

# Biçimbirimsel Analiz

Doç.Dr.Banu Diri

## Biçimbirimsel Çözümleme (Morphological Analysis)

Biçimbirimsel (Morphological) Analiz, bir cümle içerisinde yer alan her kelimenin kök ve eklerine ayrıştırılması ve görevlerinin belirlenmesi ile ilgilidir. Kök ve eklere morfem ismini de verebiliriz.

Kelimenin en küçük anlamlı birimi **morpheme**

**fox (tilki)** → 1 morpheme

**cats (kediler)** → 2 morpheme (cat ve -s çoğul eki)

Örnek: çiçekler → çiçek (kök) + ler (çoğul eki) [çiçek ve ler birer morfemdir]

Morphem 'ler

- gövdeler (stem) (kelimenin ana morpheme dir ve kelimeye anlamını verir)
- ekler (affixes) (kelimenin anlamı şekil değiştirir)

### Affix'ler

- prefix (gövdenin önünde yer alır) (impossible, natamam)
- suffixes (gövdeyi takip eder) (eats, yemekler)
- infixes (gövdenin içerisine eklenir-Filipince-Tagalog)
- circumfixes (hem gövdenin önünde hem de arkasında yer alır)  
(almancada → sagen (verb- to say) gesagst (past participle- said))

Bir kelime bir veya birden fazla affix (ek) alabilir.

**rewrites** (prefix *-re*, stem *-write*, suffix *-s*)

**unbelievably** (stem *-believe*, affix *-un*, *able*, *ly*)

İngilizce 4 veya 5 ten fazla ek almaz. Türkçe gibi *agglutinative* –sondan ekli diller 10 dan fazla da ek alabilirler.

uygarlaştıramadıklarımızdanmışsınızcasına

uygar +laş +tır +ama +dık +lar +ımız +dan +mış +sınız +casına  
civilized +BEC +CAUS +NABL +PART +PL +P1PL +ABL +PAST +2PL +AsIf

“(behaving) as if you are among those whom we could not civilize”

The various pieces of this word (the **morphemes**) have these meanings:

+BEC	“become”
+CAUS	the causative verb marker (‘cause to X’)
+NAB	“not able”
+PART	past participle form
+P1PL	1st person pl possessive agreement
+2PL	2nd person pl
+ABL	ablative (from/among) case marker
+AsIf	derivationally forms an adverb from a finite verb

Morphemlerin birleştirilerek yeni bir kelimenin yaratılması için farklı yollar mevcuttur.

Konuşma ve dil işlemede önemlidir.

- ✓ Inflection (Çekim eki) : -s, -ed (ing), -lar, -ler (tr)
- ✓ Derivation (Yapım eki) : computerize (verb),  
computerization (isim)  
göz, gözlük
- ✓ Compounding (Birleştirme) : Birden fazla gövde halindeki kelimeyi birleştirir.  
dog + house → doghouse  
pazar + yer → pazaryer-i
- ✓ Cliticization : I have → I've  
an apple  
l'opera  
-de, -ki, -mi

### Inflectional Morphology

İngilizce çekim ekleri

*noun, verb, adjective* uygulanır.

Affix → plural, possessive (sahiplik) (books, boxes, children's)

	Regular nouns		Irregular nouns	
Singular	Cat	Thrush	Mouse	Ox
Plural	Cats	Thrushes	Mice	Oxen

Morphological class	Regularly Inflected Verbs			
stem	walk	merge	try	map
-s form	walks	merges	tries	maps
-ing participle	walking	merging	trying	mapping
past form or -ed participle	walked	merged	tried	mapped

Morphological class	Irregularly Inflected Verbs		
stem	eat	catch	
-s form	eats	catches	
-ing participle	eating	catching	
preterite (geçmiş z.)	ate	caught	
past participle	eaten	caught	

### Derivational Morphology

- ✓ İngilizcede kelime türetme oldukça karışık bir yapıya sahiptir.
- ✓ İngilizcede en çok kullanılan türetme şekli fiil ve sıfatlardan isimlerin türetilmesidir. Bu işleme **nominalization** denir.

Suffix	Base Verb/Adjective	Derived Noun
-ation	computerize (V)	computerization
-ee	appoint (V)	appointee
-er	kill (V)	killer
-ness	fuzzy (A)	fuzziness

Suffix	Base Verb/Noun	Derived Adjective
-al	computation (N)	computational
-able	embrace(V) <small>kucaklamak</small>	embraceable
-less	clue (N)	clueless

### Cliticization

- proclitics (vurgu bakımından kendisinden sonra gelen kelimeye bağlıdır.  
*an* apple)
- enclitics (hem kitap hem *de* kalem aldı)

Full Form	Clitic	Full Form	Clitic
am	'm	have	've
are	're	has	's
is	's	had	'd
will	'll	would	'd

➤ Biçimbirimsel analizde kelimelerin sadece kök ve eklerinin çözümlemesi yapılmaz. Aynı zamanda kelimelerin isim, fiil, sıfat, zarf, edat gibi tipleri de belirlenir.

