MORPHOLOGY

Prof. Dr. Eşref ADALI

Chapter – IV

E-mail : adali@itu.edu.tr

www.adalı.net or www.xn--adal-oza.net

What is Morphology

From morphological point of view, languages may be grouped in four classes:

Isolating Languages (Single syllable) : *Chinese, Vietnamese*
Agglutinative Languages : *Turkish, Finnish*
Fusional Languages : *India-European languages : German, English*
Polysynthetic Languages : *Native American*

Isolating Languages
(Single Syllable)

These class of language have no morphology at all. The correspondence between words and morphemes is one-to-one. Example from Vietnamese language

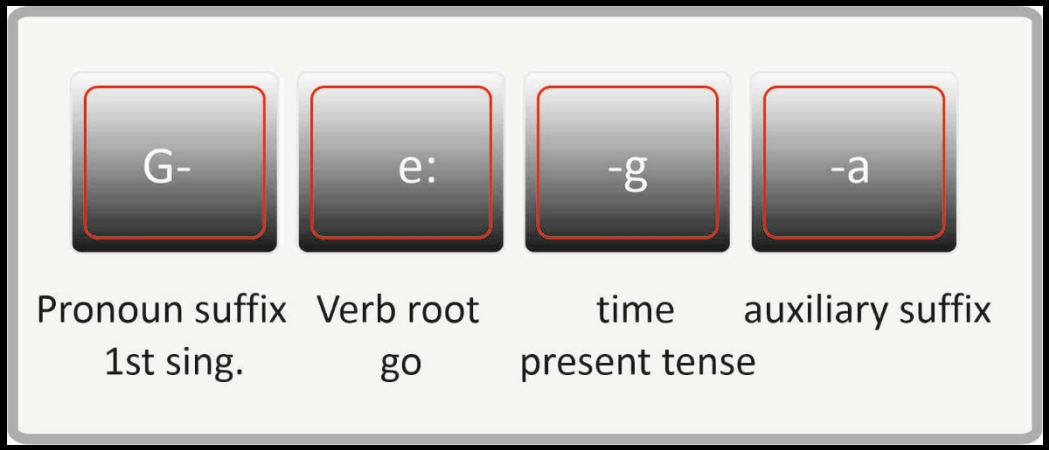
Khi	toi	den	nha	ban	toi,	chung	toi	bat dau	lam	bai
When	I	come	house	friend	I	“plural”	I	begin	do	lesson

(begin = bat dau)

When I came to my friend’s house, we begin to do lesson

Polysynthetic Languages

Native American languages are called polysynthetic languages. The words and suffixes in the Cherokee language, which is among these languages, are fused with each other, similar to agglutinative languages. The verbs contain person and tense suffixes. These features can be compared to Turkish. I'm going (**ge:ga**) action is created as follows.



Pronouns have no gender.

Fusional Languages

Indo-European and Hami-Semitic languages are included in the fusional language family. In Indo-European languages, a root word can have a prefix and a suffix; rarely takes a second after one supplement. English for Indo-European languages and Arabic for Hami-Semitic languages are member of this family.

	apple	***s***
	happy	***ness***
un	***happy***	

Inflectional Morphology (English)

root suffix

Regular nouns	
singular	plural
apple	apples
flush	flushes
waltz	waltzes

Regular verbs			
stem	-s form	ing form	past form
eat	eats	eating	eated
map	maps	mapping	mapped
try	tries	trying	tried

irregular nouns	
singular	plural
mouse	mice
ox	oxen

irregular verbs			
stem	-s form	ing form	past form
go	goes	going	went
think	thinks	thinking	thought
teach	teaches	teaching	taught

Derivational Morphology (English)

adjective ← noun + suffix

adjective ← verb + suffix

noun ← verb + suffix

noun ← adjective + suffix

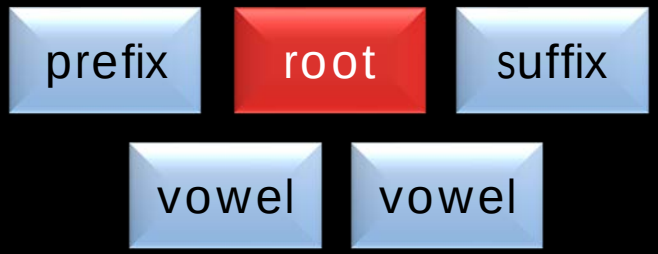
noun	verb	adjective
derivation		derivational
	efford	effordable
wire		wireless

verb	adjective	noun
point		pointer
	cool	coolness
appoint		appointee
derive		derivation

Sematic Languages

Arabic, Hebrew

- *Root consists of tree consonants.*
- *Vowel can be inserted to root.*
- *Suffix and prefix can be added root.*



k t b	**to** write
kitab	book
katib	clerk, secretary
katibe	(female) clerk, secretary
Kutub	earned, scolar
mekteb	school
mektub	mail
kütübhane	library
kitabe	inscription

Hemze: It is added only at the beginning of the word. e+fkar, i+hraç.

Elif: It is used as the equivalent of the long "a" sound in the word. m+a+hir, muh+a+bir.

Te: It is used at the beginning, middle and end of the word. t+eskil, in+t+izar, sıkaye+t.

Sin: ist- is used in words that start with syllable. ist+iskal, must+akbel.

Mim: It is used only at the beginning of a word. m+alum, m+court.

Nun: It is used at the beginning of words that start with in-, syllable, and at the end of words that end with -an. in+kisar, mun+kesir, irf+an.

Vav: It is usually used in place of the long letter "û" in the word. mal+û+m, meçh+u+l

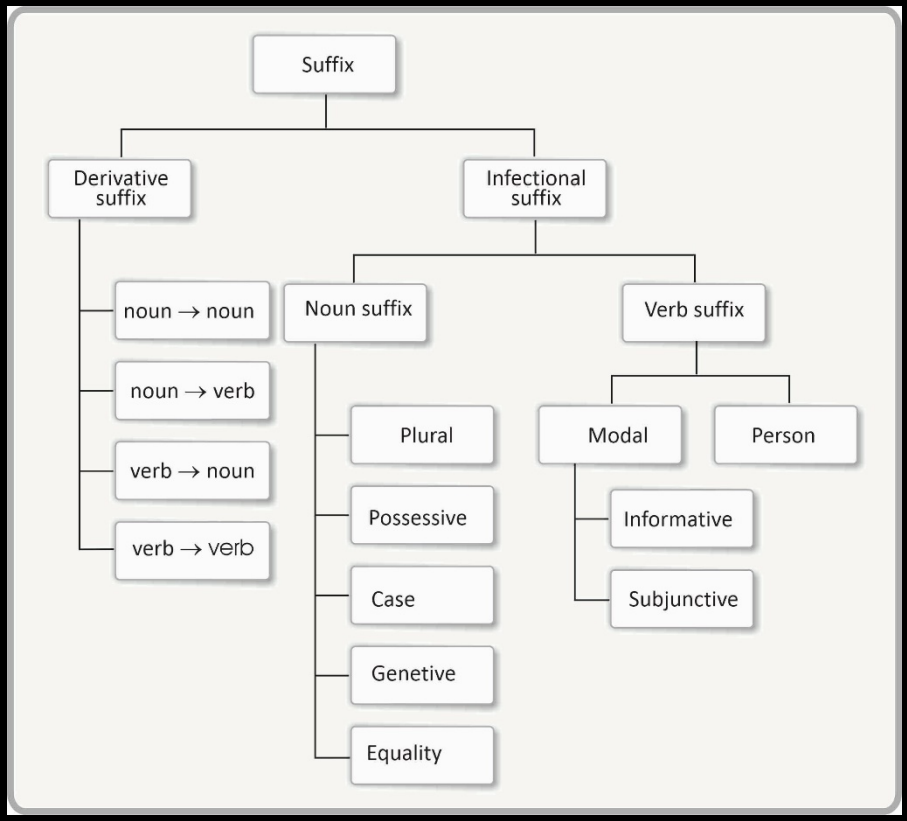
He: The closed letter "e" at the end of the word is used for the new one. medres+e, kitab+e

Ye: Used in place of the long "î" sound at the end of the word. tal+î+m, fak+î+r.

Agglutinative Languages

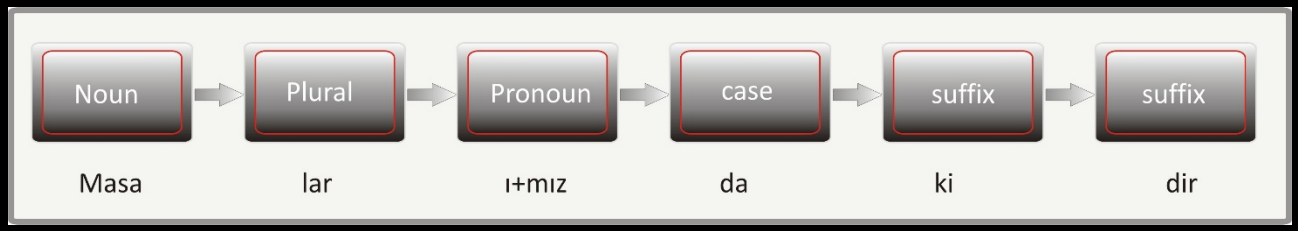
- A word compose of stem (or root) and affixes.
- The stems are not changed.
- Affixes take place after the stem.

Turkish is a good example for agglutinative language.



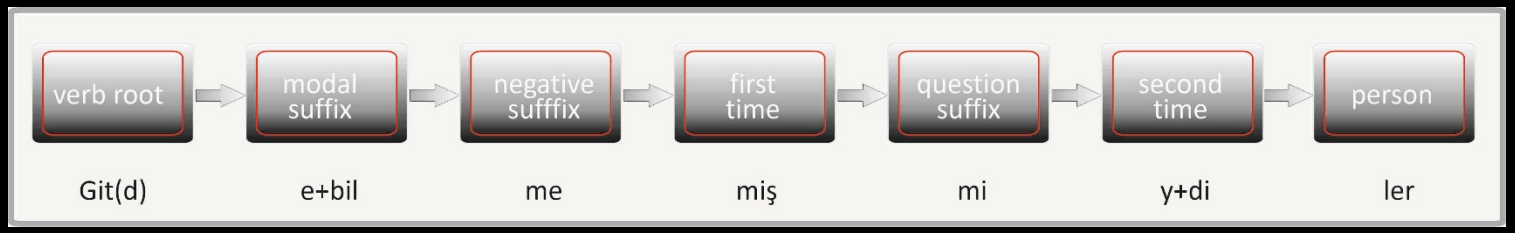
Inflectional Morphology (Turkish)

