

Yazıdan Konuşma Üretmede Kullanılan Doğal Dil İşleme Tekniklerine Genel Bir Bakış (An Overview of Natural Language Processing Techniques in Text-To-Speech Systems)

M. Oğuzhan Külekci ve Kemal Oflazer
Sabancı Üniversitesi
Mühendislik ve Doğa Bilimleri Fakültesi

kulekci@su.sabanciuniv.edu, oflazer@sabanciuniv.edu
Tel: 0216-483-9532
Faks: 0216-483-9550

Özetçe

Bu makalede doğal dil işleme tekniklerinin yazıdan konuşma üretme sistemlerinde kullanımını ile ilgili çalışmalar gözden geçirilecektir.

Abstract

This paper will present overview of the use of natural language processing techniques at various stages of TTS systems.

1 Giriş

Yazıdan konuşma üretme (YKÜ) problemi bir dilde yazılmış herhangi bir yazıda bulunan sınırlı sayıdaki harf ve karakterlerin ses işaretlerine dönüştürülmesidir [1]. Bir YKÜ sistemi iki ana kısımdan oluşur. Bunlardan ilki sistemin hedeflediği dilin özelliklerine ve seslendirilmesi hedeflenen yazıların tiplerine (gazete metni, elektronik posta, diyalog vb...) göre yapılan yazı çözümlemesi, ikincisi ise yapılan bu çözümlemenin sonuçlarından ses sinyalinin üretimidir.

Yazı çözümlemesinin ilk basamağı yazıda bulunan tüm karakterlerin seslendirilebilir harflere çevrimidir. *Yazı düzgelemesi* olarak adlandırılan bu işlemle bir yazı içerisinde bulunan kısaltma, tarih, sayı ve benzerleri gibi, sözcük olarak yazılmayan semboller konuşma karşılıkları olan sözcüklere dönüştürülür [2]. Fakat bazı diller için (Çince, Japonca, Arapça, İbranice vb.) düzgeleme işleminden önce bu dillerin yazma sistemlerinden kaynaklanan problemler

çözülmelidir. Bu çalışma dahilindeki ikinci bölüm bu tip dillerde yapılması gereken bu öncül işlemleri ele almaktadır.

Yazı çözümlemesi için yapılması gereken bir diğer işlem sözcüklerin morfolojik çözümlemelerinin bulunup gerekli hallerde tekleştirilmesidir. Doğal dil işlemede özellikle Türkçe gibi eklemeli ve türetmeli diller için çok önemli olan bu çözümlemelerin YKÜ sistemlerinde kullanımı üçüncü bölümde incelenmektedir.

Dördüncü bölüm yazı düzgelemesine ayrılmış olup, yazı içerisindeki standart olmayan (sözlükte bulunmayan) sözcüklerin gerektirdiği işlemlerin yanısıra, isimlendirilmiş varlıklar ve eşyazımlı sözcüklerle ilgili yapılması gerekenler de bu bölümde gözden geçirilmektedir.

Bir yazının insan konuşmasına en yakın şekilde seslendirilebilmesi için cümle bazında tonlamayı belirleyen tümce bürünlerinin kestirimi gereklidir. YKÜ sistemlerinde doğal konuşma üretebilmek amacıyla sözdizimsel ve anlambilimsel çözümlemeleri temel alan çeşitli yaklaşımlar geliştirilmiştir. Bu metodların genel bir değerlendirmesi de beşinci bölümde sunulmaktadır.

2 YKÜde Öncül İşlemler

Bir YKÜ sisteminin yazı düzgelemesi işlemine başlamadan önce verilen metnin içindeki tüm sözcükleri ve diğer parçaları (tarih, sayı, noktalama işareti vb.) tesbit etmesi gerekir. Temel olarak bir karakter dizisini sözcüklere ayırmak olarak tanımlanan *andaçlama* doğal dil işleme içeren

herhangi bir uygulamada ilk yapılması gereken işlemdir [3]. Bir çok dilde sözcükler boşluk karakteri veya noktalama işaretleri ile ayrılarak yazılmaktadır. Bu dillerde andaçlama yapmak nispeten daha kolay olmakla birlikte, yazı düzgelemesinin başarılı olabilmesi için yine de özel durumların tesbiti ve özen gösterilmesi gereklidir. Örneğin *TBMM'de saat 19:00'da başlayan oturum devam ediyor* cümlesinde sözcükler boşluk karakteri ile ayrılmış olmasına rağmen daha doğru bir çözümlemenin yapılabilmesi için *TBMM'de* ve *19:00'da* andaçlarının ' işaretinden tekrar ikiye ayrılmaları gerekir. Böylece *TBMM* kısaltması istenilirse uzun hali ile okunabilir ve *19:00* saati de metin düzgelemesinde 'on dokuz sıfır sıfır' şekline çevrilebilir.