➤ Biçimbirimsel Analiz'in çıktısı, Sözdizimsel (Syntax) Analizde kullanılır.

➤ Biçimbirimsel Analiz için teoride tüm kelimeler ve onların olası formları bir sözlükte tutularak gerektiğinde erişilebilir. Ancak, Türkçe ve Fince gibi sondan eklemeli dillerde bir kelimenin kökünden çok sayıda kelime türetildiğinden biçimbirimsel çözümlemenin yapılması şarttır.

Biçimbirimsel Analizde 3 temel elemana ihtiyaç vardır.

- Sözlük
- İmla Kuralları
- Biçimbirimsel Kurallar

❖ **Lexicon (Sözlük)** : Dildeki tüm gövde ve eklerin tutulduğu bir yapıdır. Ancak Türkçe gibi sondan eklemeli bir dilde tüm kelimelerin sözlükte yer alması imkansızdır. Bu yüzden kelimedeki değişimler ayırıcı işaretler ile sözlükte tutulmalıdır.

❖ **Morphotactics Rules (Biçimbirimsel Kurallar)** : Morphemlerin eklenme sırasının kuralları

❖ **Orthographic Rules (İmla Kuralları)** : Kelimeler içerisinde meydana gelen değişimleri tanımlayan kurallardır. Bazen morphemlerin arka arkaya bağlanma durumlarında morphemler değişime uğrayabilir.

- ünlü uyumu,
- ünlü düşmesi (oğul→oğlu),
- ünlü daralması (anla+yor→anlıyor),
- ünsüz sertleşmesi (meslek +daş→meslektaş),
- ünsüz yumuşaması (ağaç+a→ağaca),
- ünsüz türemesi (his+i→hissi)

gibi.

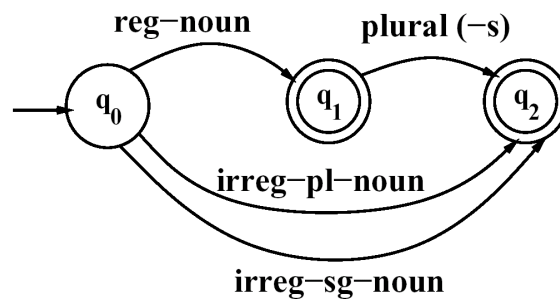
Örnek: İngilizcede bir kelimenin nasıl çoğul yazılacağını söyler.

(kelime sonundaki -y düşer, yerine -i gelir ve arkasına -es eklenir)

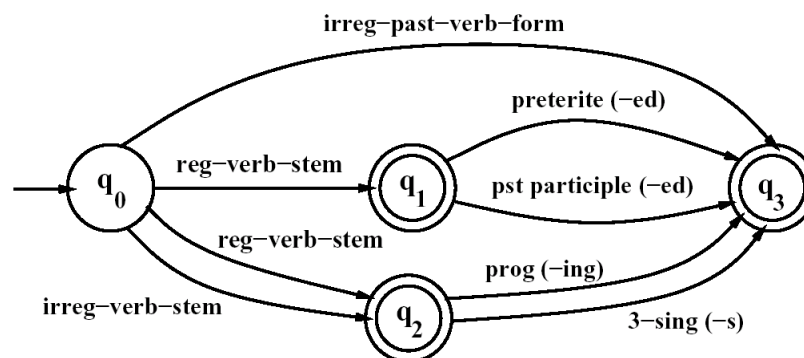
#### Parsing English morphology

Input	Morphological parsed output
cats	cat +N +PL
cat	cat +N +SG
cities	city +N +PL
geese	goose +N +PL
goose	(goose +N +SG) or (goose +V)
gooses	goose +V +3SG
merging	merge +V +PRES-PART
caught	(caught +V +PAST-PART) or (catch +V +PAST)

## Construction of a Finite State Lexicon



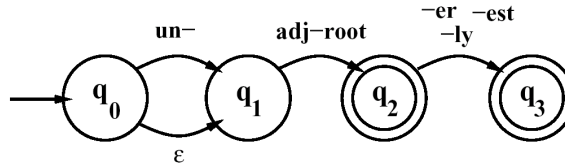
reg-noun	irreg-pl-noun	irreg-sg-noun	plural
fox	geese	goose	-s
cat	sheep	sheep	
aardvark	mice	mouse	



