

İSİMLERDE ÇEKİM EKLERİNİN OLUŞTURDUĞU SES OLAYLARININ VISUAL BASIC ORTAMINDA İNCELENMESİ

Zeynep ALTAN

İstanbul Üniversitesi, Mühendislik Fakültesi
Bilgisayar Mühendisliği Bölümü Öğretim Üyesi
zaltan@istanbul.edu.tr

Kevser AYDIN

İstanbul Üniversitesi, Mühendislik Fakültesi
Bilgisayar Mühendisliği Bölümü
akevs@yahoo.com

Anahtar Kelimeler: dilbilim, bilgisayar bilimi, biçimbilim, doğal dil işleme, yapay zeka

Özet: Bu çalışmada, Türkçe isimler Visual Basic programlama dili kullanılarak anlambilimlerine ayrıştırılmakta, aynı zamanda da istenilen ekler yaratılabilmektedir. Ayrıştırma işlemi veritabanında bulunan dillerdeki nitelik taşımayan sözlükbirimler kontrol edilerek kelime biçimbirimleri şeklinde belirlenmektedir.

1. Giriş:

Öğrenme sistemleri, varlık bilgisi, bilgi kuramı, robotik, görsel sistem, sinir ağları, arama ve planlama uygulamaları için yapay zekadan yararlanılırken; doğal dil işleme de önemli bir uygulama alanı oluşturmaktadır. Doğal dil işleme temel olarak konuşma ve yazma dilini anlama ile ilgilidir. Girilen verilerin gösteriminin ne anlama geldiğini anlamaya çalışır. Doğal dil işleme sistemleri hem dilbilgisi teorilerini araştırmak hem de doğal dil arayüzleri üretmek için geliştirilmiştir. 1950 ve 1960'lı yıllarda yapay zekanın küçük bir alt alanı olarak görülürken bu konu, araştırmacıların ve uygulamaların elde ettiği başarılar sonucunda bilgisayar bilimlerinin temel bir disiplin alanı haline gelmiştir. 20. yüzyılın bellek de en önemli sinema olaylarından biri olan Stanley Kubric'in "2001: A Space Odyssey" isimli filminin kahramanı HAL 9000 bilgisayarı, yapay bir "agent" olarak, insanlarla etkileşimini dil aracılığı ile sağlıyordu; bu sistemde insanların konuşmalarını anlamak için "doğal dil işleme" ve "konuşma tanıma"; insanlarla haberleşme için ise "doğal dil üretimi" ve "konuşma sentezi" kavramlarından yararlanılmıştı. Ayrıca "bilgi üzerinde düzeltme yapma" ve "bilgi çıkarımı" ve "sonuç çıkarma" da HAL'ın becerileri arasında idi HAL'ın bu becerileri doğal dil işleme ile ilgili olarak dili anlama bağlamında gösterilen ilerlemeleri özetlemektedir.

Chomsky 1995 de yapısal betimlemeleri hesaplamak için sonsuz sayıda yapısal betimlemeyi sonlu bir gösterim ve işleme ile ifade eden bir sistem geliştirmiştir. Biçimbilim, sözdizimi ve anlambilim farklı dilbilim alanlarına ait olmalarına rağmen, dil incelenirken bu üç kavramı birbirinden bağımsız olarak düşünmek söz konusu değildir. Çağdaş dilbilim de özellikle biçimbilim ve sözdiziminin birbirinden ayrılmayan iki kol olduğunu savunur. Özellikle Türkçe gibi bitişken dillerde biçimbilim ve sözdizimi arasındaki

etkileşim çok önemlidir. Örneğin ettirgen ekleri eylemin hem değerini hem de anlamını değiştirmektedir.

Türkçe mana bakımından iki çeşit kelime içerir:

i) nesneleri karşılayan isim türünde ise çokluk, hal, iyelik ekleri ve som ekleri,

ii) hareketleri karşılayan eylem türünde ise, eylemin kullanılırken sokulduğu kalıba veren bildirme ve/veya isteme kip(ler)i (bunlara zaman ve şahıs unsurları da eklenmektedir).

