An Outline of Turkish Morphology

Kemal Oflazer, Elvan Göçmen, Cem Bozşahin
October 1994 20th year edition, October 2014

Contents

1 Introduction

Turkish is an agglutinative language with word structures formed by productive affixations of derivational and inflectional morphemes to root words.[1] A popular—and rather exaggerated—example of a Turkish word formation is:

OSMANLILAŞTIRAMAYABİLECEKLERİMİZDENMİŞSİNİZ

which can be broken down into morphemes as follows:

OSMAN-LI-LAŞ-TIR-AMA-YABİL-ECEK-LER-İMİZ-DEN-MİŞ-SİNİZ

where the -'s indicate morpheme boundaries. This adverb can be translated into English as "(behaving) as if you were of those whom we might consider not converting into an Ottoman." For the details of Turkish grammar and word formations rules one can refer to a number of books [5, 9].

Turkish has clear but rather complex morphotactics. Morphemes added to a stem can convert the word from a nominal to a verbal structure or vice-versa, or can create adverbial constructs as above. The surface realizations of morphological constructions are constrained and modified by a number of morphophonemic rules. Vowels in the affixed morpheme have to agree with the preceding vowel in certain aspects to achieve *vowel harmony*. Under certain circumstances vowels in the roots and morphemes are deleted. Similarly, consonants in the roots words, or in the affixed morphemes undergo certain modifications, and may sometimes be deleted. Furthermore, the assimilation of a large number of words into the language from various foreign languages – most notably Arabic, Persian, and French – resulted in word formations which behave as exceptions. Turkish morphology has been investigated from a computational point of view by Köksal [4], Hankamer [2], Solak and Oflazer [7], and Oflazer [6].

[1] Turkish is an exclusively suffixing language. There are however a few very unproductive prefixes of foreign origin, such as *na- (un-)*.Words with such suffixes can be treated as separate lexical items.

	+Round		–Round	
	+Front	–Front	+Front	–Front
-High	ü	u	i	ı
–High	ö	o	e	a

Table 1: Turkish Vowels

		Bilabial	Labio-dental	Dental Alveolar	Palato-Alveolar	Palatal	Velar	Glottal
Stop								
	voiceless	p		t	ç		k	
	voiced	b		d	c		g	
Fricative								
	voiceless		f	s	ş			
	voiced		v	z	▮			
Nasal		m		n				
Liquid								
	lateral			l				
	nonlat.			r				
Glide						y	ğ	h

Table 2: Turkish Consonants

2 Current orthography of Turkish

The Turkish language has an alphabet of 29 letters in its current orthography based on the Latin characters. There are 8 vowels: `a, e, ı, i, o, ö, u,` ü, and 21 consonants: `b c ç d f g ğ h j k l m n p r s ş t v y z`. Tables 1 and 2 show the phonetic features of Turkish vowels and consonants.

There are however phonemes not covered by these. Certain long vowels are mainly used in words borrowed from foreign languages most notably Arabic and Persian. Such vowels are sometimes distinguished in older orthography by various means (such as with a ˆ on top of the vowel). In modern orthography such distinctions are almost never used. There is also a certain phoneme known as "yumuşak g" (soft g – denoted as ğ in orthography) which creates bisyllabic two-vowel sequences. At the end of a syllable, this phoneme causes the lengthening of the preceding vowel [10]. Consonants `k,g`, and `l` have palatal and non-palatal allophones. In certain cases the palatalization process has impact on the vowel harmony.

We can partition the vowels as follows to aid in the description of the vowel harmony processes.

1. Back vowels: { `a, ı, o, u` }

2. Front vowels: { e, i, ö, ü }
3. Front unrounded vowels { e, i }
4. Front rounded vowels {ö, ü}
5. Back unrounded vowels {a, ı }
6. Back rounded vowels {o, u}
7. High vowels {ı, i, u, ü}
8. Low unrounded vowels {a, e}

In Turkish, proper nouns are separated from certain suffixes by an apostrophe (’). All vowel harmony rules and some of the consonant change rules are in effect in the orthography of proper nouns.

Below we present the rules in two-level notation. In giving examples we refer to *lexical* and *surface* forms, where the former refers to the structural form of a word, and the latter refers to the phonological realization of the word. The 0’s in the examples denote the phonemes or morpheme boundary symbol which get deleted on the surface realizations.

3 Morphophonemic processes

We use the following meta-phonemes in our descriptions:

1. D : voiced (d) or voiceless (t)
2. A : back (a) or front (e)
3. H : high vowel (ı, i, u, ü)
4. R : vowel except o, ö
5. C : voiced (c) or voiceless (ç)
6. G : voiced (g) or voiceless (k)

3.1 Vowel Harmony

Turkish has vowel harmony processes that force certain vowels in suffixes agree with the last vowel in the stems or roots they are being affixed to. Some of these phenomena are exemplified below using two-level notation.

3.1.1 Resolving low-unrounded vowels

Let `A` be a vowel in a suffix, which may either be an `a` or `e`. `A` is resolved as follows: `A` is resolved as an `a` if the last vowel in the stem is a back vowel. For example:

Lexical:	`masa-lAr`	`N(table)-PLU`
Surface:	`masa0lar`	`masalar`
Lexical:	`satır-lAr`	`N(hatchet)-PLU`
Surface:	`satır0lar`	`satırlar`
Lexical:	`oto-lAr`	`N(car)-PLU`
Surface:	`oto0lar`	`otolar`
Lexical:	`kutu-lAr`	`N(box)-PLU`
Surface:	`kutu0lar`	`kutular`

`A` is resolved as an `e` if the last vowel in the stem is a front vowel. For example:

Lexical:	`ev-lAr`	`N(house)-PLU`
Surface:	`ev0ler`	`evler`
Lexical:	`kedi-lAr`	`N(cat)-PLU`
Surface:	`kedi0ler`	`kediler`
Lexical:	`göl-lAr`	`N(lake)-PLU`
Surface:	`göl0ler`	`göller`
Lexical:	`gül-lAr`	`N(rose)-PLU`
Surface:	`gül0ler`	`güller`

`A` is also resolved as an `e` if the last vowel is a long â (`â`), a long û (`û`) or an ô followed by a palatalized l (mostly in words of French origin.) The long vowels are almost always realized on the surface as their short counterparts. For example.

Lexical:	saât-lAr	N(hour)-PLU
Surface:	saat0ler	saatler

Lexical:	usûl-lAr	N(method)-PLU
Surface:	usul0ler	usuller

Lexical:	gôl-lAr	N(goal)-PLU
Surface:	gol0ler	goller

Note that between the harmonizing and the harmonized vowel, there may be one or more consonants.