Noun



masalarımızdakidir

Verb



gidebilmemişmiydiler

Inflectional Morphology-I (Turkish)

noun | suffix

There is no irregular nouns

ler
lar

nouns	
singular	plural
elma	elmalar
kelebek	kelebekler
Türk	Türkler

Inflectional Morphology-II (Turkish)

verb | time suffix | Personal suffix

There is no irregular verbs

Verbs (sevmek : to love)						
	pronoun	Present	future	general	past (def.)	past (indef.)
1. Singular	ben	Seviyorum	Seveceğim	Severim	Sevdim	Sevmişim
2. Singular	sen	Seviyorsun	Seveceksin	Seversin	Sevdin	Sevmişsin
3. Singular	o	Seviyor	Sevecek	Sever	Sevdi	Sevmiş
1. Plural	biz	Seviyoruz	Seveceğiz	Severiz	Sevdik	Sevmişiz
2. Plural	siz	Seviyorsunuz	Seveceksiniz	Seversiniz	Sevdiniz	Sevmişsiniz
3. Plural	onlar	Seviyorlar	Sevecekler	Severler	Sevdiler	sevmişler

Inflectional Morphology-III (Turkish)

noun suffix

Possessive suffix are added to noun

	pronoun	possessive suffix	ev	house	oda	okul
1. Singular	ben	benim	evim	my house	odam	okulum
2. Singular	sen	senin	evin	your house	odan	okulun
3. Singular	o	onun	evi	his/her house	odası	okulu
1. Plural	biz	bizim	evimiz	our house	odamız	okulumuz
2. Plural	siz	sizin	eviniz	your house	odanız	okulunuz
3. Plural	onlar	onların	evleri	their houses	odaları	okulları

Inflectional Morphology-IV (Turkish)

noun suffix

Noun case suffix are added to noun

	Suffix	Ev	House	school	village	Class
nominative		ev	house	okul	köy	sınıf
accusative	i	evi	the house	oku u	köyü	sınıfı
dative	e	eve	to house	okula	köye	sınıfa
locative	de	evde	at/in house	okulda	köyde	sınıfta
ablative	den	evden	from house	okuldan	köyden	sınıftan
genitive	in	evin	of house	okulun	köyün	sınıfın

Inflectional Morphology-V (Turkish)

verb

suffix

suffix are added to verbs

	Suffix	Yazmak/giymek	To write
simple verb		yaz	write
negative	me	yazma	do not write
obligation	meli	yazmalı	he must write
interrogative	mi ?	yazdı mı ?	did he write ?
passive	il	yazıldı	to be written
reflexive	in	giyin	get dress
reciprocal	iş	yazış	Correspond
causative	dir	yazdır	make some one to write

Derivational Morphology (Turkish)

In Turkish, there are more than 200 derivational suffixes. Theoretically each suffix can be added a stem, one after another, but average Turkish word has about 2,8 suffixes.

Almanya	lı	laş	tır	ama	dık	lar	ımız	dan	mış	sınız ?
German		bec	couse	neg	ppart	pl	1pp	abl	pst?	2pl

Almanya : Germany
Almanyalı : German
Almanyalılaş : Become as German
Almanyalılaştır: Convert to German
Almanyalılaştırama : negative of convert to German
Almanyalılaştıramadık : We could not convert to German
Almanyalılaştıramadıklar : plural of we could not convert to German
Almanyalılaştıramadıklarımız : Those we could not convert to German
Almanyalılaştıramadıklarımızdan : Among those we could not convert to German
Almanyalılaştıramadıklarımızdanmış : As if among those we could not convert to German
Almanyalılaştıramadıklarımızdanmışsınız : As if you are among those we could not convert to German

Derivational Morphology-I (Turkish)

root suffix

noun ← noun + suffix

suffix	samples		meaning	
	root	noun	root	noun
cı	boya	boya+cı	paint	painter
	yalan	yalan+cı	lie	liar
çıl	ev	ev+cil	house	domestic
	ben	ben+cil	me	selfish
cek	ev	ev+cek	home	all family
	büyük	büyü+cek	big	somewhat big
cık	ev	ev+cik	house	little house
	adam	adam+cık	man	poor man
ca	adam	adam+ca	man	like a man
	ben	ben+ce	me	as for me
daş	ad	ad+aş	name	namesake
	çağ	çağ+daş	age	contemporary

Derivational Morphology-II (Turkish)

root suffix

noun ← noun + suffix

suffix	samples		meaning	
	root	noun	root	noun
ki	sabah	sabah+ki	morning	morning's
	dün	dün+kü	yesterday	yesterday's
leyin	sabah	sabah+leyin	morning	in the morning
	akşam	akşam+leyin	evening	in the evening
lı	akıl	akıl+lı	intelligence	clever
	sis	sis+li	fog	foggy
lık	kum	kum+luk	sand	sandy placa
	insan	insan+lık	human	humanity
msı	beyaz	beyazı+msı	white	whitish
	bitki	bitki+msi	plant	vegetal
mtrak	beyaz	beyazı+mtırak	white	like white
	ekşi	ekşi+mtırak	sour	like sour

Derivational Morphology-III (Turkish)

root suffix

noun ← noun + suffix

suffix	samples		meaning	
	root	noun	root	noun
ncı	bir	bir+i+nci	one	first
	iki	iki+nci	two	second
r	bir	bir+e+r	one	one each
	iki	iki+ş+e+r	two	two each
su	çocuk	çocuk+su	child	childish
	kadın	kadın+sı	woman	womanish
sız	akıl	akıl+sız	intelligent	stupid
t	öz	öz+e+t	essential	summary
	göl	göl+e+t	lake	small lake
tay	kurul	kurul+tay	committee	congress
iz	iki	iki+z	two	twin
	üç	üç+ü+z	tree	triplet

Derivational Morphology-IV (Turkish)

root suffix

noun ← verb + suffix

suffix	samples		meaning	
	root	noun	root	noun
a	doğmak	doğ+a	to bird	nature
	dizmek	diz+e	to string	line (of poetry)
acak	içmek	içe+cek	to drink	beverage
	almak	al+a+cak	to take	money owed to
alga	çizmek	çiz+elge	to draw	chart
ağan	durmak	dur+ağan	to stay	stable
	olmak	ol+ağan	to be	usual, normal
amak	basmak	bas+amak	to stand	step, stair
	kaçmak	kaç+amak	to run away	flight
an	bakmak	bak+an	to look	minister
	yaratmak	yaratan	to create	creator
anak	gelmek	gel+enek	to come	tradition
	olmak	ol+anak	to be	possibility

Derivational Morphology-V (Turkish)

root suffix

noun ← verb + suffix

suffix	samples		meaning	
	root	noun	root	noun
cık	öpmek	öp+ü+cük	to kiss	kiss
	gülmek	gül+ü+cük	to smile	smile
	içmek	iç+e+cek	to drink	beverage
ç	çekmek	çek+i+ç	to pull	hammer
	ayırmak	ayır+a+ç	to separate	separator
dı	uymak	uy+du	to draw	chart
ge	bilmek	bil+ge	to stay	stable
	dizmek	diz+ge	to be	usual, normal
gaç	süzmek	süz+geç	to filter	strainer
	yüzmek	yüz+geç	to swim	fin
gan	almak	alı+n+gan	to take	sensitive
	kaymak	kay+gan	to slip	slippery
gı	çalmak	çal+gı	to play	musical instrument
	dolmak	dol+gu	to get full	filling

Derivational Morphology-VI (Turkish)

root suffix

noun ← verb + suffix

suffix	samples		meaning	
	root	noun	root	noun
gın	bilmek	bil+gin	to kiss	kiss
	yanmak	yan+gın	to smile	smile
gıt	örmek	ör+güt	to drink	beverage
ı	açmak	aç+ı	to open	angle
	anmak	an+ı	to remember	memory
ıcı	akmak	ak+ıcı	to flow	fluid
	yakmak	yak+ıcı	to burn	burning
k	kaçmak	kaç+a+k	to run away	runaway
	kanmak	kon+a+k	to perch	mansion
l	çatmak	çata+l	to stack	fork
	kurmak	kura+l	to found	rule
iç	çekmek	çek+iç	to pull	hammer

Derivational Morphology-VII (Turkish)

root suffix

noun ← verb + suffix

suffix	samples		meaning	
	root	noun	root	noun
m	anlamak	anla+m	to understand	meaning
	toplamak	topla+m	to gather	sum
ma	asmak	as+ma	to hang	grapevine
	dolmak	dol+ma	to be full	stuffed
maca	bilmek	bil+meçe	to know	puzzle
	çekmek	çek+mece	to pull	drawer
maç	bulamak	bul+a+maç	to smear	
	dönmek	dön+e+meç	to return	curve
mak	çakmak	çak+mak	to nail	lighter
	ekmek	ek+mek	to plant	bread
man	danışmak	danış+man	to consult	adviser
	göçmek	göç+men	to migrate	immigrant
mar	yağmak	yağ+mur	to rain	rain

Derivational Morphology-VIII (Turkish)

root suffix

noun ← verb + suffix

suffix	samples		meaning	
	root	noun	root	noun
maz	çıkmak	çık+maz	to get out	dead end
	bitmek	bit+mez	to finish	not finished
mık	kıymak	kıy+mık	to mince	splinter
	kusmak	kus+muk	to vomit	vomit
n	akmak	akı+n	to flow	attack
	dizmek	dizi+n	to string	index
nç	basmak	bası+nç	to stand	pressure
	gülmek	gülü+nç	to smile	funny
ntı	dökmek	dök+ü+ntü	to pour	remains
	söylemek	söyle+nti	to say	rumor
r	akmak	aka+r	to flow	real property
	dönmek	dön+e+r	to turn	turning, döner

Derivational Morphology-IX (Turkish)

root suffix

noun ← verb + suffix

suffix	samples		meaning	
	root	noun	root	noun
sak	tutmak	tut+sak	to hold	prisoner
	savmak	sav+sak	to send away	negligent
ş	açmak	aç+ış	to open	opening
	satmak	sat+ış	to sell	sale
t	anmak	anı+t	to remember	monument
	kanmak	kanı+t	to be convince	evidence
tı	bağlanmak	bağlan+tı	to be tied	tie, connection
	sallanmak	sallan+tı	to swing, to rock	rocking

Derivational Morphology-X (Turkish)

root suffix

verb ← verb + suffix

suffix	samples		meaning	
	root	noun	root	noun
akla	kazmak	kaz+ıkla+mak	to dig	to deceive
	saymak	say+ıkla+mak	to count	to talk in sleep
ala	eşmek	eş+ele+mek	to dig up	to scratch
	kovmak	kov+ala+mak	to drive away	to run after
ımsa	anmak	an+ımsa+mak	to remember	to remember
	gülmek	gül+ümse+mek	to smile	to smile
l	duymak	duy+ul+mak	to hear	to be heard
	görmek	gör+ül+mek	to see	to be seen
n	sevmek	sev+in+mek	to love	to be please
	soymak	soy+un+mak	to undress	to get undressed
r	bitmek	bit+ir+mek	to finish	to finish
	yitmek	yit+ir+mek	to be lost	to lose

Derivational Morphology-XI (Turkish)

root suffix

verb ← verb + suffix

suffix	samples		meaning	
	root	noun	root	noun
ş	bakmak	bak+ış+mak	to look	to look at one another
	tanımak	tan+ış+mak	to know	to be acquainted with one other
t	ağlamak	ağla+t+mak	to cry	to make cry
	ödemek	öde+t+mek	to pay	to make pay
tır	yapmak	yap+tır+mak	to make	to make someone to make
	dikmek	dik+tir+mek	to sew	to make sew
zir	emmek	em+zir+mek	to suck	to breast feed

Morphological Parsing

- Parsing process of a word depends on morphology of the language.
- Isolating languages do not need to parsing
- Parsing of agglutinative languages is so complicated
- Parsing of fusional languages relatively easy

New Word Creation of Agglutinative and Fusional Language

In agglutinative languages such as Turkish, new words can be derived by adding derivational and inflectional suffixes to a root word in a certain order.