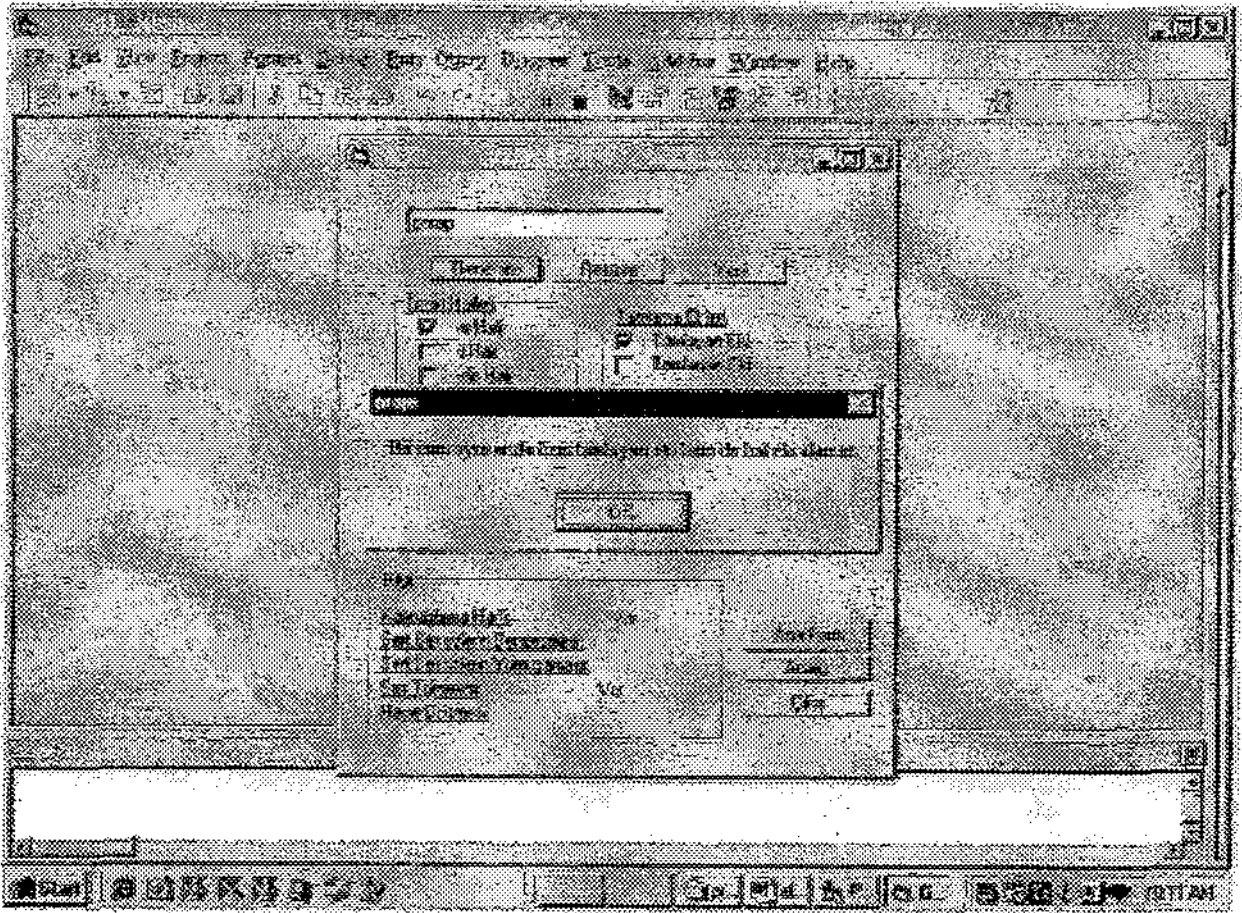
Bu işlemler için sözcük biçiminin oluşturulması, yani üretici işlemlerin gerçekleştirilmesi; aynı zamanda da sözcük biçimlerinin analiz edilmesi, yani çözümlenebilmesi gerekir. Bizim uygulamamızda yukarıdaki kelime türlerinden sadece i) durumun (som ekleri hariç) incelemekte ve kelimeler veri tabanından alınmaktadır.

