in lists like `(Human D)`, `(Case D NOM)`, and `(Occ D Optional)`. When these constraints are checked the head of argument is used. For instance, the head of the noun phrase *küçük kırmızı balık* (*little red fish*) is *balık* (*fish*) and all the constraints are applied to *balık*, unless otherwise is stated. In the definitions of constraints any number of `AND` and `OR` logical relations among the features are allowed. An integer index is used on the head of the each feature list.

- *The semantic constraints zone*, `SEM` slot contains semantic features of the argument.

- *The syntactic constraints zone*, `SYN` slot contains syntactic features of the argument.

- *The morphological constraints zone*, `MORPH` slot contains morphological features of the argument.

The `SENSES` zone consists of the senses of the verb, represented as `SENSE`, plus a number indicating the sense index. The structure of a `SENSE` slot is illustrated in Figure 3.6. Each sense consists of:

- *The constraints zone*, represented as `CONST` slot, contains a logical expression of semantic, syntactic, and morphological constraints about all the arguments. In the second step of the sense disambiguation process these constraints are checked.

```
(SENSE#                    ; # denotes the index of the sense
    (CONST   ... )         ; a logical expression of semantic,
                           ; syntactic, and morphological const.
    (V-CAT   ... )         ; semantic category of the verb
    (T-ROLE  ... )         ; thematic roles of arguments
    (C-NAME  ... )         ; concept name
    (EXAMPLE ... ))        ; examples
```

Figure 3.6. The structure of a SENSE slot.

- *The verb category zone*, represented as V-CAT slot, contains the semantic category of the verb.
- *The thematic roles zone*, represented as T-ROLE slot, describes the thematic roles of the arguments.
- *The concept name zone*, represented as C-NAME slot, describes as closely as possible the language independent concept expressed by the sense. For convenience, we describe this by an unambiguous expression in English.
- *The examples zone*, represented as EXAMPLE slot, contains one or more examples.

3.2.1 An example lexical entry

We provide the entry for *iletmek* as an example which has three senses:

- to conduct,
- to convey, and
- to tell.

These senses are kept in two argument structures. The complete entry for *iletmek* is illustrated in Figures 3.7 and in Figure 3.8.

The role of each slot is explained below:

- The ARG-ST1 slots contain the arguments and the senses of the first and second argument structures, respectively.
- The ARGS slot consists of arguments of *iletmek*. In ARG-ST1, the arguments are a subject and a direct object, although in ARG-ST2 an oblique object is also specified.
- The SENSES slot contains the senses of *iletmek*. The senses *to conduct* and *to convey* are stored in the first argument structure. ARG-ST2 only contains the most general sense of *iletmek*: *to tell, to express.*
- The LABEL slot have the label of arguments, S for the SUBJECT and D and O1 for the objects DIR-OBJ, OB1-OBJ, respectively.
- All SEM slots are filled with T to indicate that no semantic constraints specified.
- The SYN slot contains the syntactic constraints of arguments. The verb *iletmek* obligatorily (OBLIGATORY) subcategorizes the direct object of ARG-ST1. All the other arguments are optionally (OPTIONAL) subcategorized by *iletmek*.
- The MORPH slots indicate that the case of the direct objects and oblique object should be accusative (ACC) and nominative (NOM), respectively.
- In the CONST slot, it is indicated that the direct object must be an instance of POWER-ENERGY-PHYSICALOBJECT. The concept POWER-ENERGY-PHYSICALOBJECT contains power, energy, and physical objects like *ses* (*sound*), *elektrik* (*electricity*). In the first sense of second argument structure a complex logical expression is defined. The concept HUMAN-ROLE-PROFESSION is a combined concept containing humans, roles, and professions. The subject and the oblique object must be HUMAN, ROLE, or PROFESSION and direct object can be CONCEPTUALOBJECT (conceptual object), PERCEPTION, or EMOTION, but nothing else (see Section 4.3).
- The V-CAT slot contains the category of the verb. The verb *iletmek* is a PROCESS-ACTION verb in all cases.
- The thematic role of arguments are specified in T-ROLE slot. In all senses, SUBJECT and DIR-OBJ are the AGENT and the THEME, respectively. The OB1-OBJ in the ARG-ST2 is the RECIPIENT of *iletmek*.

- The `C-NAME` slot specifies the concept name of the sense.
- An example for each sense is given in the `EXAMPLE` slot.

3.3 Scope and Limitations of the Verb Lexicon for Turkish

Our lexicon is built for Turkish verbs to disambiguate verb senses in the analysis step of a machine translation system. For this reason, the lexicon is built only for verbs. Even though only the analysis information is stored, the data structure of the lexicon is augmentable for further extensions. In order to save space and simplify the disambiguation process, the semantic, syntactic, and morphological features of each senses of verbs are stored in two levels. The senses of verbs are first classified according to their possible argument structures. Let us consider the following examples:

(28) a. Elif eve giderken yolunu şaştı.
Elif confused her way home.

b. Herşeye rağmen işi bırakmamasına şaşıyorum.
I am surprised at the fact that he does not quit his job despite everything.

c. Ok hedefinden şaştı.
The arrow missed the target.

In the examples above, the verb *şaş* is used in three different senses. In (28a), the direct object *yol* (*way*) is in accusative case. In (28b), case of the direct object *iş* (*job*) is dative, and in (28c), the oblique object *hedef* (*target*) is in ablative case. It is also easily seen that there is a relation between the meaning and the case markings. The verb *şaşmak* takes a direct object in accusative case when it is used to mean *to miss* and we can say that the argument structure of this sense consists of a subject in nominative case (subject is always nominative in Turkish) and a direct object in accusative case. When it is used to mean *to be surprised about*, the argument structure contains a

```
((HEAD  . "ilet")
 (ENTRY
     (ARG-ST1
         (ARGS
             (SUBJECT
                 (LABEL  . S)
                 (SEM  . T)
                 (SYN OCC S OPTIONAL)
                 (MORPH  . T))
             (DIR-OBJ
                 (LABEL  . D)
                 (SEM  . T)
                 (SYN OCC D OBLIGATORY)
                 (MORPH
                     (OR
                         (1 CASE D NOM)
                         (2 CASE D ACC)))))
         (SENSES
             (SENSE1
                 (CONST POWER-ENERGY-PHYSICALOBJECT D)
                 (V-CAT PROCESS-ACTION)
                 (T-ROLE
                     (1 AGENT S)
                     (2 THEME D))
                 (C-NAME  . "to conduct")
                 (EXAMPLE  . "katIlar sesi en iyi iletir."))
             (SENSE2
                 (CONST  . T)
                 (V-CAT PROCESS-ACTION)
                 (T-ROLE
                     (1 AGENT S)
                     (2 THEME D))
                 (C-NAME  . "to convey")
                 (EXAMPLE  . "yardImI ilettiler."))))
```

Figure 3.7. The first argument structure of the verb *iletmek*.

```
(ARG-ST2
    (ARGS
        (SUBJECT
            (LABEL  . S)
            (SEM  . T)
            (SYN OCC S OPTIONAL)
            (MORPH  . T))
        (DIR-OBJ
            (LABEL  . D)
            (SEM  . T)
            (SYN OCC D OPTIONAL)
            (MORPH
                (OR
                    (1 CASE D ACC)
                    (2 CASE D NOM))))
        (OB1-OBJ
            (LABEL  . O1)
            (SEM  . T)
            (SYN OCC O1 OPTIONAL)
            (MORPH CASE O1 DAT)))
    (SENSES
        (SENSE1
            (CONST
                (AND
                    (1 HUMAN-ROLE-PROFESSION S)
                    (2 HUMAN-ROLE-PROFESSION O1)
                    (3
                        (OR
                            (1 CONCEPTUALOBJECT D)
                            (2 PERCEPTION D)
                            (3 EMOTION D)))))
            (V-CAT PROCESS-ACTION)
            (T-ROLE
                 (1 AGENT S)
                 (2 THEME D)
                 (3 RECIPIENT O1))
            (C-NAME  . "to tell, to express")
            (EXAMPLE  . "adama duygularInI iletti."))))
(ALIAS-LIST ))
```

Figure 3.8. The second argument structure of the verb *iletmek*.

subject and a direct object in dative case. Finally, this verb can take a subject and an oblique object in ablative case when it is used to mean *to deviate from.* In many cases, the verb has more than one sense having the same argument structures. Because of this, senses having the same argument structure are collected together and put in senses zone of each argument structure.

(29) Adamı ateşe attılar.
They put the man in a dangerous situation.
(Literally, they throw the man in fire)

Even though the structure of the lexicon is designed for sense disambiguation process, it is also suitable for use of parsers (e.g., [4, 5]), because the senses are first grouped according to their argument structures and the morphological information about each argument is stored in the argument structure slots. The parser can access the entry of the verb of the sentence. Then, the arguments of each argument structure are determined and matched with the words or phrases in the sentence while being syntactically analyzed. For example, when the parser parses (29), the correct argument structure is determined as `ARG-ST3` (argument structure 3). In this argument structure, *atmak* takes the following arguments a subject nominative case, a direct object in accusative or nominative case, and an oblique object in dative case.

(30) a.

Ben	gelmeden	hiçbir yere	gitme
I/mole	come+VN+ABL	nowhere+DAT	go+NEG+IMP+2SG

b. Don't go anywhere before I come.

c. Don't go anywhere before the mole comes.

In [4], it is stated that (30b) and (30c) are plausible interpretations of (30a) according to grammar used by the parser, though the second one is not semantically plausible. By using the semantic information stored in the lexicon, the parser can determine that the second one is not semantically plausible.

A modified system architecture proposal for the Turkish LFG parser implemented by Güngördü which uses the verb lexicon developed for Turkish is illustrated in Figure 3.9.

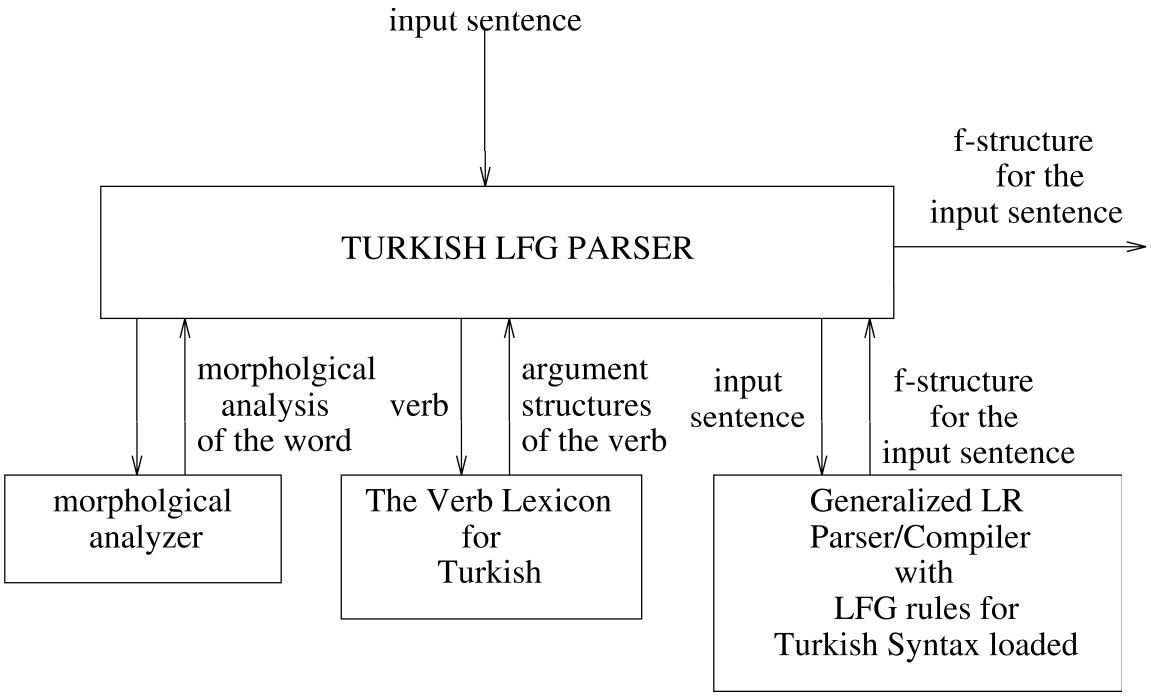

Figure 3.9. A modified system architecture for Turkish LFG Parser.

Our lexicon can also be used by a machine translation system using transfer method. After adding the information for text generation in interlingua, we can use this lexicon in the analysis process of a machine translation system using interlingua. Figures 3.10 and 3.11 illustrate the lexicon in these machine translation systems. It is also possible to use the lexicon in the translation process from English to Turkish, because the corresponding verbs in English are stored in concept name zone of each sense. The structure of verbs can be determined from arguments zone, and a sentence in Turkish can be generated by using this information.

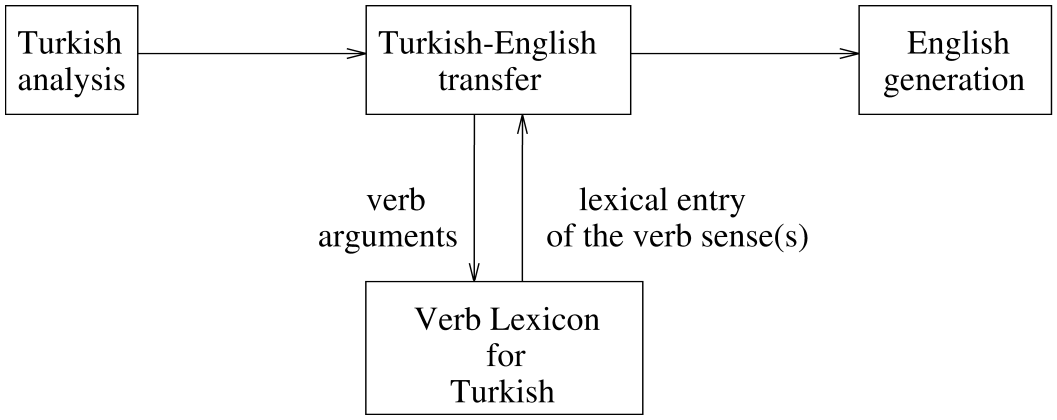

Figure 3.10. The system architecture of a Transfer-based MT system that uses the verb lexicon for Turkish.

