



24. ULUSAL DİLBİLİM KURULTAYI

17-18 Mayıs 2010

Ankara

BİLDİRİ KİTABI

Yayına Hazırlayan
Çiğdem SAĞIN-ŞİMŞEK
Çiler HATİPOĞLU

ORTA DOĞU TEKNİK ÜNİVERSİTESİ

Eğitim Fakültesi

Yabancı Diller Eğitimi Bölümü

udkodtu@metu.edu.tr

<http://www.dilbilimkurultayi.metu.edu.tr>

YUVARLAK MASA 3

YÖNETEN:

Şükriye RUHİ

**ARAŞTIRMA
KONUSU:**

**Türkçe için Genel Amaçlı
Sözlü Derlem
Oluşturmada Veri Bilgisi,
Çeviriyazı Ölçünleştirmesi
ve Derlem Yönetimi**

SÖZLÜ DERLEM İÇİN WEB TABANLI YÖNETİM SİSTEMİ

M. Güneş Can ACAR¹ ve Kerem ERYILMAZ²

1: Ankara Üniversitesi, 2: Orta Doğu Teknik Üniversitesi
gunes.acar@media.ankara.edu.tr, kerem@ii.metu.edu.tr

1. GİRİŞ

Sözlü Türkçe Derlemi Projesi Web Tabanlı Yönetim Sistemi, özellikle sözlü derlem geliştirme ve yayınlama süreçlerinde proje ekibi ve kullanıcılarına yardımcı olacak biçimde tasarlanmış, web tabanlı, çok kullanıcıli bir derlem geliştirme sistemidir. Sözlü Türkçe Derlemi Projesi kapsamında geliştirilen sistem, veri girişi, çeviriyazı ve çözümleme süreçlerinde üretilen verilerin yönetilmesini kolaylaştırmayı ve olası hataları en aza indirmeyi hedeflemektedir.

EXMARaLDA (Schmidt, 2004) araçları ile tümleşik olarak çalışacak biçimde geliştirilen sistemin öncelikli tasarım hedefleri, iş akışı denetimi, çok kullanıcı ve rol desteği, kalite denetimi, sürdürülebilirlik, kullanım kolaylığı, genişletilmeye açık olma ve diğer derlem araçları ve standartları ile uyumlu çalışabilme olmuştur. Sistemi planlarken göz önüne alınan gereklilikleri yansıtan tasarım hedeflerinden bazıları aşağıda ayrıntılı olarak ele alınmıştır.

2. DERLEM YÖNETİM SİSTEMİ TASARIM HEDEFLERİ

Kullanılacak sistemin belirlenmesi aşamasında, sistemin sahip olması gereken beş özellik belirlenmiştir. Bu beş özelliği bir arada sunabilen bir yazılımın saptanamaması üzerine, bu özellikler yeni tasarlanacak sistemin tasarım hedefleri olarak kullanılmıştır.

2.1. Çevrimiçi Erişim ve Merkezi Veritabanı

Geniş sayılabilecek bir proje ekibi ve daha geniş bir gönüllü grubunun aynı anda üzerinde çalışabileceği bir platform gereksinimi, tasarlanmakta olan sistemin web tabanlı olmasını ve merkezi bir veritabanı üzerinden çalışmasını gerektirmiştir. Sisteme çevrimiçi erişim, derlem kapsamında kaydedilen ses ve görüntü kayıtlarına, bunlarla ilgili çeviriyazı ve metaverilere kolayca erişilebilmek açısından da önemlidir.

Konuşma kaydı gibi hassas verilerin saklanacağı sistemin güvenlik açısından da özenle tasarlanması ve İnternet üzerinden gerçekleştirilebilecek herhangi bir sızmaya karşı gerekli önlemlerin alınması çevrimiçi erişimin getirdiği bir başka tasarım ölçütüdür.

2.2. İş Akışı Denetimi ve Kullanıcı Rollerini

İş akışı denetimi, sistemin, derlem geliştirme sürecindeki aşamaların doğru sırayla, doğru kişiler tarafından ve gerekli kontroller yapılarak geçildiğini garanti edebilmesi anlamına gelir. Yeni gönüllülerin sisteme kaydedilmesi, gerçekleştirilen kayıtların toplanması, sınıflandırılması ve kontrolü, çeviriyazıların çeviriyazı ekibine atanması, çeviriyazı sürecinin ve ürünlerin kontrol edilmesi gibi aşamalar, proje ekibi tarafından kolaylıkla takip edilebilir. Bu süreçte, her bir aşama bir önceki aşamanın tamamlanması şartıyla ulaşılabilir hale getirildiğinden dolayı bunların atlanması olasılığı ortadan kalkmaktadır.

Kullanıcı rollerini, her kullanıcıya bir veya daha fazla rol atanması ve bu rollerin yapabileceklerinin sınırlandırılması anlamına gelir. Kullanıcı rollerinin tanımlanmasıyla, iş

akışı denetimi bu süreçlerde kimin hangi rolleri oynayabileceğini de kısıtlayabilir. Sözelimi bir çeviriyazı kontrolü sadece proje ekibi tarafından yapılabilirken, herhangi bir götüllü kayıt bağışlayabilir.

2.3. Saydamlık ve Tutarlılık

Sistemin tasarlanmasında, sadece derlem yönetmenin kolaylaştırılması deęil, aynı zamanda derlemin çeşitli parametrelerinin anlık erişime açık olması, derlem yönetim ve geliştirilmesindeki süreçlerin açık bir şekilde takip edilebilmesi ve sistemin kendi içinde tutarlı olması temel hedefler arasında deęerlendirilmiştir.

Bu bağlamda, iş akışı denetiminde aşamalar arası geçişlerin saydam olmasına önem verilmiştir. Geçişlerin saydam olması, iki aşama arasında sistemin kalabileceği bir ara durumun mevcut olmayışıdır. Sözelimi çeviriyazı aşamasının şartları tamamlandıktan sonra çeviriyazı kontrol aşamasına geçen her kaydın temiz bir kayıt olduğundan, bir çeviriyazısı olduğundan, metaverisinin tamam olduğundan emin olunabilir. Sistemdeki bir hata dahi bu aşamadaki bir çeviriyazının bu şartları sağlamadan kontrol aşamasına gelmesine yol açmaz. Böylece kullanıcı veya yönetici, sistemdeki aşamaların derlemin gerçek halini yansıttığı ve kendi içlerinde tutarlı olduğu önkabuluyle hareket edebilir.

Aynı prensip, sistemdeki parametrelerin anlık takibine olanak sağlar. Sistem herhangi bir anda kendi içinde tutarlı olacağını garanti ettiği için, kullanıcı bilgileri, çeviriyazı istatistikleri gibi parametreler her an ulaşılabilir haldedir ve sistemin mevcut durumunu doğru şekilde yansıtır.

2.4. Genişletilebilirlik

Bir sistemin genişletilebilirliği, yeni özellikler ve prensipler eklenmesine olanak tanınması olarak tanımlanabilir. Derlem yönetim sisteminin genişletilebilir olması, sistemin yeni metaveri tipleri, yeni dosya türleri, yeni veritabanı sistemleri ve yeni akışların eklenmesine olanak sağlar. Sözelimi, yeni bir çeviriyazı biçiminin ve ilgili bir dosya türünün ortaya çıkması halinde, sisteme, derlemi bu dosya türünde de sunmaya uygun mekanizmaların eklenmesi mümkün olduğu kadar kolay hale getirilmiştir. Bunun önemi, sistemin mevcut biçimler ve dosya tipleriyle sınırlanmaması ve bunların gerektiğinde genişletilerek veya değiştirilerek derlemin kullanılabilirliğini arttırmasıdır.

2.5. Uzun Vadede Kullanılabilirlik

Bir sistemin uzun vadede kullanılabilirliğinin sağlanması, sistemin kullanıldığı alandaki deęişikliklerin sistemin kullanılabilirliğine olan etkisinin asgari düzeye indirilmesidir. Genişletilebilirlik hedefiyle yakından bağlantılı olan bu hedef, sistemin gerek kullandığı/ürettiği dosya ve veri tipleriyle, gerekse de kaynak koduyla, mümkün olan her şartta kullanımının sürdürülebilmesine bağlıdır.