2. Türkçe İsimlerin Özellikleri:

Türkçe, sonuna eklemeli bir dil olduğundan, ekler isimlerin sonuna getirilerek kullanılmaktadır (pekiştirme ekleri hariç). Ekler anlamsız oldukları için tek başlarına kullanılmazlar. Sonuna geldikleri sözcüğün anlamını veya biçimini değiştirirler. İsimlerde ekler, görevleri bakımından çekim ve yapım eki olmak üzere ikiye ayrılır. Yapım ekleri, sonuna eklendikleri isimlerin kökünde anlamsal ve yapısal değişikliğe sebep olurlar. Çekim ekleri, sözcüğün anlamını ve çeşidini değiştirmez. Yapım ekleri çekim eklerinden önce gelirler. İsimlerin sonuna eklenen dört çeşit çekim eki bulunmaktadır. Bunlar ismin

hal ekleri,
iyelik ekleri,
tamlama ekleri
çoğul eki

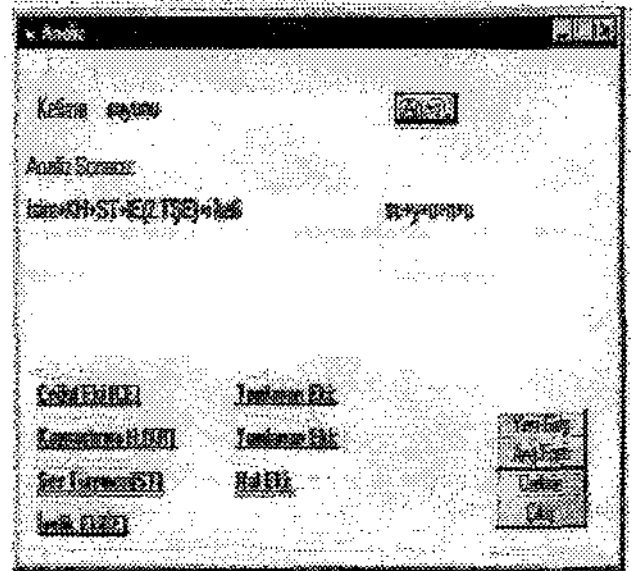
olarak sınıflanırlar.



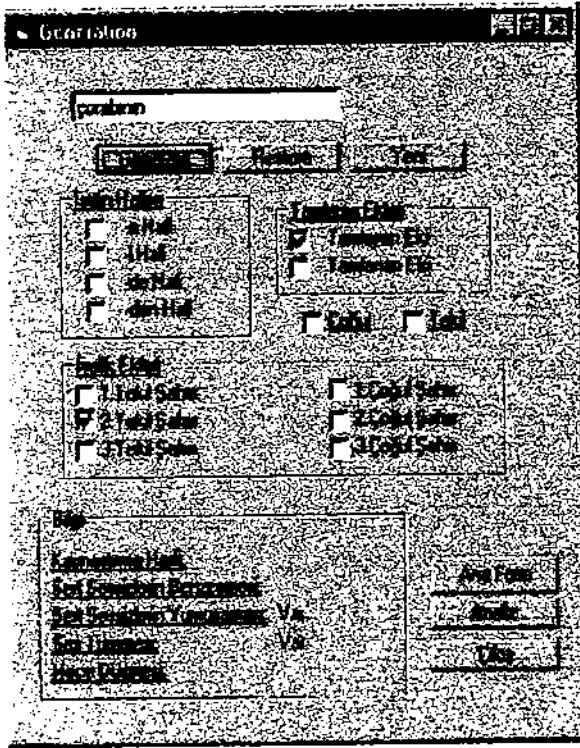
Şekil 2: İsimlere ait çekim ekleri üretici kural dışı ek üretimlerine izin vermez

Bileşik isimlerin bazıları belirtisiz isim tamlaması şeklindedir. Örneğin: "hammeli" kelimesinde "hamm" kısmı tamlayan "eli" kısmı tamlanan olarak düşünülebilir. Bileşik isimlerde, ismin -i ve -e halini aldıkları zaman -n veya -y (hammeli-n-i veya hammeli-y-e), tamlayan eki aldıklarında ise sadece -n kaynaştırma harfi (hammeli-n-in) gelir. Gelen -n harfinin 2.tekil şahısa ait iyelik eki olma olasılığı da göz önünde bulundurulmalıdır.