Çince, Japonca, Taice gibi bazı dillerin yazma sistemlerinde sözcükler ayrılmadan, bitişik yazılmaktadır. Bu durumda andaçlama işlemi zor bir problem olup, yazı üzerinde evvela *sözcük bölütlemesi* yapılması gerekir. Sözcük bölütlemesi bitişik yazılan sözcükleri ayırmak, veya bir anlamda boşluk karakterini yazının içerisinde uygun yerlere yerleştirmek olarak tanımlanabilir [4]. Çince ve benzeri dillerde sözcük bölütleme işlemi için istatistiksel, kurallara dayalı ve bu ikisinin birleşimi olan karışık yaklaşımlar denenmiştir. İstatistiki metodlar genel olarak kurala dayalı metodlardan daha kötü sonuçlar vermiş olup, son zamanlarda genellikle bu ikisinin birlikte kullanıldığı karışık sistemler üzerine çalışılmaktadır [3]. Çince sözcük bölütlemesine bir örnek olarak 日文章魚怎麼說 cümlesi 日文 (Japonlar) 章魚 (ahtapot) 怎麼 (nasıl) 說 (söyler) ,veya 日 (Japon) 文章 (yazıt) 魚 (balık) 怎麼 (nasıl) 說 (söyler) şeklinde bölütlenebilir [5]. Bu iki bölütmeden doğru olanı birincisi olup, Çince bir YKÜ sisteminden de bu cümleyi doğru şekilde andaçlaması beklenir.

Yazı sisteminden kaynaklanan bir diğer problem ise semitik diller olarak adlandırılan Arapça ve İbranice'de sözcüklerin sadece sessiz harfler kullanılarak yazılmasıdır. Bu yazım kuralı nedeniyle sadece sesli harflerde farklılıkları olan birçok sözcük aynı şekilde yazılmaktadır. Örneğin Arapça كتاب (ktb) sözcüğü hem kitaplar (kitaab) hem de sekreterler (kuttaab) sözcüklerine karşılık gelmektedir. Literatürde bu problemin çözümü için en başarılı yöntem saklı Markov modelinin kullanılması olarak görülmektedir [6]. Bu yöntemde izlenen değişken olarak sözcüklerin sessiz harflerle yazımları, saklı değişken olarakta sesli harflerle birlikte yazımları kabul edilmekte ve sözcüğün doğru çözümlemesinin kendinden önceki sözcüklere bağlı olduğu kabul edilerek sonuca gidilmektedir. Arapça ve İbranice'de dilin özelliğinden kaynaklanan bu problem, YKÜ sis-

teminin hedeflediği uygulama türüne bağlı olarak da ortaya çıkabilir. Örnek olarak Türkçe kendi yazı sisteminde böyle bir problem taşımamasına rağmen, bu dildeki bir SMS(kısa mesaj servisi) veya chat okuyucusunun da 'slm nbr' sözcüklerini 'selam naber' şekline çevirmesi beklenir.

3 Morfolojik Çözümleme

En yalın hali ile bir sözcüğün kökünü ve eklerini bulmak olan biçimbirimsel çözümleme YKÜ sistemlerinde birçok noktada kullanılmaktadır. Sözcük bölütlemesinde [7] ve semitik dillerin sesli harf üretimi probleminin [6] çözümünde de denenmiş olan biçimbirimsel çözümlemenin, YKÜ sistemlerinde en yaygın kullanımı söyleyiş veritabanlarının oluşturulması, vurgunun belirlenmesi ve yazı düzgelemesindedir.