Bu hedef çerçevesinde, sistemimizin açık kaynak kodlu olması büyük önem taşımaktadır. Sistemin mevcut çalışanları projeyi geliştirmeyi bıraksa dahi, sistem başkalarının geliştirme çabalarına açıktır. Benzer bir şekilde, sistemde kullanılan temel veri tiplerinin de açık kaynak kodlu yazılımlarla beraber kullanılabilmesine dikkat edilmiştir. Sistemin ürettiği dosya tiplerini kullanan programların artık geliştirilmemesi durumunda, sistem başka bir dosya tipi üreterek kullanılabilirliğini sağlayabilir. Sistem ayrıca, derlemi HTML dosyaları gibi çok daha genel ve ortadan kalkması ihtimali düşük dosya tipleri olarak sunabilmektedir.

3. DERLEM YÖNETİM SİSTEMİ TEMEL ÖZELLİKLERİ

Bu bölümde, sistemin temel özellikleri tanıtarak, sistemin tasarım hedefleri açısından başarımlı deęerlendirilecektir.

3.1. Web Arayüzü

Sözlü Türkçe Derlemi Projesi Web Tabanlı Yönetim Sistemi proje boyunca el edilen tüm kayıt, çeviriyazı ve metaveriler ile bunlarla ilgili süreçleri yönetmek amacıyla proje kapsamında geliştirilmiş bir özelleşmiş içerik yönetim sistemidir. Sistem, projedeki işbölümünü organize etmek, kayıtların önceden belirlenmiş bir akışa göre işlenmesini sağlamak, derlemin gelişimini çeşitli deęişkenlere göre izlemek için kullanılabilir. Sözlü Türkçe Derlemi Projesi Web Tabanlı Yönetim Sistemi proje boyunca el edilen tüm kayıt, çeviriyazı ve metaveriler ile bunlarla ilgili süreçleri yönetmek amacıyla proje kapsamında geliştirilmiş bir özelleşmiş içerik yönetim sistemidir. Sistem, projedeki işbölümünü organize etmek, kayıtların önceden belirlenmiş bir akışa göre işlenmesini sağlamak, derlemin gelişimini çeşitli deęişkenlere göre izlemek için kullanılabilir. Sözlü Türkçe Derlemi Projesi Web Tabanlı Yönetim Sistemi proje boyunca el edilen tüm kayıt, çeviriyazı ve metaveriler ile bunlarla ilgili süreçleri yönetmek amacıyla proje kapsamında geliştirilmiş bir özelleşmiş içerik yönetim sistemidir. Sistem, projedeki işbölümünü organize etmek, kayıtların önceden belirlenmiş bir akışa göre işlenmesini sağlamak, derlemin gelişimini çeşitli deęişkenlere göre izlemek için kullanılabilir.

Sisteme giriş hakkı bulunan farklı rollere sahip kullanıcılar yetki ve görevlerine göre özelleşmiş bir arayüz üzerinde çalışırlar. Kullanıcıların kendilerine dağıtılan hesaplarla yaptıkları tüm giriş, çıkış ve kritik işlemler kayıt altına alınır. Sisteme eklenen tüm kayıtlar ve bunlarla ilgili çeviriyazılar, çeşitli aşamalara ayrılmış belirli bir iş akışından geçer. Örneğin, kayıtların sisteme yüklenmesi, çeviriyazılarının yapılması, kontrol edilmesi ve (gerekli ise) düzeltilmesi iş akışında tanımlı aşamalardan bazılarıdır. Her bir aşamayı sonlandıran eylem (örneğin çeviriyazının kontrol edilerek düzeltme için sisteme yüklenmesi) bir sonraki aşamanın sorumlusuna otomatik olarak bir e-posta gönderilmesine neden olur. Otomatik olarak gönderilen bu mesajlar yoluyla sorumlulara görevlerinin bildirilmesi iş akışının sürekliliğini sağlar.

Transcriber

Showing 1 to 10 of 17 entries

Search:

Stage	Name	Transcriber	Priority	Usability	Download	City of rec.
Not assigned	103_090813_00034 065_090813_733	Assign	1			Kırklareli
Not assigned	065_091003_00072 063_091003_655	Assign	2			Ankara
Not assigned	065_091005_00074 063_091005_656	Assign	1			Ankara
Not assigned	137_090329_00113 031_090329_297	Assign	2			Kayseri
Not assigned	086_100313_00135 067_100313_810	Assign	3			Ankara

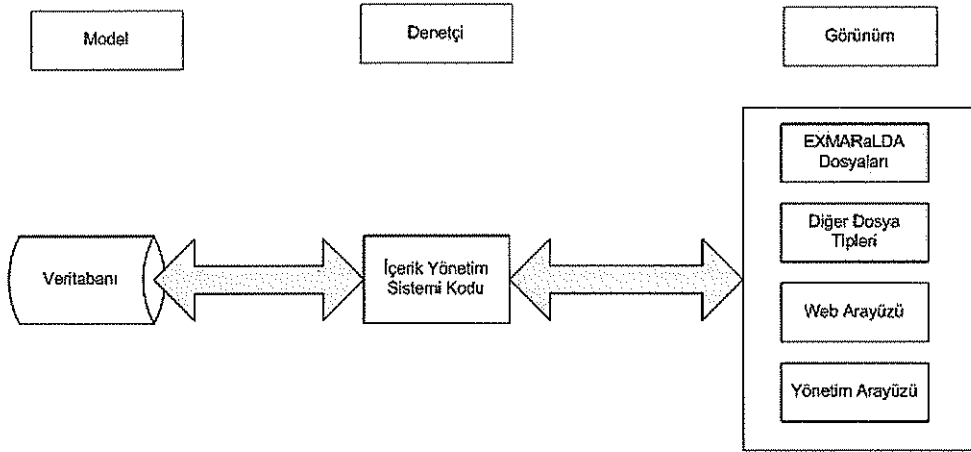
Şekil 4: Sözlü Türkçe Derlemi Projesi Yönetim Sistemi - Çeviriyazıcı atama arayüzü

Sistemin web üzerinden çalışması Linux, Windows, MacOSX gibi farklı işletim sistemlerini tercih eden kullanıcıların herhangi bir ek program kurmadan sistemi kullanabilmelerini, ayrıca, hata giderimi ve güncellemelerin kullanıcılardan bağımsız olarak yapılabilmesini sağlamaktadır.

Web üzerinden çalışan tüm sistemlerde olduğu gibi Sözlü Türkçe Derlemi Projesi Yönetim Sistemi de İnternet üzerinden gerçekleştirilebilecek saldırılar ile karşı karşıyadır. Sistemde bulunan kayıtların ve konuşmacılarla ilgili bilgilerin korunması amacıyla hem yazılım geliştirmede, hem de sunucu işletim sisteminde güvenlik açısından sağlam bir altyapı

(Debian Linux) tercih edilmiştir.

3.2. Model-View-Controller (MVC) Mimarisi



Şekil 5: MVC mimarisi

MVC (model-view-controller; model-görünüm-denetici), yazılım mühendisliğinde kullanılan bir mimari tipidir (Pawson, 2002). Sistemi model, görünüm ve denetici olarak üç ayrı öğeye ayırmaya dayanan bu mimari, kavramsal açıdan veri tipleri ve veritabanını (model), veri sunumunu (görünüm) ve veri işlemeyi (denetici) birbirinden ayırarak, bu öğeleri birbirinden bağımsız geliştirilebilmelerini sağlar. Bunun yan etkisi olarak, sistemdeki herhangi bir öğe, aynı cinsten ve uyumlu başka bir öğeyle, sistemi bozmadan değiştirilebilir. Mesela bir sistemdeki veritabanı, veri tipleri uyumlu başka bir veritabanıyla değiştirilebilir ve bu şekilde de sistem doğru çalışmaya devam eder. Böylece, sistemin eskiyen veya aksayan parçaları, sistem halen kullanımdayken değiştirilebilir.

MVC mimarisinin derlem sisteminin hedefleri açısından sundukları çok önemlidir. Öncelikle, sistemdeki tüm veri işleme tek bir merkezden (denetçi) yönetildiği için, sistemdeki tüm değişikliklerin sınırlanması ve kontrol edilmesi olanağı, mimarinin doğal bir uzantısıdır. Denetçide getirilecek kısıtlamalar sistemin tamamı için her şart altında geçerlidir. Buna bağlı olarak tüm işlemlerde sistem tutarlılığının kontrol edilmesi ve hangi kullanıcının neyi yapabileceğinin sınırlanması, yani kullanıcı rollerinin uygulanması da basitleşmektedir.

