

Heceleme Yöntemiyle Kök Sözcük Üretme

İ. Büyükkuşcu
buyukkuscu@itu.edu.tr

E. Adalı
adalı@itu.edu.tr

İstanbul Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

Anahtar sözcükler: Doğal Dil İşleme, Sentetik Sözcük Kökü Üretimi, Heceleme Yöntemi

Özetçe

Kaynaklarda da değinildiği üzere Türkçe eklemeli bir dil olduğundan, Türkçe'nin doğuşunda kök sözcüklerin tek hece olduğu düşünülmektedir. Zamanla bu tek heceli köklerin yapım ekleriyle birleştirilerek yeni sözcüklerin elde edildiği ve sık kullanılan sözcüklerin, zaman içinde değişerek yeni sözcüklerin üretilmiş olduğu söylenebilir. Türkçede var olan kök kelimeler bir ya da iki heceden oluşmaktadır. Daha fazla heceden oluşanlar bu köklerin ek almış halleridir. Ayrıca diğer dillerden alınmış sözcüklerin tek veya çift hecelik kök yapısına uymayacağı görüşü dil bilimciler arasında kabul görmektedir. Bu düşünceden kalkarak, temelde heceleme yöntemi ve Türkçe'nin biçimbirimsel kuralları göz önüne alınarak bir ve iki heceden oluşan kök sözcük üretme yöntemi, bu makalede tanıtılmıştır.

Abstract

This paper presents a synthetic root synthesis method. Turkish language is agglutinative, that is, grammatical functions are indicated by adding various suffixes to root. There are two main suffix classes: in the first class suffixes change the meaning of the root and make it a new stem via deriving the stem. New meaning calls the root's meaning but stem points to totally new thing. In the other class suffixes adds some features to the root, The thing that is pointed doesn't changes but it can marked plural or some case ending nominative, genitive, dative, accusative, locative and ablative can be added. For example, "ev" means "home". "Ev-de"

means "in home". "de and da" suffixes always give the "at that location" meaning. Again verbs have the same property. They have some time suffixes, wish and order suffixes. There are nearly 200 suffixes in Turkish. And generally each suffix gives a meaning to the root.

Lots of Turkish linguist says that when Turkish born the root words has only one syllable. And new words are created by adding suffixes to them. When we are looking for root word we should look to the word which has one or two syllable.

In this paper different from the other synthetic word synthesis we used syllables method not letter method. We applied all Turkish grammar rules while generating root words. Result root words may used as alternative root words for totally new innovative things.

1. Giriş

Türkçe sondan eklemeli ve kurallı bir dildir. Kök sözcüklerin sonuna, bu köke farklı anlamlar katan yapım ekleri eklenerek bu sözcüğü çağrıştıran ama tümüyle farklı bir şeyi anlatan gövde sözcükler elde edilir. Türkçe'de isim soylu kök veya gövdenin sonuna tekillik, çoğulluk, bulunma, yönelme gibi hal ekleri, iyelik ekleri, eşitlik ekleri eklenebilir. Eylem soylu kök ve gövdenin sonuna zaman ekleri, kip ekleri (-ayım, -eyim, -se, -sa, -mek, -mak), şahıs ekleri, ilgi ekleri eklenir. Çekim eki denilen bu ekler kökün ya da gövdenin anlamını değiştirmez, sadece ona işlerlik katar. Görüldüğü üzere Türkçe

sözcükler’de eklerle genişleyen bir yapı görünümündedir.

Besim Atalay gibi pek çok dilbilimci Türkçe’nin bir hecelik köklerden doğduğunu düşünmektedir [2]. Zamanla bu kök sözcüklere ekler getirilerek yeni sözcükler türetilmiş bunların yerleşmesi ve kökenlerinin unutulmasıyla sanki bu sözcükler kendi başlarına kök sözcükmiş gibi varsayılmaya başlanmıştır.

Bu makalede bu görüşten yola çıkmıştır. Tüm Türkçe dilbilgisi kurallarını heceler arsında var olan kuralları da içerecek şekilde uygulayarak, bir ve iki hecelik kök Türkçe sözcüklerin üretilmesi için geliştirdiğimiz yöntem tanıtılacaktır.