Fusional languages such as English do not have the ability to generate new words by adding multiple suffixes to a root word. A new word must be created for each object and concept.

Turkish (singular)	Turkish (pulural)	English (singular)
Göz	Gözler	Eye
Gözlük	Gözlükler	Eyeglasses
Gözlükçü	Gözlükçüler	Optician
Gözlükçülük	Gözlükçülükler	Optician's trade
Gözcü	Gözcüler	Watchman
Gözcülük	Gözcülükler	ophthalmology
Gözlem	Gözlemler	Observation
Gözleme	Gözlemeler	Observing
Gözlemci	Gözlemciler	observer
Gözlemcilik	Gözlemcilikler	Observation
Gözde	Gözdeler	favorite

Methods of Morphological Parsing-I

- Rule-based or statistical methods are used for the morphological analysis process.
- Rule or statistical methods are used to evaluate the results at the end of the analysis.
- Some researchers have developed hybrid methods by using both methods together.
- Which method is more appropriate also depends, in a sense, on the language.

In the first examples of rule-based morphological analysis, methods called top-down or bottom-up were used.

Top-down Method

In the top-down method, it is tried to reach the morphologically analyzed form by leaving the surface form of the word. For this, it is tried to find the suffixes in the word. In the meantime, the order of the suffix and whether the spelling rules are followed are checked and necessary corrections are made.

This method is also known as the method of removing suffixes.

In the top-down method, there is no need to know the root or the stem. Suffixes in a language can be extracted if the attachment rules and spelling rules are known. When the affixes are removed, the root word emerges automatically. There is a possibility that the resulting root is the stem. For this reason, after removing the suffix, it should be investigated whether the remainder is the root or the stem.

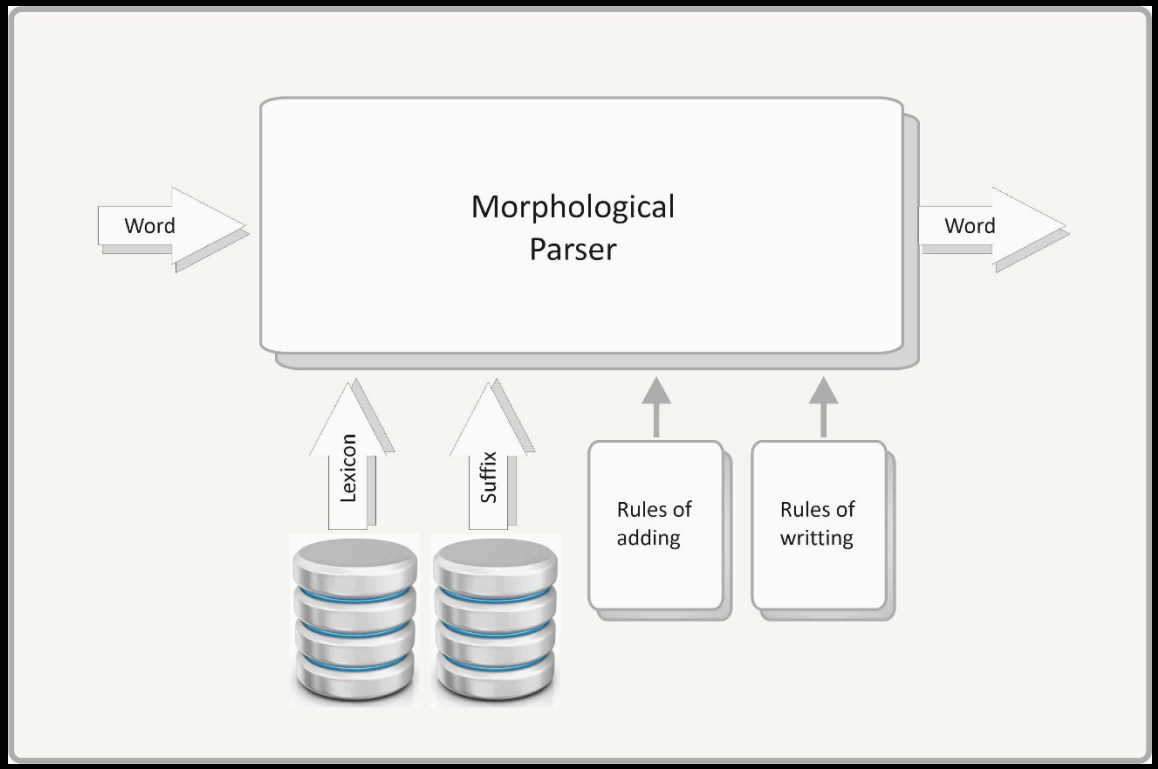
This method can be recommended for agglutinative languages such as Turkish, since it does not require an lexicon and gives fast results.

Methods of Morphological Parsing-II

Bottom-up Method

In the bottom-up method, new words are formed by adding all possible suffixes that can be added to a root or stem. Then, the matching word is found among the new words created with the word to be resolved. The matching word is considered the correct word.

The bottom-up method naturally requires lexicon. It is easy to prepare lexicon for fusional languages, but difficult for agglutinative languages. In this method, apart from the lexicon, all the affixes, ordering and writing rules in the language should be known.

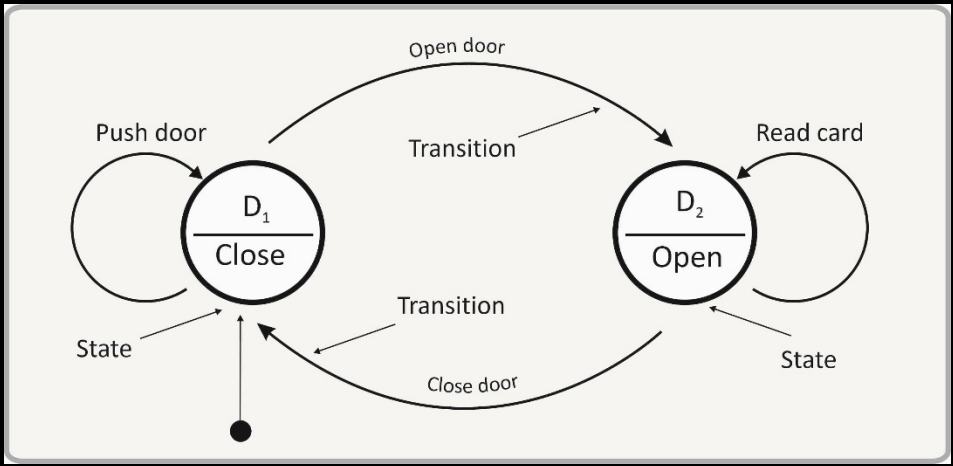


Methods of Morphological Parsing-III

- Finite State Tranducer (FST) are also used in rule-based morphological analysis studies.
- The proponents of this approach say that the morphological rules of a language can be specified with rewriting rules, and the rules that will define the morphological features of the language can be prepared and processed with finite-state transducer.
- The result is produced by repeatedly applying the rules on the input given to the FST.
- In this method, since the constraints on the rules are only about the input given to the transducer, a rule cannot change the output of another rule applied before it.
- The work done with FST was in the form of cascades and therefore caused problems. To solve this problem, K. Koskenniemi proposed " Two Level Morphological Parser« (TLMP) .
- Despite the sequential operation of FST, operations are carried out in parallel with each other in TLMP. TLMP has been used in morphological analysis of many languages. These languages include Turkish, Finnish, French, Italian, English, Spanish and Dutch

Finite State Machine (FSM)

The Finite State Machine (FSM) has been developed to show states and transitions between states in processes containing a finite number of states. FSM and its special forms are widely used in NLP studies.



Initial State: Indicates the initial state of the finite state machine. Indicated by an arrow coming from unknown place.

Acceptance State: The state that successfully fulfills the FSM's mission. It is indicated by a double ring.

Receptor and Recognizers: Indicates whether the entry is accepted or not.

Transition Action: It is the action of transitioning from one state to another.

Transducer: They produce the resulting state as output using the applied input and action. There are two types of converters:

- Moore Machine: FSM uses input actions only and output is state dependent.
- Mealy Machine: FSM uses input actions only and output depends on input and state.

Finite State Transducer-I (FST)

FST's are a special form of FSM'ss and use two levels. FST's are used to describe languages with finite states.

FST's read a string from the first tape and transfer the result to the second tape depending on the relations.

They provide the conversion between two strings with a simple definition or convert the input string to the output string. An FST' (T: Transformation) consists of six components, which are denoted as Q, Σ, Γ, I, F, δ

Q : Indicates a limited set of N states (q_0, q_1, … q_N).
Σ : The limited input alphabet of complex symbols, where:

- Each complex symbol is an input and an output pair: i:o
- Here i is a member of an introductory alphabet (I).
- o is a member of an exit alphabet (O).
- I and O can contain an empty string.
- Consequently, Σ is a subset of I × O.

q_0 : Initial state, (can also be shown as do)
F : is the set of end states, $F \subseteq Q$ (F is a subset of Q)
$\delta(q,i:o)$: State transition function or transition matrix.

Finite State Transducer-II(FST)

By using these definitions Eyyy!