3.1.2 Resolving high vowels

Let H denote a high-vowel in a suffix. It is resolved as follows:

H is resolved as a u is the last vowel in the stem is a back-rounded vowel.

Lexical:	okul-Hm	N(school)-1SG-POSS
Surface:	okul0um	okulum

Lexical:	gel-Hyor-yHm	V(come)-PROG-1SG
Surface:	gel0iyor00um	geliyorum

H is resolved as a ü if the last vowel in the stem is a front-rounded vowel, or long û or ô as defined above.

Lexical:	gün-Hm	N(day)-1SG-POSS
Surface:	gün0üm	günüm

Lexical:	göl-Hm	N(lake)-1SG-POSS
Surface:	göl0üm	gölüm

Lexical:	alkôl-Hm	N(alcohol)-1SG-POSS
Surface:	alkol0üm	alkolüm

Lexical:	usûl-Hm	N(method)-1SG-POSS
Surface:	usul0üm	usulüm

H is resolved as a ı is the last vowel in the stem is a back-unrounded vowel.

Lexical:	masal-Hm	N(tale)-1SG-POSS
Surface:	masal0ım	masalım

Lexical:	yıldız-Hm	N(star)-1SG-POSS
Surface:	yıldız0ım	yıldızım

H is resolved as a i if the last vowel in the stem is a front-unrounded vowel, or a long a.

Lexical:	ev-Hm	N(house)-1SG-POSS
Surface:	ev0im	evim

Lexical:	pir-Hm	N(master)-1SG-POSS
Surface:	pir0im	pirim

Lexical:	saât-Hm	N(watch)-1SG-POSS
Surface:	saat0im	saatim

There a very small number of special cases which presents some problems with respect to vowel harmony. These happen with the verbal roots de (say) and ye (eat), where the only vowel in the root may be deleted under certain circumstances. In these case, we assume the suffix vowel harmonizes with respect to the undropped lexical vowel.

3.2 Vowel drops

An H denoting a high vowel at the beginning of a suffix is deleted if the last phoneme of the stem is a vowel. For example:

Lexical:	masa-Hm	N(table)-1PS-POSS
Surface:	masa00m	masam

However, this drop does not occur if the high vowel is not a part of the tense suffix (-Hyor) and the verbal root stem ends with a vowel in which case this vowel drops. (This may also be viewed as the H actually dropping and the stem-final vowel becoming a high vowel if necessary)

Lexical:	kapa-Hyor	V(close)-PR-CON-3PS
Surface:	kap00ıyor	kapıyor

The last vowel in certain roots is deleted when those roots are affixed certain suffixes that either start with a vowel or a consonant that also drops. There is no uniform way of specifying such words, except possibly by explicitly enumerating them. Here we indicate such vowels by prefixing them with a certain marker.

Lexical:	bur$un-Hm	N(nose)-1SG-POSS
Surface:	bur00n0um	burnum

3.3 Consonant drops

The consonants `n,s and y` at the beginning of a suffix drop if the last phoneme of the stem is a consonant. However if the suffix is `-sHz` with `H` representing a high-vowel, then `s` does not drop even if the preceding phoneme is a consonant.

Lexical:	ev-nHn	N(house)-GEN
Surface:	ev00in	evin
Lexical:	kalem-sH	N(pencil)-3PS-POSS
Surface:	kalem00i	kalemi
Lexical:	kalem-sHz	N(pencil)-WITHOUT
Surface:	kalem0siz	kalemsiz
Lexical:	ağ$ız-yH	N(mouth)-ACC
Surface:	ağ00z00ı	ağzı

3.4 Consonant changes

Let `D` denote a suffix initial dental consonant that may resolved as either a `d` or `t`. It is resolved to a `t` is the last phoneme in the stem is resolved as one of {`ç, f, h, k, p, s, ş, t`}. Otherwise, `D` is resolved as a `d`. Some examples are:

Lexical:	kitab-DA	N(book)-LOC
Surface:	kitap0ta	kitapta

Lexical:	yulaf-DAn	N(oat)-ABL
Surface:	yulaf0tan	yulaftan

Lexical:	aç-DHk	V(open)- PERF
Surface:	aç0tık	açtık

Voiced stops b, d are realized as p, t respectively when they are stem-final or they are the last consonant in the stem which affixed a morpheme that starts with a consonant that does not drop. Some examples are:

Lexical:	kitab-lAr	N(book)-PLU
Surface:	kitap0lar	kitaplar

Lexical:	kitab-cH	N(book)-NtoN(ci)
Surface:	kitap0çı	kitapçı

Lexical:	dolab-nHn	N(closet)-GEN
Surface:	dolab00ın	dolabın

Lexical:	tad-DHk	V(taste)- PERF
Surface:	tat0tık	tattık

There are however some exceptions to this rule. These exceptions are: ab (water) kalb (heart), balad (ballad), hemoroid , önad (fore name), soyad (last name) yad (remembrance), etc.

c is another voiced obstruent like those above except that it also appears in certain suffixes as the first consonant where it gets modified to a ç, due to a reciprocal assimilation process. So the suffix-initial c is resolved as a ç if the last consonant of the stem is resolved one of {ç,f, h, k, p, s, ş} . A stem final c is resolved to a ç if it is also word final or is followed by a consonant that does not drop. Some examples are:

Lexical:	harac-cH	N(tribute)-NtoN(ci)
Surface:	haraç0çı	haraççı

Lexical:	yaş-cA	N(age)-NtoAdv(ca)
Surface:	yaş0ça	yaşça

Lexical:	harac-yA	N(tribute)-DAT
Surface:	harac00a	haraca

There are however some exceptions to this rule. These exceptions are the following monosyllabic forms, and compound forms written together using these as the second component: aç (hungry/to open) çeç (wheat pile), göç (migration), haç (cross), iç (interior), etc.

A velar stop k at the end of a stem becomes a ğ when a suffix starting with a vowel is affixed. There may be a dropping consonant before the suffix. Some examples are:

Lexical:	ayak-nHn	N(foot)-GEN
Surface:	ayağ00ın	ayağın

Lexical:	tarak-Hm	N(comb)-1PS-POSS
Surface:	tarağ0ım	tarağım

However a stem-final k preceded by a n becomes a g under the same circumstances.

Lexical:	renk-yH	N(color)-ACC
Surface:	reng00i	rengi

Lexical:	ahenk-yA	N(harmony)-DAT
Surface:	aheng00e	ahenge

However there are some exceptions to these, where the k does not change. These exceptions are the following monosyllabic forms and some polysyllabic words of foreign origin, and compound forms written together using these as the second component: afak (Arabic plural version of ufuk (horizon)), ahlak (ethics), arabesk (), ark (water canal), aşk (love) bank (chair), etc.