Mimarinin görünüm öğesi, kullanıcı arayüzünün kullanıcının rolüne göre şekillendirilmesini yine mimarinin bir yan ürünü haline getirmektedir. Bir kullanıcı arayüzü kullanmak istediğinde, denetçi, kullanıcı rollerini kontrol edip, sadece kullanıcıya ait rollerle ilgili arayüzü öğelerinin arayüze eklenmesini sağlar. Böylece her kullanıcıya, sadece yetkisi olan işlemlere ulaşma imkânı verilmiş olur. Görünüm öğesi aynı zamanda derlem çıktısının HTML dosyası, EXMARaLDA dosyası (Schmidt, 2004), düz metin dosyası, PRAAT dosyası gibi farklı biçimlerde ve dosya tiplerinde alınabilmesini ve zaman içerisinde istenen başka çıktı biçimlerinin sistemin geri kalanına dokunulmadan sisteme eklenebilmesini olanaklı kılar.

Mimarinin üçüncü öğesi olan model, veri tipleri ve veritabanının kendisinden oluşur. Sistemdeki metaveriler, konuşucu bilgileri, çeviriyazı bilgileri, işaretlemeler gibi her tür bilgi, sistemin kalanından bağımsız şekilde model kısmında bulunur. Bu model mevcut sürümde bir MySQL ilişkisel veritabanı olmakla beraber herhangi bir veri ortamı (obje veritabanı, XML dosyaları vb.) model olarak sisteme eklenebilir. Verilerin veri işleme ve görüntülemeye ayrı olması, verilerin kolayca birden fazla ortamda ve şekilde sunulabilmesini ve verinin üretilebilen herhangi bir dosya tipinden bağımsız kalmasını sağlar. Sözgelimi, yeni bir veri alanı eklenmesine karar verilirse, sistemin tutarlılığı değişmeden bu alan tüm ilgili yerlere eklenebilir ve/veya mevcut kayıtlar sistematik olarak değiştirilebilir.

3.3. Veritabanı Temelli Yaklaşım

Derlem yönetim sistemindeki model öğesi, bir MySQL veritabanında tutulmaktadır. Bütün kayıt ve konuşma bilgileri, çeviriyazılar, ilgili istatistikler ve süreç denetimine yönelik öğeler veritabanında tablolar halinde saklanır ve denetçinin talebine göre veri işleme ve veri sunumunda kullanılır. Bu yaklaşım, verinin dosyalardan okunup dosyalara yazıldığı dosya temelli yaklaşıma alternatiftir. Veriler gerektiği zaman ve gerektiği kadar işlenip, gerekli veri tiplerinde tekrar sunulabilir.

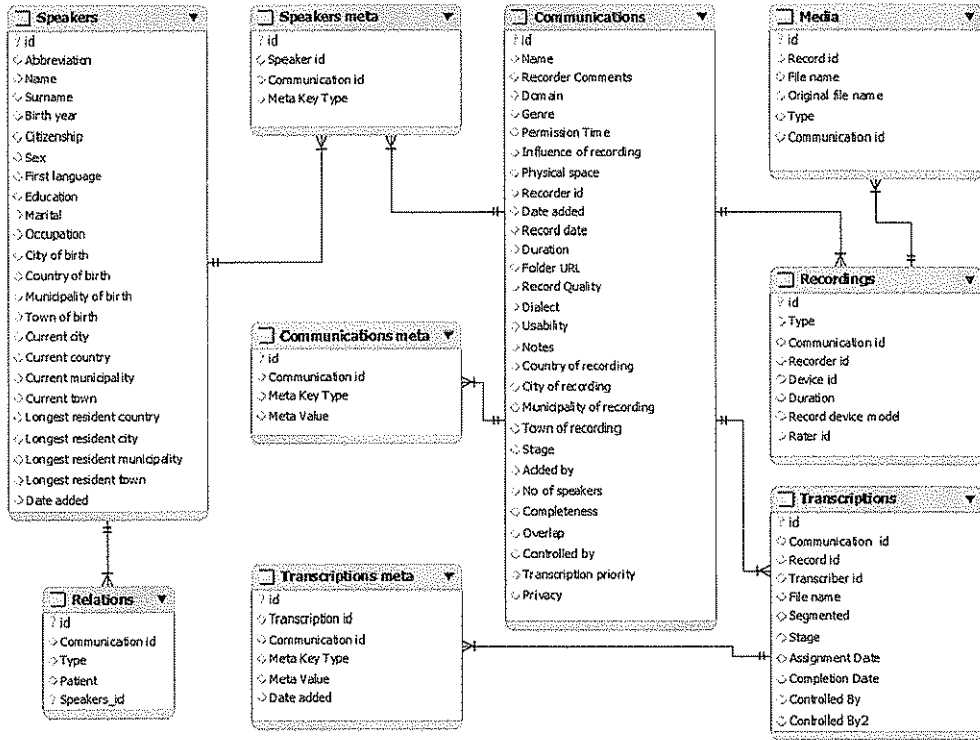
Bu yaklaşım, iş akışı denetiminin tutulan veriden ve veri biçiminden bağımsız olarak yapılabilmesine olanak sağlar. Veritabanları tutulan verinin tipi, izin verilen sınırlar içerisinde olup olmadığı gibi kontrolleri kolaylaştırır ve birçok kontrolü hazır olarak sunar. Buna bağlı olarak, tutarsız ve/veya eksik verinin girişi engellenebilir. Ayrıca ucuz veri girişinde (örn. konuşucu adı) daha önce girilen verilerin kullanıcıya seçenek olarak verimli bir şekilde sunulması da mümkündür.

Veritabanları tasarlanırken, aynı bilginin birden kez kaydedilmemesine özen gösterilir. Söz gelimi, bir çeviriyazı dosyasında her konuşucuya ilgili detayların bulunduğunu kabul edersek, konuşucuya ilgili detaylar bir dosya sisteminde on kez yer alacaktır. Bu da sistemin arama veriminin düşmesine ve verinin anlamı artmadan boyutunun artmasına sebep olur. Veritabanlarında ise birden fazla girdiye (bu örnekte kayıt girdisi) aynı konuşucu atanabilir, konuşucu detayları ise sadece bir kez tanımlanır. Aynı sebepten dolayı konuşucu detaylarındaki herhangi bir değişikliği tüm sistem için uygulamak kolaylaşır.

Veritabanlarının birçoğu son derece karmaşık veri yapılarını tutabilecek şekilde tasarlanmıştır. Bu yüzden, karmaşık ve sistematik değişiklikler dosya temelli yaklaşıma göre çok daha kolaydır. Veritabanı herhangi bir dosya şekline bağımlı olmayıp birçok veritabanı ortak matematiksel bir tabana sahip olduğu için, veritabanının eskiliğinden dolayı verilerin ulaşılamaz hale gelmesi söz konusu değildir. Benzer bir şekilde, yeni dosya tiplerinin yaratılması da sadece veritabanındaki yapının öğelerinin, yeni dosya tipindeki öğelerle eşleştirilmesine indirgenmiş olur.

Son olarak, veritabanları veri doğru şekilde girildiği sürece veritabanının büyüklüğü ne olursa olsun mümkün olduğunca verimli çalışmaları için tasarlandıkları için, metaveri, işaretleme veya çeviriyazı içinde arama yapmak dosya tabanlı sistemlere oranla çok daha hızlıdır ve gerçek zamanlı arama yapmaya daha uygundur.

Sözlü Türkçe Derlemi'nde, kayıtlarla ilgili metaverilerin saklanması için kullanılan veritabanı bölümüne ait şema aşağıda görülmektedir.



Şekil 6: Sözlü Türkçe Derleme Projesi veritabanı yapısından bir kesit

3.4. Dosya Üretimi ve Sürüm Yönetimi

Derlem yönetim sisteminde, dosyalar veriyi tutmak için değil sunmak için kullanılır. Dosyalar, her talep edildiklerinde en güncel veri kullanılarak tekrar üretilir ve sunulur. Dosya üretilirken, dosyaya giren veri kullanıcıya göre filtrelenebilir. Söz gelimi özel hayatın gizliliği adına konuşucu adları otomatik gizlenebilir veya proje ekibinin çeviriyazıcıya notları metaveriye eklenebilir.

Üretilen dosyalar SVN (Subversion) adlı sürüm yönetim sistemi ile ilk üretimden itibaren otomatik olarak kayda alınır. Sürüm yönetimi, dosyadaki her değişikliğin nerede, kim tarafından, ne zaman ve nasıl yapıldığını kayıt altına alır. Ayrıca sistem ve/veya kullanıcının değişiklik hakkında notlar düşmesine olanak sağlar. Böylece üretilen dosyalardaki bir değişiklik kaynağına kadar takip edilebilir ve hatalı üretimler daha kolay fark edilip düzeltilebilir veya eski haline döndürülebilir. Ayrıca hatalı üretimlerin çeviriyazıcılara göre dağılımı takip edilebilir.