Reg-verb-stem	Irreg-verb-stem	Irreg-past-verb	past	Past-part	Pres-part	3s g
walk	cut	caught	-ed	-ed	-ing	-s
fry	speak	ate				
talk	sing	eaten				
impeach	sang					
	spoken					

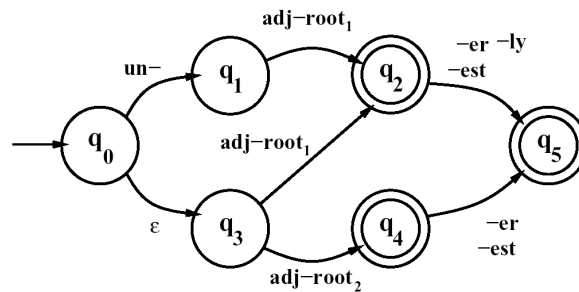
İngilizcede kelimeye yapılan türetme işleminin çekim işlemine göre kontrolü daha zordur.

İngilizcede sıfatlar için oluşturulmuş basit bir yapı.



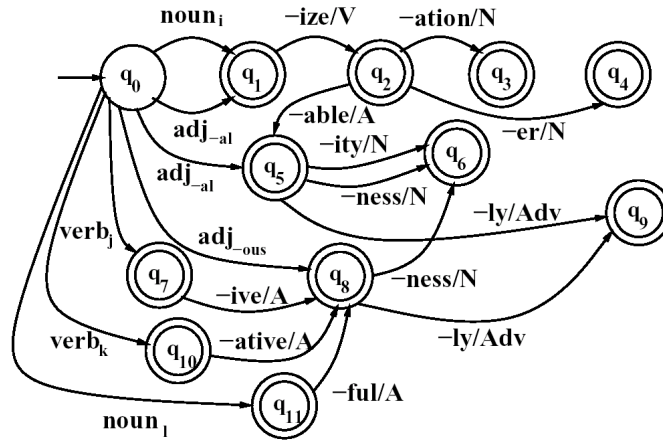
*big, bigger, biggest*  
*cool, cooler, coolest, coolly*  
*red, redder, reddest*  
*clear, clearer, clearest, clearly, unclear, unclearly*  
*happy, happier, happiest, happily*  
*unhappy, unhappier, unhappiest, unhappily*  
*real, unreal, really*

- FSA#1 listelenen tüm sıfatları tanır ancak hatalı yazılmış *unbig*, *redly*, ve *realest* kelimelerini de tanır.
- #1 üzerinde değişiklik yaparak #2 oluşturalım.
- root1 → *-un*, *-ly* ile kullanılan sıfatlardan
- root2 → diğerlerinden oluşan bir sözlük olsun.



*An FSA for a fragment of English adjective Morphology #2*





*An FSA for another fragment of English derivational morphology*



### **Biçimbirimsel Analizin Adımları**

Bir kelimenin Biçimbirimsel Analiz'inin yapılabilmesi için ilgili dile ait bir

- sözlüğe,
- eklerin listesine,
- eklerin bir biri ardı sıralanış kurallarına

ihtiyaç vardır.

**Kelimenin Kökünün Bulunması.** Kelimenin kökünü bulmak için bir sözlüğe ihtiyaç vardır. Bu sözlük, içerisinde ilgili dilde yer alan :

- tüm kök ve gövdeleri,
  - bunların tür bilgileri,
  - kelimedeki oluşabilecek ses düşmeleri
- gibi değişimler hakkında bilgiler de tutulmalıdır.

➤ Bazı çalışmalarda kelimenin sonundan başlanarak teker teker harf atarak kelimenin köküne ulaşılırken, bazı çalışmalarda ise kelimenin başından başlayarak ve teker teker harf ekleyerek kelimenin köküne ulaşılır.

➤ Tek harfe sahip tek bir kök vardır (o)

➤ Kelimenin kökünün türü isim, fiil, sıfat, zamir, zarf, edat, ünlem ve bağlaç olabilir. Eğer kökün türü edat, bağlaç ve ünlem dışında bir tür ise, kökten arta kalan eklerin çalışılan dil için geçerli olup olmadığının kontrol edilmesi gerekir.

➤ İsim soylu, fiil soylu kelimeler için Sonlu Durum Makineleri (SDM) tasarlanır. Bunların dışında sayılar (tarih, zaman, rakam, yazı ile gösterilen sayılar) içinde SDM'lerinin tasarlanması gerekir.

➤ Yapım eki almış kelimelerin, ek fiillerin, aykırı durumların ve sözlükte olmayan kelimeler için de çözümleme yapılması gerekmektedir.