At stem-final g (in words of foreign origin) also becomes a ğ when a suffix starting with a vowel is affixed. There may be a dropping consonant before the suffix.

Lexical:	radyolog-yA	N(radiologist)-DAT
Surface:	radyoloğ00a	radyoloğa

However under the circumstances above if the stem-final g is preceded by another consonant (only n and r seem to be such consonants) then the g does not become a ğ. Some examples are:

Lexical:	brifing-Hm	N(briefing)-1PS-POSS
Surface:	brifing0im	brifingim

Lexical:	aysberg-HnHz	N(iceberg)-2PP-POSS
Surface:	aysberg0iniz	aysberginiz

There are again some exceptions to these rules. These are: demagog, füg (fugue) gag, lig (league), pedagog, sinagog.

3.5 Words ending with (su)

Turkish has a large number of nominal roots ending with *su* (water) e.g., *akarsu* (river). *Su*, along with *ne* (what) does not obey the standard inflection rules. For example su-sH (water -3PS-POSS) is suyu and not susu and su-nHn (water-GEN) is suyun and not sunun.[2] Thus a stem final y inserted to such stems when a suffix starting with a vowel or a dropping consonant is affixed. Here are some examples:

Lexical:	akarsu0-yHnHz	N(river)-2PP-POSS
Surface:	akarsuy00unuz	akarsuyunuz

Lexical:	akarsu-lar	N(river)-PLU
Surface:	akarsu0lar	akarsular

3.6 Gemination

Certain nominal forms in of Arabic or Persian origin, there is a gemination process whereby the last consonant is duplicated when certain suffixes are added. Some examples are:

[2] For ne the normal inflections are also valid.

Lexical:	üs0-sH	N(base)-3SG-POSS
Surface:	üss00ü	üssü

Lexical:	hak0-yH	N(right)-ACC
Surface:	hakk00ı	hakkı
Lexical:	hak-lAr	N(right)-PLU
Surface:	hak0lar	haklar

The suffixes that cause this gemination are those that start with a dropping consonant. The words that undergo this gemination process are: hak, tıb, med, hal (solution) şık, ad (recognition) had, üs, zam, af, sır, hat.

3.7 s-drop

The first consonant 3SG-POSS suffix (-sH) when added to certain words of Arabic origin ending with a vowel, drops in exception to the general rule. Words with s dropping and not dropping are considered legal, though two vowel sequences are not at all common in Turkish.

Lexical:	cami-sH	N(mosque)-3SG-POSS
Surface:	cami-0i	camii
Surface:	cami-si	camisi

The following words have this property: bayi, cami, çıma, enva, filvaki, ibda, içtima, ifşa, ihtira, ikna, imtina, indifa, inkıta, intiba, irca, irtica, irtifa, ıttıla, kablelvuku, kanı, maktu, mani, matbu, mayi, mebde, mecmu, memba, menşe, merci, meta, mevdu, mevki, mevzi, mevzu, mısra, mudi, murabba, mürteci, muti, muttali, müvezzi, niza, rükü, sanayi, şayi, şeci, şema, şua, şüyu, tabı, teberru, terfi, teşci, teşri, teşyi, tetebbu, tevabi, tevazu, tevdi, tevessü, tevsi, tulu, vaki, vasi, veda, vuku, zayi, zıya, zürra.

4 Affix Inventory

In this section we present the set of suffixes that are available in Anatolian Turkish for word formation via derivational or inflectional means. Many words derived from derivational suffixes are lexicalized in the sense that their meaning composition is no longer related to the meaning of the stem in a predictable way. The suffixes marked with ¶ in the tables are the ones which are relatively more productive and compositional in this sense.

4.1 Coding scheme

Upper-case letters in morphemic representations denote meta-phonemes[3]. Parentheses indicate insertion/deletion depending on the previous segment.

Codes for morphemes are of the form

$P_0P_1P_2\_P_3\ldots$

where

P_0: Position 0. Final grammatical category.

P_1: Source grammatical category.

P_2: Type of process. I for inflection and D for derivation. A N means not applicable.

remainder is the mnemonic name of the morpheme.

Grammatical category codes are :

N	Noun
V	Verb
A	Adverb
J	Adjective
R	Pronoun
P	Postposition
C	Conjunction
X	Exclamation

For instance, NVD_xyz means the xyz affix produces a noun from a verb via a derivation.

4.2 Noun Inflections (NNI_xxxxxx)

Elements—in order— are given below. All except N are optional.

1. Noun stem (N)
2. Plural (NNI_PLU)

[3] see the section on morphophonemic processes

3. Possessive (NNI_POSSxx)

4. Case (NNI_xxx)

5. Relative (NNI_REL)

Morphemic Representation	Code	Gloss	Examples
-lAr	¶ NNI_PLU	Plural	arabalar, evler
-(H)m	¶ NNI_POSS1s	1st person singular possessive	arabam, evim
-(H)mHz	¶ NNI_POSS1p	1st person plural possessive	arabamız, evimiz
-(H)n	¶ NNI_POSS2s	2nd person singular possessive	araban, evin
-(H)nHz	¶ NNI_POSS2p	2nd person plural possessive	arabanız, eviniz
-(s)H	¶ NNI_POSS3s	3rd person singular possessive	arabası, evi
-lArH	¶ NNI_POSS3p	3rd person plural possessive	arabaları, evleri
-(y)H	¶ NNI_OBJ	Objective (accusative) case	arabayı, evi
-nH	¶ NNI_OBJ3	Objective case (after 3P poss)	masasını
0	NNI_NOM	Nominative case	araba, ev
-(n)Hn	¶ NNI_GEN	Genitive case	arabanın, evin
-(y)A	¶ NNI_DAT	Dative case	arabaya, eve
-nA	¶ NNI_DAT3	Dative case (after 3P poss)	masasına
-DA	¶ NNI_LOC	Locative case	arabada, evde
-nDA	¶ NNI_LOC3	Locative case	masasında
-DAn	¶ NNI_ABL	Ablative case	arabadan, evden
-nDAn	¶ NNI_ABL3	Ablative case	masasından
-(y)lA	¶ NNI_INC	Instrumental/comitative case	arabayla, evle
-ki	¶ NNI_REL	Relative	evdeki, arabadakilerinki

4.3 Derivations producing nouns (NxDxxxxx)

The 'adjective' and noun distinction in Turkish is a difficult one. [4] Most adjectives can be used as nouns, and undergo the derivations from a noun. Nouns can perform the function of an adjective as noun modifier in noun-noun groups (*izafet*). See the chapter on syntax.