Dosyaların otomatik üretilmesi, dosyaların yapılarında hata olması ihtimalini en aza indirger. Yapısal bir hata olsa dahi hatalar sistematik olacağı için fark edilmeleri ve düzeltilmeleri daha kolaydır. Dosyaların üretiliyor olması, yeni dosya tiplerinin eklenip çıkarılmasına sistemin kalanından bağımsız bir şekilde olanak tanır. İstenirse,

veritabanındaki tüm veri bir dosya sistemine aktarılarak tüm işlemler dosyalar üzerinden de yapılabilir (örn. EXMARaLDA araçları). Sistem, bunu kolaylaştırmak için dosyaları sistematik bir dizin yapısı ile üretir.

Derlemede çeviriyazıları tamamlanmış kayıtların altderlemeler halinde web üzerinden diğer araştırmacılarla paylaşılması veritabanı ve EXMARaLDA araçları kullanılarak otomatik olarak gerçekleştirilmektedir.

4. SONUÇ

Sözlü Türkçe Derleme Projesi kapsamında geliştirilen yazılımlar ve bu yazılımlarla tümleşik biçimde çalışan araçlardan oluşan Web Tabanlı Yönetim Sistemi, MVC gibi güncel yazılım geliştirme mimarileri ve AJAX gibi güncel web teknolojileri kullanılarak, sistemin gizlilik, güvenlik, sürdürülebilirlik, genişletilebilirlik gibi özel gereksinimleri dikkate alınarak tasarlanmış ve gerçekleştirilmiştir.

Veri girişlerinde insan kaynaklı hataları en aza indireyecek biçimde tasarlanan sistem, iş akışı özelliği sayesinde yüklenen kayıtların tüm süreçlerden sorunsuz ve eksiksiz biçimde geçmesini garantilemektedir. Sistem tarafından hesaplanan istatistikler ise derlem gelişiminin izlenmesi yoluyla dengeli ve temsil gücü yüksek bir derlem elde etmek açısından önemlidir.

Derlemin bu yazıda anlatılan geliştirmeye yönelik özelliklerine ek olarak ayrıntılı arama ve çözümlenmelerin yapılabileceği bir araştırmacı arayüzü de geliştirilme aşamasındadır.

Kaynakça

- Pawson, R. (2002). "Naked Objects", *IEEE Software*, 19(4), 81-83, doi:10.1109/MS.2002.1020291
- Schmidt, T. (2004). Transcribing and Annotating Spoken Language with EXMARaLDA. In Proceedings of the LREC-Workshop on XML based richly annotated corpora, Lisbon 2004, Paris: ELRA. http://www1.uni-hamburg.de/exmaralda/Daten/4D-Literatur/Paper_LREC.pdf

SÖZLÜ DERLEM ÇEVİRİYAZISINI STANDART DİL VE AĞIZA GÖRE ÖLÇÜNLEŞTİRME

Ciler HATİPOĞLU ve Özlem KARAKAŞ

Orta Doğu Teknik Üniversitesi

ciler.hatipoğlu@gmail.com, ozlm.krks@gmail.com

1. GİRİŞ

Bu çalışmanın üç ana amacı vardır:

- (1) Standart Türkçe ile ağız çalışmaları alanında yapılmış ve halen yapılmakta olan araştırmaları gözden geçirmek.
- (2) Büyük ölçekli derlemlerde (örn. BNC, ANC) sözlü örneklemelerin nasıl toplandığını ve konuşma dili ile ağızların bu derlemlerde nasıl temsil edildiğini belirtmek.
- (3) Ortam değiştikçe değişebilen (Ergenç 2002) ve ağızların sınıflandırılmasında da kullanılabilen (Demir 1996; Günşen 2006; Korkmaz 1976; Tulu 2008), -AcAk, -(I)yor, -(A/I)r ve -(y)A biçimbirimlerinin Sözlü Türkçe Derleminde (STD) nasıl temsil edildiğini ve ölçünleştirme aşamalarını anlatmak.

2. DİL

Bilindiği gibi dilin iki cephesi vardır: konuşma dili ve yazı dili. Konuşma dili 'bir ulusun, bir dil birliğinin dilinin yazıyla ilişkili olmayan ve çeşitli söyleyiş özellikleri taşıyan yönüdür ve bu nedenle ... temel olma niteliği taşır' (Aksan 1990:85). Yazı dili ise, 'sözlü bildirişim aracı dili görsel ve tek boyutlu bir düzen içinde sunan... bir dizgedir' (Ergenç 1995:11). Yazılı dil, yani alfabeler, sesleri anlatmak için kullanılan araçlardır. Bu yüzden bir dilin sesleri farklı alfabelerle yazıya aktarılabilir, 'bir ses için kimi zaman birden fazla simge kullanılabilir ya da bir ses hiçbir harf ile gösterilmeyebilir' (Hatipoğlu 2006:16). Türk dili ilk önce *Göktürk (Orhon)* daha sonra *Uygur ve Arap*, en son da *Latin* alfabeleriyle yazılmıştır. Örneğin, Göktürk alfabesinde 38 harf/işaret bulunmaktadır ve bazı ünsüzler için (örn. [b], [d], [g], [k], [l], [n], [r], [s], [t], [y]) ikişer harf oluşturulmuştur (Tekin 1988). Ancak, Türkçedeki 8 ünlü için Göktürk alfabesinde sadece 4 im kullanılmıştır (yani, [a ve e], [ı ve i], [o ve u], [ö ve ü] aynı harflerle temsil edilmiştir). 1 Kasım 1928'de '*Türk Yazı İnkılabı*' ile kabul edilen ve günümüzde de kullanılan Latin alfabesinde ise Türkiye Türkçesinin sesleri 29 harfle gösterilmektedir. Kullanılan bu harflerin 21'i ünsüzleri, 8'i de ünlüleri temsil etmektedir. Yeni alfabenin oluşturulması sırasında Türkçedeki seslerin sayı ve nitelikleri dikkate alınarak, Türk dilinin ses yapısına uygun bir alfabe yaratılması amaçlanmıştır (yani, *sesçil/fonetik bir alfabe* yaratılmıştır). Yeni alfabede, Türkçedeki her ses ayrı bir harf ile ve her harfin de sadece bir ses ile eşleştirilmesine dikkat edilmiştir. Bir dilin yazısı oluşturulurken genelde *yazı lehçesine* göre şekillenir. *Yazı lehçesi*, o dilin lehçeleri veya ağızları arasından seçilen bir lehçedir/ağızdır. Türkiye Türkçesi yazı lehçesi Batı Türk Dili'nin Anadolu lehçesidir (yani, 'Yeni Türkçe'de ses özellikleri ve çekim yönlerinden İstanbul ağzı esas alınır). Yeni alfabenin oluşturulmasında İstanbul Türkçesinin temel almasının iki önemli sebebi vardır: 1) İstanbul o dönemde Türkiye Cumhuriyetinin başkentidir ve okur-yazar nüfusunun ve aydınların çoğu orada yaşamaktadır; 2) Bununla birlikte İstanbul yazılı basının da merkezidir (Hatipoğlu 2006:16). Bu iki etken İstanbul ağzının kültür dili olarak kabul edilmesini ve yeni alfabe yazılırken esas alınmasını

sağlamıştır. Ancak İstanbul ağzı dışında Türkiye'de başka ağızlar da vardır (örn. Anadolu ağızları, Rumeli ağızları, Ercilasun 2000) ve onların yazı dilinde nasıl gösterildiği aşağıdaki bölümde kısaca anlatılacaktır.

3. AĞIZ ÇALIŞMALARI

'Ağız araştırmalarının dil çalışmalarında özel bir yeri vardır' çünkü bu çalışmalar bir taraftan dillerin daha gerçekçi bir şekilde betimlenmesine katkıda bulunurken diğer taraftan da dil atlasların oluşturulmasına katkı sağlar (Özsoy & Taylan 2000:i). Alanyazına bakıldığında, Türkçe ağızları üzerine yapılan çalışmaların hem uzun bir geçmişi olduğunu (örn. daha 1867 yılında A. Maksimov, Hüdavendigâr ve Karamanlı ağızları üzerine bir çalışma yapmıştır) hem de bu alanda bugüne dek çok sayıda araştırma yapıldığını görmek mümkündür. Bu çalışmaların sonucunda özellikle Anadolu ve Rumeli ağızları alt-alanının bir hayli ilerlemiş olduğu görülmektedir ve bu 'ağızların hemen hemen tamamından derlenmiş metinler vardır' (Ercilasun 2000:1). Bunların yanında Kara Deniz, Kıbrıs, Irak Türk ağızlarını inceleyen birçok çalışma bulunmaktadır.