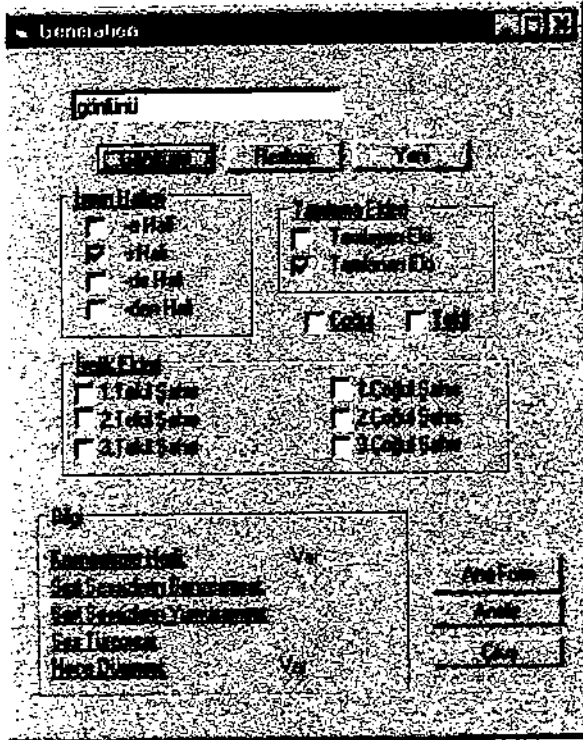
"Çay" sözcüğü çay-ı-m, çay-ı-nız gibi sessiz harfle başlayan bir ekle türetildiğinde, kelimenin kökünden sonra -ı seslisini aldığı için ses türemesi olayını oluşturur. "Su" kelimesi ise Türkçe'nin dilbilgisinde ilginç bir örnek oluşturur. Sesli ile bitmesine rağmen sessizle başlayan iyelik eki aldığında hem -y kaynaştırma harfini almaktadır, hem de ses türemesi meydana gelmektedir.



Şekil 3: Türkçe'de ses türemesi olayı



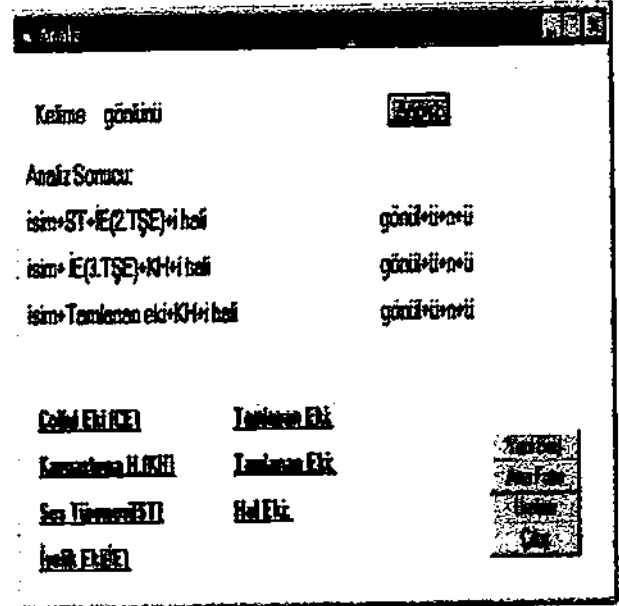
Şekil 4: Giriş "çorap" kelimesi olmak üzere sert sessizlerin yumuşaması kuralı örneği



Şekil 5: Türkçe'de ses düşmesini oluşturan üreteç

Eğer, isim sert sessiz ile bitiyor ve b,c,d,g/ğ sessizlerinden herhangi biri ile başlayan bir ek alıyorsa, bu sessizler p,c,t,k sünekli sert sessizlerine dönüşerek *sert sessizlerin benzeşmesi* olayını oluşturmalar. Türkçe isimler, bu sessizlerden sadece -d harfi ile başlayan çekim ekleri aldıkları için, ses olayı sadece -d den -t ye dönüşümü olarak gerçekleşir. Bu temel dilbilgisi kuralı Bölüm 3' de iki düzeyli ağ modeli ile genelleştirilmektedir.