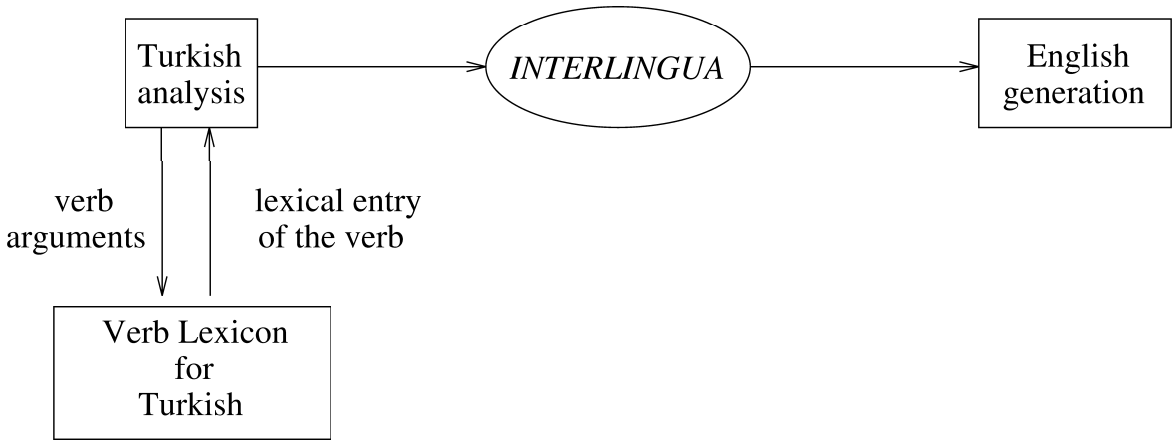

Figure 3.11. The system architecture of an Interlingua MT system that uses the verb lexicon for Turkish.

Chapter 4

Operational Aspects of the Lexicon

As we have already stated, the verb lexicon can be used in various applications:

- syntactic analysis,
- transfer and verb sense disambiguation in machine translation.

This chapter presents the verb sense disambiguator which uses the verb lexicon for Turkish. This tool is developed to illustrate the operational aspects of the verb lexicon. The verb sense disambiguator is designed to be used in a number of applications such as parsing, MT transfer with suitable and programmatic interfaces, etc. The main function of this tool is to resolve the sense of verbs in Turkish using the semantic, syntactic, and morphological information available in the context of the verb. In this process, an input text (a sentence or a phrase) is taken and an output text containing semantic information about the context of the verb is generated. As we stated, the input to the disambiguator is not a raw text but syntactically and morphologically analyzed structure involving the basic concept. This system is designed to be easily used as an internal stage of any machine translation process. Figure 4.1 illustrates the architecture of our verb sense disambiguator.

In this chapter, we will first explain the sense disambiguation process, and the constraints used in this process. After that, the ontology, in which the hierarchy and inheritance among the semantic concepts are defined as semantic features, will be presented. Finally, we will present the limitations and

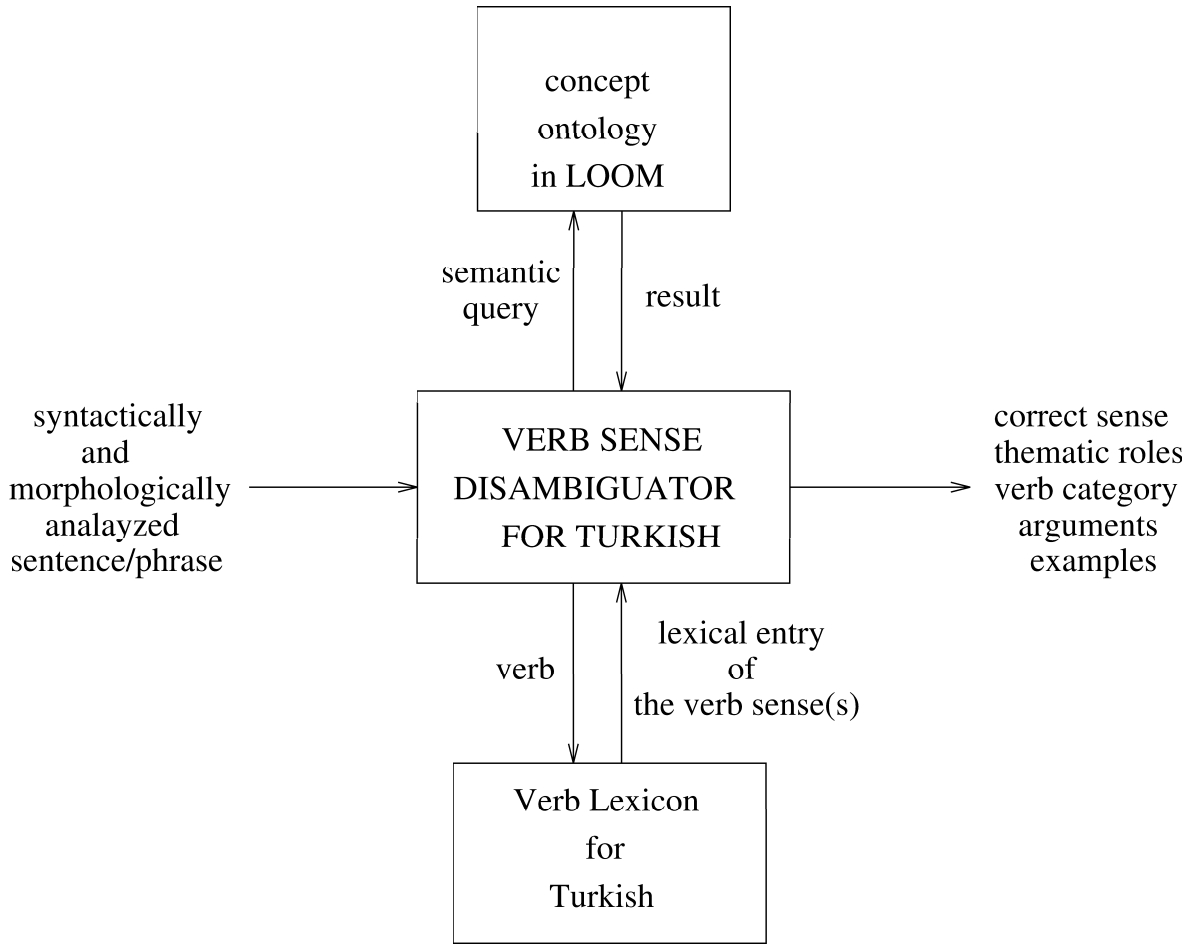

Figure 4.1. The system architecture of the Sense disambiguator

functionality of the verb sense disambiguator.

4.1 The Sense Disambiguation Process

The lexicon for Turkish verbs is designed mainly for sense disambiguation process. By using the structure of the lexicon, we built a two-level semantic network to resolve sense of verbs. The disambiguation process is applied to all sentences/phrases given in the input structure. For each sentence/phrase, after determining its head (predicate), we search the entry of the predicate in the lexicon. The two-level mapping process is started if an entry for the verb exists in the lexicon. The sense disambiguation process is handled in two main steps:

1. The argument structure that matches the arguments of the sentence is determined.
2. The correct sense is determined by checking the constraints of each sense.

In the first step, arguments of the sentence are matched with the argument structures (`ARG-ST`) of the entry (`ENTRY`) for the predicate of the sentence. In this process, the morphological cases of the arguments have important roles because the objects are *bound* to arguments in the sentence if the case of arguments are matched. Here *bind* refers to giving the label of the argument (e.g., `D` for a direct object (`DIR-OBJ`), `I` for an indirect object (`IND-OBJ`), etc.). We then check semantic, syntactic, and morphological constraints of the objects and the subject. If all constraints are satisfied *argument structural mapping* process is successful.

The second step starts after all the constraints of arguments of the sentence are satisfied. In this step, the constraints of each sense (`SENSE`) in `SENSES` slot of the `ARG-ST` are checked. If the constraints in the `CONST` zone are satisfied, the correct sense is determined.

In some cases the sentences might be ambiguous and they may have many interpretations. Because of this, the other senses are also checked and the disambiguation process stops after all the senses and argument structures are

tested. Semantic, syntactic, and morphological constraints along with the ontology for semantic concepts are discussed in the following sections.

4.2 Constraints

The features of each argument are stored in three different slots containing semantic, syntactic, and morphological features. In the first step of the sense disambiguation process, these features are used as constraints to map the argument structure having these arguments. The constraints used while mapping to correct sense also contain semantic, syntactic, and morphological constraints of the arguments, but not in separate zones. The second value in the constraint list specifies the label of the argument that the constraint is applied. The root of the noun phrase or prepositional phrase is tested for the constraints, because, they generally give the meaning of the NP or PP. For example, in *yeşil ördek* (*green duck*), *ördek* is the head of the NP.

The logical operators `AND` and `OR` are used to construct more complex constraints. (31a) and (31b) illustrate `AND` and `OR` operations between two constraints. Each of the constraints are preceded by an integer denoting the head of the association list. These integers are omitted in constraint satisfaction process. In the logical expression in (31a), "X" and "Y" represent arguments and "const1" and "const2" mean that constraints should be applied to "X" and "Y", respectively. If one of the arguments does not exist, the constraints applied to it are assumed to be true, and the other one is checked. If any of the constraints can not be satisfied then the value of the expression is assigned to false. In the logical expression in (31b), "X" and "Y" represent arguments and "const1" and "const2" mean constraints should be applied to "X" and "Y", respectively. If one of the arguments does not exist, the constraint applied to it is assumed to be false and the other one is checked. If any of the constraints is satisfied, the value of the expression is assigned to true. After argument structural mapping is done, constraints of each sense are checked and concept name of the sense, category of the verb, thematic roles of each argument, and some examples demonstrating verb usage in that sense are output.

(31) a.

```
(OR
    (1 const1 X)
    (2 const2 Y))
```

b.

```
(AND
    (1 const1 X)
    (2 const2 Y))
```

4.2.1 Syntactic Constraints

Syntactic constraints are for checking features of the arguments of the analyzed sentences relevant to a syntactic point of view. The syntactic features of the arguments are determined by a syntactic analyzer, and the ones that can disambiguate verb senses are used as syntactic constraints. For example, word category, number person, definiteness, optionality of arguments are features of this kind. In (32a) the verb *ak* in sense *to move* can not be accompanied by a subject in singular form ((32b) is not valid). The design tries to be more generic so that more features can be added easily by adding new case frames to the constraint satisfaction network.

(32) a. İnsanlar Antalya'ya aktı.
People flowed into Antalya.

b. ? İnsan Antalya'ya aktı.

4.2.2 Morphological Constraints

These constraints are for checking morphological features of the arguments of the analyzed sentences. The morphological features of the arguments are not determined by our sense disambiguator, but are determined previously by a morphological analyzer (e.g., [12]). Constraints on the morphology of the lexical forms that play a role in ambiguity resolution are stored as morphological

constraints. For example, surface case, causative, possessive markers are features of this kind. The case of an argument has an important role in the object matching process. In (28) checking case of the object is enough to determine the correct sense of the verb with which it is used. New features can be tested by adding a new case frame to the constraint satisfaction network.

4.2.3 Semantic Constraints

Semantic constraints play the most important role in the disambiguation process. Let us consider the following sentences:

(33) a. Balondan gaz kaçtı.
Gas has leaked out of the balloon.

b. Hapisten kaçmışlar.
They escaped from the prison.

In (33a) and (33b), *kaç* is used in two different senses. The main difference among these senses can be determined when the sentences are semantically analyzed. The subject of *kaç* should be a liquid or a gas when it is used in the sense *to leak out.* However, the subject must be an animal when *kaç* is used in the second sense, i.e., *to escape from.* These senses of *kaç* can only be disambiguated by analyzing the role fillers and semantic features.

In the semantic analysis process, the semantic features of the role fillers must be known to check the semantic constraints. For this reason, an auxiliary noun lexicon has been developed. Each item is stored in the lexicon with its semantic features. The semantic features of the lexical elements of the noun lexicon are stored in an ontological knowledge base.