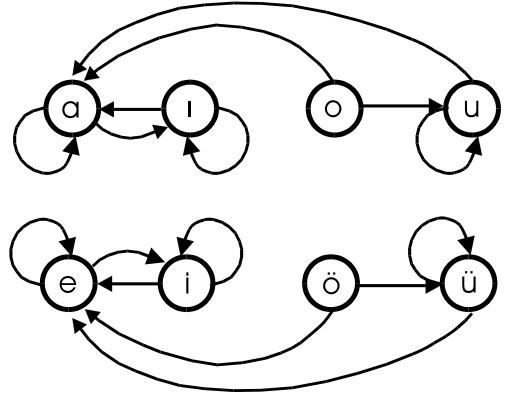
G. Eryiğit ve arkadaşlarının makalesinde tanıtılan yöntem, bir sözcüğü oluşturan harflerin Türkçe’nin dilbilgisi kurallarına uygun olarak dizilmesi olarak açıklanabilir [2]. Bu makalede tanıtılan yöntem ise hece temelli olup heceleme yönteminin sağladığı diğer kurallarla birlikte tüm dilbilgisi kurallarına da dikkat etmektedir. Bu makalede sunulan heceleme yöntemi ile G. Eryiğit ve arkadaşlarının Sentetik Türkçe Sözcük Kökleri Üretimi başlıklı makalesinde kullanılan yöntem daha ayrıntılı biçimde karşılaştırılmıştır.

2. Kullanılan Türkçe Dilbilgisi Kuralları

Kök sözcükler üretilirken aşağıdaki kurallara dikkate alınmıştır [3,4,5]:

- 3- S harfin sesli, Z sessiz olduğunu göstermek üzere, Türkçe’de S, SZ, SZZ, ZS, ZSZ, ZSZZ biçiminde hece türleri vardır.
- 4- Üç sessiz sözcük başında ve sonunda yana gelmez.
- 5- İki sesli yanyana gelmez.
- 6- Türkçe sözcükler büyük ünlü uyumuna uyar.

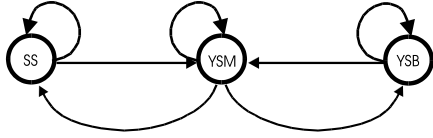
- 7- Türkçe sözcükler küçük ünlü uyumuna uyar. Sesli harflerin diziliş biçimi Şekil-1’de gösterilmiştir.



Şekil-1: Türkçe’de sesli harflerin diziliş kuralı

- 8- “j” Türkçe bir harf değildir. Türkçe kökenli sözcüklerde bulunmaz; o yüzden sessizler içine alınmamıştır.
- 9- “ğ” ile başlayan sözcük yoktur.
- 10- Sert sessizlerin benzeşmesi hece içlerinde dikkate alınmıştır.
- 11- Sözcüklerin “b,c,g,d” ile bitmemesi dikkate alınmıştır.
- 12- Türkçe’de sessizlerin sözcük içinde dizilişleri de kurallara bağlanmıştır. Türkçe’de sessiz uyumu sözcüklerde yan yana gelen sessizlerin ton bakımından birbirine uygun olmasına dayanır. Bu uyum kuralına göre :
 - Yumuşak sessizlerden sonra yumuşak sessiz veya yumuşak karşılığı bulunmayan bir sert sessiz gelebilir.
 - Yumuşak sessizlerden sonra tüm sessizler gelebilir.
 - Yumuşak karşılıkları bulunan sert sessizlerden sonra sert sessizlerden biri gelebilir.

Türkçe’de sessiz uyum kuralı Şekil-2’de gösterilmiştir.



Şekil-2: Sessiz uyum kuralı

Türkçe sözcüklerde, hece birleştirme kuralları da vardır. Bu kurallar da çalışmamızda göz önüne alınmıştır. Böylece geçerliliği daha yüksek sonuçlar elde edilmiştir.

- 12- Eğer sessiz ile biten bir heceye sessiz ile başlayan bir ek geliyorsa o zaman ikinci hece de de “ğ” yoktur
- 13- Aynı hecede sonda iki sessiz harf varsa ve sondan ikinci sessiz harf “l” harfi ise sonuncu harf “ç,k,p,t” harflerinden biridir.
- 14- Aynı hecede sonda iki sessiz harf varsa ve sondan ikinci sessiz harf “r” harfi ise sonuncu harf “ç,k,p,s,t” harflerinden biridir.
- 15- Aynı hecede sonda iki sessiz harf varsa ve sondan ikinci sessiz harf “s” harfi ise sonuncu harf “t” harfidir.
- 16- Aynı hecede sonda iki sessiz harf varsa ve sondan ikinci sessiz harf “ş” harfi ise sonuncu harf “t” harfidir.