Türkçe'de orta hece vurgusuz olduğuna için iki heceli, son harfi sessiz olan ve son hecesindeki sesli harf dar olan kelimenin sonuna sesli ile başlayan bir çekim eki aldığı zaman orta hecedeki dar sesli harf düşerek *hece düşmesi* ses olayını meydana getirir. Örneğin: "Sumanın gönlünü aldık." cümlesinde "Sama'nın gönlü" ismi tamlamasıdır. "gönlü" kelimesi -ü tamlanan eki alarak son hecesindeki sesli düşmüş ve "gönlü" haline gelmiştir. Daha sonra kaynaştırma harfi alarak ismin -e haline dönüşmüştür.



Şekil 6: Ses düşmesinin çözümlenmesi

Ses Düşmesi, iki heceli son hecesindeki dar sesli olan bütün kelimelerde (aydın - aydını türemesi gibi) meydana gelmeyebilir. Bu kelime de yapısal olarak iki heceden oluşmaktadır ve son hecesindeki sesli ise dar seslidir; ayrıca ismin -i hal ekini almıştır. Fakat son hecedeki seslide düşme olayı meydana gelmemiştir.

3. İki Düzeyli Biçimbirimsel Modelleme

Biçimbirim dilin biçimini araştırma bilimidir. Bölüm 2 'de incelediğimiz gibi Türkçe'de kelime formları sıralanmış birimlerden oluşmaktadır, bunlar başlangıçta temel simgeleniş ve üyinsel şeklinde birbirini takip eden diğer eklerdir. Bazı olağanüstü durumların dışında Türkçe'nin biçiminde sesli ve sessiz uyumu mutlaka görülür. Bu şekilde eklerin incelenmesi, araya pek çok parçanın girmesi nedeni ile, epey karışıktır. Biçimbirim aynı zamanda, kelimenin anlamlı en küçük parçası olarak tanımlanır, fakat bu her zaman doğru olmayabilir. Biçimbirimi sözlüksel en küçük fonksiyon olarak da tanımlayabiliriz. İki düzeyli biçimbirimde gerçek gelişme Kimmo Koskeniemi'nin 1983'te geliştirdiği sistemi ile görülmektedir. Finlandiyalı bu bilim adamının geliştirdiği sistem kelime tanıma ve üretimini tek model üzerinde simgelediği için, farklı dillerde yapılan biçimbirimsel çalışmaların artışı gözlenmektedir. Kimmo'nun geliştirdiği bu sistem, biçimbirimsel kurallara ilave olarak sesbilimsel kuralları da içermektedir.

Koskeniemi geliştirdiği iki düzeyli bu modeli PC-KIMMO sistemi olarak adlandırmış ve aşağıda da görülebileceği gibi "sözlüksel form" ve "görünüş formu" şeklinde simgelemiştir. Türkçe'nin biçimbirimlerini incelemeye bir örnek vermek istersek, "çorabım" kelimesi biçimbirimsel analiz sonucunda "çorap" haline dönüşecektir.



Şekil 7: PC_KIMMO sistemi

Biz de çalışmamızda, Bölüm 4'de açıklanan üretici ve çözümleyici sistemimizin geliştirilmesine yardımcı olmak üzere, PC-KIMMO sistemine benzer iki-düzeyli biçimbirim modelini Türkçe'nin önemli ses olayları için tasarladık.