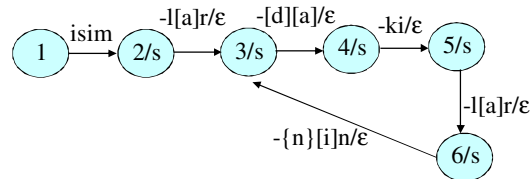
Vasif  
Nabiyey

Genel olarak Türkçe kelimeler şu şekilde biçimbirimlerine ayrılır

İsim: isim kökü + çoğul eki + iyelik eki + durum eki + kı° eki

Fiil: fiil kökü + çatı eki + olumsuzluk eki + bileşik fiil eki + ana zaman eki + soru eki + yardımcı zaman eki + kişi eki

Türkçe için kısıtlı sayıdaki isim biçimbirimlerini tanıyabilen bir yapı

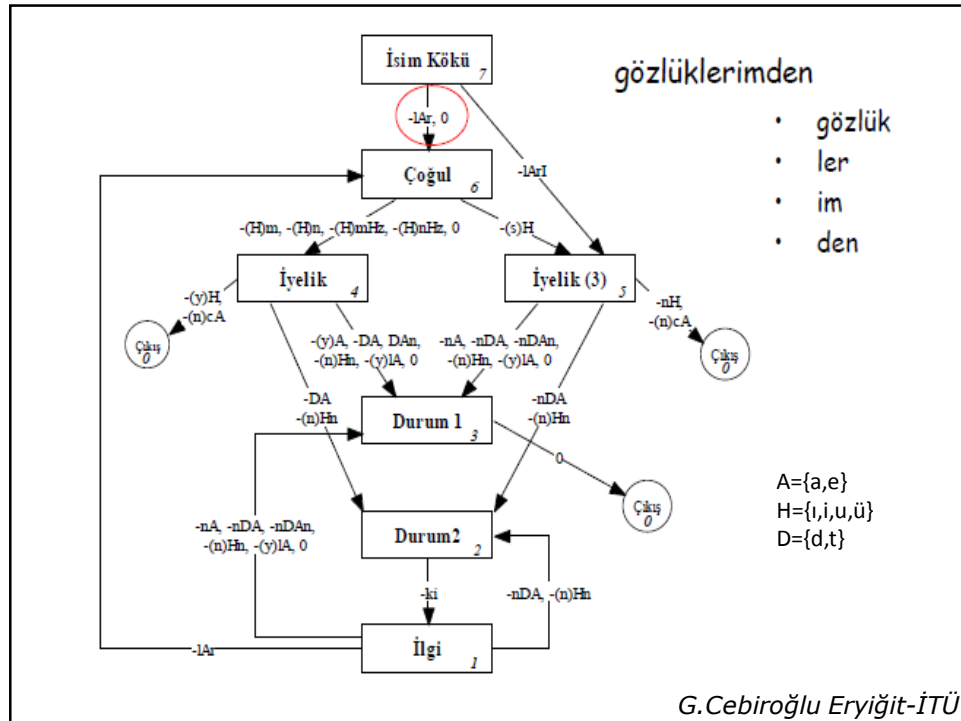


akıllardakilerin--> isim + l[a]r + [d][a] + ki + l[a]r + {n}[i]n

akıl + lar + da + ki + ler + in

evlerdeki lerde--> isim + l[a]r + [d][a] + ki + l[a]r + [d][a]