Ancak her alanda olduğu gibi ağız çalışmalarında da bazı sorunlar mevcuttur. Bu sorunlar arasında, STD projesi çalışmaları sırasında en sık karşımıza çıkan sorun 'çeviriyazı sistemindeki çok seslilik' (Karahana 2000:23) olmuştur. Ağızlarda var olan sesler yazı dilindekinden farklı olabilir. Bölüm 1'de belirtildiği gibi günümüzde kullanılan Türkiye Türkçesi alfabesinde 29 harf bulunmaktadır, bunların 21'i ünsüzleri, 8'i de ünlüleri temsil etmektedir. Oysa Sağır, 1997 yılında yaptığı çalışmasında Türkiye Türkçesi ağızlarında insan kulağının ayırt edebildiği 41 ünlü ve 63 ünsüzün varlığından bahis eder. Kullanılan alfabe ağızlarda var olan sesleri karşılamaya yetmediğinden ağız çalışmaları yapan araştırmacılar bu 'yeni' sesleri gösterebilmek için farklı işaretler kullanmaktadırlar (Demiray 2008). 'Ancak işaretler konusundaki farklı tercihler, metinleri anlamayı zorlaştırmakta ve karışıklığa sebep olmaktadır' (Karahana 2001:24). Farklı ağızlar ile ilgili çalışmalar karşılaştırıldığında ya 'aynı işaretlerin farklı sesleri' göstermek için kullanıldıklarını ya da 'aynı sesin farklı işaretler' ile gösterildiğini açıkça ortaya çıkmaktadır (Delice 1995; Demiray 2008:203) (bkz. Tablo 3.1). Bu da her ağız bölgesi için ayrı çeviriyazı alfabesi kullanıldığını gösterir (Akar 2006).

Tablo 3.1

Kapalı a: a ile ı arasında söylenen ancak a'ya daha yakın olan bir ses

	Kaynak	Ağız	Simge
1	Korkmaz, Zeynep (1963)	Nevşehir ve Yöresi Ağızlar	á
2.	Buluç, Saadetin (1973)	Telafer Türkçesi Üzerine	ā
3.	Günay, Turgut (1978)	Rize İli Ağızları	á
4.	Gülensoy, Tuncer (1988)	Kütahya ve Yöresi Ağızlar	â
5.	Sinan, Ahmet Turan (1992)	Akçadağ, Doğanşehir ve Darende Ağızları	â

3. BÜYÜK ÖLÇEKLİ DERLEMLERDE STANDART KONUŞMA DİLİ VE AĞIZLARIN TEMSİLİ

Bu bölümde İngiliz Ulusal Derlemi (BNC) ve Amerikan Ulusal Derlemi (ANC) gibi büyük kapsamlı ve genel amaçlı derlemelerin ilk önce kısa bir tanıtımı yapıp daha sonra bu derlemlerde konuşma dilinin ve ağızların nasıl temsil edildiği anlatılacaktır.

3.1. İngiliz Ulusal Derlemi (BNC)

BNC, 1960 yılından sonra yazılan ve/veya konuşulan İngiliz İngilizcesini temsil etmek amacıyla ile çok çeşitli kaynaktan alınmış örneklerin bir araya getirilmesi ile oluşturulan 100 milyon sözcüklü bir derlemdir. BNC içinde 4124 örneklem bulunmaktadır. Bu örneklemelerin %90 yazılı, %10'u da konuşma dilinden alınmıştır. BNC'nin sözlü kısmındaki veriler iki aşamada toplanmıştır (Crowdy 1993). Veri toplamanın ilk aşamasında, projede yer alan araştırmacılar farklı bağlamlardaki dil kullanımını ortaya çıkarabilmek için İngiltere'nin 12 bölgesine dağılıp (i) eğitimsel ve bilgilendirici ortamlarda (örn. sempozyumlarda, okullarda) verilen sunumları ve dersleri, (ii) farklı türden radyo ve televizyon programlarını (örn. haber, tartışma ve söyleşi), (iii) resmi iş toplantılarını (örn. iş/danışmanlık görüşmeleri), (iv) resmi ve halka açık toplantıları (örn. vaizler, siyasi konuşmalar), (v) eğlenceli ortamda geçen konuşmaları (örn. spor karşılaşmaları sırasından yapılan yorumlar) sistematik bir şekilde kayıt etmişlerdir. Sözlü veri toplamanın ikinci aşamasında ise İngiltere'nin 38 farklı bölgesinde ikamet eden, dört farklı sosyo-ekonomik sınıfa ait olan, yaşları 15 ile 60+ arasında değişen, 124 kadın ve erkek gönüllü katılımcı yer almıştır. Bu gönüllü katılımcıların her birine taşınabilir ses kayıt cihazı verilip iki gün boyunca buldukları farklı ortamlarda yaptıkları konuşmaları kayıt etmeleri istenmiştir. Bu şekilde BNC'ye 2000 saatlik kayıt daha eklenmiştir.

BNC'nin büyüklüğünden ve oluşturma amacından dolayı (yani, derlemin içinde yer alan örneklemelerin İngiliz İngilizcesinin herhangi bir bilgi alanı, türü veya kesiti ile sınırlı olmamalı, ana dilleri İngiliz İngilizcesi olan insanların kullandıkları dilin tüm yönleri ile ortaya çıkarılmalıdır, Kennedy 1998) derlemin içinde standart konuşma dilinden farklılık gösteren kullanımlar ve değişik ağızlardan örnekler bulunmaktadır. Ancak Burnage ve Dunlop (1992) tarafından yayınlanan *İngiliz Ulusal Derleminin İşlenmesi* isimli makalede yazarlar, BNC projesinin ölçünlü İngilizceden sapan sesletim ve kullanışları belirtmek için yeterince kaynağa sahip olmadığının altını çizerek. Bunun yanında, ölçünlü dil ile bağdaşmayan sesletimlerin belirlenip, incelendikten sonra yazılı dile geçirilmesinin hem çok zor hem de öznel bir iş (yani, çeviri yazı yapanların kişisel yargılarını kullanmak zorunda oldukları bir işlem) olduğundan dolayı BNC bünyesine dâhil edilen konuşmaların tümünün ölçünlü yazıyla temsil edildiğini, ve özel bir şekilde sesletilen sözcük ve/veya ifadelerin sesbilgisel veya bürünsel özelliklerinin belirtilmediğini yazarlar. Bu kurallara tek istisnanın BNC'nin ağız ve bölgesel sesletim listesindeki sözcüklerin olduğunu belirtirler. Ağızlar ise BNC'de kodlar ile belirtilirler. Daha önce yapılan dilbilimsel çalışmalarda var oldukları ortaya çıkarılan ağızlar için kodlar belirlenmiş ve bunlar BNC'ye dahil edilen her konuşmanın başında belirtilmiştir. Örneğin *İskoçya ağızı* 'XSD' kodu ile, *Galler ağızı* 'XWA' kodu ile, İngiltere'nin İç kısımlarının kuzey doğu bölümlerinde kullanılan ağızlar için de 'XME' kodu kullanılmaktadır. Yani, BNC'nin %10'unu oluşturan sözlü bölümde konuşucular standart İngilizce dışında bir aksan veya bir ağız kullanmış olsalar bile bu özellikler çeviri yazıda gösterilmemiştir.

3.2. Amerikan Ulusal Derlemi (ANC)

ANC projesi 1998 yılında başlatılan, 1990 ve sonrasında kullanılan yazılı, sözlü ve elektronik Amerikan İngilizcesini temsil etmeyi amaçlayan, yapı ve örneklem olarak genellikle BNC'yi takip eden 100 milyon sözcüklü bir derlem projesidir. BNC'de olduğu gibi, ANC'nin %90'nını yazılı, %10'unu da konuşma örneklemi oluşturmaktadır. ANC'deki sözlü kısım dört kaynaktan gelen örneklemeler içermektedir (<http://www.americannationalcorpus.org/SecondRelease/contents.html>):

- (i) **Callhome** bölümü, ana dili Amerikan İngilizcesi olan konuşmacılar arasında gerçekleştirilen, her biri yaklaşık 10 dakika süren (50,494 sözcük içeren), 24 telefon konuşmasından oluşmaktadır.
- (ii) **Switchboard** bölümünde her biri yaklaşık 6 dakika süren 2320 (yaklaşık 3 milyon sözcük) doğal konuşma yer almaktadır. Bu konuşmalar her iki cinsin, farklı yaş (20-69 arasında değişmekte) ve eğitim düzeyi olan (eğitimsiz ile doktoralı arasında değişmekte) 500'ün üstünde Amerikan İngilizcesinin ana lehçelerini ana dilleri olarak konuşan (örn. Batı (Northern), Güney (Southern), Kuzeydoğu Amerika eyaletleri lehçesi (New England), New York Şehri Lehçesi) muhataplar arasında gerçekleştirilmektedir.
- (iii) **Charlotte Narratives and Conversation Collection (CNCC)** kısmı Mecklenburg İlçesi (Mecklenburg County), Kuzey Karolina (North Carolina) ve civarında ikamet eden toplulukların konuşmalarını temsil eden 95 anlatı, sohbet ve röportajdan oluşmaktadır. Konuşmacılar ile ilgili bilgiler her çeviri yazının üstbilgi kısmında verilmiştir.
- (iv) **Micase**'de Michigan Akademik Sözlü İngilizce Derlemi'nden (Michigan Corpus of Academic Spoken English) alınmış 50 çeviri yazı bulunmaktadır. Konuşmacıların yaşı, cinsiyeti ve rolü/görevi ile ilgili bilgiler her çeviri yazının üstbilgi kısmında verilmiştir.