3- Yapay Sözcük Üretme Yöntemleri ile Karşılaştırma

Sözcük üretme sırasında, sadece harflerin dizilişi ile ilgilenildiğinde, bazı durumların değerlendirilmediği görülmektedir. G. Cebiroğlu ve arkadaşlarının çalışmasında, yedi harfe kadar olan sözcükler üretilmiştir. Doğal olarak bu sözcükler, bir iki ya da üç heceden oluşmaktadır. Bu çalışmada gözlenen eksiklikler iki başlık altında toplanabilir:

- 1- Sözcük kapsamı
- 2- Sözcük doğruluğu

3.1 Sözcük Kapsamı

Harflerin diziliş kuralları göz önüne alınarak üretilen sözcüklerin bazılarında, Türkçe’nin hece yapısı kuralı bozulmaktadır. Ayrıca, yedi harften oluşan sözcükler üç heceli olabilmektedir.

Bizim çalışmamızda hece yöntemini temel alındığından, sözcükler üretilirken harf sayısı değil hece sayısı temel alınmıştır. Üretilen sözcük kümelerinde, her hece türünün tüm olası durumları bulunmaktadır. Tablo-1’de hece yöntemi ile harf diziliş yöntemleri karşılaştırmalı biçimde gösterilmiştir.

Tablo-1: Hece ve harf diziliş yöntemlerinin karşılaştırılması

1.hece	2.hece	3.hece	4.hece	1.	2.	3.	4.	Harf Sayısı	Hece Sayısı
S				o				1	1
SZ				ok				2	1
SZZ				alt				3	1
S	-ZS			a	-c1			3	2
SZ	-ZS			as	-k1			4	2
S	-ZSZ			a	na			4	2
SZ	-ZSZ			aş	sız			5	2
S	-ZS	-ZS		a	-h	-c1		5	3
S	-ZSZZ			a	ğa			5	2
SZ	-ZS	-ZS		as	-k1	-c1		6	3
SZ	-ZSZZ			is	ten			6	2
S	-ZS	-ZSZ		ü	-tü	sü		6	3
S	-ZSZ	-ZS		ü	zü	-lü		6	3
SZ	-ZS	-ZSZ		as	-k1	sız		7	3
SZ	-ZSZ	-ZS		an	la	-h		7	3
S	-ZS	-ZSZZ		a	-na	yu		7	3
S	-ZS	-ZS	-ZS	a	-ra	-ba	s1	7	4
S	-ZSZ	-ZSZ		a	yar	sız		7	3

Tablo-1’den şu sonuçlar çıkarılmaktadır:

Bir ve iki heceli sözcükler en çok altı harften oluşabilir. Beş harften oluşan üç heceli sözcükler de vardır. Bizim çalışmamız en çok iki heceli sözcükleri kapsadığından bu durumlar dışlanmıştır. Bu çalışma kapsamında olan yapılar, Tabloda taralı olarak gösterilmiştir.

Bu çalışmada kullandığımız yöntem aynen kullanılarak üç ve dört heceli olası tüm sözcükler elde edilebilir. Ancak biz kök sözcüklerin bir ya da iki heceli olduğu varsayımından yola çıktığımız için buna gerek duymadık.

Tablo-2’de, üretmiş olduğumuz bir ve iki heceli sözcüklerin yapısı ve sayısı gösterilmiştir.

Tablo-2: Bir ve iki heceli sözcüklerin sayıları

Hece	Hecelere Ayrılmış	Hece Sayısı	Olası Sayı
S	S	1	8
SZ	SZ	1	128
SZZ	SZZ	1	112
ZS	ZS	1	152
ZSZ	ZSZ	1	2312
ZSZZ	ZSZZ	1	2128
SZS	SZS	1	342
SZSZ	S-ZSZ	2	5202
SZSZZ	S-ZSZZ	2	4788
SZZS	SZ-ZS	2	4914
SZSZS	SZ-ZSZ	2	74574
SZSZSZ	SZ-ZSZZ	2	68796
ZSZS	ZS-ZS	2	6498
ZSZSZ	ZS-ZSZ	2	98838
ZSZSZZ	ZS-ZSZZ	2	90972
ZSZZS	ZSZ-ZS	2	88794
ZSZSZS	ZSZ-ZSZ	2	1347534
ZSZSZSZ	ZSZ-ZSZZ	2	1243116

Bir ve iki heceli sözcük toplamı 3.039.208

G. Eryiğit ve arkadaşlarının makalesinde tanıtılan yöntemde bir ve iki heceli sözcükler, Türkçe’de var olan olası tüm sözcükleri kapsamakta ama üretilen üç ve dört heceli sözcükler olası üç heceli tüm sözcükleri kapsamamaktadır. Örneğin yedi harften oluşan sözcükler üretildiğinde, sadece S-ZS-ZS-ZS kalıbındaki dört heceliler dahil edilmiştir. ZS-ZS-ZS-ZS, ZSZ-ZS-ZS-ZS veya S-ZS-ZSZ-ZSZ kalıbındaki dört heceli sözcükler dahil edilmemiştir.