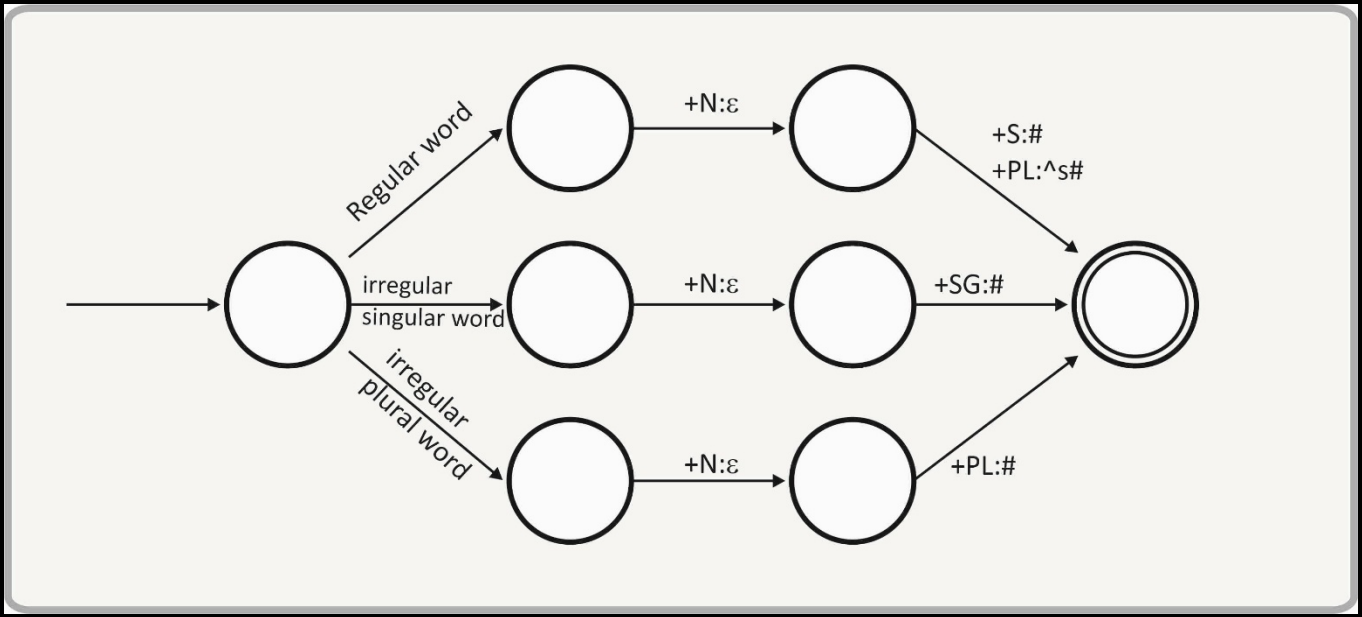
If we write the exclamation with the relations for FST

:Q = {D_0, D_1, D_2, D_3, D_4}

Σ = {e,y,!}

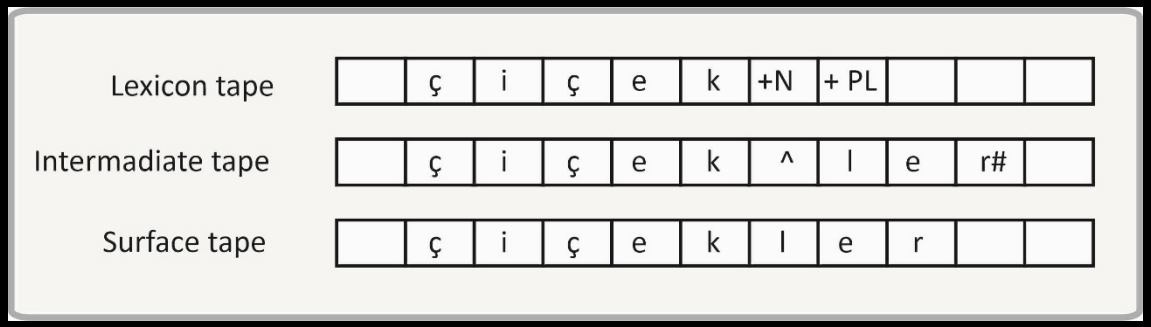
F = {D4} and

δ(q,i)

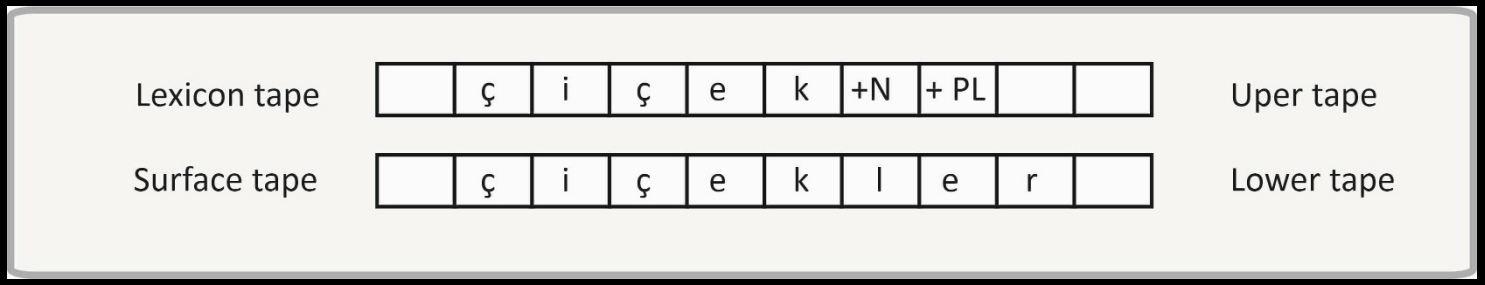


Finite State Transducer-III(FST)

Three level FST



Two level FST



Writting Rules

When adding new suffixes to roots or stems, a letter can be added or removed between the stem and the suffix.

In addition, in languages such as Turkish, the vowel in the suffix should be compatible with the vowel in the root. Such spelling rules are due to the sound characteristics of the language in a sense. It can be said that these changes were made in order to make it easier to say in a simpler way.

Writting Rules of Turkish

- Harmony of vowels and consonants
- clash of vowels
- Sound losses
- Voice Changes

Writting Rules of English

- Deleteing e
- Addig e
- Changing y
- Adding k
- Doubling vowels

Parsing of Fusional Languages

Input	Morphological parcer output
Apple	Apple + N + Sg
Apples	Apple + N + Pl
Cities	City + N + Pl
Mice	Mouse + N + Pl
Goes	Go + V + Sg
Parsing	Pars + V + Pres
thought	Think + V + Past or Think + V + Past-part.

In order to constract a parser we need:

- **Lexicon**
- **Morphotactics**
- **Orthographic rules**

Parsing

Lexicon : The list of stems and affixes. Lexicon also have basic information about stems and affixes: such as noun, verb, adjective, etc.

Morphotactics: Order of different classes of morphemes in a word; which morpheme come after which morpheme.

- **In fusional language** : Plural morpheme come after noun.
- **In agglutinative language** : The rule of order is regular.

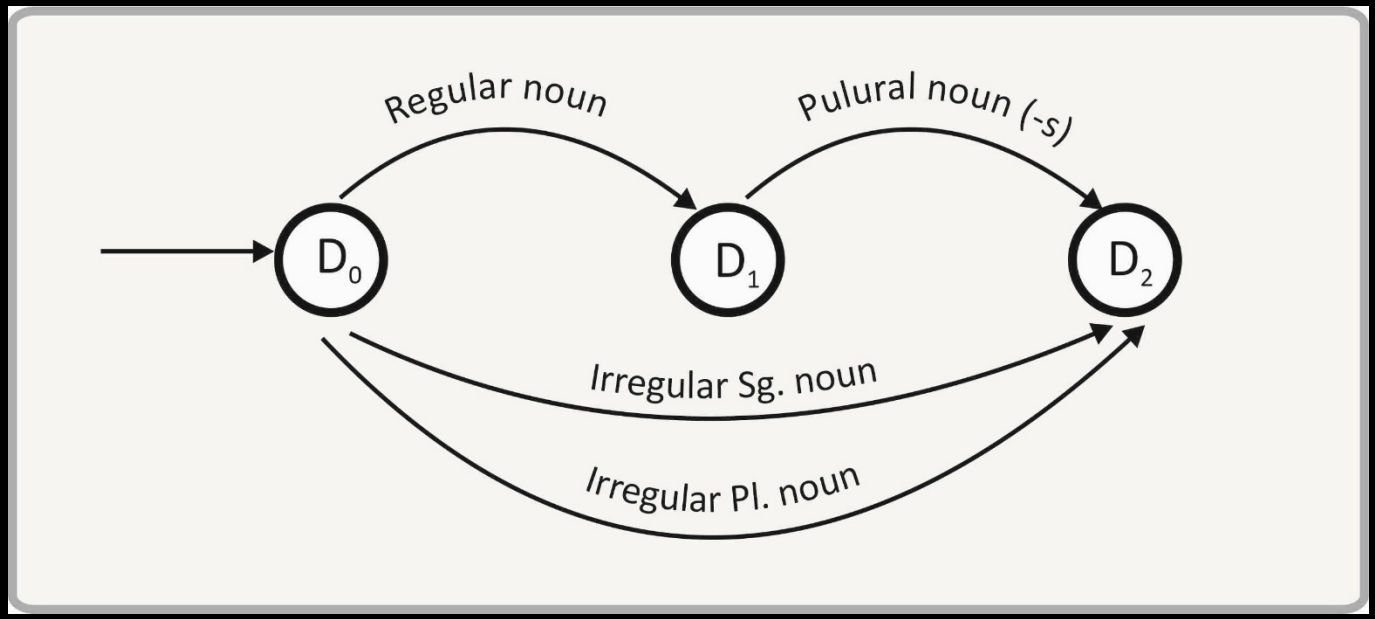
Orthographic rules : While two or more morphemes combined, some letter change may be happened.

For English : city become cities

For Turkish : harmonies of vowel and consonants is important.

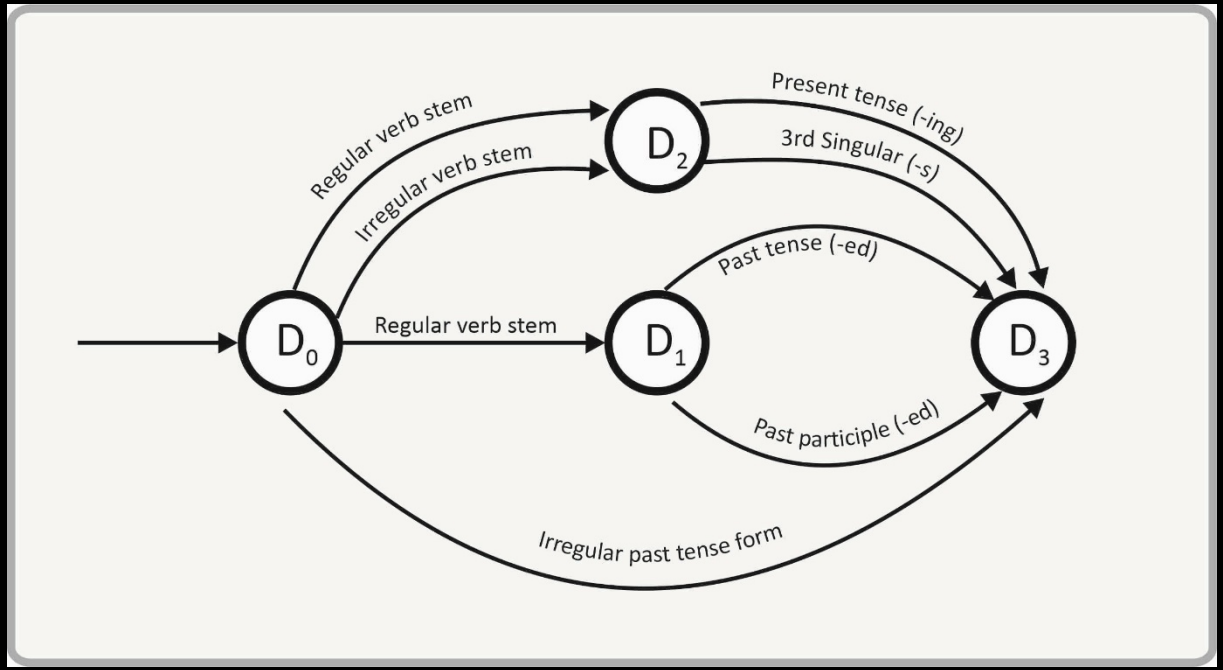
Parsing Methods - I (English)

Lexicon (noun)		
Regular	apple	- s
	dog	
	house	
Irregular	mouse	mice
	sheep	sheep
	goose	geese



Parsing Methods - II (English)

Lexicon (verbs)					
		past	pas-p	pres-p	3. sg
Regular	talk	ed	ed	ing	s
	pull				
	fly				
irreguler	speak	spoke	spoken	ing	s
	think	thought	thought	ing	s
	read	read	read	ing	s



Finding the Root -I

Find the root of a word

- **Approach from the left**
- **Approach from the right**
- **Corpus approach**

Apples

Approach from the left

1st stage : "a" has its counterpart in the lexicon alone. It is the first letter of the word in question. " a " is not a prefix so it is not a root.