3.1. Türkçe'nin temel dilbilgisi kurallarının iki-düzeyli biçimbirimsel modeli

Süreksiz sert sessiz harflerin yumuşaması kuralı iki-düzeyli biçimbirim yapısına benzer bir terminoloji ile aşağıdaki gibi simgelenebilir:

$$SSz : YSz \Leftrightarrow (Se_1^{n-1} // Se_2^{n-1})_n : _ (v // a)$$

$$S.F. : k i t a p \phi , k e r p i \phi \phi$$

$$G.F. : k i t a b a , k e r p i c e$$

Görüldüğü gibi süreksiz sert sessiz ve yumuşak sert sessiz harfler verilmiş sırasına göre karşılıklı olarak eşleşmektedir.

Bir başka iki düzeyli kural;

$$\phi : d // t \Leftrightarrow YSz // SSz _ : _ e/a$$

ve sınırlayıcısı

$$\phi : (e/a) \Leftrightarrow (Se_1 // Se_2) _ : d/t _$$

şeklinde sessizlerin benzeşmesi kuralını simgeler.

$$S.F. : k i t a p \phi \phi , k e r p i \phi \phi \phi$$

$$G.F. : k i t a p t a , k e r p i \phi t e$$

Yukarıdaki her iki simgelenişlerde sesli harfin 'e' veya 'a' olarak alınması sözlüksel (temel) formdaki son sesli harfe göre büyük sesli uyumundan (veya diğer hal ekinin alınması durumunda 'ı, i, u, ü' dar seslilerinden herhangi birinin alınması da yine temel formdaki son sesli harfe göre küçük ses uyumundan) çıkarılacaktır.

Hece düşmesini simgeleyen iki-düzeyli biçimbirimsel model aşağıdaki şekilde verilmektedir. Gömül ve akıl kelimeleri örnek olarak alınacak olursa;

$$S.F. : g ö n ü t \phi , a k t k \phi$$

$$G.F. : g ö n \phi l ü , a k \phi l i$$

haline gelecektir. Bu ses olayına ait iki-düzeyli kural:

$$Se_4 : \phi \Leftrightarrow _ Sz : _ Sz Se_3$$

olarak yazılabilir.

4. Çözümleyici ve Üreticinin Tasarınlanması

Çözümleyici, kendi kullanıcı arayüzüne giren yapılan kelimeyi kök ve eklerine ayırabilen olabilecek tüm ihtimalleri gözönünde bulundurarak bilgisayarı programdır. Bu program biçimbirim araştırmamızı genelleştirmek amacı ile iki-düzeyli biçimbirim oluşturacak şekilde tasarlanmıştır.

Geliştirilen model 'Koskenniemi' nin geliştirdiği biçimbirimsel dönmüşlüci modeline benzer soyut bir modeldir; ayrıca hem çözümleyici hem de üretici için iki yönlü olarak izlenebilmektedir. Üreteç ise, yine kullanıcı arayüzünde girilen kök halindeki bir isme istenilen çekim eklerini ekleyen nesneye yönelik bir bilgisayar programıdır. Bu çalışmada üretici ve çözümleyici MS Visual Basic 6.0 versiyonu kullanılarak tasarlanmış ve modelin veritabanı ise MS Access97'de hazırlanmıştır.

Programda iki adet senaryo bulunmaktadır. İlk senaryoya göre kullanıcı üretici arayüzündeki giriş alanına bir isim kökü girip, bunun o ismin istenilen hal eki ve şahıs ekini almasını ister. Örneğin giriş olarak "gönlü" kelimesi girildiğinde, bu ismin i-hal eki ve 3. T.Ş. kişi eki için "gönlünü" çıktısı verdiği gözlenir. Bu form Şekil 5'dekinden farklı bir isteği yansıtmaktadır. İkinci senaryoya göre de kullanıcı çözümleyici arayüzüne giriş olarak bir kelime, örneğin "gönlünü" yazar. Çözümleme sonucu Şekil 6'dan da görüleceği gibi bir tane olmayacaktır. Başka bir ifade ile birbirinden farklı 3 üretici girdisi için aynı sonuç alınacaktır.