This ontological knowledge base is based on semantic networks. In a semantic network, there are facts and first kind of rules which describe the hierarchy of individuals and classes (e.g., *concepts*) that make up the world. The second kind of rules are the ones that the classes applies to. Some of the first kind of rules are *is-a, is-part-of, has-property*, etc. By *is-a* relation properties of a more general class is inherited to a less general one. For example, "humans are

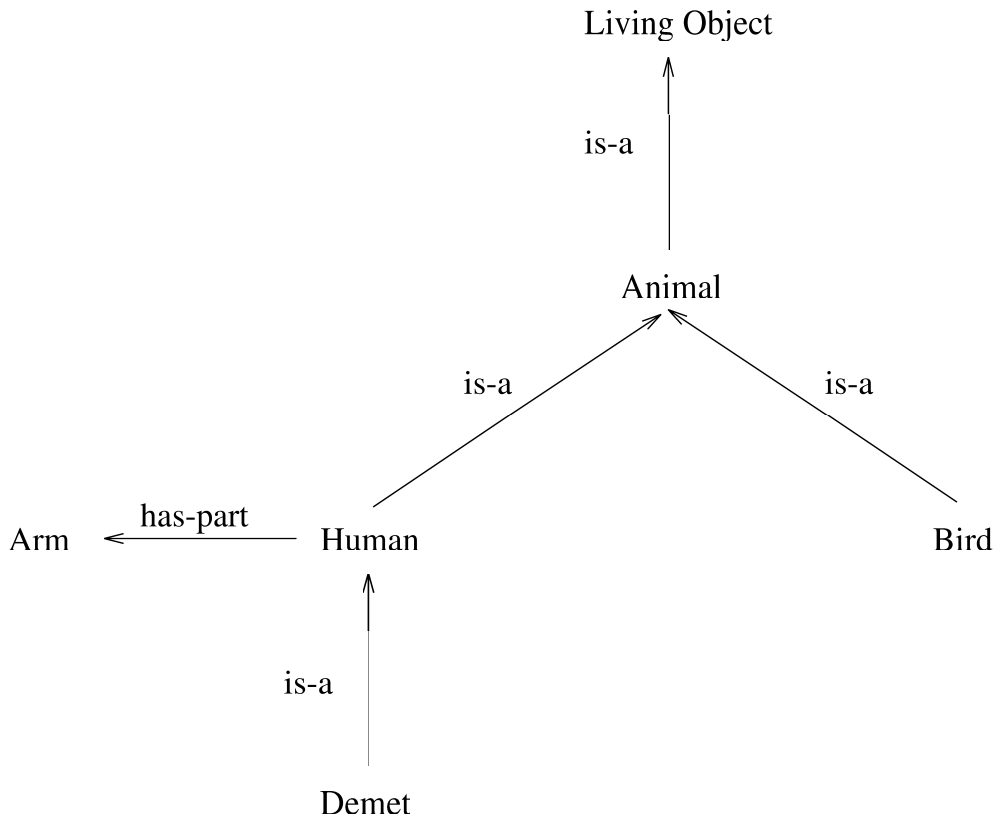

Figure 4.2. A simple semantic network

animals" represented as `(is-a human animal)`. Thus, humans inherit all the properties of animals. The sentence "humans have mouths" can be represented as `(is-part-of mouth human)`. The *is-part-of* relation is used to define part relation between classes. The *has-property* relation is used to state properties of concepts. Semantic networks are shown by directed graphs, where each node denotes a class and the isa-hierarchy between classes is shown by arcs. Figure 4.2 illustrates a simple semantic network.

Semantic analysis part in sense disambiguation process starts with finding the lexeme of the argument. Its semantic constraints are found in the verb lexicon, and whether these constraints are satisfied is determined by querying the knowledge base.

4.3 Ontology

The semantic categories are defined in an *ontology* which is a model of the world providing definitions of semantic categories, such as, human, thing, non-living object, living object, etc. These semantic categories are domains for the semantic features of the entries in machine traceable lexicons for natural language processing. In the process of building the ontology, an interconnected network of ontological units is defined. Storage, access, and update procedures become available by this organization.

Our ontology is based on the ontology in [10]. Semantic markers for nouns are defined in ten major concepts:

- `Thing-Object` containing such as things and objects,
- `Commodity-Ware` containing artificial matters useful to humans,
- `Idea-Abstraction` containing non-matters which results from intellectual activities in the human brain, such as ideas and abstractions,
- `Part` containing structural parts, elements, and contents of things and matters,
- `Attribute` containing properties, qualities, or features which are representatives of things,
- `Phenomenon` containing physical, chemical, and social actions,
- `Doing-Action` containing human actions,
- `Sentiment-MentalActivity` containing humans' mental activities,
- `Measure` containing measures, and
- `Time-Space` containing time, space, and topologies.

Each of these groups consist of subconcepts. The interconnected network of `Thing-Object` concept and its subconcepts are illustrated in Figure 4.3. The semantic categories and their subconcepts are detailed in Appendix A.

The ontology is implemented in LOOM [1]. By using the reasoning mechanisms of this tool, an extensible inheritance mechanism providing a general facility for default reasoning is defined. Figure 4.3 illustrates the hierarchy between sub-concepts of `Thing-Object` which is one of the major concept containing concrete matters. The nodes indicate *is-a* relations between concepts on the nodes. The feature inheritance is determined by *is-a*. For example, in the concept hierarchy illustrated in Figure 4.3, all `men` are also `living objects`. When we assign a semantic feature to *Ali* as `man`, it is inserted into LOOM as man, by the reasoning mechanism of this tool; also it is inserted as `human`, `human-role-profession` which contains `human`'s, `role`'s, and `profession`'s, `animal`, `living object`, and `thing-object`. Furthermore, accessing mechanism lets us ask whether *Ali* is a human and get the answer "yes", or ask if it is a woman and get the answer "no". For each word more than one feature can be given, for example, *İnanç* is a `man` and also a `belief`.

4.4 Limitations of the Sense Disambiguation Process

The sense disambiguation process has some limitations:

- The sense disambiguation process analyzes the texts in sentence level and so it can not use the information which can only be determined by discourse analysis. This limits the reliability of the disambiguation process.
- We did not deal with the events as arguments. For example, in (34) the gerund clause *kızının sigaraya başlaması* denotes an event and this event can also be analyzed by the verb sense disambiguator by adding a special processing node to the constraint satisfaction network. We will incorporate this feature to the lexicon later.

  (34) Kızının sigaraya başlaması, Ali Bey'i çok üzdü.
  His daughter's starting smoking upset Mr. Ali.

- The nouns stored in the noun lexicon have different senses according to the context they are used in. We add all of their features to the lexicon. For example, both the constraints (35a) and (35b) about *devrim* are satisfied by LOOM.

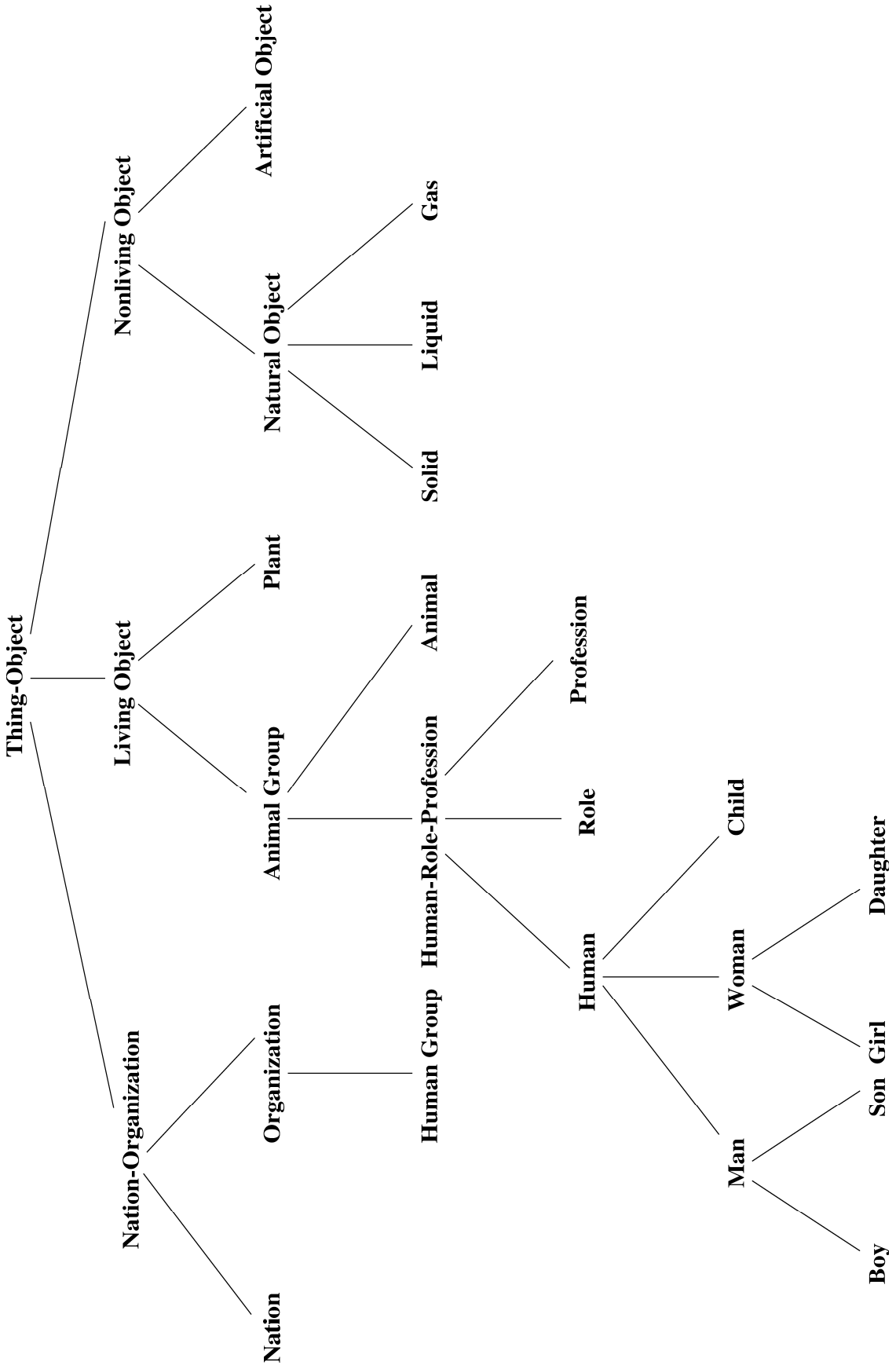

Figure 4.3. The hierarchical in the `Thing-Object` category

(35) a. `(Human devrim)`

b. `(SocialPhenomenon devrim)`

- In Turkish, verbs can be accompanied by adverbial complements, such as *karşı* (*against*), *taraf* (*by*), *kez* (*times*), *anlamsız* (*meaningless*), *iyi* (*fine*), etc. These components sometimes determine the sense of verbs. Since these components are not arguments, they are not considered. We will again incorporate this into the lexicon later.

4.5 Functionality of the Sense Disambiguator

Let us consider the example below:

(36) a. Kalem alındı.
pen buy+PASS+PAST
or
? pen take offense+PAST

b. The pen was bought.

c. ? The pen was offended.

Sentence (36a) has two syntactically plausible interpretations ((36b) and (36c)), but the second one is not semantically plausible. An ordinary parser can not resolve this semantic ambiguity. The sense disambiguator is specially designed for resolving such semantic ambiguities in Turkish. When these two interpretations are tested by the sense disambiguator, the second one is determined as semantically implausible.

In MT systems, it can be used in the analysis process of Turkish sentences. In the analysis process of both transfer based systems and interlinguas, the meaning of sentences must be determined. Multiple meaning interpretation of verbs, and thus the sentences, can be resolved by the verb sense disambiguator.

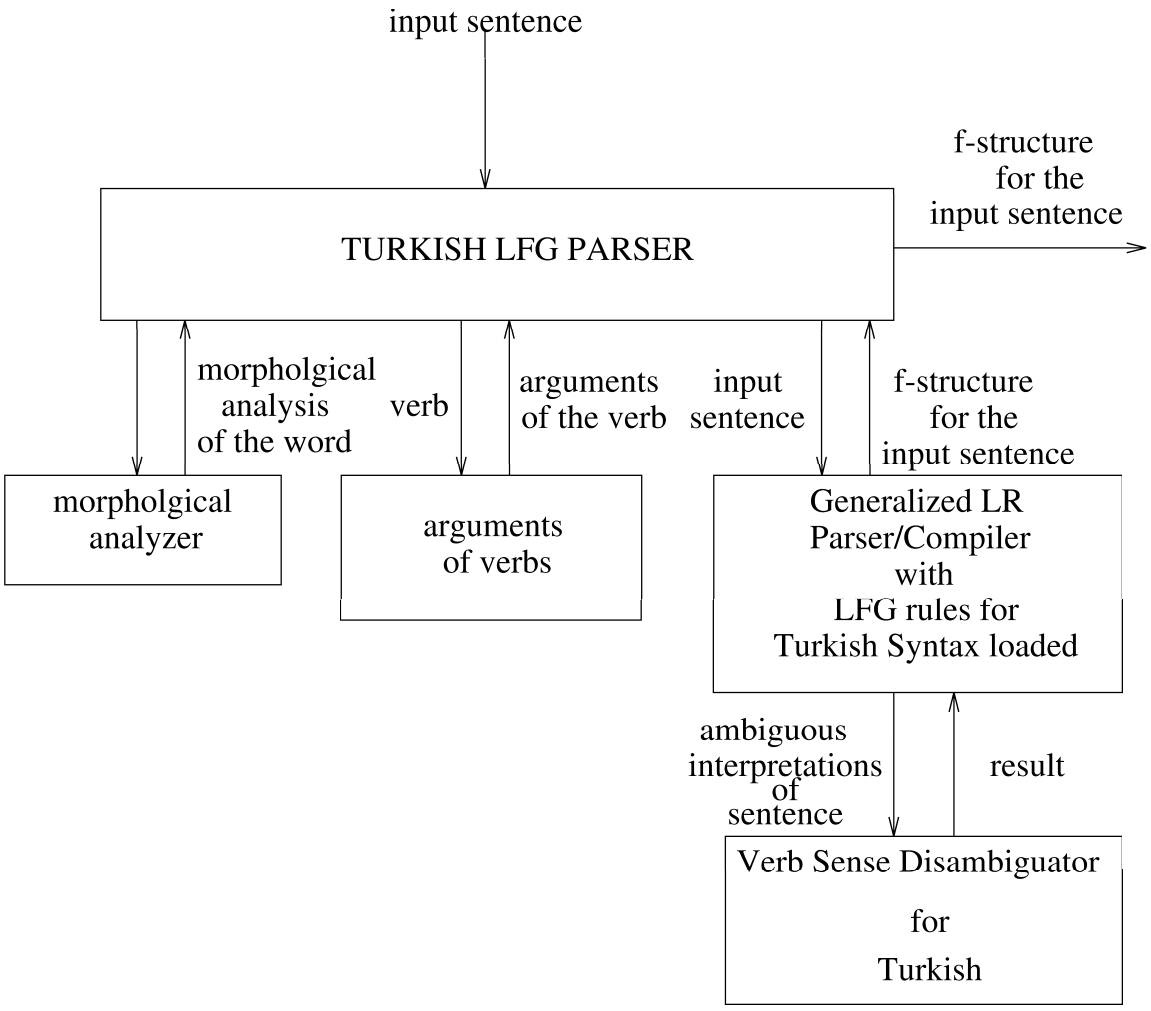

Figure 4.4. The system architecture for Turkish LFG parser that uses the Verb sense disambiguator

Chapter 5

Implementation

The verb lexicon and the verb sense disambiguator have been implemented using the *Lucid Common Lisp* (*LCL*) and the *LispView* [13]. Special access, update and delete functions are defined on the lexicon.