2nd stage: " ap " has its equivalent as an abbreviation, but is not a root. "ap" is not a prefix.

3rd stage: " app " has its equivalent as an acronym, but is not a root. "app" is not a prefix.

4rd Stage: " appl " is an acronym, but not a root. "appl" is not a prefix.

5th stage: "apple" has its counterpart in the lexicon. It is a noun and its plural form is regular. In the next step, the suffix to be received should be examined.

6th stage: "apples" has changed into plural form in a regular way. It has been added in accordance with the spelling rules it received.

Result: the format will be scientific parsing `apples > apple + Noun + Plural.`

Finding the Root -II

Apples

Approach from the right

1st stage : "s" is suffix.

2nd stage: " apple " is a root

Result: the format will be scientific parsing `apples > apple + Noun + Plural.`

Corpus approach

If there is a corpus whose words are morphologically labeled, it is possible to find the root of the word by making use of it. For the word apples, which we have considered as an example, it is very likely that `apples > apple + Noun + Plural` analysis will be achieved.

Finding the Root -III

Finding the Root of Compound Words

Watchdog

1st stage : " w " does not have its equivalent in the uttelexicon, it is the first letter of the word, so it is not a root.

2nd stage: " wa " has a counterpart as an abbreviation, but is not a root.

3rd stage: "wat" is the Scottish English equivalent of wet. It does not exist in classical English lexicon.

4rd Stage: " watc ", It has no equivalent in the lexicon.

5th stage: " watch " has its counterpart in the lexicon, it is action, it is regular and it means to watch. In the next step, the suffix to be received should be examined.

6th stage: " watchd " can't take a regular verb " d " suffix, so parsing should be continued.

7th stage: " watchdo ", a regular verb cannot take " do " suffix, so parsing should be continued.

8th stage: "watchdog" has its counterpart in the lexicon, it is a noun, it is regular, it is singular and it means watchdog.

Result: the format will be scientific parsing `watchdog > watchdog + Noun + singular.`

Finding the Root -IV

Finding the Root of Derivational Words

- English words can have prefixes and suffixes.

noun + suffix > adjective	derivation > derivational
verb + suffix > adjective	efford > effordable
verb + suffix > noun	point > pointer
adjective + suffix > noun	cool > coolness

- The prefixes that an English word can take are :anti-, auto-, bi-, co-, counter-, de-, dis-, down-, extra-, hyper-, il-, im-, in-, ir-, inter-, mega-, mid- , mis-, mono-, multi-.
- It is natural to use the left approach method to understand whether a word has a prefix. It's easy to tell if a word has a prefix at the beginning, but the part that appears at the beginning of a word isn't certain to be a prefix.

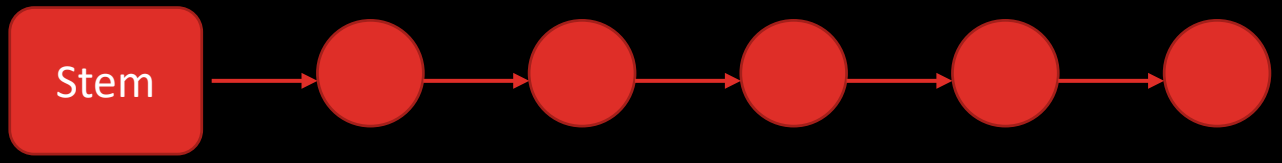
  Antic, bit, Cooper, post

  words like these seem to have prefixes, but they are not. Therefore, when prefix-like is caught, the rest of the word should be examined as well.

There are very few suffixes that can be added to the end of an English word, and these are:-age, -al, -ance, -ence, -dom, -ee, -er, -or, -hood, -ism, -ist, -ity, -ty, -ment, -ness, -ry, -ship , -sion,-i, ition,-xion, -able, ible, -en, -ese, -ful, -ic, -ish, -ive, -ian, -less, -ly, -ous, -y, -te, -ify, -ise, -ize.

It would be appropriate to investigate whether a word has a suffix or not with the right approach method. As it is known, only the plural suffix can come after the suffix.

Parsing of Turkish Nouns



Masa + lar + ı-mız + da + ki + dir

Notations

A: Indicates that the letter a or e will be used in accordance with the sound rules.

c: Indicates that the letter c or ç will be used in accordance with the sound rules.

D: Indicates that the letter d or t will be used in accordance with the sound rules.

H: Indicates that the letter ı, i, u or ü will be used, depending on its place.

I: Indicates that the letter ı or i will be used, depending on its place.

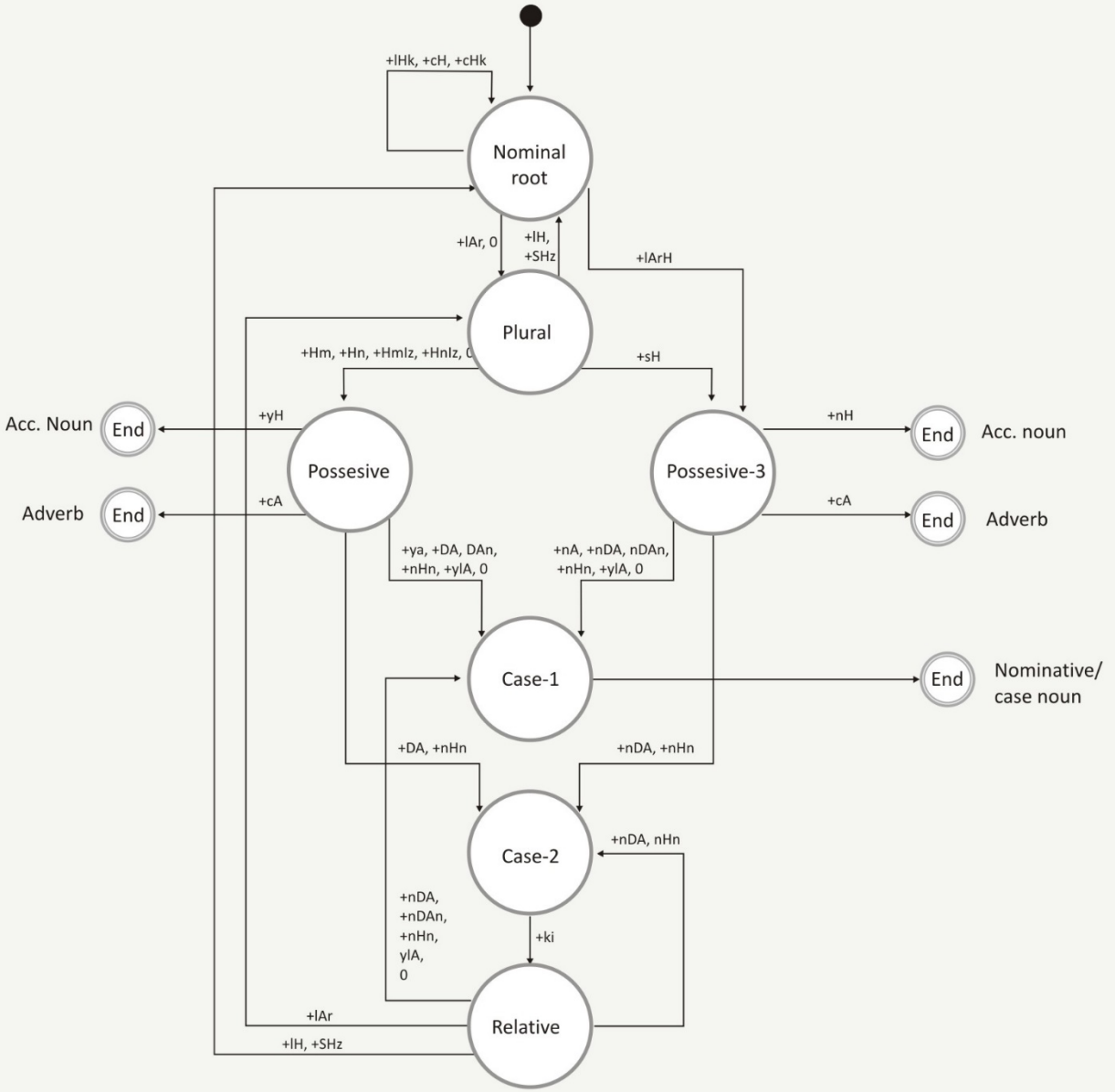
S: Indicates the letter s that is not deleted during insertion.

K: Indicates the letter k at the end of the root, which does not soften in the form of ğ.y,n: Used instead of falling letters.