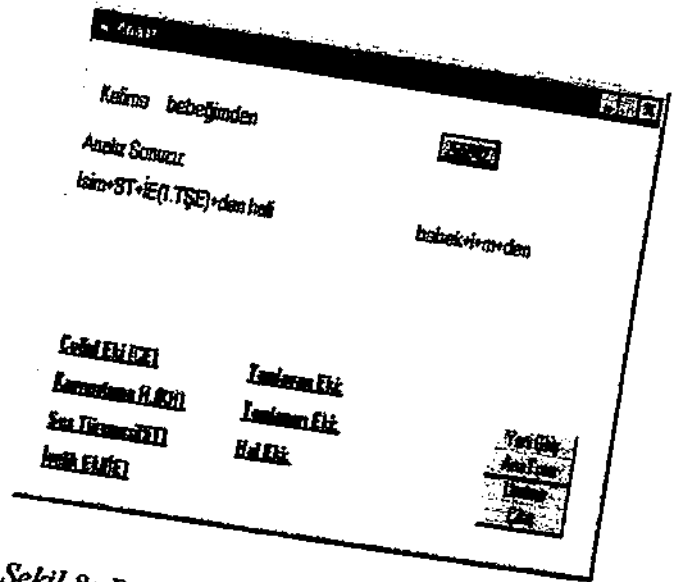
5.Sonuç:

Türkçe, Ural-Altay dil grubuna giren bir dildir. Türkçe'nin doğal dil işlemlerinde sorunlar yaratan bazı özellikleri vardır. Türkçe sözcük yapısı ve üretimi açısından bitişken bir dildir. Bu özelliği ile Fince ve Macarca'ya benzemektedir. Bu tip dillerde sözcükler bir kök ve bu köke eklenen biçimbirimlerden meydana gelirler. Bu biçimbirimler eklendikleri kök veya gövdenin anlamını, sözcük türünü veya sözdizimsel işlevini değiştirebilir. Türkçe'de bu şekilde eklemelerle oluşturulan bir sözcük ile ifade edilebilen herhangi bir kavram, bazen başka bir dilde ancak bir cümle ile ifade edilebilir.

Bu çalışma, küçük bir doğal dil işleme uygulamasıdır ve de başarılı bir doğal dil uygulamasının ilk şartlarından biri olan biçimbirimsel ayrıştırmayı isimler üzerinde gerçekleştirmektedir. Türkçe cümlelerin içindeki çekim sözcükleri sözdizimini doğrudan etkiledikleri

an, cümlelerin anlamını belirlemede biçimbilimin de artmaktadır. Fakat çalışmamız şu aşamada sadece biçimbirim ile ilgilenmekte, derin anlamsal analizini göz önüne almamaktadır.

Doğal dil işleme uygulamaları, laboratuvar ortamından çıkıp gerçek eğitim gerekse endüstri ortamında kullanılmasıyla, dil çalışmalarını kolaylıklardan öte büyük değişimlere de uğatacaktır.



Şekil 8: Bir başka çözümleme örneği

Kaynakça:

- [1] Jurafsky D. & Martin J.H., 2000, "An Introduction to Natural Language Processing, Computational Linguistic and speech recognition", Prentice Hall,
- [2] Chomsky N., 1995, "The Minimalist Program", Cambridge, MIT Press.
- [3] Vardar B., 1998, "Dilbilimin Temel Kavram ve İlkeleri", Multilingual.
- [4] Roche E. and Shabes Y.(editors), 1997, Finite State Language Processing, MIT press.
- [5] Martin J.C., 1991, "Introduction to Languages and The Theory of Computation", pp.279-282, McGraw-Hill, Inc.
- [6] <http://www.fxtc.xerox.com/research/mltt/fst>
- [7] <http://www.nyu.edu/pages/linguistics/anltbk.html>
- [8] http://crl.nmsu.edu/Zajac_R_Morphological_Processing_and_Finite_State_Techniques
- [9] <http://agora.leeds.ac.uk/nti-kbs/ai5/atn.html>

Chapter 3 of the Gazdar & Mellish text for a full description of this topic.