For the graphical user interface, LispView (a generic application programmer's interface package) is used. Object-oriented programming approach is adopted, in order to have a maintainable and structured software. Hence, the functions provided by CLOS (Common Lisp Object System) package are used. Each interface component is handled as an object just like the basic LispView interface tools.

The concept ontology for semantic concepts and world knowledge is represented in LOOM, which is a high-level programming language and environment intended for use in intelligent application programs [1]. The concepts are defined in a knowledge base and inheritance among these concepts is defined in LOOM. An editor is implemented for extension of the world knowledge when yet undefined words are encountered in the sense disambiguation process.

In the next sections, we describe the graphical interface tool and explain the functionality of each menu item. We also give sample runs of our sense disambiguator.

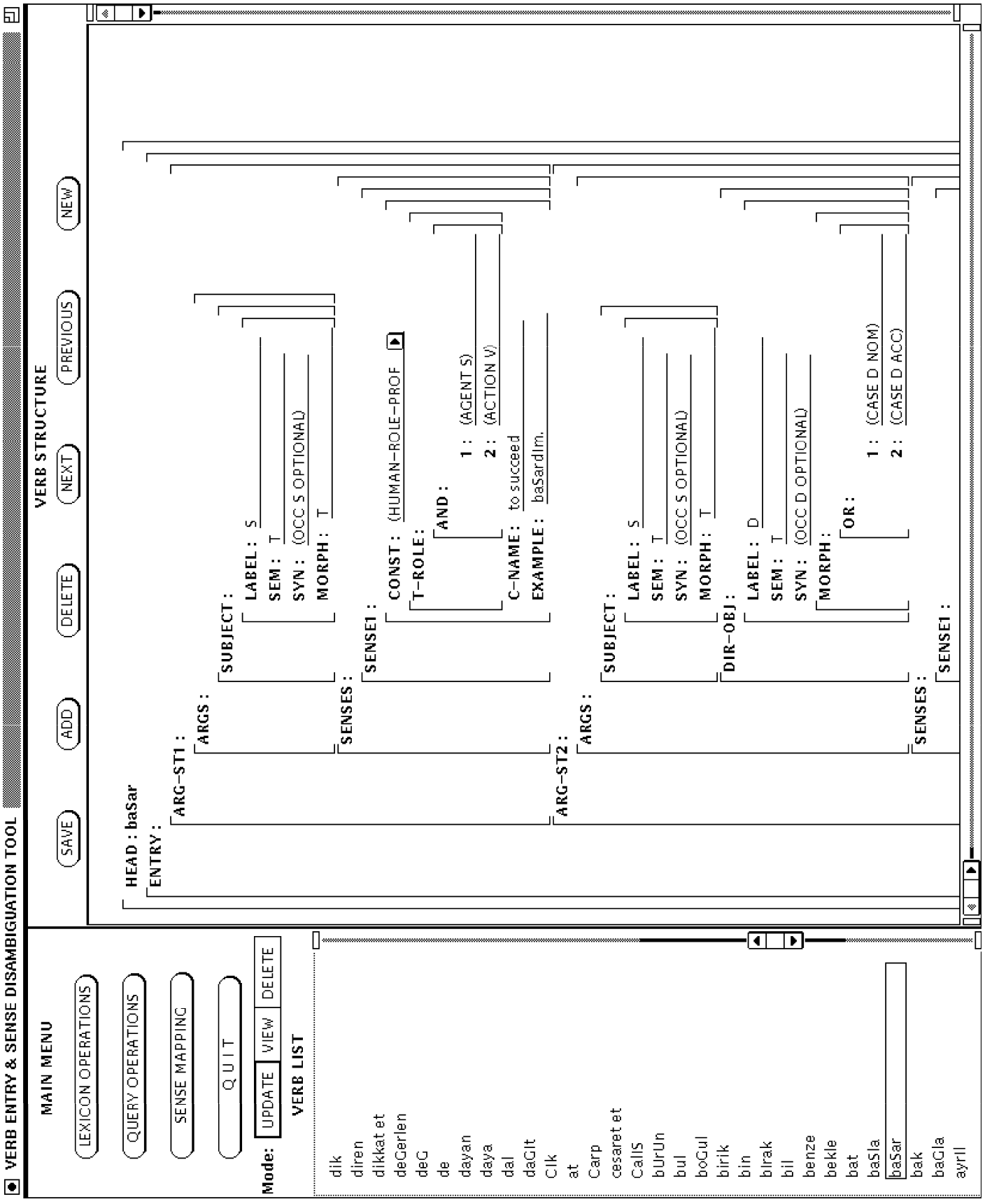

Figure 5.1. The Verb entry and Sense disambiguation tool

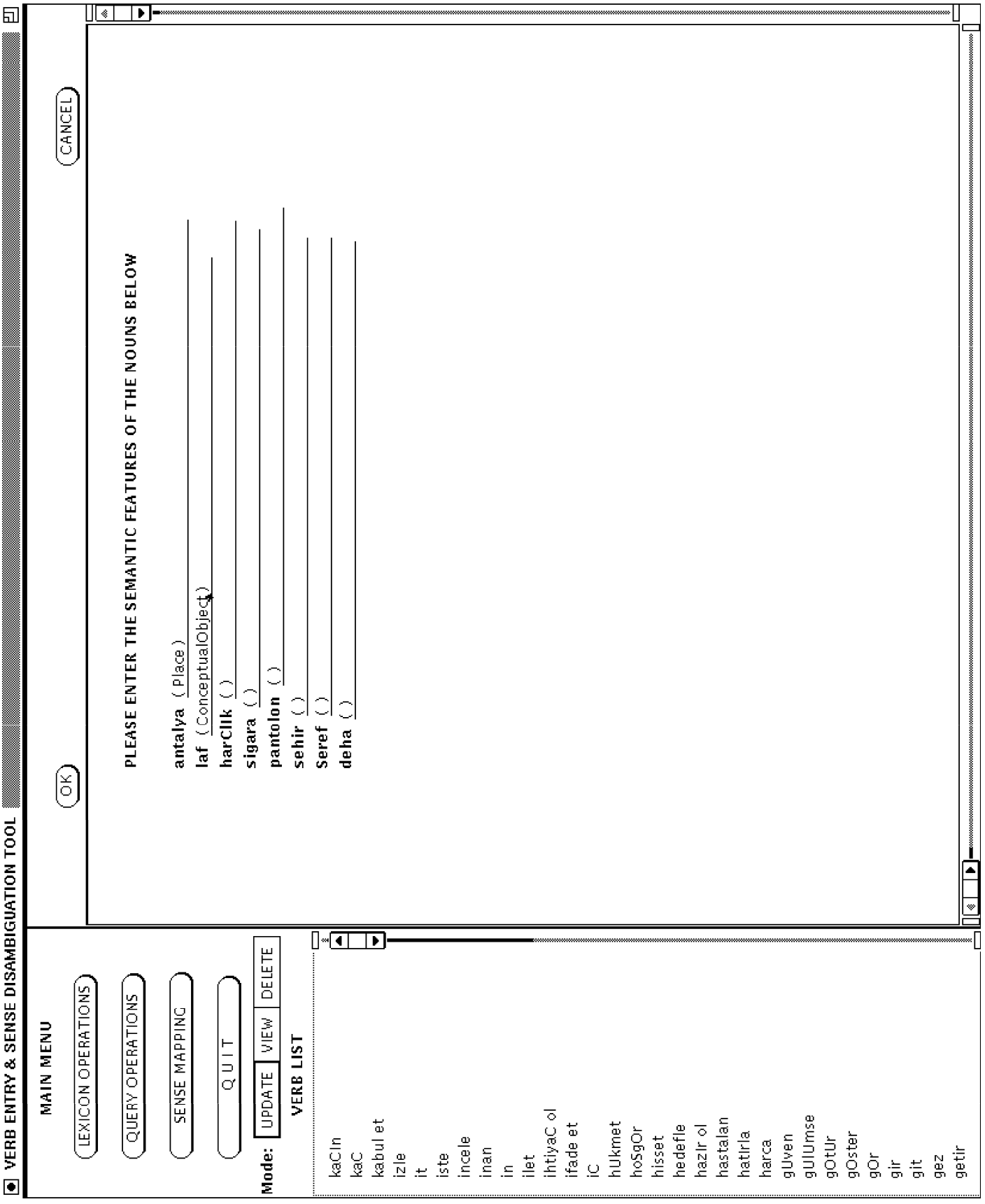

Figure 5.2. The semantic feature editor for nouns

5.1 The Verb Entry and Sense Disambiguation Tool

The tool consists of five parts:

- *Main Menu:*
  Main menu contains four menu buttons.

  - The *LEXICON OPERATIONS* activates the command window *LEXICON*. The pathname and filename of the lexicon file are specified in text fields of this window. The push buttons *SAVE* and *LOAD* are for saving and loading the lexicon file, respectively. The lexicon can be saved in a pretty-printed format, which can be edited and modified by a text editor if necessary. *PRETTY SAVE* button does this operation. The *CANCEL* button is for canceling the operation.
  - The *QUERY OPERATIONS* is for query operations in the lexicon. For example, all intransitive verbs or the verbs having direct object can be determined by using this functionality of the tool. However, it has not been implemented yet.
  - The *SENSE MAPPING* button activates the command window *SENSE MAPPING*. The pathname, test filename, and output filename are specified in text fields of the window. Users can see the trace of run by activating a flag. The push buttons *OK* and *CANCEL* are for continuing and cancelling the sense disambiguation operation, respectively.
  - The *QUIT* button is for aborting the execution as usual.

- *Modes:*
  There are three modes in lexicon operation:

  - In the *UPDATE* mode, the entry of the verb chosen is displayed in the editor field. This structure can then be edited.
  - In the *VIEW* mode, the structure of the chosen verb is displayed, but no editing is allowed.
  - In the *DELETE* mode, the chosen verb is removed from the lexicon.

- *Scrolling verb list:*
  This scrolling list contains the verbs in the lexicon. Structure of any of these verbs can be seen by choosing name of that verb.

- *Verb structure menu buttons:*
  This menu consists of six menu buttons.

  - The *SAVE* button is used to update the edited verb in the lexicon.
  - The functionality of the *ADD* button is to add a new verb to the lexicon. This button activates a window to read the root of the verb. The buttons *OK* and *CANCEL* are for continuing and cancelling the operation, respectively.
  - The *DELETE* button is for deleting the edited verb from the lexicon.
  - The *NEXT* button is for accessing the next verb in the lexicon.
  - The *PREVIOUS* button is for accessing the previous verb in the lexicon.
  - The *NEW* button is for clearing the structure of a verb, without deleting the head of the verb.

- *Editor*
  The structure of the verb is displayed in this window. Each headword, i.e., `ENTRY`, `ARGS`, `SENSE1`, etc., has its own menu and these menus are popped up when left button of the mouse is clicked on that word. Each part of the structure can be zoomed in by double-clicking mouse on left parenthesis covering that part. Viewport of the editor can be moved by scrollbars.

5.2 Sample Runs

The input to the sense disambiguator is a syntactically analyzed text. Even though this text is syntactically disambiguated by a parser, the ambiguities of verb senses must be resolved.

The sense disambiguator gets the data from an input file. Sentences are presented as frames with various slots for predicate, subject, and objects of

the sentence syntactically and morphologically analyzed. For example, the sentence *Sıcakların artmasıyla insanlar Antalya'ya aktı* is represented as:

```
((SENTENCE "sIcaklarIn artmasIyla insanlar antalya'ya aktI.")
 (VERB (*CAT* V) (*R* "ak") (*ARG* 3SG) (TENSE PAST))
 (SUBJECT ((*CAT* N) (*R* "insan") (*AGR* 3PL) (*CASE* NOM)))
 (OBJECT-1 ((*CAT* N) (*R* "antalya") (*AGR* 3SG) (*CASE* DAT)))
 (ADJUNCT ("sIcaklarIn artmasIyla")))
```

For the present, these frames are generated manually. The `SENTENCE` slot contains the whole sentence. The `ADJUNCT` slot has adjunctive complements of the verb. The `VERB`, `SUBJECT`, and `OBJECT` slots contain the verb, the subject, and objects of the sentence. the category of arguments are given in `*CAT*` slot. The `*R*` slot contains the root of the argument. The `TENSE` slot represents the tense of the verb. Finally, agreement of arguments are given in `*AGR*` slot [1].

The sense disambiguation process is explained step by step in the output file. Besides, the semantic information about the resolved sense and all the constraints that are tested are given. Some disambiguation examples are illustrated below:

Example 1:

Input Sentence:
Sıcakların artmasıyla insanlar Antalya'ya aktı.[2]

Input List:

```
((SENTENCE "sIcaklarIn artmasIyla insanlar antalya'ya aktI.")
 (VERB (*CAT* V) (*R* "ak") (*ARG* 3SG) (TENSE PAST))
 (SUBJECT ((*CAT* N) (*R* "insan") (*AGR* 3PL) (*CASE* NOM)))
 (OBJECT-1 ((*CAT* N) (*R* "antalya") (*AGR* 3SG) (*CASE* DAT)))
 (ADJUNCT ("sIcaklarIn artmasIyla")))
```

[1] These are the output of the morphological analysis.