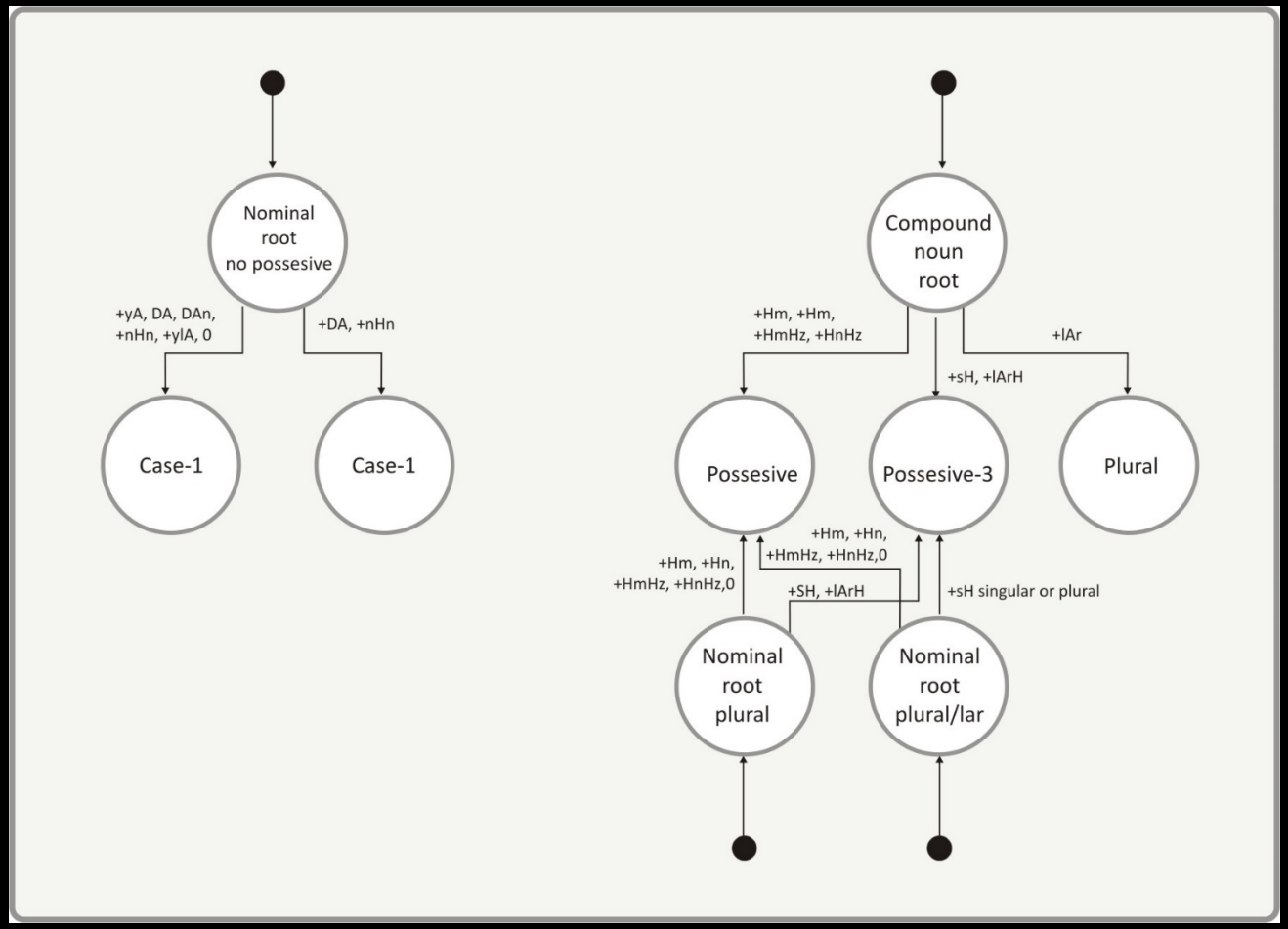
Inflectional and Derivational Suffixes for Nouns

Suffix No	Suffix	Explanation	Example
1	+lAr	Pulural	dede+ler
2	+Hm	1st person singular possessive	dede+m
3	+Hn	2nd person singular possessive	dede+n
4	+sH	3rdt person singular possessive	dede+si
5	+HmHz	1st person plural possessive	dede+miz
6	+HnHz	2nd person plural possessive	dede+niz
7	+lArH	3rd person plural possessive	dede+leri
8	+yH	-i case (accusative)	dede+yi
9	+yA	-e case (dative)	dede+ye
10	+Da	-de case (locative)	dede+de
11	+DAn	-den case (ablative)	dede+den
12	+nHn	-in case (complement)	dede+nin
13	+nH	-i case (after 3rd sg person suffix)	dede+si+ni
14	+nA	-e case -i case (after 3rd sg person suffix)	dede+si+ne
15	+nDA	-de case (after 3rd sg person suffix)	dede+si+nde
16	+nDAn	-den case (after 3rd sg person suffix)	dede+si+nden
17	+yIA	Instrumental/together	dede+yle
18	+ki	Interest	dede+m+de+ki
19	+ncA	Relative	dede+n+ce
20	+lH	lı suffix	dede+li
21	+sHz	sız suffix	dede+siz
22	+lHk	lık suffix	dede+lik
23	+cH	ci suffix	dede+ci
24	+cHk	Diminutive suffix	dede+lik

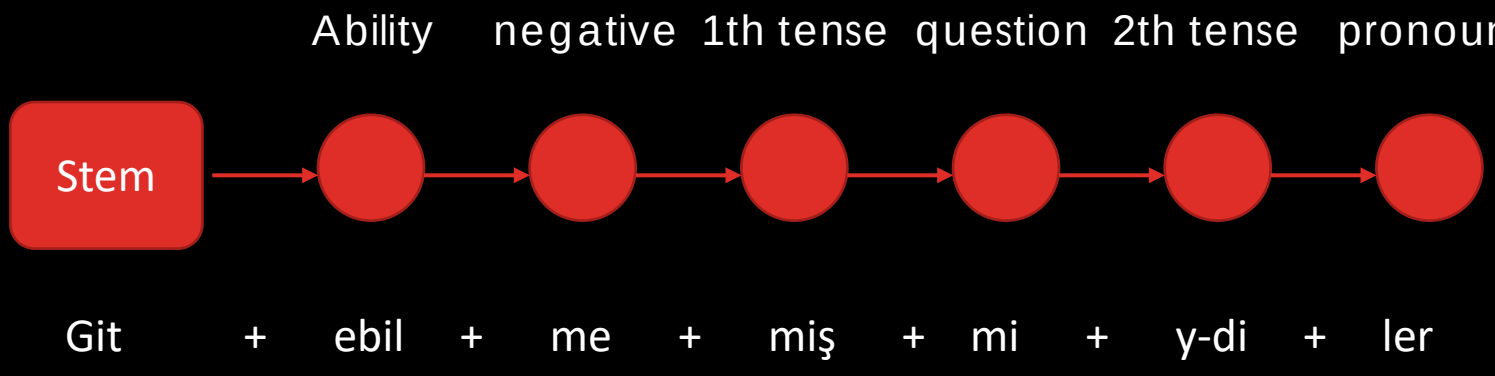
FSM for Morphological Parsing of Turkish nouns



FSM for Morphological Parsing of Turkish compound nouns



Parsing of Turkish Verbs



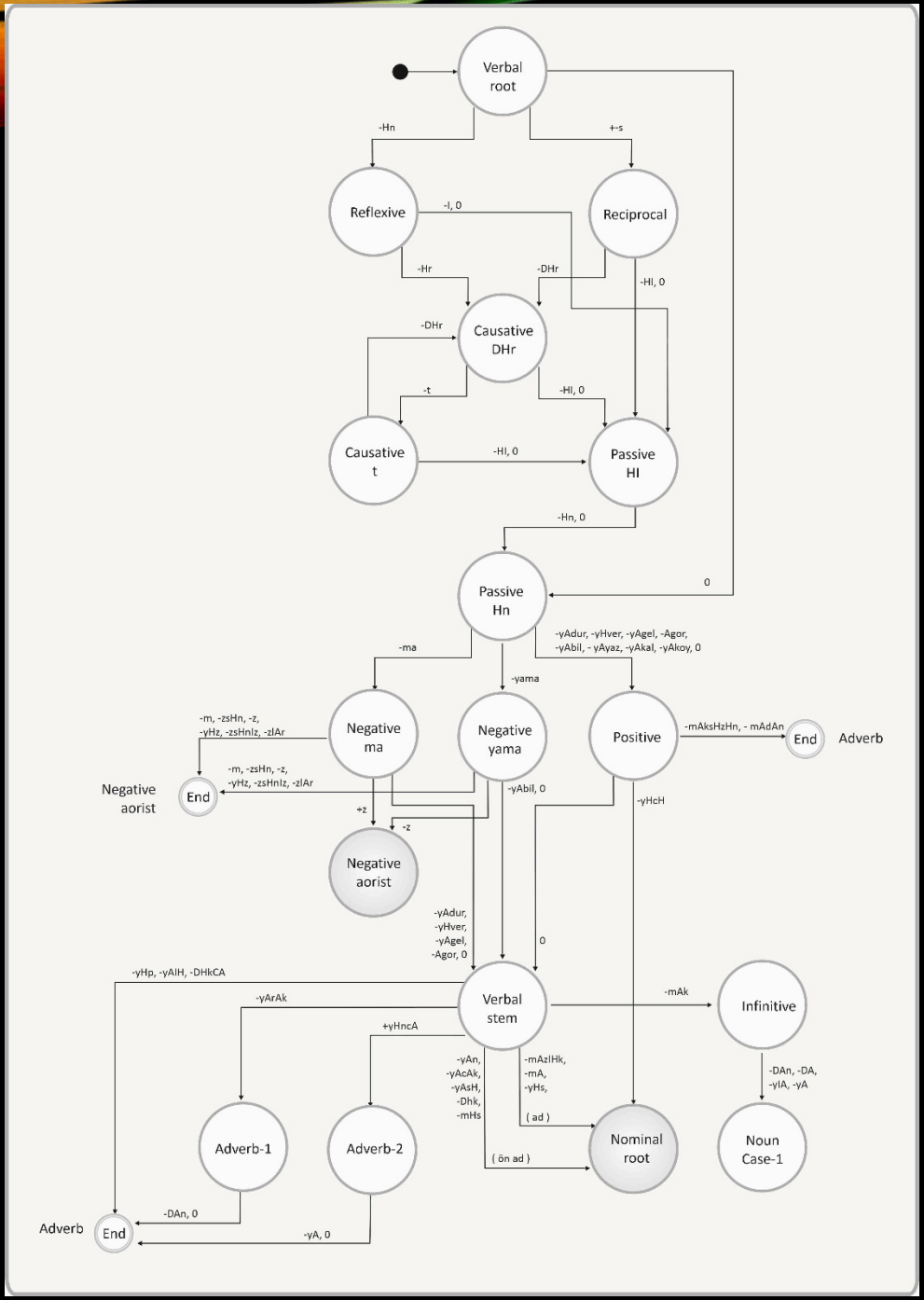
Inflectional and Derivational Suffixes for Verbs-I

Suffix No	Suffix	Explanation	Example
1	-Hn	reflexive	giy-in
2	-Hs	colloborative	öp-üş
3	-DHr	causative	giy-dir
4	-r	zausative	iç-ir
5	-t	causative	taşı-t
6	-Hr	ergative	piş-ir
7	-Hl	passive reflexive	öp-ül
8	-mAk	infinitive	öp-mek
9	-yAdur	sustained compaund verb	oku-yadur
10	-YAgel	sustained compaund verb	oku-yagel
11	-yAgör	sustained compaund verb	oku-yagör
12	-yAkal	sustained compaund verb	oku-yakal
13	yAkoy	sustained compaund verb	oku-yakoy
14	-yAyaz	approach compaund verb	oku-yayaz
15	-YAbil	competence compaund verb	oku-yabil
16	-yHver	swiftness compaund verb	oku-yuver
17	-ma	negative	oku-ma
18	-yama	negative	oku-yama
19	-m	1st singular person negative	okuma-m
20	-zsHn	2nd singular person negative	okuma-zsın
21	-z	3rd singular person negative	okuma-z
22	-yHz	1st plural person negative	okuma-yız
23	-zsHnHz	2nd plural person negative	okuma-zsınız