ANC'de lehçelerin temsil edilmesinde BNC'de kullanılan stratejiler benimsenmiş. Yani, her konuşmanın başında konuşma ile ilgili veri bilgisi verilmektedir ve bu metabilgi içinde konuşucuların kullandıkları ağız ve aksanlar da belirtilmektedir ancak ağız farklılıkları çeviri yazıya yansıtılmamaktadır. Örneğin North Midland (Ağız: orta eyaletlerin kuzey kısımlarında konuşulan) lehçesi kullanılan bir sohbet için lehçe bilgisi çeviri yazının üstbilgi kısmında verilir ancak bu lehçenin özellikleri çeviri yazıya yansıtılmaz.

4. SÖZLÜ TÜRKÇE DERLEMİ (STD)

Derlemler yapılandırılmış metinler bütünüdür. Onlar bir dildeki farklı sözlü ve/veya yazılı metinlerin elektronik ortamda bir araya getirilmesi ve genellikle bilgisayar yardımıyla işlenmesi sonucunda oluşturulan veritabanlarıdır. STD, Türkiye'nin çeşitli bölgelerinde, farklı ortamlarda ve farklı yaş, cinsiyet, eğitim seviyesi ve iş gruplarına dâhil konuşmacılar tarafından yüz yüze ya da çeşitli iletişim araçlarıyla (örn. telefon/skype konuşmaları) gerçekleştirilen Türkçe konuşmaların bir araya getirilmesi ile oluşturulan bir veri tabanıdır (STD aresi: <http://std.metu.edu.tr/>). STD'nin amacı günümüz Türkçesinin bilgisayar ortamında izlenmesini ve taranmasını sağlayacak bilimsel bir kaynak oluşturmaktır. STD'nin daha önce Bölüm 3'te bahis edilen büyük ölçekli derlemlerden önemli bir farklı vardır, bu da derlemin tamamının sözlü metinlerden oluşmasıdır oysa ANC ve BNC derlemlerinin sadece %10 sözlüdür. Bu açıdan STD hem Türkçe adına yapılan çalışmalar arasında hem de diğer dünya dilleri üzerine yapılan derlem çalışmaları arasında özel bir yere sahiptir.

STD'nin oluşturulması aşamasında iki önemli konu dikkate alınmıştır:

- (1) Derleme dâhil edilen çok sayıdaki çeviriyazı arasında tutarlılık olmalıdır. Bundan dolayı Bölüm 2'de anlatılan ve ağız çalışmalarında kullanılan 'her ağız bölgesi için ayrı çeviriyazı alfabesi' yöntemi yerine derleme dahil edilen tüm çeviriyazılarda kullanılacak 'ortak çeviriyazı kurallarının' belirlenmesi gerekmektedir;
- (2) 'Tutarlı' çeviriyazı kurallarını oluştururken derlemelerin, genellikle geniş bir araştırmacı kitlesi tarafından kullanıldığını ve daha önceden belirli bir alanda eğimi olmayan kullanıcıların da derlemeden yararlanmak isteyebileceğinin göz önünde bulundurularak kurallar dizinin oluşturulması gerekmektedir.

Bu sebeplerden dolayı da STD'nin oluşturulması ve geliştirilmesi aşamasında Türkiye Türkçesi üzerine yapılan sesbilim, sesbilgisi ve ağız çalışmaları taranmış, ilgili edebi eserler gözden geçirilmiş (örn. Haldun Taner'in *Gözlerimi Kaparım Vazifemi Yaparım*) ve Türkçe'de sıkça kullanılan, ağızların sınıflandırılmasında da ölçüt olarak alınabilen (Demir 1996; Günşen 2006; Tulu 2008), ancak ortam değiştikçe değişebilen (Ergenç 2002) **-AcAk**, **-(I)yor**, **-(A/I)r** ve **-(y)A** biçimbirimlerinin STD'de nasıl temsil edilmesi gerektiği konusunda aşağıda sıralanan kurallar belirlenmiştir.

STD çeviriyazılarının oluşturulmasında EXMARaLDA çeviriyazı programı kullanılmıştır (http://www.exmaralda.org/en_index.html). EXMARaLDA çeviriyazı programında, konuşmacılar tarafından üretilen dilin çeviriyazıya daha ayrıntılı bir şekilde yansıtılabilmesi için konuşma zamanına/sırasına bağlantılı, farklı seviyede satırlar oluşturulmaktadır. STD'nin kullandığı sistemde, konuşmacıların söylediklerini standart/ölçünlü dildeki şekli ile yansıtan satır **ÇEVİRİYAZI SATIRI**'dir (**verbal (v)**, bkz. Örnek 4.1). Sesbilim/Sesbilgisel kurallar ile açıklanamayan standart sesletim dışına çıkan özel/kişisel sesletimlerinin, daha önce yapılan ağız çalışmalarında belirli bir bölgede/topluluk tarafından kullanıldığı tespit edilen sesletim biçimlerinin, konuşmacıların herhangi biri tarafından kullanılan ve konuşmayı etkileyen konuşma dışındaki seslerin/ses olaylarının (örn. gülme, kekeleme) ve/veya konuşmanın dikkat çeken dilötesi özelliklerinin gösterildiği satır **AÇIKLAMA SATIRI**'dir (**comment (c)**, bkz. Örnek 4.1).

Örnek 4.1.: (STD Derleme Dosya No: 012_090128_00002)

52 [00:54.9]

BUR000030 [v]	geleceğim.	ÇEVİRİYAZI SATIRI (VERBAL TIER)
BUR000030 [c]	gelicem	AÇIKLAMA SATIRI (COMMENT TIER)

4.1. [-(I)yor] Biçimbirimi

Sesbilim ve sesbilgisi alanında yapılan çalışmalar tarandıktan sonra [-(I)yor] biçimbiriminin STD'de temsil edilmesi ile ilgili aşağıdaki iki kurların takip edilmesi kararlaştırılmıştır:

Kural 1

Türkçede ölçünlü dil kullanımında bile

i) [-(I)yor + 2./3. şahıs ekleri (bkz. Tablo 4.1. örnek 1 ve 2)

ii) [-(I)yor + DIK (bkz. Tablo 4.1. örnek 3 ve 4)

yapılarında [-(I)yor] ekinin sonsesinde bulunan /r/ sesinin düşürüldüğü/söylenilmediği ortaya çıkmıştır (Ergenç 1995). Dilbilim alanında yapılan çalışmalar bir sözcük içinde yer alan seslerden birinin farklı sebeplerden dolayı 'atılabileceğini' göstermiştir. Bu nedenler

arasında başta geleni, 'dilde kolay söyleyişe olan eğilimdir' (Aksan 1978: 44). Türkiye Türkçesinde, özellikle konuşma sırasında, [-(I)yor] ekinin sonunda bulunan /r/ ünsüzü, sözcük içinde bulunduğu (örn. **söylüyoRlar** gibi durumlarda) söyleyişi kolaylaştırma amacı ile düşürülür. Aksan (1978), bu gibi ünsüz düşmelerine Türkiye Türkçesinde sıkça rastlandığını belirtir. Bu sebepten dolayı STD için geliştirilen sistemde çeviriyazı satırına(yani, v satırına) sadece standart/ölçünlü dilde kullanılan yazı kullanılacaktır. Açıklama satırına [-(I)yor] ekinin 'r-siz' sesletimi yazılmayacaktır.