[2] The special Turkish letters are represented using uppercase ASCII code of the nearest character.

Output:

```
**************************************************************
The sentence to test is:

 sIcaklarIn artmasIyla insanlar antalya'ya aktI.
**************************************************************

==============================================================

The lexicon verb "ak" has 2 different argument structures

==============================================================

          ARGUMENT STRUCTURAL MAPPING

--------------------------------------------------------------

.. testing constraints of ARGUMENT STRUCTURE#1
.... ARGUMENT STRUCTURE#1 has 3 different SENSES.
.... Verb has no OBJECTS in this ARGUMENT STRUCTURE.

.... * Analyzing the given Sentence *

....... "insan" is the SUBJECT
......... The verb has object/s.
......... No mapping possible!
--------------------------------------------------------------

.. testing constraints of ARGUMENT STRUCTURE#2
.... ARGUMENT STRUCTURE#2 has 2 different SENSES.
.... Verb can be used with 1 optional OBJECT.

.... The OBJECT is:
....... A OBLIQUE-OBJECT

.... * Analyzing The given Sentence *
```

```
....... "insan" is the SUBJECT
....... "antalya" is an OBLIQUE-OBJECT

.... Testing constraints of the SUBJECT "insan"

....... Testing SEMANTIC constraints
....... Satisfied:
           ((OR
                 (1 HUMAN-ROLE-PROFESSION S)
                 (2 LIQUID S)))

....... Testing SYNTACTIC constraints
....... Satisfied:
           (OCC S OPTIONAL)

....... Testing MORPHOLOGICAL constraints
....... Satisfied:
           T

.... Testing constraints of the OBLIQUE OBJECT "antalya"

....... Testing SEMANTIC constraints
....... Satisfied:
           T

....... Testing SYNTACTIC constraints
....... Satisfied:
           (OCC D OPTIONAL)

....... Testing MORPHOLOGICAL constraints
....... Satisfied:
           (CASE D DAT)

.... Constraints of the ARGUMENT STRUCTURE#2 are satisfied.

....... Searching the correct SENSE in this ARGUMENT STRUCTURE
```

```
....... Checking constraints of SENSE#1
....... Satisfied:
          ((AND
                (1 HUMAN-ROLE-PROFESSION S)
                (2 PLACE D)))

oooooooooo CONCEPT-NAME of this sense is "to move, to flow".
oooooooooo An EXAMPLE : "insanlar istanbul'a aktI".
oooooooooo AGENT of "ak" is the SUBJECT: "insan".
oooooooooo GOAL of "ak" is the OBLIQUE-OBJECT: "antalya".
oooooooooo The predicate "ak" is an ACTION-PROCESS verb.

....... Checking constraints of SENSE#2
....... Not satisfied:
          (LIQUID S)
```

This example illustrates a simple sense disambiguation process. The Turkish verb *akmak* (*to flow*) has two structures of usage. In the first structure, there can be an object in nominative or accusative case and in the second, the verb can only be used with a direct object having a dative case marker. This difference in morphological structure between the usage of verbs in senses allows us to eliminate the senses *to wear out* (*kumaş*), *to mix up* (*boya*), *to slip away* (for humans). Although, there is no structural difference between the usages of *ak* in sense *to move* (somewhere) and *to flow* (for liquids), *insan* is not a liquid and mapping to this sense fails.

After determining the correct sense of the usage of *akmak* (*to flow*), the concept name of that sense, an example, thematic roles of arguments, and semantic category of the verb are output.

Example 2:

Input Sentence:
Gözleri elaya kaçıyordu.

Input List:

```
((SENTENCE "gOzleri elaya kaCIyordu")
 (VERB ((*CAT* V) (*R* "kaC") (*AGR* 3SG) (TENSE PS-CONT)))
 (SUBJECT ((*CAT* N) (*R* "gOz") (*CASE* NOM) (*AGR* 3PL)))
 (OBJECT-1 ((*CAT* N) (*R* "ela") (*CASE* DAT) (*AGR* 3SG))))
```

Output:

```
**********************************************************************
The sentence to test is:

 gOzleri elaya kaCIyordu.
**********************************************************************

======================================================================

The lexicon verb "kaC" has 3 different argument structures

======================================================================

          ARGUMENT STRUCTURAL MAPPING

----------------------------------------------------------------------

.. testing constraints of ARGUMENT STRUCTURE#1
.... ARGUMENT STRUCTURE#1 has 5 different SENSES.
.... Verb has no OBJECTS in this ARGUMENT STRUCTURE.

.... * Analyzing the given Sentence *

....... "gOz" is the SUBJECT
......... The verb checked has object/s.
......... No mapping possible!
----------------------------------------------------------------------

.. testing constraints of ARGUMENT STRUCTURE#2
.... ARGUMENT STRUCTURE#2 has 4 different SENSES.
.... Verb can be used with 1 optional OBJECT.
```

```
.... The OBJECT is:
....... AN OBLIQUE-OBJECT

.... * Analyzing The given Sentence *

....... "gOz" is the SUBJECT
....... "ela" is the first OBLIQUE-OBJECT

.... Testing constraints of the SUBJECT "gOz"

....... Testing SEMANTIC constraints
....... Satisfied:
          T

....... Testing SYNTACTIC constraints
....... Satisfied:
          (OCC S OPTIONAL)

....... Testing MORPHOLOGICAL constraints
....... Satisfied:
          T

.... Testing constraints of the first OBLIQUE OBJECT "ela"

....... Testing SEMANTIC constraints
....... Satisfied:
          T

....... Testing SYNTACTIC constraints
....... Satisfied:
          (OCC O1 OPTIONAL)

....... Testing MORPHOLOGICAL constraints
....... Satisfied:
          (CASE O1 DAT)
```

```
.... Constraints of the ARGUMENT STRUCTURE#2 are satisfied.

....... Searching the correct SENSE in this ARGUMENT STRUCTURE
....... Checking constraints of SENSE#1
....... Not satisfied:
          ((AND
                (1
                     (OR
                          (1 ANIMATEOBJECT S)
                          (2 LIQUID O1)))
                (2 HUMANPART S)))

....... Checking constraints of SENSE#2
....... Not satisfied:
          ((AND
                (1 INANIMATEOBJECT S)
                (2 PREPOSITIONAL O1)))

....... Checking constraints of SENSE#3
....... Satisfied:
          ((OR
                (1 CONCEPTUALOBJECT O1)
                (2 COLOR O1)))

ooooooooo CONCEPT-NAME of this sense is "to tend toward, to be
          more like".
ooooooooo An EXAMPLE : "ceket yeSile kaCIyordu. AdamIn sOzleri
          mUbalaya kaCIyordu.".
ooooooooo THEME of "kaC" is the SUBJECT: "gOz".
ooooooooo GOAL of "kaC" is the first OBLIQUE-OBJECT: "ela".
ooooooooo The predicate "kaC" is a STATE verb.

....... Checking constraints of SENSE#4
....... Not Satisfied because more specific sense exists
-----------------------------------------------------------------

.. testing constraints of ARGUMENT STRUCTURE#3
```

```
.... ARGUMENT STRUCTURE#3 has 4 different SENSES.
.... Verb can be used with 1 optional OBJECT.

.... The OBJECT is:
....... AN OBLIQUE-OBJECT

.... * Analyzing The given Sentence *

....... "gOz" is the SUBJECT

......... Morphological constraints are not satisfied
......... Because:
........... Case of the OBJECT "ela" is DAT and it does not
........... satisfy morphological  constraints of any of the
........... OBJECTS of ARGUMENT STRUCTURE#3.
```

This example illustrates the usage of the semantic constraints. The verb *kaç* (*to escape from*) has three argument structures. The first argument structure has no objects, but in the input sentence, *ela* (*hazel*) is subcategorized as the object. Thus, it is not possible to map to this structure. In the third argument structure, *kaç* does not subcategorize an object in dative case. However, *ela* (*hazel*) is in dative case. Hence, mapping to this argument structure is not realized. On the other hand, all the constraints of the second argument structure are satisfied. Since some of semantic constraints are not satisfied, mapping to the senses *to run into* and *to slip to* (for inanimate objects) is also impossible. The correct sense is determined as *to tend toward, to be more like*, and hence mapping to "default" sense is rejected.

Example 3.

Input Sentence:
Hergün dayak yiyen çocuk sonunda evden kaçtı.

Input List:

```
((SENTENCE "hergUn dayak yiyen Cocuk sonunda evden kaCtI")
```

```
(VERB ((*CAT* V) (*R* "kaC") (*AGR* 3SG) (TENSE PAST)))
(SUBJECT ((MODIFIED
            ((*CAT* N) (*R* "Cocuk") (*AGR* 3SG) (*CASE* NOM))
          (MODIFIER "hergUn dayak yiyen")))
(OBJECT-1 ((*CAT* N) (*R* "ev") (*AGR* 3SG) (*CASE* ABL)))
(ADJUNCT ("sonunda"))))
```

Output:

```
**********************************************************************
The sentence to test is:

 hergUn dayak yiyen Cocuk sonunda evden kaCtI.
**********************************************************************

======================================================================

The lexicon verb "kaC" has 3 different argument structures

======================================================================

          ARGUMENT STRUCTURAL MAPPING

----------------------------------------------------------------------

.. testing constraints of ARGUMENT STRUCTURE#1
.... ARGUMENT STRUCTURE#1 has 5 different SENSES.
.... Verb has no OBJECTS in this ARGUMENT STRUCTURE.

.... * Analyzing the given Sentence *

....... "Cocuk" is the SUBJECT
......... The verb has object/s.
......... No mapping possible!
----------------------------------------------------------------------

.. testing constraints of ARGUMENT STRUCTURE#2
```

```
.... ARGUMENT STRUCTURE#2 has 4 different SENSES.
.... Verb can be used with 1 optional OBJECT.

.... The OBJECT is:
....... AN OBLIQUE-OBJECT

.... * Analyzing The given Sentence *

....... "Cocuk" is the SUBJECT

......... Morphological constraints are not satisfied
......... Because:
........... Case of the OBJECT "ev" is ABL and it does not
........... satisfy morphological  constraints of any of the
........... OBJECTS of ARGUMENT STRUCTURE#2.
-----------------------------------------------------------------

.. testing constraints of ARGUMENT STRUCTURE#3
.... ARGUMENT STRUCTURE#3 has 4 different SENSES.
.... Verb can be used with 1 optional OBJECT.

.... The OBJECT is:
....... AN OBLIQUE-OBJECT

.... * Analyzing The given Sentence *

....... "Cocuk" is the SUBJECT
....... "ev" is the first OBLIQUE-OBJECT

.... Testing constraints of the SUBJECT "Cocuk"

....... Testing SEMANTIC constraints
....... Satisfied:
          T

....... Testing SYNTACTIC constraints
....... Satisfied:
```

```
          (OCC S OPTIONAL)

..... Testing MORPHOLOGICAL constraints
..... Satisfied:
          T

.... Testing constraints of the first OBLIQUE OBJECT "ev"

..... Testing SEMANTIC constraints
..... Satisfied:
          T

..... Testing SYNTACTIC constraints
..... Satisfied:
          (OCC O1 OPTIONAL)

..... Testing MORPHOLOGICAL constraints
..... Satisfied:
          (CASE O1 ABL)

.... Constraints of the ARGUMENT STRUCTURE#3 are satisfied.

..... Searching the correct SENSE in this ARGUMENT STRUCTURE
..... Checking constraints of SENSE#1
..... Not satisfied:
          ((AND
                (1 HUMAN-ROLE-PROFESSION S)
                (2
                     (OR
                          (1 CONCEPTUALOBJECT O1)
                          (2 DOING-ACTION O1)
                          (3 SENTIMENT-MENTALACTIVITY O1)))))

..... Checking constraints of SENSE#2
..... Not satisfied:
          ((OR
                (1 GAS S)
```

```
                (2 LIQUID S)))

....... Checking constraints of SENSE#3
....... Not satisfied:
          ((AND
                (1 HUMAN-ROLE-PROFESSION S)
                (2 HUMAN-ROLE-PROFESSION O1)))

....... Checking constraints of SENSE#4
....... Satisfied:
          T

oooooooooo CONCEPT-NAME of this sense is "to escape from".
oooooooooo An EXAMPLE : "hapishaneden kaCmayI baSardIlar.".
oooooooooo AGENT of "kaC" is the SUBJECT: "Cocuk".
oooooooooo SOURCE of "kaC" is the first OBLIQUE-OBJECT: "ev".
oooooooooo The predicate "kaC" is a PROCESS-ACTION verb.
```

Here, *kaçmak* (*to escape from*) means something different than that of Example#2. In this case, the object *ev* (*home*) has an ablative case marker. Therefore, the constraints of the third argument structure are satisfied. Constraints of first three senses (*to avoid*, *to leak out*, and *to stay away from*) can not be satisfied and the correct sense is determined as the "default" sense of this structure: *to escape from.*

Example 4.

Input Sentence:
Levent'in tüm harçlığı sigaraya gidiyor.