[4]Nouns and adjectives are sometimes collectively called *substantives*

4.3.1 Nouns from nouns (NND_xxxxxx) or adjectives (NJD_xxxx)

Morphemic Representation	Code	Example
-CA	NJD_CA	akça, karaca, Hintçe
-CA	NND_CA	Türkçe, Arapça
-cAğIz	NND_CAGZ	adamcağız, köyceğiz
-cAk	NND_CAK	oyuncak, yavrucak
-CH	NND_CI	ekmekçi, odacı, işçi, çiftçi
		yolcu, öncü, yabancı, aracı, konuşmacı
		kaderci, akılcı, milliyetçi, Atatürkçü
-CHk	NND_CIK	bademcik, kızılcık, maymuncuk
-CHl	NND_CIL	öncül, balıkçıl
-dAş	NND_DAS	sırdaş, arkadaş, meslekdaş, gönüldeş
-gen	NND_GEN	üçgen, altıgen, köşegen
-lHk	NND_LIK	günlük, gözlük, sabunluk, salatalık, gecelik
		kitaplık, fidelik, kömürlük, odunluk
		ebelik, doktorluk, taşçılık
		halkçılık, maddecilik, ırkçılık
-lHk	¶ NJD_LIK	bolluk, güzellik, titizlik, sıkışıklık

4.3.2 Nouns from verbs (NVD_xxxxxx)

Morphemic Representation	Code	Example
-AcAk	NVD_ACAK	alacak, verecek, içecek, yakacak, çekecek, kıracak
-Ak	NVD_AK	durak, yatak, batak, konak, yunak, sığınak, tapınak ölçek, kayak, kaydırak, uçak, yutak, saçak
-amak	NVD_AMAK	basamak, kaçamak, tutamak
-An	¶ NVD_AN	bakan, kapan, kalan, çağlayan, bölünen, bölen
-AnAk	NVD_ANAK	gelenek, görenek, yetenek, olanak, tutanak sağanak, ödenek
-cA	NVD_CA	sakınca, düşünce, eğlence, dinlence, söylence
-gA	NVD_GA	dalga, bilge, süpürge, önerge, bildirge, gösterge
-(G)Aç	NVD_GAC	kıskaç, süzgeç, sayaç, büyüteç
-GAn	NVD_GAN	ısırgan, sergen, yelken, ergen
-GH	NVD_GI	silgi, sargı, atkı, keski, süngü çatkı, dolgu, çizgi, içki, bitki sevgi, saygı, ilgi, etki, görgü
-gHç	NVD_GIC	dalgıç
-gHn	NVD_GIN	yangın, salgın, düzgün, bilgin, bozgun
-H	NVD_I	ölü, dolu, soru, korku, sanı yapı, dizi, sürü, batı, doğu, koşu
-(y)HcH	NVD_ICI	satıcı, yüzücü, okuyucu ısıtıcı, susturucu, uyuşturucu, taşıyıcı
-(H)k	NVD_IK	tanık, delik, kırık, göçük, bölük, katık ayrık, konuk, öksürük, buyruk, sarık, açık
-(H)m	NVD_IM	doğum, ölüm, yudum, atım, sayım, seçim dönüm, yarım, ekim, pişirim, içim, tadım bağlam, kavram, sağlam, uçurum, oturum, bitirim, kaldırım
-(H)n	NVD_IN	ekin, yığın, tütün, akın, ışın, gelin
-(Hn)ç	NVD_INC	kazanç, ilenç, sapınç, bilinç
-HntH	NVD_INTI	akıntı, kesinti, döküntü, kuruntu
-(y)Hş	NVD_YIS	dalış, geçiş, uçuş, yürüyüş
-(H)t	NVD_IT	ayırt, geçit, umut, yoğurt, yakıt, kesit, taşıt
-mA	NVD_MA	korunma, bekleme, araştırma
-mAcA	NVD_MACA	bilmece, kandırmaca, koşmaca, çekmece, atmaca
-mAk	¶ NVD_MAK	yapmak, yemek
-mAn	NVD_MAN	sayman, göçmen, okutman, eğitmen
-mAzlHk	¶ NVD_MAZLIK	aldırmazlık, dinlemezlik, uyuşmazlık
-tH	NVD_TI	belirti, kızartı

4.3.3 Adjectives from nouns (JND_xxxxxx) or adjectives (JJD_xxxxx)

Morphemic Representation	Code	Example
-CA	JJD_CA	mertçe, güzelce, yaşlıca
-CH	JND_CI	şakacı, inatçı, uykucu, karahaberci
-cHk	JND_CIK	incecik, ufacık, küçücük
-cHl	JJD_CIL	evcil, bencil, ölümcül, insancıl
-(H)msH	JJD_IMSI	tatlımsı, mavimsi, sertimsi, hamurumsu
-(H)mtrak	JJD_MTRAK	ekşimtrak, yeşilimtrak
-(H)ncH	JND_INCI	birinci, ikinci, üçüncü, onuncu
-(H)z	JND_IZ	ikiz, üçüz
-lH	JND_LI	tatlı, sesli, uslu, türlü, Asyalı, Çinli
-lHk	JND_LIK	yemeklik, kiralık, turşuluk
-mAn	JJD_MAN	kocaman, küçümen, şişman, delişmen
-sAl	JND_SAL	anayasal, sorunsal, biçimsel
-msAr	JJD_MSAR	iyimser, kötümser, karamsar
-(m)sH	JND_MSI	erkeksi, yılansı, çocuksu, budalamsı
-sHz	¶ JND_SIZ	tatsız, evsiz, sonsuz, köksüz
-(ş)Ar	JND_SER	birer, ikişer, dokuzar