Söyleyiş veritabanları bir dildeki tüm sözcüklerin söyleyişlerinin ses sinyaline dönüştürülebilir şekilde kodlandığı veritabanlarıdır. Sözcük çekim ve türetmelerinin sınırlı sayıda olduğu diller için uygulanması kolay olan bu yöntem, teorik olarak sınırsız sayıda sözcük içeren Türkçe gibi eklemeli ve türetmeli diller için yetersiz kalmaktadır. Bu gibi dillerde bir sözcüğün doğru okunuşunun oluşturulması için biçimbirimsel çözümlemesinin yapılması gerekir [8]. Örneğin Türkçe *karın* sözcüğünün biçimbirimsel çözümlemeleri ve bunlara karşılık gelen okunuşları şu şekildedir:

1. (ca:"-r)kar+Noun+A3sg(1n)+P2sg+Nom
2. (ka-"r)kar+Noun+A3sg(1n)+P2sg+Nom
3. (ca:"-r)kar+Noun+A3sg+Pnon(1n)+Gen
4. (ka-"r)kar+Noun+A3sg+Pnon(1n)+Gen
5. ("ka-r)kar+Verb+Pos(1n)+Imp+A2sg
6. (ka-"r1)karı+Noun+A3sg(n)+P2sg+Nom
7. (ka-"r1n)karın+Noun+A3sg+P+Nom

Yukarıdaki yedi çözümmeden sözcüğün kullanıldığı yerin içeriğine göre en uygun olanı seçilip (biçimbirimsel çözümleme tekleştirme) karşılık gelen söyleyiş üretilir. Biçimbirimsel çözümleme ile aynı zamanda sözcük etiketlenmesi de yapılmış olur. Bu etiketler cümle içi vurgunun oluşturulabilmesi için çok gerekli olan sözcük öbeklerinin ve tümce bürünlerinin kestiriminde ve sözdizimsel çözümlemelerde temel bir rol oynar. Cümle bazındaki vurgunun yanı sıra, bir sözcük içindeki baskın hecenin tesbitinde de yine biçimbirimsel çözümleme kullanılır [9].

4 Yazı Düzgelemesi

Yazı düzgelemesini bir yazı içerisinde geçen standard dışı andaçların, eşyazımlı sözcüklerin ve isim-

lendirilmiş varlıkların belirlenerek, söyleyişlerinin üretilmesi olarak üç kısımda incelemek mümkündür. Herhangi bir metinde standart sözcüklerin yanısıra sayılar, tarihler, Romen rakamları, kısaltmalar, kısaadlar, web ve e-posta adresleri gibi birçok standart dışı (sözlükte olmayan) andaçların da bulunması çok olasıdır. Örneğin Türkçe bir yazıda ortaya çıkabilecek bu tip durumların bazıları şu şekilde sıralanabilir: *01.10.1999, 1/10/99* gibi tarih belirteçleri okunuşlarının oluşturulması için içeriğe göre "bir kasım bindokuzyüzdoksandokuz" veya "bir on bindokuzyüzdoksandokuz" şekline dönüştürülmelidir. *1/4* kesri yine içeriğe göre "bir bölü dört", "çeyrek" veya "dörtte bir" olarak düzgelelenmelidir. *Cad.* kısaltması "caddesi", *IV. Murad* "dördüncü Murad" olarak değiştirilmelidir. Kısaadların söyleyişlerinin oluşturulması ise daha detaylı bir çözümleme gerektirir. Zira kısaadların okunmasında standard bir yol olmayıp, birkaç çeşit yaklaşımdan uygun olanı tercih edilmelidir. Örneğin *TBMM* genellikle "Türkiye Büyük Millet Meclisi" şeklinde uzun haliyle okunurken, AIDS İngilizce'deki gibi telaffuz edilir ve hiçbir zaman uzun hali ile okunmaz. *TÜBİTAK* sanki normal bir sözcük imiş gibi okunur. Birçok kısaad sözcüğü oluşturan harfleri tek tek okuyarak söylenirken (*BDDK*), bazen de karışık yaklaşımlar gerektiren istisnalarla (*RTÜK* – ilk harf okunur, son üç harf tek sözcük gibi seslendirilir) karşılaşılabılır.

YKÜde isimlendirilmiş varlıkların (şahıs adları, yer adları, özel isimler vb.) söylenişlerinin oluşturulması dildeki diğer standart sözcüklerden farklıdır. Bu sözcükler aynı zamanda cümle tonlamasını belirleyen sözcük öbeklerinin bulunması açısından da önem taşırlar. Bir dildeki isimlendirilmiş varlıkların sayısı genellikle bir veritabanına toplanamayacak kadar çok ve çeşitli olabilir. Bu nedenle sistem, bir cümle içerisindeki bu tip sözcükleri belirleme ve söylenişlerini oluşturma yeteneğine sahip olmalıdır. İsimlendirilmiş varlıkların cümle içinde tesbiti için hem kural tabanlı sistemler [10] hem de saklı Markov modelleri önerilmiş olup, saklı Markov modelleri ile İngilizce için başarı oranı yüzde 96 olarak belirtilmiştir [11].