Tablo 4.1

-Iyor + 2./3. şahıs ekleri + -Iyor + DIK Örnekler

Standart yazı	Söylenen/Sesletilen
(1) geliyorsun(uz)	geliyosun(uz)
(2) geliyor(lar)	geliyo(lar)
(3) geliyordum	geliyodum
(4) geliyordu	geliyodu

Ancak Türkiye'nin bazı bölgelerinde kullanılan ağızlarında ölçünlü dilden farklı kuralların geçerli olduğu belirlenmiştir. Nevşehir ve civarında yaptığı incelemelerde Korkmaz (2007:500), şimdiki zaman eki [-(I)yor]'un bazen sadece "r" sesinin düşürüldüğünü, bazen de (Derinkuyu'nun Çakıllı, Doğala ve Avanos'un Genezin köyü ağızlarında olduğu gibi) 'y' erimesi ve hece kaynaşması yoluyla, ekin **-or** veya **o'ya**' dönüştüğünü ortaya çıkarmıştır (bkz. Tablo 4.2). Karakeçili ağızları üzerine yaptıkları çalışmalarında, İlhan ve Yavuz (2006) da şimdiki zaman eki [-(I)yor]'un "r" ünsüzünün düşmesi sonucu bölgede konuşulan ağızda oluşan ünlü uzunluğunun bölgenin 'en karakteristik özelliği' olduğunu belirtirler (bkz. Tablo 4.2). Bu gibi durumlar için STD siteminde Kural 2'nin takip edilmesine karar verilmiştir.

Tablo 4.2

Sesbilgisi kuralları ile açıklanamayan sesletimler/Bölgesel sesletim biçimleri örnekleri

Standart yazı (Çeviriyazı satırı)	Söylenen/Sesletilen (Açıklama satırı)	Ağız
(1) geliyorum	geliyom/ gelyom /celiyorum	
(2) geliyor	geliy	
(3) bilmiyorum	bilmiyom	
(4) ne yapıyorsunuz?	napıyoñuz?	
(5) biz geliyoruz	biz gediyoñ	Nevşehir ve yöresi ağızları (Korkmaz 2007:500)
(6) niye almıyorsun	niye almõñ	
(7) durmuyor dediler	durmõr dediler	
(8) bazen biniyor	bazım biniyõ	
(9) kutuluyor	gurtuluyõ	Karakeçili Ağızı (İlhan ve Yavuz 2006:116)
(10) tutuyor	dutyõ	

Kural 2

Sesbilim/Sesbilgisi kuralları ile açıklanamayan ve standart sesletim dışına çıkan (örn. bir biçimbirim düşürülmesi), ve/veya daha önce yapılan ağız çalışmalarında bir bölgede kullanıldığı ortaya çıkan sesletim biçimleri açıklama satırında yer alacaklardır (bkz. Tablo 4.2). Çeviriyazı satırına sadece standart/ölçünlü dilde kullanılan yazı kullanılacaktır (bkz. Örnek 4.2).

Örnek 4.2.: (STD Derlemi Dosya No: 012_090128_00002)

218 [03:56.8]

RUK000029 [v]	diyorum	ÇEVİRİYAZI SATIRI
RUK000029 [c]	diyom	AÇIKLAMA SATIRI

4.2. [-AcAk] Biçimbirimi

[-AcAk] eki Türkçede ilk önce 13. yüzyılda bir sıfatfiil olarak kullanılmaya başlanmıştır. Bugünkü anlamları olan gelecek zaman eki olarak kullanımına ise ilk önce 14. yüzyılın ikinci yarısında rastlanmaktadır ve bu dönemde kullanımı genellikle sadece 3. kişilerle sınırlıdır (Demir & Yılmaz 2006:101). Bu dönemden sonra ek halk tarafından benimsenmiş ve kullanımı konuşma dilinde gittikçe yaygınlaşmıştır, fakat [-AcAk] edebi üslup tercih edenler tarafından uzunca bir süre kaba bulunduğundan yazı dilindeki yerini ancak 17. yüzyılda almıştır (Adamović 1985:92-112).

Bugün Türkiye Türkçesinin ölçünlü dilinde, eylem köküne gelecek zaman biçimbirimi [-AcAk] eklendiğinde köklerin sesletiminde kimi değişimler yaşanmaktadır. /y/ka, bekle/ örneklerinde olduğu gibi, eylem kökü ünlüyle bitiyorsa kök ve ardıl ünlüleri arasındaki ünlü çatışmasının önüne geçilebilmesi için araya /y/ geçiş sesi getirilir. Oluşturulan yapıdaki yeni /y/ sesinin daraltıcı özelliğinden dolayı eylem kökündeki geniş ünlü daralır ve /yıkayacak, bekleyecek/ sözcükleri /yiki:cak, bekle:cek/ olarak sesletilir (Ergenç 1995). [-AcAk] biçimbirimi /gel-, al-/ gibi ünsüz ile biten eylem köklerine eklendiğinde koruyucu ünsüzü /y/'ye gereksinim duyulmaz. Ancak bu yapılarda [-AcAk] ekinin ilk ünlüsü daraltılarak sesletilir. Yani, ölçünlü dil kullanıcıları bile [-AcAk] biçimbirimiyle oluşturulan çekimli eylemlerinin 1. ünlüsünü 'i' olarak sesletirler (bkz. Tablo 4.3).

Tablo 4.3

-AcAk Biçimbiriminin sesletimi ile ilgili örnekler

Standart yazı	Söylenen/Sesletilen
(1) geleceksin	geliceksin
(2) gelecek	gelicek
(3) geleceksiniz	geliceksiniz

Kural 1

Çeviriyazı satırına sadece standart yazı dilinde kullanılan yapılar kullanılacaktır (yani, gelecek, alacağım). Açıklama satırında 1. ünlüsü (uzun) 'i' olarak sesletilen çeşitler belirtilmeyecektir.

Kural 2

Yukarıda sıralanan sesbilim/sesbilgisel kuralları ile açıklanamayan ve standart sesletimin dışına çıkan (bkz. Tablo 4.4), ve/veya daha önce yapılan ağız çalışmalarında belirli bir topluluk tarafından kullanıldığı tespit edilen sesletim biçimleri açıklama satırında belirtilecektir.

Tablo 4.4

Sesbilgisi kuralları ile açıklanamayan sesletimler/Bölgesel sesletim biçimleri örnekleri

Çeviriyazı satırı	Açıklama satırı
(1) geleceğim	gelecem/gelicem/ gelcem/ geleceem/gelceem
(2) geleceksiniz	gelceksiniz ('e' sesletilmemiştir)/ gelceeniz ('e' ünlüsü uzun sesletildiğinde 'ee')
(3) geleceksin	gelicen/gelcen/gelceyn
(4) gelecek	gelcek ('e' sesletilmemiş)
(5) gelecekmış	geleceemiş ('k' sesletilmemiş, 'e' uzun sesletilmiş)
(6) geleceğiz	gelicez/ gelcez

4.3. [-(I)r/(A)r] Biçimbirimi

[-(I)r/(A)r] eki/ekleri hem zaman hem de şekil ifade eden eklerdendir. Zaman eki olarak geniş zamanı, şekil eki olarak da bildirme ifade ederler (Ergin 1993). Türkçede bugün iki geniş zaman eki vardır: [-(A)r] ve [-(I)r]. Bunlardan ilki, /l/ ve /r/ sesi ile bitenler hariç, tüm tek heceli ve bazı çok heceli fiillerin sonunda kullanılır (örn. ara-r, koru-r). Geniş zaman eklerinden ikincisi ise çok heceli (örn. aktar-ır), türetilmiş tek heceli (örn. yen-ir), ve /l/ ve /r/ sesleri ile biten tek heceli (örn. ver-ir) fiiller ile kullanılır (Yavuz ve diğ. 2004). Ölçümlü Türkçede de bu geniş zaman eklerinin kullanılması beklenmektedir.

Oysa Türkiye'nin farklı bölgelerinde yapılan çalışmalarda bazı ağızlarda geniş zaman ekinin "r" ünsüzünün sesletilmediği veya başka sesler ile değiştirildiği ortaya çıkarılmıştır. Örneğin, Karaman İli ağızları üzerine yaptıkları incelemelerde, Karpuz ve Uysal (2010), geniş zaman ekinin kullanımı ile ilgili iki önemli bulgu elde etmişlerdir:

- çekimli bazı fiillerde fiil kök veya gövdesinde yer alan /r/ sesinden sonra gelen zaman eki, ses uyumsuzluğu nedeniyle /y/ sesine dönüştürülür (örn. "kız evine varırız" yerine "göz evine varıyız", "sütlü tarhanayı da pişiririz" yerine "sütlü tahranayı da bişiriyiz");
- kimi örneklerde ise geniş zaman eki tamamen düşürülür (örn. "azığını yanına alırsın" yerine "azzını yanına alın", "ineğin varsa inek sağarsın" yerine "ineñ varsa inek sañ", Karpuz ve Uysal 2010:8).