4.3.4 Adjectives from verbs (JVD_xxxx)

Morphemic Representation	Code	Example
-dHk	JVD_MADIK	görülmedik, olmadık, ummadık
-AcAk	¶ JVD_ACAK	oturacak, içecek, söylenecek, olmayacak
-AgAn	JVD_AGAN	olagan, süregen, duragan, gezegen
-(A)k	JVD_AK	korkak, sarsak, dişlek, yuvarlak, uzak
-An	¶ JVD_AN	yaratan, gören, sevilen, tüten
-gAç	JVD_GAC	utangaç, üşengeç
-GAn	JVD_GAN	çalışkan, alıngan, çekingen, dövüşken
-giç	JVD_GIC	bilgiç
-GHn	JVD_GIN	dalgın, yorgun, üzgün, bezgin, şaşkın yetişkin, alışkın
-H	JVD_I	sıkı, ölü, katı, dolu, duru
-(y)HcH	JVD_ICI	yırtıcı, bölücü, geçici, üzücü
-(H)k	JVD_IK	batık, kesik, çizik, tutuk, yanık, çatık işlek, oynak, patlak, büyük, soğuk, bulanık
-(H)lH	JVD_ILI	ekili, sarılı, örtülü, kurulu
-Hn	JVD_IN	uzun, sayın
-(Hn)ç	JVD_INC	kıskanç, iğrenç, korkunç, gülünç
-(H)ntH	JVD_INTI	uyuntu, özenti, süprüntü
-Hr (Ar)	JVD_IR	olur, okur, düşünür, bilir, içilir akar, bakar, keser, döner
-mA	JVD_MA	dökme, yapma, kaplama, serpme
-mAz	¶ JVD_MAZ	görünmez, yılmaz, utanmaz
-mHş	¶ JVD_MIS	geçmiş, okumuş, pişmiş

4.4 Verb Inflections (VVI_xxxxxxx)

Elements are given below. VVI_Txxx and VVI_PERS are required if the verb is finite.

1. Verb stem (V)
2. Reflexive (VVI_REFX)
3. Reciprocal/Collective (VVI_RECP)
4. Causative (VVI_CAUSx)
5. Passive (VVI_PASSx)

6. Impossible (VVI_IMP)
7. Negative (VVI_NEG)
8. Tense-aspect (VVI_Txxxx)
9. Auxiliary (VVI_Xxxxx)
10. Person (VVI_PERSxxx)

Morphemic Representation	Code	Gloss	Examples
-(H)n	VVI_REFX	Reflexive	kapan, kaçın, örtün, vurun, edin
-Hş	VVI_RECP	Reciprocal/Collective	kaçıştır, büzüştür, koşuşmak
-DHr	¶ VVI_CAUSD	Causative	kaldır, arttır, güldür, sektir
-t	¶ VVI_CAUST	Causative	çıkart, küçült
-(H)r	¶ VVI_CAUSR	Causative	çıkar, batır
-Hl	¶ VVI_PASSL	Passive	yapılmış, küçüldü
-(H)n	¶ VVI_PASSN	Passive	vidalandı
-(y)AmA	¶ VVI_IMP	Impossible	geleme, kalama
-mA	¶ VVI_NEG	Negative	gelme, kalma
-(H)r	¶ VVI_TAORSH	Aorist tense	kalır, bulur büyür, gelir
-(A)r	¶ VVI_TAORSA	Aorist tense	geçer, kaçar
-(H)yor	¶ VVI_TPROG	Progressive	geçiyor, kalıyor, buluyor, gülüyor
-DH	¶ VVI_TPAST	Past tense	kaldı, geçti, buldu, güldü
-mHş	¶ VVI_TNARR	Narrative past	kalmış, bulmuş, ölmüş
-(y)AcAk	¶ VVI_TFUTR	Future	kalacak, gelecek, isteyecek
-(y)A	¶ VVI_TOPTA	Optative	gelmiyeydi, kazmıyaydı
-mAlI	¶ VVI_TNECE	Necessitative	gelmeli, bulmalı, bilmeli
-sA	¶ VVI_TCOND	Conditional	gelse, vursa, bulasa
-yAbil	¶ VVI_TABIL	Abilitative	gidebil, kalamayabil
-yAmA	¶ VVI_ANEG	Negative abilitative	gideme, okuyama
-yAdur	¶ VVI_TDUR	Continuous	gidedur, çalışadur
-yAkal	¶ VVI_TKAL		bakakal
-yAyaz	¶ VVI_TYAZ		düşeyaz, unutayaz
-yAgör	¶ VVI_TGOR		yapagör
-yAgel	¶ VVI_TGEL		yapagel
-yAkoy	VVI_TKOY		alıkoy
-(y)DI	¶ VVI_XPAST	Past aux	yapsaydı, gelmişti, gelecekti
-(y)mHş	¶ VVI_XDUBT	Dubitative aux	tembelmiş, gitmişmiş, buradaymış
-(y)sA	¶ VVI_XCOND	Conditional aux	buradaysa, bulduysa, gelmişse
-(y)ken	¶ VVI_XADV1	Adverbial aux	gelmişken, buradayken
-ArAk	¶ VVI_XADV2	Adverbial aux	bakarak, gelerek
-cAsInA	¶ VVI_XADV3	Adverbial aux	bilmişcesine, uçarcasına
-(H)m	¶ VVI_PERS1s	1st person singular	geldim, bulmuşum
-(H)z	¶ VVI_PERS1p1	Type I 1st person plural	geliriz, bulmuşuz
-k	¶ VVI_PERS1p2	Type II 1st person plural	geldik, baksak
-(sH)n	¶ VVI_PERS2s	2nd person singular	gelsen, bulursun
-(sH)nHz	¶ VVI_PERS2p	2nd person plural	gelseniz, bulursunuz
-DHr	¶ VVI_DHR	copula	buradadır, gelmişizdir
-0	¶ VVI_PERS3st1	Type I 3rd singular	okurlar, gelmiş
-z	¶ VVI_PERS3st2	Type II 3rd singular	yapamaz, gelemez

4.5 Derivations producing verbs

4.5.1 Verbs from nouns (VND_xxxxx) or adjectives (VJD_xxxx)

Morphemic Representation	Code	Example
-dA	VND_DA	parıldamak, höpürdemek
-A	VND_A	yaşamak, kanamak, dilemek, türemek, kocamak
-(A)l	VJD_AL	azalmak, incelmek, düzelmek, daralmak, körelemek yükselmek, ufalmak
-Ar	VJD_AR	ağarmak, kararmak, göğermek, yeşermek
-et	VND_ET	gözetmek, yönetmek
-HmsA	VJD_IMSE	küçümsemek
-lA	VND_LA	sepetlemek, çuvallamak, kalaylamak, kazıklamak tuzlamak, katlamak, öğütlemek, sabunlamak patlamak, gürlemek, melemek, havlamak atlamak, saplamak, yoklamak
-lA	VJD_LA	sollamak, genişlemek, serinlemek, ucuzlamak
-lAn	VND_LAN	evlenmek, yaşlanmak, uslanmak, ayaklanmak, kibirlenmek
-lAş	¶ VJD_LAS	güzelleşmek, yoksullaşmak, başkalaşmak
-sA	VND_SA	susamak, tavsamak, kapsamak, umursamak, önemsemek,
-sA	VJD_SA	garipsemek, ıraksamak, yakınsamak