Input List:

```
((SENTENCE "levent'in tUm harClIGI sigaraya gidiyor.")
 (VERB ((*CAT* V) (*R* "git") (*AGR* 3SG) (TENSE PS-CONT)))
 (SUBJECT ((MODIFIED
              (*CAT* N) (*R* "harClIk") (*CASE* NOM) (*AGR* 3SG))
           (MODIFIER "levent'in tUm")))
```

```
 (OBJECT-1 ((*CAT* N) (*R* "sigara") (*CASE* DAT) (*AGR* 3SG))))
```

Output:

```
*****************************************************************
The sentence to test is:

 levent'in tUm harClIGI sigaraya gidiyor.
*****************************************************************

=================================================================

The lexicon verb "git" has 4 different argument structures

=================================================================

          ARGUMENT STRUCTURAL MAPPING

-----------------------------------------------------------------

.. testing constraints of ARGUMENT STRUCTURE#1
.... ARGUMENT STRUCTURE#1 has 3 different SENSES.
.... Verb has no OBJECTS in this ARGUMENT STRUCTURE.

.... * Analyzing the given Sentence *

....... "harClIk" is the SUBJECT
......... The verb has object/s.
......... No mapping possible!
-----------------------------------------------------------------

.. testing constraints of ARGUMENT STRUCTURE#2
.... ARGUMENT STRUCTURE#2 has 5 different SENSES.
.... Verb can be used with 1 optional OBJECT.

.... The OBJECT is:
....... AN OBLIQUE-OBJECT
```

```
.... * Analyzing The given Sentence *

....... "harClIk" is the SUBJECT
....... "sigara" is the first OBLIQUE-OBJECT

.... Testing constraints of the SUBJECT "harClIk"

....... Testing SEMANTIC constraints
....... Satisfied:
          T

....... Testing SYNTACTIC constraints
....... Satisfied:
          (OCC S OPTIONAL)

....... Testing MORPHOLOGICAL constraints
....... Satisfied:
          T

.... Testing constraints of the first OBLIQUE OBJECT "sigara"

....... Testing SEMANTIC constraints
....... Satisfied:
          T

....... Testing SYNTACTIC constraints
....... Satisfied:
          T

....... Testing MORPHOLOGICAL constraints
....... Satisfied:
          (CASE O1 DAT)

.... Constraints of the ARGUMENT STRUCTURE#2 are satisfied.

....... Searching the correct SENSE in this ARGUMENT STRUCTURE
```

```
....... Checking constraints of SENSE#1
....... Not satisfied:
          ((AND
                (1 CONCEPTUALOBJECT S)
                (2
                     (OR
                          (1 PLACE O1)
                          (2 CONCEPTUALOBJECT O1)))))

....... Checking constraints of SENSE#2
....... Not satisfied:
          ((AND
                (1
                     (OR
                          (1 CLOTHING S)
                          (2 COLOR S)))
                (2 HUMAN-ROLE-PROFESSION O1)))

....... Checking constraints of SENSE#3
....... Satisfied:
          ((OR
                (1 MONETARY S)
                (2 COMMODITY-WARE O1)))

oooooooooo CONCEPT-NAME of this sense is "to be spend on".
oooooooooo An EXAMPLE : "paranIn coGu benzine gidiyor.".
oooooooooo THEME of "git" is the SUBJECT: "harClIk".
oooooooooo GOAL of "git" is the first OBLIQUE-OBJECT: "sigara".
oooooooooo The predicate "git" is an PROCESS verb.


....... Checking constraints of SENSE#4
....... Not satisfied:
          ((AND
                (1
                     (OR
                          (1 HUMAN-ROLE-PROFESSION S)
```

```
                              (2 HUMANGROUP S)
                              (3 ORGANIZATION S)))
                    (2 ARTIFICIALPHENOMENON 01)
                    (3 OCC 01 OBLIGATORY)))

....... Checking constraints of SENSE#5
....... Not satisfied:
           ((AND
                 (1
                      (OR
                           (1 MONETARY S)
                           (2 COMMODITY-WARE S)))
                 (2 HUMAN-ROLE-PROFESSION 01)))

-----------------------------------------------------------------

.. testing constraints of ARGUMENT STRUCTURE#3
....  ARGUMENT STRUCTURE#3 has only 1 SENSE.
.... Verb can be used with 1 optional OBJECT.

.... The OBJECT is:
....... AN OBLIQUE-OBJECT

.... * Analyzing The given Sentence *

....... "harClIk" is the SUBJECT

......... Morphological constraints are not satisfied
......... Because:
.......... Case of the OBJECT "sigara" is DAT and it does not
.......... satisfy morphological  constraints of any of the
.......... OBJECTS of ARGUMENT STRUCTURE#3.
-----------------------------------------------------------------

.. testing constraints of ARGUMENT STRUCTURE#4
....  ARGUMENT STRUCTURE#4 has only 1 SENSE.
.... Verb can be used with 2 optional OBJECTS.
```

```
.... The OBJECTS are:
....... 2 OBLIQUE-OBJECTS

.... * Analyzing The given Sentence *

....... "harClIk" is the SUBJECT
....... "sigara" is the second OBLIQUE-OBJECT

.... Testing constraints of the SUBJECT "harClIk"

....... Testing SEMANTIC constraints
....... Satisfied:
          T

....... Testing SYNTACTIC constraints
....... Satisfied:
          (OCC S OPTIONAL)

....... Testing MORPHOLOGICAL constraints
....... Satisfied:
          T

.... Testing constraints of the second OBLIQUE OBJECT "sigara"

....... Testing SEMANTIC constraints
....... Satisfied:
          T

....... Testing SYNTACTIC constraints
....... Satisfied:
          (OCC O2 OPTIONAL)

....... Testing MORPHOLOGICAL constraints
....... Satisfied:
          (CASE O2 DAT)
```

```
.... Constraints of the ARGUMENT STRUCTURE#4 are satisfied.

....... Searching the correct SENSE in this ARGUMENT STRUCTURE
....... Checking constraints of SENSE#1
....... Not satisfied:
           (ANIMATEOBJECT S)
```

Since *istemek* (to want) does not subcategorize any objects in dative case when it is used in senses *to last* (for products), *to work* (for products like watch), *to go on* (for events), and *to be sold*, mapping to these senses is automatically eliminated. For the senses *to lead*, *to suit*, *to perform*, *to last*, and *to go from* some constraints can not be satisfied. Therefore, the correct sense is determined as *to be spend on*.

Example 5.

Input Sentence:
Ben senden bir ekmek parası istemiştim.

Input List:

```
((SENTENCE "ben senden bir ekmek parasI istemiStim.")
 (VERB ((*CAT* V) (*R* "iste") (*AGR* 1SG) (TENSE PAST)))
 (SUBJECT ((*CAT* N) (*R* "ben") (*CASE* NOM) (*AGR* 1SG)))
 (OBJECT-1 ((*CAT* N) (*R* "sen") (*CASE* ABL) (*AGR* 2SG)))
 (OBJECT-2 ((MODIFIED
                ((*CAT* N) (*R* "para") (*CASE* NOM) (*AGR* 3SG)))
            (MODIFIER "bir ekmek"))))
```

Output:

```
**********************************************************************
The sentence to test is:

 ben senden bir ekmek parasI istemiStim.
**********************************************************************
```

```
================================================================

The lexicon verb "iste" has 2 different argument structures

================================================================

          ARGUMENT STRUCTURAL MAPPING

----------------------------------------------------------------

.. testing constraints of ARGUMENT STRUCTURE#1
.... ARGUMENT STRUCTURE#1 has 4 different SENSES.
.... Verb can be used with 1 optional OBJECT.

.... The OBJECT is:
....... A DIRECT-OBJECT

.... * Analyzing The given Sentence *

....... "ben" is the SUBJECT
....... "para" is the DIRECT-OBJECT

......... Morphological constraints are not satisfied
......... Because:
.......... Case of the OBJECT "sen" is ABL and it does not
.......... satisfy morphological  constraints of any of the
.......... OBJECTS of ARGUMENT STRUCTURE#1.
----------------------------------------------------------------

.. testing constraints of ARGUMENT STRUCTURE#2
....  ARGUMENT STRUCTURE#2 has only 1 SENSE.
.... Verb can be used with 2 optional OBJECTS.

.... The OBJECTS are:
....... A DIRECT-OBJECT
....... AN OBLIQUE-OBJECT
```

```
.... * Analyzing The given Sentence *

....... "ben" is the SUBJECT
....... "para" is the DIRECT-OBJECT
....... "sen" is the first OBLIQUE-OBJECT

.... Testing constraints of the SUBJECT "ben"

....... Testing SEMANTIC constraints
....... Satisfied:
           T

....... Testing SYNTACTIC constraints
....... Satisfied:
           (OCC S OPTIONAL)

....... Testing MORPHOLOGICAL constraints
....... Satisfied:
           T

.... Testing constraints of the DIRECT OBJECT "para"

....... Testing SEMANTIC constraints
....... Satisfied:
           T

....... Testing SYNTACTIC constraints
....... Satisfied:
           (OCC D OPTIONAL)

....... Testing MORPHOLOGICAL constraints
....... Satisfied:
           ((OR
                 (1 CASE D NOM)
                 (2 CASE D ACC)))
```

```
.... Testing constraints of the first OBLIQUE OBJECT "sen"

....... Testing SEMANTIC constraints
....... Satisfied:
           T

....... Testing SYNTACTIC constraints
....... Satisfied:
           (OCC O1 OPTIONAL)

....... Testing MORPHOLOGICAL constraints
....... Satisfied:
           (CASE O1 ABL)

.... Constraints of the ARGUMENT STRUCTURE#2 are satisfied.

....... Searching the correct SENSE in this ARGUMENT STRUCTURE
....... Checking constraints of SENSE#1
....... Satisfied:
           T

oooooooooo CONCEPT-NAME of this sense is "to ask something from".
oooooooooo An EXAMPLE : "komSudan bir ekmek ister misin?".
oooooooooo AGENT of "iste" is the SUBJECT: "ben".
oooooooooo THEME of "iste" is the DIRECT-OBJECT: "para".
oooooooooo SOURCE of "iste" is the first OBLIQUE-OBJECT: "sen".
oooooooooo The predicate "iste" is a PROCESS-ACTION verb.
```

Since the verb takes an oblique object in ablative case, the senses *to ask for marriage*, *to ask for someone*, *to require* and *to necessitate* are automatically eliminated and "default" sense is determined as the correct sense.

Example 6.

Input Sentence:
Biz dün akşam biraz içtik.

Input List:

```
((SENTENCE "biz dUn akSam biraz iCtik")
 (VERB ((*CAT* V) (*R* "iC") (*AGR* 1PL) (TENSE PAST)))
 (SUBJECT ((*CAT* N) (*R* "biz") (*CASE* NOM) (*AGR* 1PL)))
 (ADJUNCT ("dUn akSam biraz")))
```

Output:

```
**********************************************************************
The sentence to test is:

 Biz dUn akSam biraz iCtik.
**********************************************************************

======================================================================

The lexicon verb "iC" has 3 different argument structures

======================================================================

          ARGUMENT STRUCTURAL MAPPING

----------------------------------------------------------------------

.. testing constraints of ARGUMENT STRUCTURE#1
....  ARGUMENT STRUCTURE#1 has only 1 SENSE.
.... Verb has no OBJECTS in this ARGUMENT STRUCTURE.

.... * Analyzing the given Sentence *

....... "biz" is the SUBJECT
.... Testing constraints of the SUBJECT "biz"

....... Testing SEMANTIC constraints
....... Satisfied:
```

```
          T

....... Testing SYNTACTIC constraints
....... Satisfied:
          (OCC S OPTIONAL)

....... Testing MORPHOLOGICAL constraints
....... Satisfied:
          T

.... Constraints of the ARGUMENT STRUCTURE#1 are satisfied.

....... Searching the correct SENSE in this ARGUMENT STRUCTURE
....... Checking constraints of SENSE#1
....... Satisfied:
          T

oooooooooo CONCEPT-NAME of this sense is "to get alcohol".
oooooooooo An EXAMPLE : "akSam biraz iCtik.".
oooooooooo AGENT of "iC" is the SUBJECT: "biz".
oooooooooo The predicate "iC" is an ACTION-PROCESS verb.


-----------------------------------------------------------------

.. testing constraints of ARGUMENT STRUCTURE#2
.... ARGUMENT STRUCTURE#2 has 3 different SENSES.
.... Verb can be used with 1 optional OBJECT.

.... The OBJECT is:
....... A DIRECT-OBJECT

.... * Analyzing The given Sentence *

....... "biz" is the SUBJECT

.... Testing constraints of the SUBJECT "biz"
```

```
....... Testing SEMANTIC constraints
....... Satisfied:
          T

....... Testing SYNTACTIC constraints
....... Satisfied:
          (OCC S OPTIONAL)

....... Testing MORPHOLOGICAL constraints
....... Satisfied:
          T

.... Constraints of the ARGUMENT STRUCTURE#2 are satisfied.

....... Searching the correct SENSE in this ARGUMENT STRUCTURE
....... Checking constraints of SENSE#1
....... Not satisfied:
          ((AND
                (1 HUMAN-ROLE-PROFESSION S)
                (2
                     (OR
                          (1 IS D "nargile")
                          (2 IS D "sigara")
                          (3 IS D "pipo")
                          (4 IS D "puro")))))

....... Checking constraints of SENSE#2
....... Satisfied:
          ((AND
                (1 HUMAN-ROLE-PROFESSION S)
                (2 LIQUID D)))

oooooooooo CONCEPT-NAME of this sense is "to drink".
oooooooooo An EXAMPLE : "meyva suyu iCelim mi?".
oooooooooo AGENT of "iC" is the SUBJECT: "biz".
oooooooooo The predicate "iC" is an ACTION-PROCESS verb.
```

```
....... Checking constraints of SENSE#3
....... Not Satisfied because more specific sense exists
-------------------------------------------------------------

.. testing constraints of ARGUMENT STRUCTURE#3
....  ARGUMENT STRUCTURE#3 has only 1 SENSE.
.... Verb can be used with 2 optional OBJECTS.

.... The OBJECTS are:
....... A DIRECT-OBJECT
....... AN OBLIQUE-OBJECT

.... * Analyzing The given Sentence *

....... "biz" is the SUBJECT

.... Testing constraints of the SUBJECT "biz"

....... Testing SEMANTIC constraints
....... Satisfied:
          T

....... Testing SYNTACTIC constraints
....... Satisfied:
          (OCC S OPTIONAL)

....... Testing MORPHOLOGICAL constraints
....... Satisfied:
          T

.... Constraints of the ARGUMENT STRUCTURE#3 are satisfied.

....... Searching the correct SENSE in this ARGUMENT STRUCTURE
....... Checking constraints of SENSE#1
....... Not satisfied:
```

```
((AND
      (1 HUMAN-ROLE-PROFESSION S)
      (2 HUMAN-ROLE-PROFESSION D)
      (3 CONCEPTUALOBJECT O1)
      (4 OCC O1 OBLIGATORY)))
```

In this example, *içmek* (*to drink*) is used as an intransitive verb. Since the object is not specified, two ambiguous senses *to get alcohol, to drink* can not be resolved. Both of them refer to *to drink some liquid.*

Example 7.