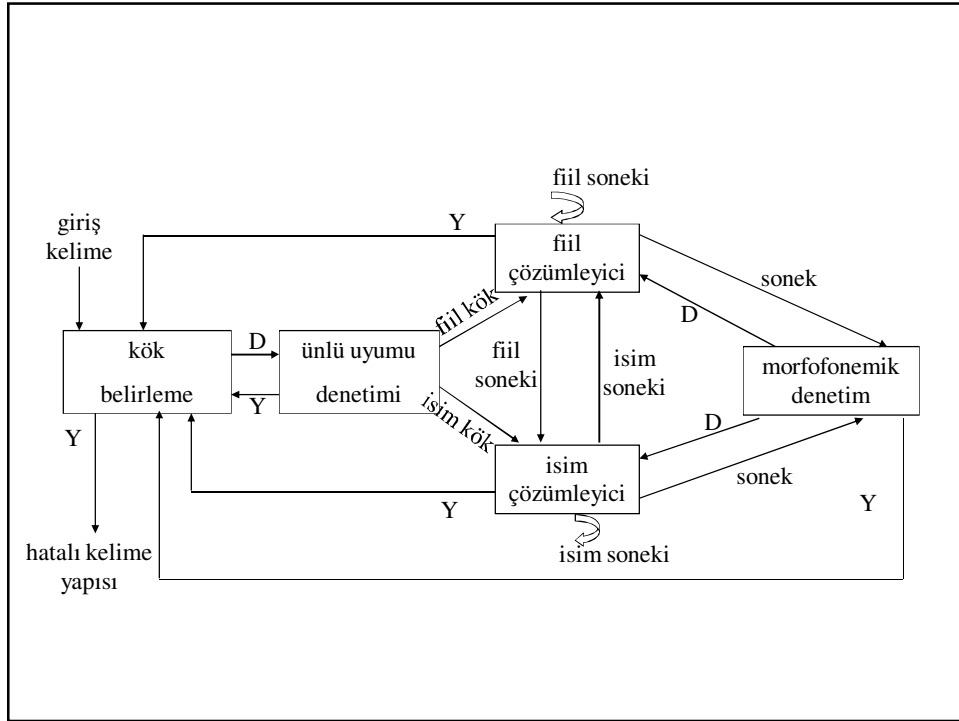
ev + ler + de + ki + ler + de



Türkçe’de morfolojik analiz 3 adımdan oluşur.

1. Kök belirleme
2. Morfolojik testler
3. Morfolojik ayrıştırma

Türkçe'nin morfolojik analizi için *K.Oflazer* ve *C.Bozşahin* tarafından önerilen bir yapı bu alandaki çalışmalarda kullanılmaktadır.



### Kök Belirleme

Türkçe bir kelimenin biçimbirimsel ayrıştırması yapılmadan önce kelimenin kökünün doğru olarak belirlenmesi gerekir. Öncelikle Türkçe'deki kelimeleri ve dildeki düzensiz yapıları içerisinde barındıran bir sözlük oluşturulur. Bu sözlükte, her kelimenin yapısal özelliklerini gösteren bayrak alanları bulunmalıdır.

1. Maksimum eşleme (maximal match) algoritması ile kelimenin kökü aranır.
2. Önce kelimenin tamamı sözlükte aranır. Kelime sözlükte bulunuyorsa ek almamış olarak belirlenir ve herhangi bir ayrıştırma yapılmaz.
3. Kelime bulunmaz ise her seferinde kelimenin sağından bir harf atılarak, kalan kısım sözlükte tekrar aranır.

4. Bu işlem herhangi bir kök bulununcaya kadar tekrar edilir. Kelime tek harf olarak kaldığında, kelimenin hatalı yazılmış olduğuna karar verilir.

5. Bu algoritma ile her zaman doğru kök bulunmayabilir. Kök bulunduktan sonra kelimenin geri kalan kısmı kurallara uygun bir şekilde ayrıştırılamıyorsa kelimenin yanlış bulunduğuna karar verilir ve yeni bir kök aranır.

Örnek:

*Yazıldın* kelimesine bakalım. Sağdan harf atarak gittiğimizde *yazı* kelimesi kök olarak bulunur. Fakat kelimenin geri kalanı doğru olarak ayrıştırılamaz. *Yazı* kökü geçersiz sayılır ve yeni kök aranır. Sonra *yaz* kelimesi kök olarak bulunur.

6. Türkçe’de bazen kelime kökleri aldıkları eklere göre değişebilir. Bu durum kelimenin kökünün bulunmasını zorlaştırır. Değişen bu köklerin bulunabilmesi için oluşturulmuş olan sözlükte ilgili kelimenin yanına bir bayrak yardımıyla işaret konulur.

Örnek: *oğlumuz* kelimenin kökü *oğul* dur.

Kök hece düşmesine uğramış *oğl* olmuştur.

***Oğl*** kelimesi sözlükte bulunamaz. Kullanılan algoritmada değişiklikler yapılarak hece düşmesine uğrayan kelimelere bir işaret konulur. Son iki harfi sessiz olan ve bir sonraki harfi sesli olan kelimelerde, son iki sessiz harfin arasına bir sonraki sesli harf yazılarak kelime sözlükte aranır. Böyle bir kelime sözlükte bulunur ve hece düşmesine uğrayacağı işaretlenmiş ise bu kelime kök olarak bulunur.

Oğulumuz yazılan kelime de oğul kök olarak bulunur ancak hece düşmesine uğramadığı için yanlış yazılmış kabul edilir.

7. Türkçe’de kökte oluşan bir başka değişimde hece yumuşamasıdır.

b,c,d, g/ğ → p, ç, t, k

**Örnek:** *bacağım* kelimesinin asıl kökü *bacak* ‘tır. *Bacağ* kelimesi sözlükte bulunamayacaktır. Hece yumuşaması kuralını uyguladığımızda *bacağ*, *bacak* olarak değişecek ve sözlükte hece yumuşaması bayrağı ile birlikte bulunacağından kök olarak kabul edilecektir.