4.5.2 Verbs from verbs (VVD_xxxxx)

Morphemic Representation	Code	Example
-DAr	VVD_DAR	aktarmak, kaytarmak, kotarmak
-AklA	VVD_AKLA	duraklamak, iteklemek, tartaklamak
-AlA	VVD_ALA	eşelemek, oğalamak, kovalamak, şaşalamak, silkelemek
-(H)klA	VVD_IKLA	uyuklamak, didiklemek, dürtüklemek, sayıklamak
-mAk	VVD_MAK	yapmak, gelmek
-HştHr	VVD_USTUR	itiştirmek, veriştirmek, atıştırmak, oğuşturmak

4.5.3 Adverbs from nouns (AND_xxxxx) or adjectives (AJD_xxxxx)

Morphemic Representation	Code	Example
-CA	AND_CA	dostça, sınıfça
-CA	AJD_CA	usulca, ayrıca, böylece
-cAk	AND_CAK	evcek
-(y)A	AND_YA	beriye, uca, yarına, aşağıya
-(y)A	AJD_YA	temize, ucuza
-Hn	AND_IN	kışın, güzün
-Hn	AJD_IN	ilkin, ansızın
-lA	AND_LA	hızla, güçlükle, zamanla, öncelikle
-leyin	AND_LAYIN	sabahleyin, akşamleyin

4.5.4 Adverbs from verbs (AVD_xxxxx)

Morphemic Representation	Code	Example
-dHkçA	AVD_DIKCA	oldukça, gittikçe
-ArAk(tAn)	AVD_ARAK	olarak, giderekten, bakaraktan, gülerek, bilmeyerek

4.5.5 Adverbs from adverbs (AAD_xxx)

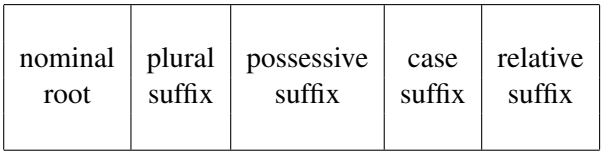
Morphemic Representation	Code	Example
-cek	AAD_CEK	demincek

5 Morphotactics

5.1 Paradigms

Turkish has two main paradigms for word formation. The *nominal paradigm* applies to nouns and adjectives and describes the order of the inflectional suffixes. This paradigm is described in Figure 1. The verbal paradigm applies to verbs and describes the order of the inflectional suffixes that are applicable to verbal roots. It is shown in Figure 2. These paradigms do not however describe cross-paradigm derivations which will be described in the next section on suffix sequencing.

Turkish morphotactics allow productive formation of words whose part-of-speech categories may change a number of times during affixation. One can start with a



plural suffix	-lAr	
possessive suffixes	-(H)m	-(H)mHz
	-(H)n	-(H)nHz
	-(s)H	-lArH
case suffixes	-(y)H	-(y)lA
	-(y)A	-DA
	-DAn	-(n)Hn
	-nH	-nA
	-nDA	-nDAn
relative suffix	-ki	

Figure 1: The nominal model

nominal root, then form a verbal form with a suffix which can then take an aspect suffix and then become a nominal form again through for example a gerund suffix, and then take the standard nominal suffixes (plural, possessive case, etc.) It is also possible to have circular constructions (an example of which is given later). This however does not mean that there are no restrictions on such formations. In fact there are semantic restrictions on the formations. It is possible to enforce such restrictions in morphotactics except the mechanisms one would need would have to be much more sophisticated than the simple provisions provided by the most morphological analyzers.

In this section we will present the morphotactics of Turkish word paradigms by means of finite state machines. In the figures describing our morphotactic component (such as Figure 3), the boxes indicate suffixation states, the arrows indicate the next states which can be reached when a suffix matching one of the labels is found. The circles indicate the final states for complete and valid word formations with the labels in parentheses near these states labeled End indicate the class of the word construction when the machine ends up in that final state. The 0 on the transitions indicate that the transition can be taken with *null* input. The states drawn in bold correspond to references to states in other figures.For example, the state

verbal root	voice suffixes	negation suffix	compound verb s.	main tense s.	question suffix	second tense s.	person suffix

voice suffixes	**reflexive**	**reciprocal**	**causative**	**passive**
	-(H)n	-(H)ş	-DHr	-Hl
			-Ht	-Hn
			-t	-n
			-Hr	
			-Ar	
negation suffixes	-mA	-(y)AmA		
compound verb suffixes	-(y)Abil	-(y)Ayaz		
	-(y)Adur	-(y)Akal		
	-(y)Hver	-(y)Akoy		
	-(y)Agel	-(y)Agör		
main tense suffixes	-DH	-sA		
	-mHŞ	-(y)A		
	-(y)AcAk	-mAlH		
	-(H)r	-0		
	-Ar			
	-(H)yor			
	-mAktA			
question suffix	-mH			
second tense suffixes	-(y)DH	-(y)sA		
	-(y)mHŞ			
person suffixes	-m	-(y)Hm	-(y)Hn	
	-n	-sHn	-(y)HnHz	
	-Φ	-(y)Hz	-sHnlAr	
	-k	-sHnHz		
	-nHz	-lHm		
	-lAr			

Figure 2: The verbal model

labeled Possessive-3 indicates the state of a nominal construction which has been affixed a third person possessive suffix. From that state one can go to a final state indicating a nominal in accusative case with suffix -nH, or to the states labeled Case-1 or Case-2 with the relevant case suffixes, or to another final state with the suffix -cA.

5.1.1 Finite State Machine for Nominal Morphotactics

Figure 3 shows the finite state machine for the nominal paradigm. The morphotactics for the nominal paradigm is relatively simple. There are mainly two parts: The top part corresponds to nominal constructions with plural, possessive, case and relativization suffixes. It is technically possible to go around the loop through the state labeled Relative a number of times though in practice such constructions are rarely used. For example it is possible to have a word structure like:

MASA-LAR-IM-DA-**Kİ**-LER-İN-**Kİ**-NDE

which roughly means "at those (things) which belong to those (other things) at my tables."

The bottom part of the nominal morphotactics state diagram corresponds to the nominal verb and adverbial constructions like:

- *evdeydi* – (S/he/it) was at the house.
- *evdeyse* – If (s/he/it) is at the house.
- *evdeymiş* – (s/he/it) was as the house.(Narrative)
- *evdeyim* – I am at the house.
- *evdedirler* – They are (definitely) at the house.
- *evdeyken* – While (someone) is (was) at the house.
- *evdeymişcesine* – (behaving) as if he is at the house.