Input Sentence:
Biz kimin Serefine içiyoruz?

Input List:

```
((SENTENCE "biz kimin Serefine iCiyoruz")
 (VERB ((*CAT* V) (*R* "iC") (*AGR* 1PL) (TENSE PR-CONT)))
 (SUBJECT ((*CAT* N) (*R* "biz") (*AGR* 1PL) (*CASE* NOM)))
 (ADJUNCT "kimin Serefine"))
```

Output:

```
**************************************************************
The sentence to test is:

 biz kimin Serefine iCiyoruz.
**************************************************************

==============================================================

The lexicon verb "iC" has 3 different argument structures

==============================================================
```

```
          ARGUMENT STRUCTURAL MAPPING

-------------------------------------------------------------

.. testing constraints of ARGUMENT STRUCTURE#1
....  ARGUMENT STRUCTURE#1 has only 1 SENSE.
.... Verb has no OBJECTS in this ARGUMENT STRUCTURE.

.... * Analyzing the given Sentence *

....... "biz" is the SUBJECT
......... The verb has object/s.
......... No mapping possible!
-------------------------------------------------------------

.. testing constraints of ARGUMENT STRUCTURE#2
.... ARGUMENT STRUCTURE#2 has 3 different SENSES.
.... Verb can be used with 1 optional OBJECT.

.... The OBJECT is:
....... A DIRECT-OBJECT

.... * Analyzing The given Sentence *

....... "biz" is the SUBJECT

......... Morphological constraints are not satisfied
......... Because:
........... Case of the OBJECT "Seref" is DAT and it does not
........... satisfy morphological  constraints of any of the
........... OBJECTS of ARGUMENT STRUCTURE#2.
-------------------------------------------------------------

.. testing constraints of ARGUMENT STRUCTURE#3
....  ARGUMENT STRUCTURE#3 has only 1 SENSE.
.... Verb can be used with 2 optional OBJECTS.
```

```
.... The OBJECTS are:
....... A DIRECT-OBJECT
....... AN OBLIQUE-OBJECT

.... * Analyzing The given Sentence *

....... "biz" is the SUBJECT
....... "Seref" is the first OBLIQUE-OBJECT

.... Testing constraints of the SUBJECT "biz"

....... Testing SEMANTIC constraints
....... Satisfied:
          T

....... Testing SYNTACTIC constraints
....... Satisfied:
          (OCC S OPTIONAL)

....... Testing MORPHOLOGICAL constraints
....... Satisfied:
          T

.... Testing constraints of the first OBLIQUE OBJECT "Seref"

....... Testing SEMANTIC constraints
....... Satisfied:
          T

....... Testing SYNTACTIC constraints
....... Satisfied:
          T

....... Testing MORPHOLOGICAL constraints
....... Satisfied:
          (CASE O1 DAT)
```

```
.... Constraints of the ARGUMENT STRUCTURE#3 are satisfied.

....... Searching the correct SENSE in this ARGUMENT STRUCTURE
....... Checking constraints of SENSE#1
....... Satisfied:
           ((AND
                 (1 HUMAN-ROLE-PROFESSION S)
                 (2 HUMAN-ROLE-PROFESSION D)
                 (3 CONCEPTUALOBJECT O1)
                 (4 OCC O1 OBLIGATORY)))

ooooooooooo CONCEPT-NAME of this sense is "to drink for".
ooooooooooo An EXAMPLE : "Serefine iCmek. saGlIGIna icmek.".
ooooooooooo AGENT of "iC" is the SUBJECT: "biz".
ooooooooooo GOAL of "iC" is the first OBLIQUE-OBJECT: "Seref".
ooooooooooo The predicate "iC" is an ACTION-PROCESS verb.
```

In this case, the object having a dative case marker plays the most important role. The constraints for the sense *to drink for* are satisfied and this sense is determined as the correct sense.

Example 8.

Input Sentence:
Adam eve girdi.

Input List:

```
((SENTENCE "adam eve girdi")
 (VERB ((*CAT* V) (*R* "gir") (TENSE PAST) (*AGR* 3SG)))
 (SUBJECT ((*CAT* N) (*R* "adam") (*AGR* 3SG) (*CASE* NOM)))
 (OBJECT-1 ((*CAT* N) (*R* "ev") (*CASE* DAT) (*AGR* 3SG))))
```

Output:

```
**************************************************************
```

```
The sentence to test is:

 adam eve girdi.
**************************************************************

==============================================================

The lexicon verb "gir" has 2 different argument structures

==============================================================

          ARGUMENT STRUCTURAL MAPPING

--------------------------------------------------------------

.. testing constraints of ARGUMENT STRUCTURE#1
....  ARGUMENT STRUCTURE#1 has only 1 SENSE.
.... Verb has no OBJECTS in this ARGUMENT STRUCTURE.

.... * Analyzing the given Sentence *

....... "adam" is the SUBJECT
......... The verb has object/s.
......... No mapping possible!
--------------------------------------------------------------

.. testing constraints of ARGUMENT STRUCTURE#2
.... ARGUMENT STRUCTURE#2 has 10 different SENSES.
.... Verb can be used with 1 optional OBJECT.

.... The OBJECT is:
....... AN OBLIQUE-OBJECT

.... * Analyzing The given Sentence *

....... "adam" is the SUBJECT
....... "ev" is the first OBLIQUE-OBJECT
```

```
.... Testing constraints of the SUBJECT "adam"

....... Testing SEMANTIC constraints
....... Satisfied:
           T

....... Testing SYNTACTIC constraints
....... Satisfied:
           (OCC S OPTIONAL)

....... Testing MORPHOLOGICAL constraints
....... Satisfied:
           T

.... Testing constraints of the first OBLIQUE OBJECT "ev"

....... Testing SEMANTIC constraints
....... Satisfied:
           T

....... Testing SYNTACTIC constraints
....... Satisfied:
           (OCC O1 OPTIONAL)

....... Testing MORPHOLOGICAL constraints
....... Satisfied:
           (CASE O1 DAT)

.... Constraints of the ARGUMENT STRUCTURE#2 are satisfied.

....... Searching the correct SENSE in this ARGUMENT STRUCTURE
....... Checking constraints of SENSE#1
....... Not satisfied:
           ((AND
                 (1 HUMANPART S)
                 (2
```

```
                    (OR
                         (1 CLOTHING 01)
                         (2 MEANS 01)))))

....... Checking constraints of SENSE#2
....... Not satisfied:
          ((AND
                (1
                    (OR
                         (1 HUMAN-ROLE-PROFESSION S)
                         (2 HUMANPART S)))
                (2 CONCEPTUALOBJECT 01)))

....... Checking constraints of SENSE#3
....... Not satisfied:
          ((AND
                (1 HUMANGROUP S)
                (2 PLACE 01)))

....... Checking constraints of SENSE#4
....... Not satisfied:
          ((AND
                (1 PERCEPTION S)
                (2 HUMANPART 01)))

....... Checking constraints of SENSE#5
....... Not satisfied:
          (IS 01 "akIl")

....... Checking constraints of SENSE#6
....... Not satisfied:
          (IS 01 "birbiri")

....... Checking constraints of SENSE#7
....... Not satisfied:
          ((AND
                (1 HUMAN-ROLE-PROFESSION S)
```

```
                (2 NUMBERS O1)))

..... Checking constraints of SENSE#8
..... Not satisfied:
          ((AND
                (1 HUMAN-ROLE-PROFESSION S)
                (2 ORGANIZATION O1)))

..... Checking constraints of SENSE#9
..... Not satisfied:
          (IS O1 "mide")

..... Checking constraints of SENSE#10
..... Satisfied:
          T

oooooooooo CONCEPT-NAME of this sense is "to enter, to get in".
oooooooooo An EXAMPLE : "Once iCeri gir, sonra konuSuruz.".
oooooooooo AGENT of "gir" is the SUBJECT: "adam".
oooooooooo GOAL of "gir" is the first OBLIQUE-OBJECT: "ev".
oooooooooo The predicate "gir" is an ACTION-PROCESS verb.
```

In this example, intransitive usage is automatically eliminated since the verb is used with objects. The constraints of senses *to fit*, *to capture*, *to get* (for perceptions), *to understand* (with "akıl"), *to fight*, *to enroll*, and *to eat* can not be satisfied, and hence "default" sense is determined as the correct one.

Chapter 6

Conclusions and Suggestions

In this thesis we presented a verb lexicon and a verb sense disambiguator for Turkish. The lexicon is developed for 100 Turkish verbs and in the resolution process of ambiguous senses of verbs we dealt only with the accompanying arguments of verbs.

The lexicon can be improved by adding new verbs and the words in grammatical categories other than verbs, as well. Furthermore, as the size of lexicon grows a faster accessing algorithm can be used. A small modification in the structure of the lexical items lets us to add nouns, adjectives, adverbs, etc., i.e., an `ENTRIES` zone can be added as the parent of the `ENTRY` zone and the `ENTRY`s in different grammatical categories may be collected together in several `ENTRY` zones indexed by an integer. The category of each entry can be specified in a `CAT`egory slot as a child of the `ENTRY` zones. As a result, the structure illustrated in figure 6.1 can be obtained.

In our lexicon, we only stored semantic, syntactic, and morphological information about subject and objects. However, in order to resolve ambiguities, we sometimes need to analyze adverbs of the verbs. This problem can be eliminated by adding the information about these accompanying words. Although the lexicon contains concept name, examples, and semantic analysis information of the senses the target language (e.g., English) generation information can be added, too. As we already stated, we will incorporate adverbial complements and events into the lexicon later.

```
((HEAD . "at")
 (ENTRIES
    (ENTRY1
       (CAT . VERB)
       (ARG-STRS
           (ARG-ST1
               ...
           (ARG-ST2
               ...
                  )))
    (ENTRY2
       (CAT . NOUN)
       (SENSES
           (SENSE1
               ...
                  ))))
```

Figure 6.1. The Structure of the Lexicon having words in all grammatical categories.

We have developed the noun lexicon for about 500 Turkish nouns, adding new nouns will improve performance. As we stated, in sense disambiguation process, we did not deal with the events as arguments. The events can be analyzed by adding a special processing node to the constraint satisfaction network. Considering adverbial complements will increase the accuracy of the sense disambiguator, too.

Bibliography

[1] D. Brill. *LOOM Users Guide.* ISX Corporation, 1.4 edition, August 1991.

[2] D. Farwell and Y. Wilks. Ultra: a Multi-lingual Machine Translator. Technical report, Memoranda in Computing and Cognitive Science, Computing Research Laboratory, New Mexico State University, 1990.

[3] W. Frawley. *Linguistic Semantics.* Lawrange Erlbaum Association Publishers, New Jersey, 1992.

[4] Z. Güngördü. A Lexical-Functional Grammar for Turkish. Master's thesis, Department of Computer Engineering and Information Sciences, Bilkent University, Ankara, Turkey, July 1993.

[5] Z. Güngördü and K. Oflazer. Parsing Turkish using the lexical-functional grammar formalism. In *Proceedings of COLING-94, the 15th International Conference on Computational Linguistics*, Kyoto,Japan, 1994.

[6] W. J. Hutckins and H. L. Somers. *An Introduction to Machine Translation.* Academic Press, London, 1992.

[7] Information Sciences Institute at the University of Southern California, Marina del Rey. *The Penman Reference Manual. The Penman Users Guide. The Penman Primer.*

[8] I. Meyer, B. Onyshknevych, and L. Carlson. Lexicographic principles and design for knowledge-based machine translation. Technical report, Carnegie Melon University, 1990.

[9] M. Nagao. Some dictionary information for machine translation. In C. wing Guo ed., editor, *Machine Tractable Dictionaries-Design and Construction*, pages 82–94. Alex Pub., Manchester, July 30-31 1992.

[10] M. Nagao, J. Tsujii, and J. Nakamura. The Japanese Government Project for Machine Translation. In *Computational Linguistics*, volume 11. April-September 1985.

[11] S. Nirenburg and C. Defrise. Practical Computational Linguistics. In M. R. R. Johnson, editor, *Computational Linguistics and Formal Semantics*. Cambridge University Press, 1991.

[12] K. Oflazer. Two-level description of Turkish morphology. In *Proceedings of the Sixth Conference of the European Chapter of the Association for Computational Linguistics*, April 1993. A full version appears in *Literary and Linguistic Computing*, Vol.9 No.2, 1994.

[13] Sun Microsystems, Inc, California. *LispView Programming Manual*, 1.1 edition, October 1991.

[14] N. Yalçın. The Deep Surface Case Relations in Turkish. Master's thesis, Boğaziçi University, 1989.

Appendix A

Ontology

- `Thing-Object`
  This conceptual category contains concrete matters, such as things and objects. The semantic slots of this concept are as listed below:
  - `Thing-Object` slot containing things and objects, e.g., *kalem* (*pencil*), *adam* (*man*), *akasya* (*acacia*).
  - `Nation-Organization` slot containing nations and organizations, e.g. *Türkiye* (*Turkey*), *Bilkent Üniversitesi* (*Bilkent University*).
  - `Living Object` slot containing living things, e.g., *çocuk* (*child*), *kanarya* (*canary*), *bitki* (*plant*).
  - `Non-living Object` slot containing non living things, e.g., *bilgisayar* (*computer*), *kitap* (*book*).
  - `Animal group` containing animals and humans, e.g., *kaplumbağa* (*turtle*), *Halil*.
  - `Plant` slot containing living organisms that are not animals, e.g., *çiçek* (*flower*), *karanfil* (*carnation*).
  - `Natural Object` slot containing non-living objects produced by nature, e.g., *demir* (*iron*), *kaya* (*rock*).
  - `Artificial Object` slot containing artificial objects produced by man, *çelik* (*steel*), *tahta* (*wood*).
  - `Nation slot` containing people associated with a particular country under one government, e.g., *Türkiye* (*Turkey*), *İsveç* (*Sweden*).