8. Kelimenin kökünde oluşabilecek diğer değişiklikler de benzer çözümlerle algoritmaya eklenerek çözülebilir.

9. Kök belirleme işleminde bazen çok fazla işlem yapılabilir.

okula → okul

aldığımız → aldığımızı, aldığım, aldığım, aldığım, aldık, aldı, ald, alıd, alıt, alt, al

### Biçimbirim testler

Kelimenin köküne ulaştıktan sonra kelimenin geri kalanı ek dizisi olarak düşünülür, eklerin dizilişi ve ses uyumlarının Türkçe dil yapısına uygun olup olmadığı kontrol edilir.

1. İlk test olarak ünlü uyumu testi yapılır.

2. Kelimenin aldığı eklerde olabilecek değişikliklerin doğruluğu kontrol edilir. Bu testler kelime eklerine ayrıştırıldıktan sonra yapılır.

**Örnek:** *yapdıkça*, *yaptığça*, *yaptıkca*, *yaptığca* ve *yaptıkça* kelimelerinin hepsi *yap* köküne sahiptir.

Hepsi ünlü uyumundan geçer.

*yap+dık+ca* yapısına göre ayrıştırılır. Fakat sadece *yaptıkça* doğru yazılmıştır. *Yap* kökünden sonra gelecek *-dık* ekinin sert sessiz uyumundan dolayı *-tık* olarak değişmesi gerekmektedir.

Aynı sert sessiz uyumdan dolayı da *-ca*, *-ça* olarak değişir.

Kaynaştırma harfleri ve harf düşmeleri de bu kısımda incelenir

### Bıçimbirim ayrıştırma

Kelimelerin aldığı ekler belirlenir. Dizilişleri ve kelimelerin bu ekleri alıp alamayacağına bakılarak morfolojik ayrıştırma yapılır.

Ekler, m[a]l[ı], {y}[a]c[a]k, l[a]r gibi tanımlanırlar.

Bu bölümde kelimenin kökü bulunduktan sonra peşi sıra gelen eklerin hangi yapılarda oldukları incelenir.

Yapmелisin kelimesi **yap+m[a]l[ı]+s[ı]n** morfolojik yapısına uygun düşer ve ayrıştırılır ancak ünlü uyumuna uygun düşmediği için hatalı olarak işaretlenir.

Bazı kelimeler birden fazla morfolojik yapıya uygun düşer.

Hangi kelimenin hangi ekleri alabileceği sözlükte tanımlanan bayraklar ile ifade edilir.

Bir sonraki slide'da yer alan şekilde kullanılan sembollerin açıklaması şu şekilde yapılabilir:

- Daire içerisine alınan ifadeler, makinenin geçebileceği durumları
- Çift çerçeve içerisinde yer alan ifadeler, başlangıç durumunu
- Siyah çerçeve ile çizilmiş olan ifadeler de sonlanma durumunu
- Gri ile boyanmış durumlar ise, isim soylu bir köke gelen ek ile türün fiil olarak değiştiğini (veya tam tersi durum)

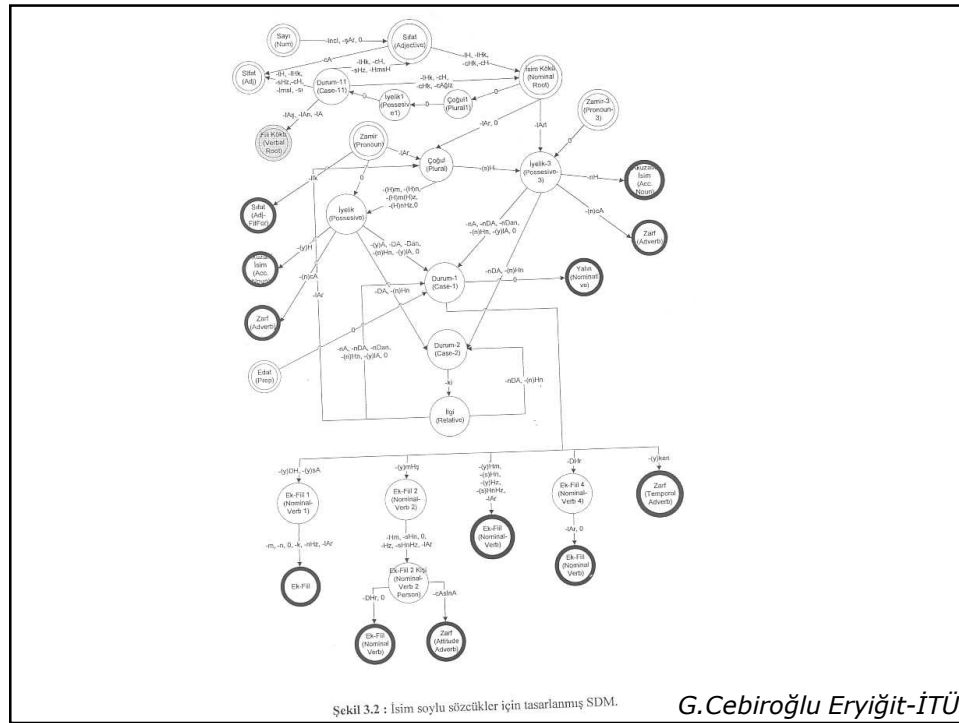
göstermektedir.