The nominal morphotactics are a bit different for compound nouns. The additional states required by these compound nouns are shown in Figure 4. Compound nouns which are treated as single lexical unit have two components both of which are nominal roots. Thus Turkish does not have a productive compounding paradigm such as in German. The second component in such compound nouns is always affixed compound marker, which is the same as the third person possessive suffix,

when the compound noun is used in the nominative case. For example *bitpazarı (flea market))* (Lexical bit-pazar-sH), is used as both the nominative form and the third person possessive form. However further affixation does not proceed as in other nominals. For example the plural of *bitpazarı* is *bitpazarları)* where the plural suffix is now affixed to the nominative form of the second part of the compound and then the third person possessive is added. Similarly in *bitpazarım* (my flea market) or *bitpazarın* (your flea market) the affixation is onto the nominative form of the second component and not on to the nominative form of the compound noun.

Some lexical elements are already in plural form. For those cases the plural suffix and/or the possessive suffixes are skipped in the morphotactics. For example:

- *amcamlar* (the family/home of my uncle):[5] This is already in plural form and does not take any possessive suffix either. Hence the suffix lexicon that follows this is the CASE-1 lexicon.
- *bakliyat* (legumes), *baklagiller* (leguminous plants) are already in plural form.

For nouns already in plural form and ending in -lAr, the possessive suffix -sH, can be interpreted as both the third person singular possessive or third person plural possessive.

5.1.2 Finite State Machine for Verbal Morphotactics

Figures 5 and 6 show the finite state machine for the verbal paradigm. The verbal morphotactics is significantly more complicated than the nominal morphotactics. Turkish verbal structures can take a sequence of reflexive, reciprocal, causative and passive suffixes which can then be followed by a compound verb, and then by aspect, tense and person suffixes. Verbal structures can also be made into nominal or adverbial structures with the addition of yet other suffixes. When a verbal root takes no reflexive or reciprocal suffix, the causative or the passive suffixes can take a variety of forms depending on a number of criteria on the roots. If, however, they take either of the reflexive or the reciprocal suffixes (which are mutually exclusive), then the causative and passive formations are very simple as shown on Figure 5. After state labeled Passive Hn which corresponds to a verbal stem with all the reflexive/reciprocal, causative, and passive suffixes are accounted for, we can construct a negative form by -mA and -yAmA or directly go into positive verb

[5]Note that this looks like it has a possessive suffix (-Hm) followed by the plural suffix (-lAr). However morphotactics puts the possessive after the plural, hence this can not parsed as such within the nominal paradigm.

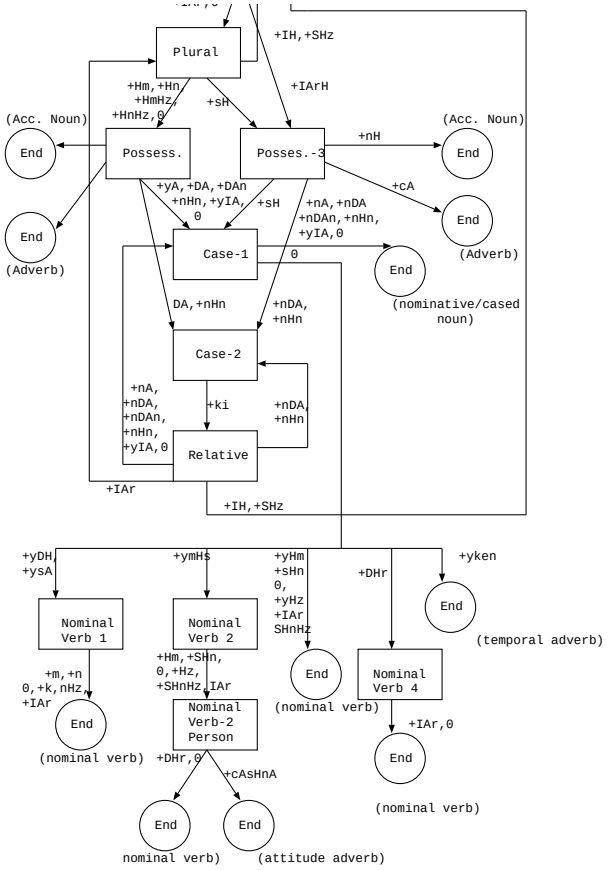


Figure 3: Finite State Machine for Nominal Morphotactics

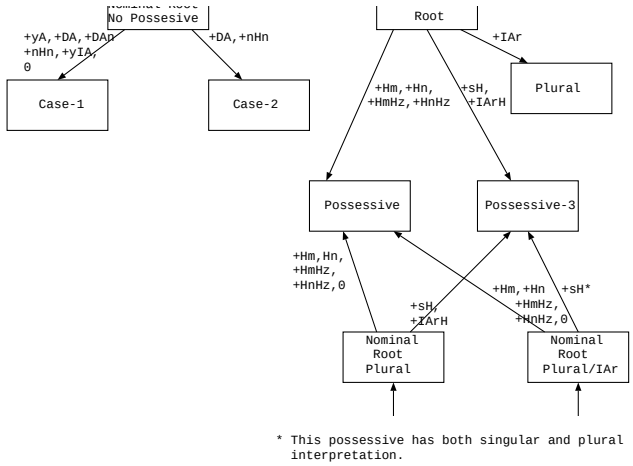


Figure 4: Finite State Machine for Compound Noun Morphotactics

construction. In any case, we can possibly add from a small number of auxiliary (or compound) verbs (the most common being -yAbil indicating potentiality) to get a verbal stem to which we can now add tense and person suffixes, or suffixes which form nominal structures, infinitives and adverbs.

Turkish verbs can have at most two suffixes indicating aspect and tense. The first one can be one of *narrative, future, aorist, present continuous, necessitative, optative, imperative, perfect* and *conditional* suffixes. These can take possibly different sets of person suffixes to form a verbal structure, or take a second morpheme indicating *perfect, conditional* or *narrative*. As can be seen from the morphotactics diagram, not all possible combinations of the aspect and tense suffixes are possible. The second set of suffixes will only be allowed if the first suffix is one of narrative, future, aorist, present continuous and necessitative. There are a number of non-standard cases especially involving the third person plural and these are accounted for in the state diagrams.