Inflectional and Derivational Suffixes for Verbs-II

Suffix No	Suffix	Explanation	Example
24	-zlAr	3rd plural person negative	okuma-zlar
25	-mAksHzHn	Adverb suffix	oku-maksızın
26	-mAdAn	Adverb suffix	oku-madan
27	-yHp	Adverb suffix	oku-yup
28	-yAlH	Adverb suffix	oku-yalı
29	-DHkCA	Adverb suffix	oku-dukça
30	-yArAk	Adverb suffix	oku-yarak
31	-yHncA	Adverb suffix	oku-yunca
32	-DAn	Adverb suffix (after-yArAk)	okuyarak-dan
33	-yA	Adverb suffix (after -yHncA)	okuyunca-ya
34	-yAn	suffix used for adjective from verb	oku-yan
45	-yAsH	suffix used for adjective from verb	oku-yası
36	-DHk	suffix used for adjective from verb	oku-duk
37	-mHs	suffix used for adjective from verb	oku-muş
38	-yAcAk	suffix used for adjective from verb	oku-yacak
39	-mAzlHk	suffix used for noun from verb	oku-mazlık
40	-mA	suffix used for noun from verb	oku-ma
41	-yHs	suffix used for noun from verb	oku-yuş
42	-yHcH	task suffix	oku-yucu
43	-yA	-e case (after infinitive)	okuma-ya
44	-DA	-de case (after infinitive)	okumak-da
45	-DAn	-den case (after infinitive)	okumak-dan
46	-ylA	togetherness (after infinitive)	okuma-yla

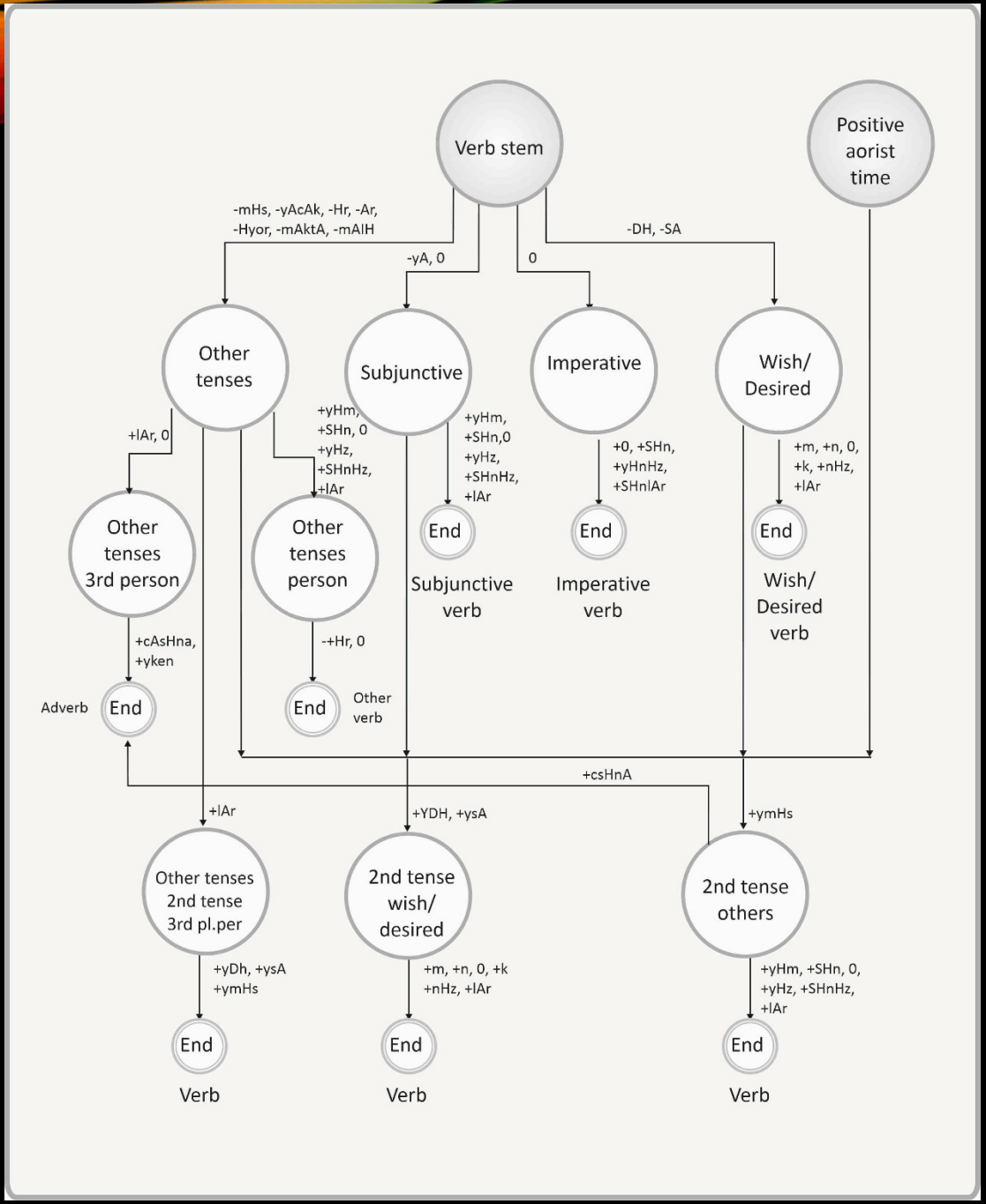
FSM for Morphological Parsing of Turkish verbs



Tense Suffixes of Turkish Verbs

Suffix No	Suffix	Explanation	Example
1	-mHs	reported past tense	gel-miş
2	-yAcAk	future tense	gel-ecek
3	-Hr	aorist Tense	gel-ir
4	-Ar	aorist Tense	gel-ir
5	-Hyor	present continuous tense	gel-iyor
6	-mAktA	continuum tense	gel-mekte
7	-mAlH	imperative tense	gel-meli
8	-yA	subjunctive tense	gel-e
9	-DH	past tense	gel-di
10	-SA	conditional tense	gel-se
11	+yHm	1st. singular person (conditional tense)	gel-ir-im
12	+SHn	2nd. singular person (conditional tense)	gel-ir-sin
13	+yHz	1st. plural person (conditional tense)	gel-ir-iz
14	+SHnHz	2nd. plural person (conditional tense)	gel-ir-siniz
15	+lAr	3rd. plural person (conditional tense)	gel-ir-ler
16	+DHr	notification suffix	gel-iyor-dur
17	+cAsHna	attitude adverb	gel-i-cesine
18	+yken	time adverd belirteci	gel-ir-ken
19	+yDH	story conjoined tense	gel-miş-ti
20	+ysA	narrated conjoined tense	gel-miş-se
21	+ymHş	narrated conjoined tense	gel-meli-ymiş
22	+m	1st. singular person	gel-di-m
23	+n	2nd. singular person	gel-di-n
24	+k	1st. plural person	gel-di-k
25	+nHz	2nd. plural person	gel-di-niz
26	+yHn	1st. singular person	gel-in
27	+lHm	1st. plural person	gel-e-lim
28	+yHnHz	2nd. plural person	gel-iniz
29	+sHnHAr	3rd. plural person	gel-sinler