Eşyazımlı sözcükler çoğunlukla farklı şekillerde okunurlar. Örneğin *Aydın* sözcüğü Türkçe'de bir il veya bir şahıs adı şeklinde isimlendirilmiş varlık olarak görülebildiği gibi; entellektüel insan anlamında, normal bir isim olarak da kullanılabilir. Bunların herbirinin söyleniş farklıdır. *Aydın* sözcüğünü içeren bir cümle için seslendirilebilmesi için isimlendirilmiş varlık tesbitinin yapılarak eşyazımlı bu sözcüklerden hangisinin kastedildiği belirlenmelidir. Eşyazımlı sözcüklerin çözümlenmesinde biçimbirimsel çözümleme sonucu

elde edilen ve sözcüğün tipini içeren bilgilerden oluşan sözcük etiketleri de önemli rol oynar.

5 Vurgu ve Cümle Tonlamaları

Sentetik bir konuşma ile bir insanın konuşması arasındaki en temel fark vurgudan kaynaklanır. Günümüzde YKÜ sistemlerinin *anlaşılabilirlik* açısından yeterli bir seviyede oldukları, fakat *doğallık* bakımından aynı düzeyde başarılı olmadıkları belirtilmektedir [12]. İnsanlar bir metni okurken, okuduklarını anladıkları için gerekli vurguyu başarmakta, buna karşın bilgisayarlar seslendirdikleri bir yazı üzerinde hiçbir anlam bilgisine sahip olmadıklarından, sentezlenen konuşma yapay ve tek düze bir halde duyulmaktadır. Bu sebeple, doğal dil işleme kapsamında biçimbirimsel, sözdizimsel ve anlambilimsel çözümlerler kullanılarak, konuşma üretiminde daha doğal ses sentezlerinin elde edilmesi araştırılmaktadır.

Bir cümle için vurgusunu hem o cümleyi oluşturan sözcükler, hem de bu sözcüklerin cümle içindeki dizilimleri belirler. Örnek olarak *'Bugün ben Ankara'ya gidiyorum'* ve *'Ankara'ya bugün ben gidiyorum'* cümleleri aynı sözcüklerden oluşmalarına rağmen tonlamaları farklıdır. Doğru vurgunun üretilmesi için ilk aşamada her sözcüğün baskın hecesinin biliniyor olması, bunun sonrasında ise sözcük dizilimine bağlı olarak hangi sözcüklerin vurgulu söyleneceği, hangi sözcüklerin ulama ile birleştirileceği ve hangi sözcükler arasında daha uzun duraklar verileceği bilgisinin elde edilmesi gereklidir [13]. Bir sözcüğün baskın hecesi veya heceleri biçimbirimsel çözümleme ile bulunabilir [9]. Sözcük öbeklerinin ve tümce bürünlerinin kestirilmesine bağlı olan diğer etkenlerin tesbiti için ise genellikle sözdizimsel çözümleme temel olarak alınmaktadır [14]. Sözdizimsel çözümleme sonucunda bir cümle içerisindeki sözcüklerin görevleri ve birbirleri ile olan ilişkileri belirlenmiş olup, bu bilgiler sözcük öbeklerinin belirlenmesini sağlar.

Cümleyi oluşturan sözcüklerin biçimbirimsel çözümlenmelerinin yapıldığı, cümle için sözdizimsel çözümlenmesinin gerçekleştirildiği ve bunlara ek olarak cümle tipinin (soru, diyalog vb.) verildiği kabulü ile bir YKÜ sisteminde vurgunun oluşturulması üç basamakla açıklanabilir [15]. Bu yaklaşıma göre ilk adımda sözdizimsel çözümleme ışığında tümce bürünleri bulunur. İkinci adımda hangi sözcüklerin baskın söyleneceği, hangilerinin ulanacağı belirlenir. Üçüncü adımda ise tesbit edilen tümce bürünlerinin ne şekilde tonlanacağı, hangi sözcük grubunun daha baskılı söyleneceğine, yani cümle için genel ahengine, cümle tipine göre karar verilir.

Cümle tonlamasında olan farklılıkların bazı durumlarda anlamı da değiştireceği de dikkate alınmalıdır. Nitekim *Adam ol abin gibi asi olma.* cümlesi *gibi* sözcüğünden iki tümce bürününe ayrılırsa farklı, *ol* sözcüğünden ayrılırsa tamamen farklı manaya gelmektedir. Vurgulamayı, sözcük ve cümle bazındaki yaklaşımların ötesinde, cümleler arası etkileşimlerin ve yazının bir bütün olarak manasının da etkilediği unutulmamalıdır. Daha derin incelemeyi gerektiren bu tip bir yaklaşım için ise biçimbirimsel ve sözdizimsel çözümlemenin yanı sıra anlambilimsel bir çözümlemeden de yararlanan araştırmalar mevcuttur [16].