Ayrıca G. Eryiğit ve arkadaşlarının Sentetik Türkçe Sözcük Kökleri Üretimi makalesinde [1] en çok yedi harfli 11.837.848 tane sözcük elde edilmiştir. Bunların 3.039.208 tanesi, kurallara uygun bir ve iki heceli sözcüklerdir. Geri kalanlar üç ve dört heceli sözcüklerdir. Bizim çalışmamızda elde ettiğimiz sözcük sayısı yaklaşık, tüm olası sözcüklerin ancak 1/4ünü oluşturmaktadır.

3-2 Sözcük Doğruluğu

Çalışmamızda, G. Eryiğit ve arkadaşlarının makalesindeki yöntemden farklı olarak Bölüm-

2’de tanıtılmış olan sadece heceleme yapısına özgü 12-17 kuralları da kullanılmıştır. Bu kuralların uygulanmış olması elde edilen sözcük doğruluğunu artırmıştır. Örneğin *Kar-ğa* örneğinde görüldüğü gibi sessiz ile biten bir hecenin sonuna ğ ile başlayan hece gelmez (Kural 13) *SZZ-ZS*, *ZSZZ-ZS*, *SZZ-ZSZ*, *ZSZZ-ZSZ* sözcüklerinde yanyana olan Z’lerle ilgili kurallar uygulanmıştır. (Kural 14-15-16-17)

4. Türk Dil Kurumu Sözcükleriyle Karşılaştırma

Türk Dil Kurumunun sözlüğünde bulunan sözcükler, geliştirdiğimiz bir program kullanılarak hecelerine ayrılmıştır. Sonuç Tablo-3’te gösterilmiştir.

Tablo-3: TDK sözlüğündeki sözcük sayısı ile bu çalışmada üretilen sözcük sayısı

HECE SAYISI	TDK SÖZCÜK SAYISI	PROGRAM SÖZCÜK SAYISI
1 heceli	3370	5182
2 heceli	19239	3034026

Tablo-3’ten görüldüğü gibi, bir ve iki heceli sözcükler, aslında 1,2,3,4,5,6 ve 7 harfli sözcükleri kapsamaktadır.

Türkçe dilbilgisi kurallarına uygun kullanılmayan pek çok tek ve iki heceli kök sözcük vardır ve bunlar yeni sözcükler üretmek için kullanılabilir.

5. Sonuçlar ve Gelecek Çalışmalar

Makalemizde tanıtılan çalışma sonucunda dil bilimciler için olası Türkçe sözcük kökleri üreten, doğruluğu yüksek bir yazılım geliştirilmiştir. Bu şekilde yabancı sözcüklerin dilimize yerleşmeden karşılıklarının bulunması daha kolay olacaktır. Dilbilimciler tarafından bir sözcük kökleri sözlüğünün oluşturulması, bu konuda ileride yapılacak çalışmalar için çok yararlı olacaktır.

Elde edilmiş olan kök sözcükler kullanılarak bu sözcüklerin sonuna düzenli bir şekilde yapım ve çekim ekleri getirilerek yine yapay bir şekilde Türkçe’deki tüm sözcükler (kök ve gövde) üretilebilir. Uyguladığımız yöntemde anlamsız sözcük sayısı olabildiğince aza indirgenmiştir

Kaynaklar

- [1] Cebirođlu, G. Tantuđ, A.C. Adalı, E. Erenler, "*Y.Sentetik Trke Szck Kkenleri retimi*", International XII. Turkish Symposium on Artificial Intelligence and Neural Networks TAINN 2003
- [2] Atalay B. "*Trk Dilinde Ekler ve Kkler zerine Bir Deneme*", Trk Dil Kurumu Yayınları , 1942
- [3] Hengirmen, M., "*Trke Dilbilgisi*", Engin Yayıncılık, Ankara, 2002.
- [4] Banguođlu, T., "*Trkenin Grameri*", Trk Dil Kurumu Yayınları : 528, Ankara, 2000.
- [5] Cebirođlu, G. Adalı E., "*Szleksz Kke Ulađma Yntemi*", TBD 19. Biliřim Kurultayı, İstanbul, 2002.