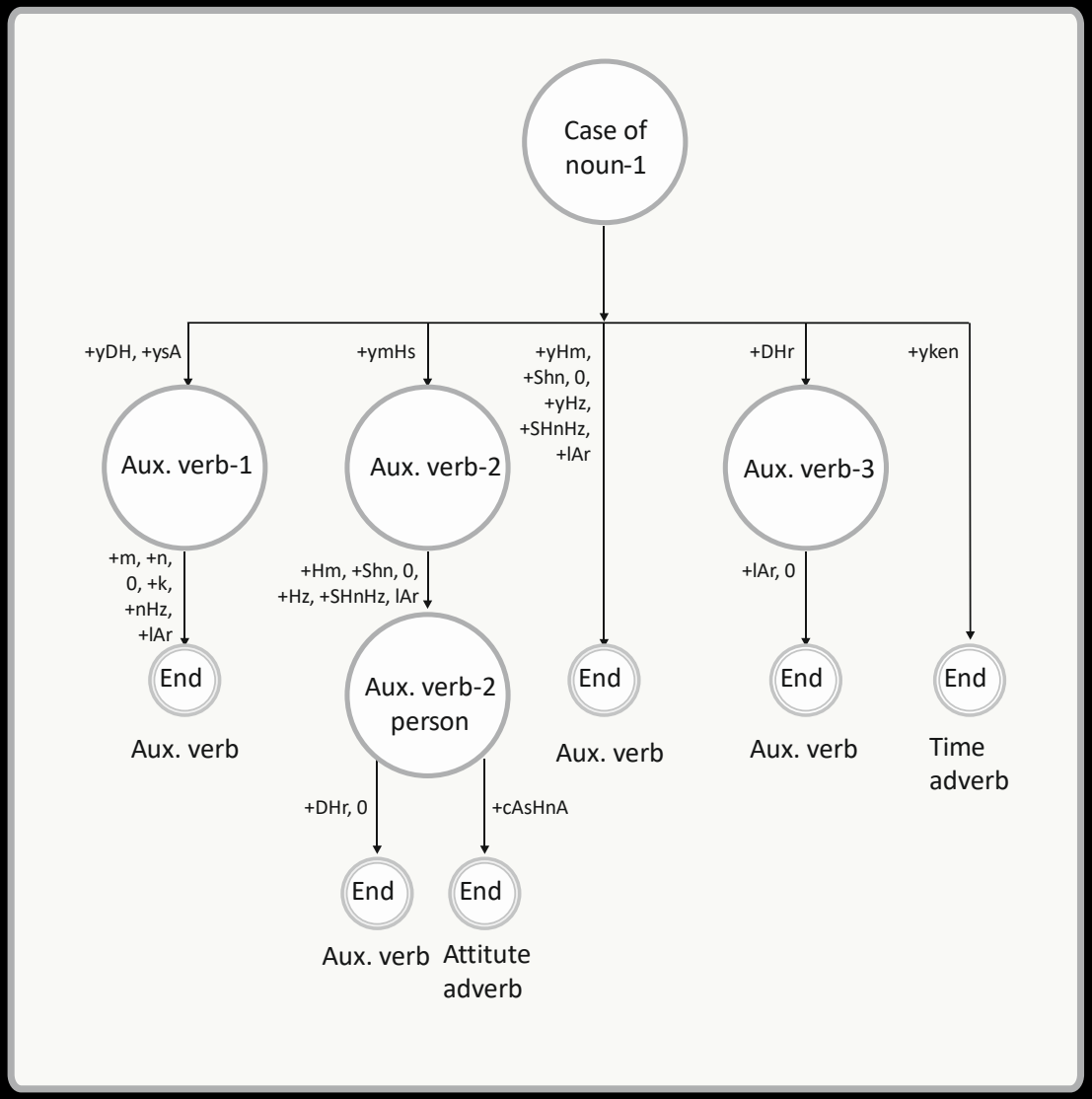
FSM for time suffixes of Turkish verbs



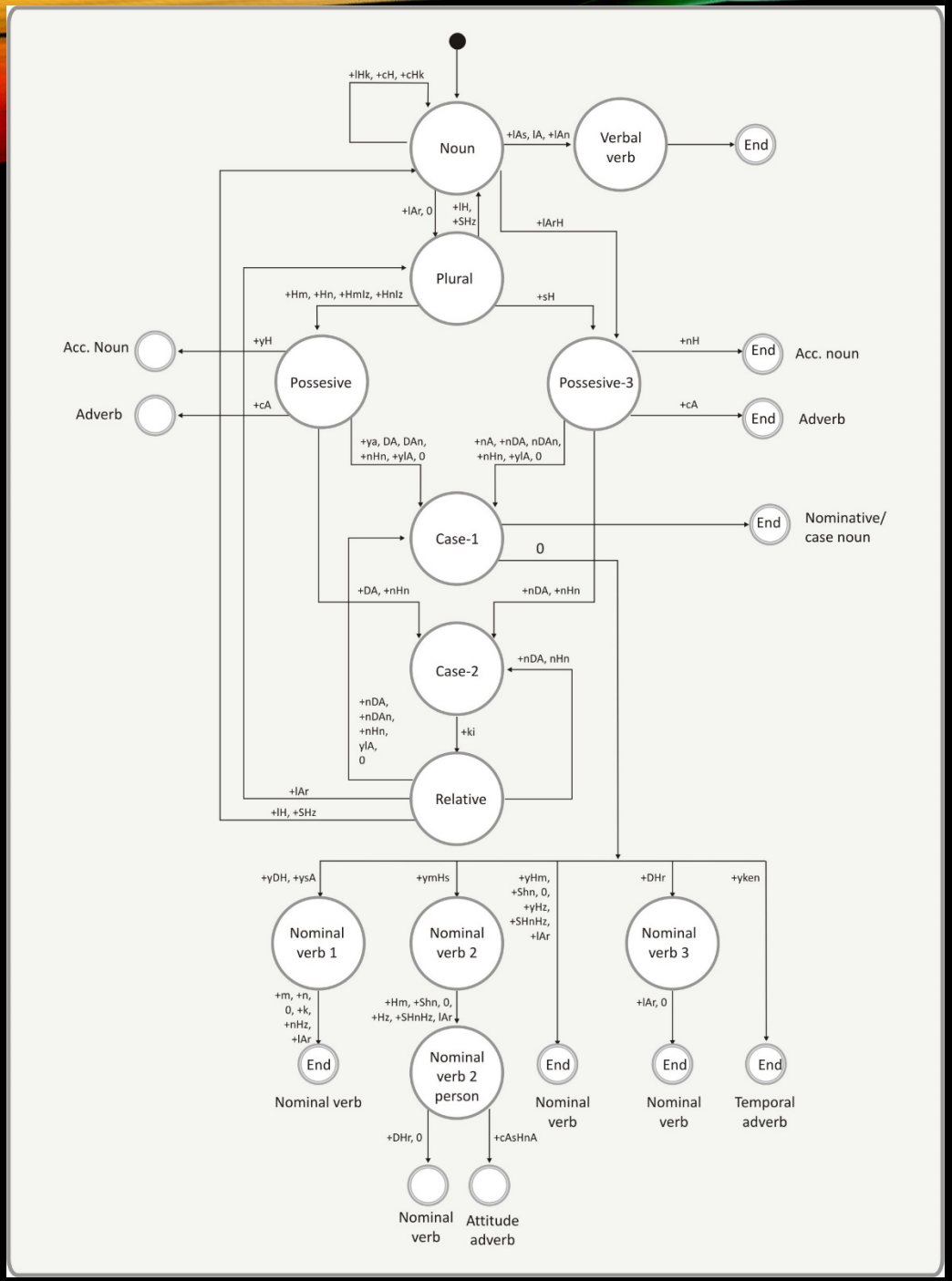
Tense Suffixes of Turkish Verbs

Suffix No	Suffix	Explanation	Example
1	+yHm	1st. singular person	sınıfta+yım
2	+SHn	2nd. singular person	sınıfta+sın
3	+DHr	3rd. singular person	sınıfta+dır
4	+yHz	1st. plural person	sınıfta+yız
5	+sHnHz	2nd. plural person	sınıfta+sınız
6	+lAr	3rd. plural person	sınıfta+lar
7	+m	1st. singular person, after +yDH and ysA	sınıfta+ydı+m
8	+n	2nd. singular person, after +yDH and ysA	sınıfta+ydı+n
9	+k	1st. plural person, after +yDH and ysA	sınıfta+ydı+k
10	+nHz	2nd. plural person, after +yDH and ysA	sınıfta+ydı+nız
11	+yDH	past tense	sınıfta+ydı
12	+ymHs	reported past tense	sınıfta+ymış
13	+ysA	conditional tense	sınıfta+ysa
14	+yken	time adverb	sınıfta+yken
15	+cAsHnA	attitute adverb	sınıfta+ymış+casına

FSM for Morphological Parsing of Turkish auxiliary verbs



FSM for Morphological Parsing of Turkish verbs and nouns



Morphological Disambiguation

One of the main issues of natural language processing studies is the labeling of words with their classes. Determining the nature of a word in a sentence is easy for someone who knows grammar, but it is not so easy to do by computer. This is because every language has more than one attribute of a word.

1- Koyun: noun (sheep which is animal)

2- Koyun : verb (imperative of the verb to put)

3- Koy+un: noun, pertaining to the bay (small recess in the sea)

1- Alın: noun (the part of the head above the eyes, under the hair)

2- Alın : action (imperative of the verb to take)

3- Al+ın: adjective, related to ‘al’ (red color)

References

[1] B. Atalay, Türk Dilinde Ekler ve Kökler Üzerine Bir Deneme, TDK Yayınları, İstanbul, 355-378, 1942.

[2] Cherokee language, http://cherokeeregistry.com/language.pdf

[3] TDE205U, Osmanlıca Türkçe Grameri-1 , adp.anadolu.edu.tr

[4] M. Kay, Nonconcatenative Finite-State Morphology, https://pdfs.semanticscholar.org /3234/467ab4ca6ebb46c28bd607e29eb728271b69.pdf

[5] Gencan, T. N., Dilbilgisi, Türk Dil Kurumu Yayınları, Ankara, 4.Baskı, 195-309, 1979.

[6] T. Banguoğlu, Türkçenin Grameri, Atatürk ve Türk Dil Kurumu, TDK Yayınları:528, 1986

[7] B. Brodda, K. Fred, An Experiment with Automatic Morphological Analysis of Finnish. Helsinki: University of Helsinki. 1980

[8] H. J. Kaalep, An Estonian Morphological Analyzer. Computers and Humanities , 31 (2), 115-133, 1997

[9] G. Cebiroğlu, E. Adalı, An Affix Stripping Morphological Analyzer For Turkish. International Conference on Artificial Intelligence and Applications, (s. 209)304). Innsbruck. 2004

[10] J. Hankamer, Finite State Morphology and Left to Right. Fifth West Coast Conference on Formal Linguistics, (s. 29)34). Stanford. 1986

[11] Kimmo Koskenniemi, Two-Level Morphology: A General Computational Model for Word-Form Recognition and Production, University of Helsinki, Dept. of General Linguistics, Finland, Publications No:11, 1983

[12] K. Koskenniemi,Two-Level Model for Morphological Analysis, Proceedings of the Eighth international joint conference on Artificial intelligence - Volume 2, Jan. 1983

[13] L. Karttunen, K. R. Beesley, A Short History of Two-Level Morphology, http://www.ling.helsinki.fi/~koskenni/esslli-2001-karttunen/[14.05.2017 12:54:40]

[14] Cherokee Language, https://en.wikipedia.org/wiki/Cherokee_language

[15] L. Karttunen, K. R. Beesley, Twenty-five years of finite-state morphology https://web.stanford.edu/group/cslipublications/cslipublications/koskenniemi-festschrift/8)karttunen-beesley.pdfL.

[16] A. Pirkola, Morphological Typology of Language for IR, Journal of Documentation 57 (3), 330-348

[17] G. Cebiroğlu, A. C. Tantuğ, E. Adalı, Synthetic Turkish Word Root Generation, Proceedings of the Turkish Artificial Intelligence and Neural Networks, TAINN 2003, Canakkale,Turkey

[18] G. Cebiroğlu, E. Adalı, Probabilistic Turkish Word Root Generation, Proceedings of the 3rd Asia Pacific International Symposium on Information Technology, Istanbul, Turkey, 2004

[19] Crystal,D. A Dictionary of Linguistics and Phonetics, Blackwell, 3rd Edition. (1991).

[20] Sinclair,J. Corpus Concordance, Collocation". OUP. (1991).

[21] Nadas, A. “Estimation of Probabilities in the Language Model of the IBM speech recognition system”. IEEE Transactions on Acoustics, Speech, and Signal Processing, 32:4, pp. 859-861, (1984).

[22] Kukich K. “Technique for Automatically Correcting Words in Text”. Periodical Issue Article of ACM Press, pp.377-439. (1992).

[23] Church, K. & Gale, W. “Probability Scoring for Spelling Correction”. Statistics and Computing, pp.93-103. (1991).

[24] Shannon, C.E. “Prediction and Entropy of Printed English”. The Bell System Technical Journal, 30:1, pp. 50-64. (1951).

[25] Jurafsky, D. & Martin, J.H. “Speech and Language Processing”, Prentice Hall, pp. 193-199. (2000).

[26] Church, K. & Mercer, R. Introduction to the Special Issue on Computational Linguistics Using Large Corpora. Computational Linguistics, 19:1, pp. 1-24. (1993).

[27] W. N. Francis, H. Kucera, Brown Corpus Manual, Brown Univ. 1964,1971, 1979

[28] ODTÜ Türkçe Derlem (OTD), Computational Linguistics, 19:1, pp. 1-24. (1993).

[29] Türkçe Ulusal Derlemi, http://www.tnc.org.tr

[30] Á. Corral, G. Boleda, R. Ferrer-i-Cancho, Zipf’s Law for Word Frequencies: Word, Forms versus Lemmas in Long Texts, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4497678/[24.12.2017 18:32:57]

[31] D. Yüret, Discovery of Linguistic Relations Using Lexical, Doktora tezi MIT, 1998

[32] S. Kazkılınç, Türkçe Metinlerin Etiketlenmesi, İTÜ Fen Bilimleri Ens. 2012

[33] K. Oflazer, Ö. Çetinoğlu, O. Bilgin, Building a Wordnet for Turkish, Sabancı Üniversitesi, İstanbul. 2004

[34] M. F. Amasyalı, Türkçe Wordnet’in Otomatik Oluşturulması, SIU, Kayseri, 2005.

[35] H. Yavuz, A. Balcı, Turkish Phonology and Morphology, Anadolu Üniv. Yayın No:2290

[36] N. Demir, E. Yılmaz, T.N. Gencan, Türkçenin Biçim Bilgisi, Anadolu Üniv. Yayın No:2370

[37] E. Adalı,Türkçenin Temelleri, İTÜ, İstanbul, 2010.

[38] Y. Göknel, Modern Türkçe Dilbilgisi, Hür Efe Matbaası, İzmir, 60-105, 1974.

[39] Z. Korkmaz, Türkiye Türkçesi Grameri (Şekil Bilgisi), TDK: 827, 2003