Farklı kaynaklarda kullanılan notasyonlar değişebilir. Ayrıca durumlar arasındaki oklar, köke veya gövdeye gelen ekin ne olduğunu gösterir.

Oklar üzerinde yer alan "0" ise, durumlar arası geçişte kelimenin hiçbir ek almadığını bildirir.

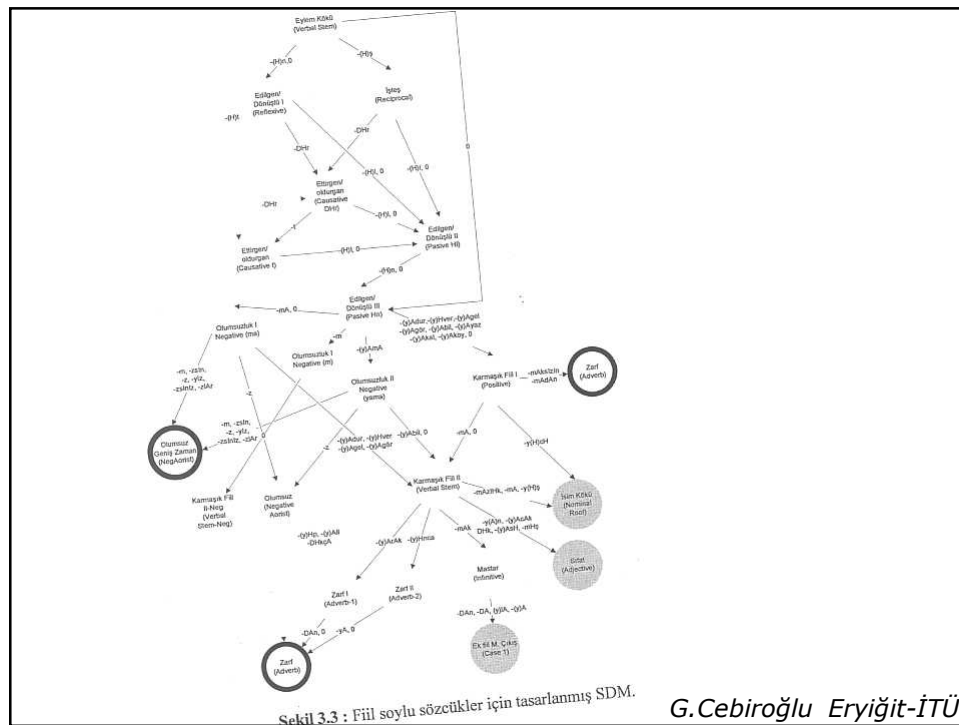
Makinelerde birden fazla başlangıç durumu kullanılmıştır. Sıfat ve zamir türündeki kelimelerde mantıksal sonuçların türetilmediğinin gözlenmesi üzerine birden fazla başlangıç durumu kullanılarak sorunun giderilmesine çalışılmıştır.

*G.Cebiroğlu Eryiğit-İTÜ*



Şekil 3.2 : İsim soylu sözcükler için tasarlanmış SDM.