6 Sonuç

Literatürde *yazı çözümleme* kapsamında olan çalışmalar gözden geçirilmiştir. Sonuç olarak bir YKÜ sisteminin tasarımında yapılması gerekenler en genel haliyle aşağıdaki beş basamaklı modelle özetlenebilir:

1. Dil veya uygulamadan kaynaklanan ihtiyaç doğrultusunda öncül işlemler (sözcük bölütleme – sesli harf ekleme) gerçekleştirilir.
2. Seslendirilebilir harf dizisi olarak yazılmamış olan andaçlar (tarih, sayı vb.) sözcüklere çevrilir.
3. Her sözcüğün biçimbirimsel çözümlenmeleri bulunur.
4. İsimlendirilmiş varlıkların tesbiti, eşsesli sözcüklerin çözümlenmeleri ve söyleyişlerin oluşturulmasına yönelik biçimbirimsel çözümleme tekleştirilmeleri yapılır.
5. Sözdizimsel ve anlambilimsel çözümlenmeler yapılarak doğru cümle tonlaması seçilir.

References

- [1] Ksenia Shalanova and Roger Tucker. South asian languages in multilingual TTS-related database. Technical Report HPL-2003-9, HP Mobile and Media Systems Laboratory, 2003.
- [2] Richard Sproat, Alan W. Black, Stanley Chen, Shankar Kumar, Mari Ostendorf, and Christopher Richards. Normalization of non-standard words. *Computer Speech and Language*, 15(3):287–333, July 2001.
- [3] Jonathan J. Webster and Chunyu Kit. Tokenization as the initial phase in NLP. In *Proceedings of COLING-92*, volume 4, pages 1106–1110, 1992.
- [4] Jin Guo. Critical tokenization and its properties. *Computational Linguistics*, 23(4):569–596, 1997.
- [5] Richard Sproat, Chilin Shih, William Gale, and Nancy Chang. A stochastic finite-state word-segmentation algorithm for Chinese. *Computational Linguistics*, 22(3):377–404, 1996.
- [6] Ya’akov Gal. An HMM approach to vowel restoration in arabic and hebrew. In *Proceedings of ACL’02 Workshop on Computational Approaches to Semitic Languages*, University of Pennsylvania, Philadelphia, July 2002.
- [7] M. Oğuzhan Külekci and Mehmed Özkan. Turkish word segmentation by using morphological analyzer. In *Proceedings of Eurospeech 2001*, 2001.
- [8] Kemal Oflazer and Sharon Inkelas. A finite state pronunciation lexicon for Turkish. In *Proceedings of EAACL workshop on finite state methods in NLP*, 2003.
- [9] Kenneth Church. Morphological decomposition and stres assignment for speech synthesis. In *Proceedings of ACL’86*, pages 156–164, 1986.
- [10] Daniel M. Bikel, Scott Miller, Richard Schwartz, and Ralph Weischedel. Nymble: A high performance learning name-finder. In *Proceedings of Fifth Conference on Applied Natural Language Processing*, pages 194–201, 1997.
- [11] GuoDong Zhou and Jian Su. Named entity recognition using an hmm-based chunk tagger. In *Proceedings of the ACL 2002*, pages 473–480, July 2002.
- [12] M. Beutnagel, A. Conkie, J. Schroeter, Y. Stylianou, and A. Syrdal. The AT&T Next-Gen TTS system. Joint Meeting of ASA, EAA and DAGA, Berlin, Germany, March 1999.
- [13] Julia Hirschberg. Pitch accent in context: Predicting intonational prominence from text. *Artificial Intelligence*, 63(1–2):305–340, 1993.
- [14] J. Bachenko and E. Fitzpatrick. A computational grammar of discourse-neutral prosodic phrasing in english. *Computational Linguistics*, 16(3):155–170, 1990.
- [15] Alan W. Black and Paul Taylor. Assigning intonation elements and prosodic phrasing for english speech synthesis from high level linguistic input. In *Proceedings of ICSLP’94*, volume 3, Yokohama, Japan, 1994.
- [16] Scott Prevost and Mark Steedman. Specifying intonation from context for speech synthesis. *Speech Communication*, 15(1–2):139–153, 1994.