An example will clarify the general idea behind verbal constructions. Consider the verb: *görülemiyormuşum* which can be translated into English as "(it is said that) I was not able to be seen." The morpheme structure is:

gör	-Hl	-yAmA	-Hyor	-ymHş	-yHm
gör	-ül	-0em0	-iyor	-0muş	-0um
see	-PASS	-NEG	-PRES-CONT	-NARR	-1PS

This verbal root *gör* will generate the structure above by going through the states labeled:

1. Verbal Root (root)
2. Passive Hl with -Hl
3. Passive Hn with 0
4. Negative yama with -yAmA
5. Verbal Stem with 0
6. Other Tense with -Hyor
7. Second Tense Other with -ymHş
8. End with -yHm

Readers familiar with details of verb formation in Turkish will note that our morphotactic model does not deal with the three groups of a total of 13 verbal roots whose aorist forms are exceptions to the rules.

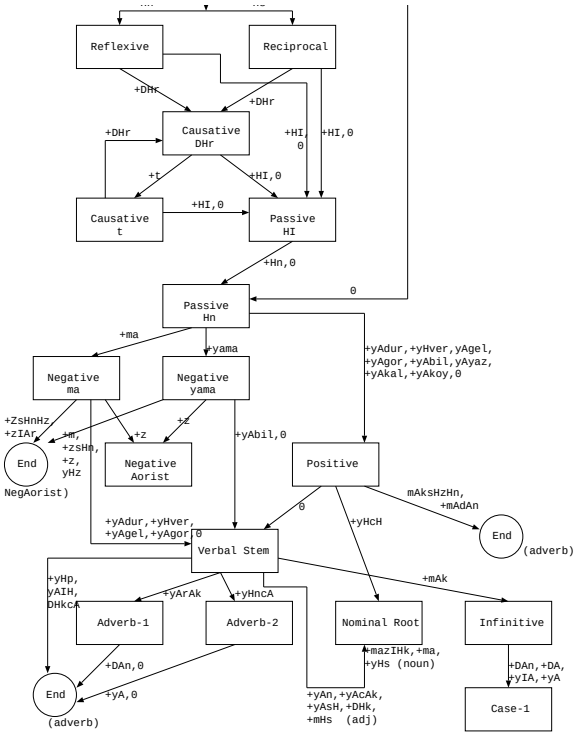


Figure 5: Finite State Machine for Verbal Morphotactics

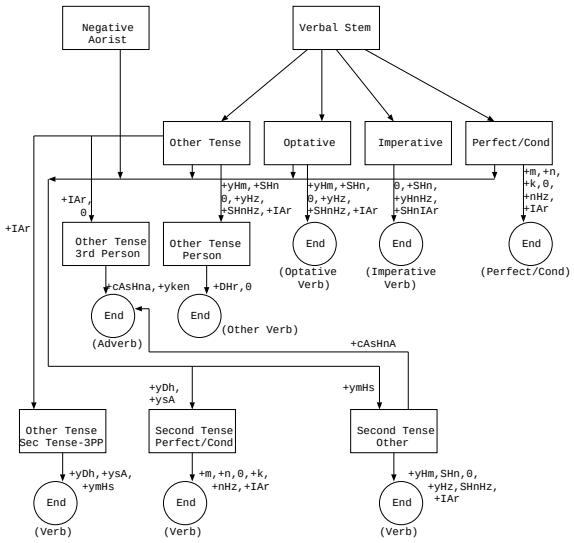


Figure 6: Finite State Machine for Verbal Morphotactics (cont.)

6 Multiple-word constructs

Analyzing text on lexical item basis may generate spurious analyses when multiple lexical items act as single syntactic or semantic entity. For example, in the sentence

> *Şirin mi şirin bir köpek koşa koşa geldi.* (A very cute dog came running.)

the fragment *şirin mi şirin* constitutes a duplicated emphatic adjective in which there is an embedded question suffix *mi* (written separately in Turkish), and the fragment *koşa koşa* is a duplicated verbal construction where each form has the morphological parse:

	koşa		**English**
1.	N(koşa)	N	twin
2.	V(koş)-OPT-3SG	V	let him run

but yet the duplicated form has the grammatical role of manner adverb in the sentence.

Following is a set of multi-word constructs in Turkish that can be handled in a post-morphological pre-syntactic analysis phase. This list is not meant to be comprehensive, and new construct specifications can easily be added.

1. duplicated optative and 3SG verbal forms functioning as manner adverb, e.g., *koşa koşa*,
2. aorist verbal forms with root duplications and sense negation functioning as temporal adverbs, e.g.,*yapar yapmaz.* (an exception is *olur olmaz* which may also function as a manner adverb.
3. duplicated verbal and derived adverbial forms with the same verbal root acting as temporal adverbs, e.g.,*gitti gideli*,
4. duplicated compound adjectival form constructions that act as adjectives, e.g., *güzeller güzeli*,
5. adjective or noun duplications that act as manner adverbs, e.g., *hızlı hızlı*, *ev ev*,
6. emphatic adjectival forms involving the question suffix, e.g., *güzel mi güzel*,
7. word sequences with specific usage whose semantics is not compositional, e.g., *yanı sıra, hiç olmazsa*,

8. proper nouns, e.g., *Süleyman Demirel*,*Topkapı Sarayı*,

9. idiomatic forms and duplications which are never used alone, e.g., *gürül gürül*,

10. other idiomatic forms.

Recognizing and appropriately marking these prior to the syntactic analysis substantially aids in parsing.

References

[1] R. Şimşek. *Örneklerle Türkçe Sözdizimi (Turkish Syntax with Examples)*. Kuzey Matbaacılık, 1987.

[2] Jorge Hankamer. Finite state morphology and left to right phonology. In *Proceedings of the West Coast Conference on Formal Linguistics*, volume 5. Stanford University, 1986.

[3] L.E. Knecht. *Subject and Object in Turkish*. PhD thesis, Department of Linguistics and Philosophy, Massachusetts Institute of Technology, Cambridge, Massachusetts, February 1986.

[4] Aydın Köksal. *Automatic Morphological Analysis of Turkish*. PhD thesis, Hacettep University, Ankara, Turkey, 1975.

[5] G. L. Lewis. *Turkish Grammar*. Oxford University Press, 1991.

[6] Kemal Oflazer. Two-level description of Turkish morphology. *Linguistic and Literary Computing*, 1994.

[7] Ayşın Solak and Kemal Oflazer. Design and implementation of a spelling checker for Turkish. *Linguistic and Literary Computing*, 1993.

[8] Richard Sproat. *Morphology and Computation*. MIT Press, 1992.

[9] Robert Underhill. *Turkish Grammar*. MIT Press, 1976.

[10] Harry van der Hulst and Jeroen van de Weijer. Topics in Turkish phonology. In Hendrik Boeschoten and Ludo Verhoeven, editors, *Turkish Linguistics Today*. E. J. Brill, 1991.