- `Organization` slot containing organized body of people, e.g., *İstanbul Festivali* (*Istanbul Festival*), *Bilkent Üniversitesi* (*Bilkent University*).
- `Solid` slot containing strong materials or constructions, e.g., *bardak* (*glass*), *dosya* (*file*).
- `Liquid` slot containing objects in the form of liquid, e.g., *su* (*water*), *deniz* (*sea*), *ayran* (*yoğurt drink*).
- `Gas` slot containing substance in the form of gas, e.g., *hava* (*air*), *oksijen* (*oxygen*).
- `Human Group` contains group of human, e.g., *ahali* (*people*).
- `Human-Role-Profession` slot containing humans, roles, and professions, e.g., *Demet*, *baba* (*father*), *aşçı* (*cook*).
- `Animal` slot containing living things that can feel and move, e.g., *kelebek* (*butterfly*), *köpek* (*dog*).
- `Human` slot containing human beings, e.g., *kadın* (*woman*), *Kemal*.
- `Role` slot containing human roles, e.g., *anne* (*mather*), *bebek* (*baby*).
- `Profession` slot containing human professions, e.g., *kitapçı* (*bookseller*), *doktor* (*doctor*).
- `Man` slot containing male humans, e.g., *Yılmaz*, *Serkan*, *Berkant*.
- `Woman` slot containing female humans, e.g., *anne* (*mother*), *Kübra*, *Özlem*.
- `Child` slot containing young human beings, e.g., *çocuk* (*child*).
- `Boy` slot containing male children, e.g., *genç* (*young*).
- `Son` slot containing male children of parents, e.g., *oğul* (*son*).
- `Girl` slot containing female children of parents, e.g., *kız* (*girl*).
- `Doughter` slot contains male children of parents, e.g., *kız* (*doughter*).

- `Commodity-Ware`
  This conceptual category contains artificial matters useful to humans. The semantic slots of this concept are as listed below:
  - `Commodity-Ware` slot containing commodities and wares, e.g., *elbise* (*clothes*), *ayakkabı* (*shoe*).

- `Material` slot containing the materials of which something is made, e.g., *süt* (*milk*), *pancar* (*beet*).
- `Means-Equipment` slot containing means and equipments, e.g., *maşa* (*table*), *makina* (*machine*).
- `Products` containing the things produced, e.g., *gözlük* (*glasses*), *şeker* (*sugar*).
- `Means` containing methods of doing production, e.g., *taşlama* (*stoning*).
- `Equipment` containing tools which is used in production, e.g., *matbaa makinası* (*press*).
- `Clothing` slot containing materials made by weaving, e.g., *elbise* (*clothes*), *kazak* (*pullover*).
- `Vehicle` slot containing carriages, e.g., *araba* (*car*), *kamyon* (*lorry*).

• `Idea-Abstraction`
This concept contains non-matters which results from intellectual activities in the human brain, such as ideas and abstractions. The semantic slots of this concept are listed below:

- `Idea-Abstraction` slot containing ideas of human and abstractions, e.g., *fikir* (*idea*), *sözcük* (*word*).
- `Theory` slot containing reasoned accounts offered to explain facts or events, e.g., *Darwin teorisi* (*theory of Darwin*).
- `Conceptual` Object slot containing concepts, e.g., *uyku* (*sleep*), *yaşam* (*life*).
- `Sign-Symbol` slot containing abstractions like signs and symbols, e.g., *harf* (*letter*), *işaret* (*sign*).
- `Name` slot containing names of objects and things, e.g., *Karabaş*.
- `Words` slot containing words, e.g., *defter* (*notebook*).
- `HumanName` slot containing names of human beings, e.g., *Devrim*, *Serdar*.
- `ManName` slot containing names of male humans, e.g., *Hakan*.
- `WomanName` slot containing names of female humans, e.g., *Pınar*, *Nalan*.

- `Part`
  This concept contains structural parts, elements, and contents of things and matters. The semantic slots of this concept are as follows:
  - `PartGroup` slot containing structural parts, elements, and contents of things and matters, e.g., *direksiyon* (*steering-wheel*), *tekerlek* (*wheel*).
  - `Part` slot containing components of things and matters, e.g., *disket sürücü* (*disk driver*), *kart* (*card*).
  - `Element` or Content slot containing elements and contents of things and matters, e.g., *nesne* (*object*), *futbolcu* (*football player*).
  - `Element` slot containing elements of things and matters, e.g., *eleman* (*element*).
  - `HumanPart` slot containing parts of humans, e.g., *el* (*hand*), *ayak* (*foot*).
  - `Content` slot containing the things contained by things and matters, e.g., *oda* (*room*).
- `Attribute`
  This semantic concept contains properties, qualities, or features which are representative of things. The semantic slots of this concept are as follows:
  - `Attribute` slot containing properties, qualities, and features which are representative of things, e.g., *hızlı*.
  - `Property-Characteristic` slot containing properties and characteristics of things, *güzel* (*fine*), *hızlı* (*fast*).
  - `Status-Figure` slot containing status of things, e.g., *bakımlı* (*well-kept*).
  - `Relations` slot containing the acts of relating between things and matters, e.g., *evlilik* (*marriage*).
  - `Structure` slot containing structure of things, e.g., *şekil* (*shape*), *yapı* (*structure*).
  - `Form-Shape` slot containing form and shape of things and matters, e.g., *üçgen* (*triangular*), *kare* (*square*).
  - `State-Condition` slot containing state or condition of things or matters, e.g., *bozuk* (*out of order*), *kirli* (*dirty*).

- `Color` slot containing sensations produced in the eye by rays of light of different wavelengths, e.g., *kırmızı* (*red*), *mavi* (*blue*).

- `Phenomenon`
  This concept contains physical, chemical, and social actions without human activity. The semantic slots of this concept are listed below:

  - `Phenomenon` slot containing physical, chemical, and social phenomena without human activity, e.g., *patlama* (*explosion*).
  - `NaturalPhenomenon` slot containing natural things are known to exist, e.g., *yangın* (*fire*).
  - `ArtificialPhenomenon-Experiment` slot containing artificial phenomena or experiments, e.g., *deney* (*experiment*).
  - `SocialPhenomenon` slot containing social things that are known to exist by the senses, e.g., *ayaklanma* (*revolt*).
  - `Power-Energy-Physical` Object slot containing power and energy, e.g., *patlama* (*explosion*).
  - `ArtificialPhenomenon` slot containing artificial things which are known to exist by the senses, e.g., *yangın* (*fire*).
  - `Experiment` slot containing tests carried out to study something, e.g., *deney* (*experiment*).
  - `Event-Happening` slot containing something happening, e.g., *kavga* (*fight*).
  - `Political-Economical` actions slot containing political and economical actions, e.g., *savaş* (*war*), *kriz* (*crisis*).
  - `Custom-SocialConvention` slot containing customs and social conventions, e.g., *düğün* (*wedding*).
  - `Political` actions slot containing political events, e.g., *suikast* (*criminal attempt*).
  - `Economical` actions slot containing economical events, e.g., *devalüasyon* (*devaluation*).
  - `Custom` slot containing usual behaviors among members of a social group, e.g., *nişan* (*engagement*).
  - `SocialConvention` slot containing social conventions, e.g., *ahlak* (*ethics*).

- `Doing-Action`
  This concept contains human actions. The semantic slots of this concept are listed below:
  - `Doing-Action` slot containing actions and doings done by humans, e.g., *kavga* (*fight*), *oyun* (*game*).
  - `Action-Deed` slot containing actions and deeds, e.g., *koşu* (*race*).
  - `Movement-Reaction` slot containing movements and reactions, e.g., *gitme* (*going*).
  - `Effect-Operation` slot containing effects and operations done by humans, e.g., *etki* (*effect*).
  - `Action` slot containing process of doing things by the humans, *yürümek* (*to walk*), *vurmak* (*to hit*).
  - `Deed` slot containing acts of humans, e.g., *iş* (*work*).
  - `Movement` slot containing acts of changing position taken by human beings, e.g., *göç* (*emigration*), *gitme* (*going*).
  - `Reaction` slot containing oppositions to progress, e.g., *tepki* (*reaction*).
  - `Effect` slot containing outcome of the actions, e.g., *etki* (*effect*).
  - `Operation` slot containing operations, e.g., *arama* (*search*).
- `Sentiment-MentalActivity`
  This concept contains humans' mental activities. The semantic slots of this concept are listed below:
  - `Sentiment-MentalActivity` slot containing humans' mental activities, e.g. *düşünme* (*thinking*).
  - `Perception` slot containing abilities to perceive, e.g., *görme* (*seeing*).
  - `Emotion` slot containing excitement of feelings, e.g., *duygu* (*feeling*), *kızgınlık* (*anger*).
  - `Recognition-Thought` slot containing recognitions and thoughts of humans, e.g., *anlayış* (*understanding*), *fikir* (*idea*).
  - `Recognition` slot containing process of knowing that a person or thing has seen, heard, etc before, e.g., *tanıma* (*recognition*).

- `Thought` slot containing human ideas and intensions, e.g., *görüş* (*opinion*).
- `See` slot containing voluntary visual process, e.g., *görmek* (*seeing*).
- `Feeling` slot containing physical or emotional states that humans can be in, e.g., *kızgın* (*angry*), *küskün* (*resentful*).
- `Belief` slot containing voluntary mental process of holding a belief, e.g., *inanç* (*belief*).
- `Know` slot containing mental process describing the involuntary state of knowing, e.g., *bilgili* (*knowledgeable*).
- `Think` slot containing mental process of thinking, e.g., *düşünme* (*thinking*).
- `Liking` slot containing involuntary favorable mental/emotional reactions to some entity or state of affairs, or a process that presupposes a favorable reaction, e.g., *sevgi* (*love, affection*), *aşk* (*love*).
- `Striving` slot containing mental reactions attempting to bring about some state of affairs or an event, e.g., *cesaret* (*courage*).
- `Wanting` slot containing mental reactions towards some objects or state of affairs, e.g., *istek* (*desire*).
- `Disliking` slot containing that are negative towards some objects or state of affairs, e.g., *nefret* (*disgust*).
- `Fearing` slot containing mental reactions that are feelings of alarm or disquite caused by awareness or expectation of danger, e.g., *korku* (*fear*).

- `Measure`
  This concept contains measures. The semantic slots of this concept are as follows:
  - `Measure` slot containing sizes, quantities, etc., e.g., *bir* (*one*), *santimetre* (*centimeter*).
  - `Number` slot containing quantities and amount, e.g., *bin* (*thousand*), *dokuz* (*nine*).
  - `Unit` slot containing units, e.g., *metre* (*meter*), *kilo* (*kilo*).
  - `Standard` containing standards of measurement , e.g., *standart* (*standard*), *metre-kilogram-saniye* (*meter-kilogram-second*).

- `Time-Space`
  This concept contains time, space, and topologies. The semantic slots of this concept are listed below:
  - `Time-Space` slot containing spatial and temporal nouns, i.e *saat* (*hour*), *alan* (*arena*).
  - `Space-Topography` slot containing the nouns indicating space or topography, e.g., *Yalova*, *uzay* (*space*).
  - `Time` slot containing portions or measures of time, e.g., *yarın* (*tomorrow*), *hafta* (*week*).
  - `Space` slot containing areas, e.g., *uzay* (*space*).
  - `Topography` slot containing places and locations, e.g., *burası* (*here*), *İstanbul* (*Istanbul*).
  - `TimePoint` slot containing nouns indicating point of time, e.g., *şimdi* (*now*), *dün* (*yesterday*).
  - `TimeDuration` slot containing nouns indicating a period of time, e.g., *gün* (*day*) *saat* (*hour*), *ay* (*month*), *yıl* (*year*).
  - `TimeAttribute` slot containing time attributes, e.g., *önce* (*before*), *sonra* (*after*).
  - `Area` slot containing the nouns indicating area, e.g., *tarla* (*field*).
  - `Place` slot containing the places, e.g., *Bursa*.
  - `Location` slot containing the positions, e.g., *köşe* (*corner*).
  - `Past` slot containing nouns refering to past, e.g., *dün* (*yesterday*).
  - `Present` slot containing nouns refering to present, e.g., *bugün* (*today*).
  - `Future` slot containing nouns refering to future, e.g., *yarın* (*tomorrow*).

Appendix B

List of Verbs in the Lexicon

The verbs stored in the lexicon are given in Table B.1.

VERBS				
itiraz et	kaçın	kaç	kabul et	izle
it	iste	incele	inan	in
ilet	ihtiyaç ol	ifade et	iç	hükmet
hoşgör	hisset	hedefle	hazır ol	hastalan
hatırla	harca	güven	gülümse	götür
göster	gör	gir	git	gez
getir	gerek	gel	geç	fark et
etkile	eski	ertele	emin ol	elde et
duyur	duy	düşün	düş	dön
dokun	doğ	dik	diren	dikkat et
değerlen	değ	de	dayan	daya
dal	dağıt	çık	at	çarp
cesaret et	çalış	bürün	bul	boğul
birik	bin	bırak	bil	benze
bekle	bat	başla	başar	bak
bağla	ayrıl	ayır	atla	aşağıla
ara	al	anla	an	altüst et
aldır	açık ol	aktar	ak	kat
kapa	kap	kanıtla	kaldır	kur
kır	kaybol	kaybet	kay	kazan

Table B.1. The verbs having an entry in the lexicon.