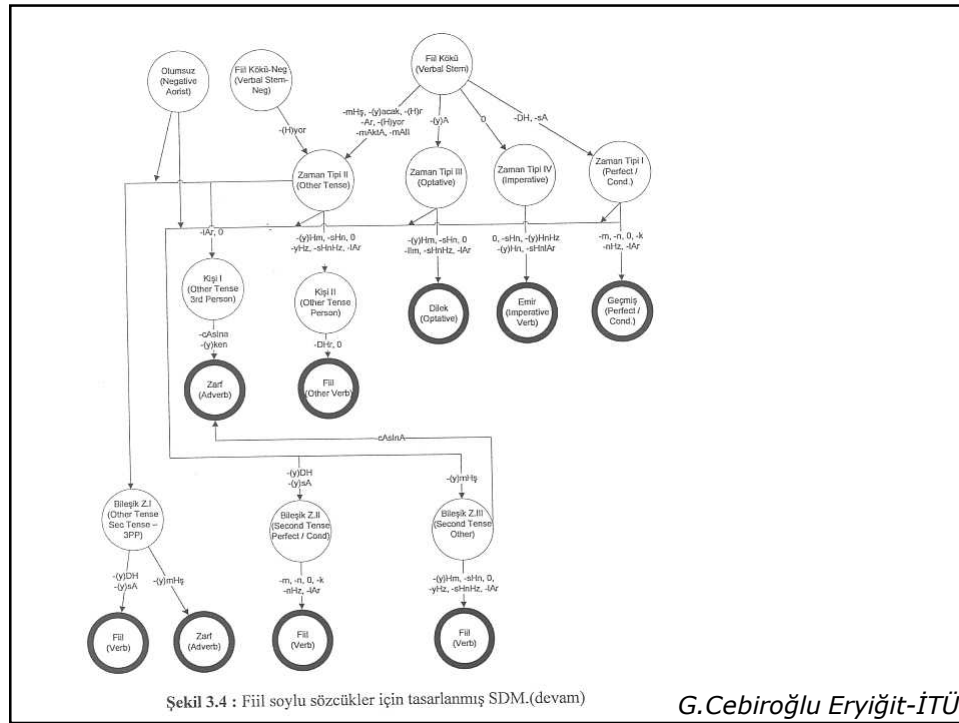
*G.Cebiroğlu Eryiğit-İTÜ*



Sekil 3.3 : Fiil soylu sözcükler için tasarlanmıř SDM.

*G.Cebiroğlu Eryiğit-İTÜ*





## Stem (Kök / Gövde) Tanımı

**Stem**, bir kelimenin **yüzeysel kökü** veya **gövdesidir**. Kelimeden ekler (çoğul, zaman, kişi vb.) kaldırılarak elde edilir, ancak her zaman anlamlı bir kelime olmak zorunda değildir.

Stemming işlemi, kelimenin yalnızca biçimini sadeleştirir, anlamını tam olarak korumaz.

### ✂ Örnek:

- *koşuyorum* → *koşu*
- *studies* → *studi*
- *happiness* → *happi*

“Stem” biçimbilimsel olarak kelimenin köküne en yakın yüzeysel biçimdir, fakat her zaman dilin sözlüğünde bulunmaz.

## Lemma (Sözlük Biçimi) Tanımı

**Lemma**, bir kelimenin **sözlükte yer alan temel biçimidir**. Lemmatization, dil bilgisi kuralları ve sözlük yardımıyla kelimeyi anlamlı en temel hâline indirger.

Bu biçim, genellikle fiiller için *master hâli* (–mek / –mak), isimler için ise *tekil yalın hâldir*.

### ✂ Örnek:

- *koşuyorum* → *koşmak*
- *studies* → *study*
- *went* → *go*
- *wolves* → *wolf*

“Lemma”, kelimenin dilbilgisel kurallar çerçevesinde yorumlanmış, anlamlı ve sözlükte tanımlı biçimdir.

## Kıyasla Özet

Özellik	Stem	Lemma
Amaç	Ekleri kaldırmak	Anlamlı kökü bulmak
Kullanılan bilgi	Yalnızca biçimsel kurallar	Dil bilgisi + sözlük
Sonuç anlamlı mı?	Her zaman değil	Evet
Türkçe örnek	<i>koşuyorum</i> → <i>koşu</i>	<i>koşuyorum</i> → <i>koşmak</i>
İngilizce örnek	<i>studies</i> → <i>studi</i>	<i>studies</i> → <i>study</i>

## Türkçe’de Stemming ve Lemmatization Örnekleri

Kelime	Stem (Yüzeysel Kök)	Lemma (Sözlük Biçimi)	Açıklama
koşuyorum	koşu	koşmak	“-yor” geniş zaman eki, fiil çekimi yapılmış.
gittim	git	gitmek	Fiil çekimli, lemma mastar hâli.
okuyucu	okuyu	okumak	“-cu” yapım eki, lemma fiil kökü.
yapmadım	yapma	yapmak	Olumsuzluk eki “-ma/-me”, lemma kökü olumlu hâlidir.
kitaplarımdan	kitaplar	kitap	Çoğul, iyelik ve hâl ekleri var.
evimizde	evimiz	ev	İyelik “-imiz” ve yer hâli “-de”.
oğlundan	oğlun	oğul	Ünlü düşmesi (oğul → oğlundan).
güzelliğinden	güzelliğ	güzel	Yapım eki “-lik” isim türetmiş; lemma sıfat hâli.
gidemeyecektim	gidemeyec	gitmek	Bileşik zaman, olumsuzluk, kişi ekleri var.
güzelleşti	güzelleş	güzel	Yapım eki “-leş” sıfattan fiil türetmiş.

- *Stemming* yalnızca biçimsel kesme işlemi yapar, anlamı korumaz.
- *Lemmatization* ise dil bilgisi ve sözlük tabanlı analiz yaparak kelimenin temel anlamını ortaya çıkarır.
- Türkçe sondan eklemeli olduğu için lemma, çoğunlukla **mastar** (–mek/–mak) ya da **temel isim** biçimindedir.