Nevşehir ve yöresindeki köylerde yaptığı çalışmalarda Korkmaz (2007) geniş zaman çekiminde olumlu ve olumsuz teklik 2. şahıs çekimi ile 1. şahıs çokluk çekiminin yazı dilindekinden farklı olduğunu belirtir. Bu yörede "gidersin" yerine "gideñ, getmeñ", "okursun/okumazsın" yerine de "ohuñ/ohumañ" kullanıldığının altını çizer.

Türkiye'nin Kara Deniz bölgesinde kullanılan Trabzon ağızları üzerine yaptığı çalışmalarda sonra Brendemoen (2000) bu bölgede geniş zaman ekinin dağılımının tutarsız olduğunu sonucuna varır. Bayburt İl sınırına yakın olan bazı yerlerde "bilmek" eyleminin geniş zamanının sürekli "bilir" şeklinde kullanıldığını ancak diğer dar ünlülü geniş zaman biçimlerinin ise hep yuvarlak ünlüyle sesletildiğini ortaya çıkarır (örn. gelur, verur) (Brendemoen 2000:38).

Ölçünlü Türkçe ve ağızlar arasında gözlemlenen bu farklılıklardan dolayı geniş zaman eklerinin STD sisteminde temsili ile aşağıda belirtilen kural benimsenmiştir:

Kural

Çeviriyazı satırına standart sesletim yazılırken, açıklama satırına duyulan sesletim yazılacaktır.

Tablo 4.5

[-(Dr)/(Ar)] biçimbiriminin sesbilgisi kuralları ile açıklanamayan bazı sesletimleri

<u>Çeviriyazı satırı</u>	<u>Açıklama satırı</u>
(1) yapar	yapaa
(2) koşar	koşaa
(3) gidersin	gideñ
(4) varırız	varıyız

4.4. İstek kipi [-Allm, -Ayım, -A]

Türkçe gramer kitaplarında istek kipinin kişilere göre belirlenen çekimleri Tablo 4.6'da gibi verilmektedir ve genellikle 1. tekil ekinin statüsü tartışılmaktadır. Türkçede 1. tekil emir ve istek kiplerinin formlarının da aynı olmasından dolayı 'otur-ayım, gel-eyim, başla-yayım' gibi eylemlerin sonundaki ekin emir kipi mi yoksa istek kipi olarak mı sınıflandırması konusunda tartışmalar devam etmektedir. Bazı araştırmacılar birinci kişi kendi kendine emredemez düşüncesinden hareketle [-Allm, -Ayım, -A] ekinin istek kipi başlığı altında incelenmesinin doğru olduğunu iddia etmektedirler. Başka bir grup araştırmacıya göre ise birinci şahıs da kendi kendisine emredebilir ve bu sebepten dolayı da [-Allm, -Ayım, -A] ekleri istek görevinde kullanılsalar dahi aslında emir ekidirler (Ergin 1993). Belki de bu sebepten dolayı Demircan (1977:110) istek kipi gibi kip eklerinin eylemin ne gibi bir ruh durumu ile ilgili olduğunu anlatan ekler olduğunu altını çizmektedir.

Tablo 4.6

İstek kipinin çekimi

	<u>OTUR</u>	<u>GEL</u>		<u>OTUR</u>	<u>GEL</u>
1. tekil	otur-ayım	gel-eyim	1. çokluk	otur-alım	gel-elim
2. tekil	otur-asın	gel-esin	2. çokluk	otur-asımız	gel-esiniz
3. tekil	otur-a	gel-e	3. çokluk	otur-alar	gel-elər

İstek kipi biçimbirimi [-Allm]'ın günlük hayattaki kullanımını incelendiğinde hem sesbilgisel hem de biçimbirimsel açıdan ilginç değişimlere uğradığı görülmektedir. Ölçünlü Türkiye Türkçesi konuşanların bile [-Allm] ekinin ilk ünlüsünü genellikle *i/i* olarak (örn. **veriyim, ahıym**) seslettikleri ortaya çıkarılmıştır (Ergenç 1995). Bu sesletimin sebebi olarak da /y/ sesinin daraltıcı etkisi gösterilmektedir. Ergenç (1995:39), /y/ sesinin etkisiyle eylemsiyle eylem kökünde bulunan geniş ünlünün daraldığını, bu durumun da sesletimi etkileyerek değiştirdiğinin altını çizer. Bunun yanında istek kipinin 1. kişi ile ilgili olan ve gramer kitaplarında sıralana ekleri bazı ağızlarda kullanılmamaktadır. Belki de tarihsel gelişmelerin de etkisiyle bazı ağızlarda 1. tekil ve çokluk biçimbirimleri "gelelim, oynayalım" yerine "gelem, oynayam, gelek, oynayak" olarak kullanılmaktadır (Demir & Yılmaz 2006:193-194).

Kural 1

Çeviriyazı satırına sadece standart/ölçünlü dilde kullanılan sesletim belirtilecektir. Açıklama satırında 1. ünlüsünü 'i/i' olarak sesletilen varyasyonlar belirtilmeyecektir.

Kural 2

Çeviriyazı satırına sadece standart/ölçünlü dilde kullanılan sesletim belirtilecektir. Açıklama satırında ağızlarda kullanılan ve sesbilgisel kurallar ile açıklanamayan sesletimler belirtilecektir (bkz. Tablo 4.7).

Tablo 4.7

[-Allm] biçimbiriminin sesbilgisi kuralları ile açıklanamayan bazı sesletimleri

<u>Çeviriyazı satırı</u>	<u>Açıklama satırı</u>
(1) gidelim	gidek
(2) yapalım	yapak
(3) alayım	alam

5. SONUÇ

Bu çalışmada STD projesi ekibinin sözlü dili yazıya geçirirken yaşadığı zorlukları, geçilen aşamaları ve şu anda varılan sonuçları ve alınan kararlar anlatılmıştır. STD projesi çerçevesi içinde yapılan çalışmalarda, Türkçe gibi birçok ağza sahip dillerin ölçünleştirilmesinin ve çeviriyazıda temsil edilmesinin ne kadar zor olduğu ortaya çıkmıştır. BNC ve ANC gibi büyük ölçekli derlemlerde bile ağızların temsil edilmesine teşebbüs edilmezken, STD projesi ekibinin yaptığı çalışmalar sonucunda STD'den sadece sözdizimi, biçimbilimi ve sözcükbilimi çalışanları değil ağız çalışmaları, anlambilimi, sesbilim ve sesbilgisi, toplumbilim, yabancı dil eğitimi (örn. Yabancı Dil olarak Türkçe) gibi alanlarda Türkiye'de ve Türkiye dışında çalışan dilbilimcilerin de yararlanabileceği bir kaynak oluşturulmuştur. Bunun yanında çeviriyazı kurallarının mümkün olduğu kadar basit/anlaşılır tutulması sayesinde de dilbilim eğitimi almamış ancak insan ve toplum olguları ile ilgilenen eğitim bilimleri, sosyoloji, psikoloji ve mühendislik alanlarında çalışan bilim insanlarının da STD'den sıkça yararlanacağı kanaatindeyiz. STD'yi kullanmak isteyen ve çeviriyazı kurallarını daha yakından incelemek isteyen uzmanlar Ruhi ve diğ. (2010) kaynağından yararlanabilirler.

Kaynakça

- Adamović, Milan. (1985). *Konjugationsgeschichte der türkischen Sprache*. Leiden: E. J. Brill.
- Akar, Ali. (2006). Ağız araştırmalarında yöntem sorunları. *Turkish Studies /Türkoloji Dergisi*, 1 (2), 41-53.
- Aksan, Doğan. (1978). *Türkiye Türkçesi Gelişmeli Sesbilimi*. Ankara: Türk Dil Kurumu Yayınları 451.
- Aksan, Doğan. (1990). *Her Yönü İle Dil: Ana Çizgileriyle Dilbilim I*. Ankara: Türk Dil Kurumu Yayınları.
- American National Corpus: Second Release. (2010). Retrieved from <http://www.americannationalcorpus.org/SecondRelease/contents.html>
- Brendemoen, Bernt. (2000). Trabzon ağızlarındaki ünlü uyumu ve yönetsel açıdan düşündürdükleri. A. Sumru Özsoy ve Eser E. Taylan (Haz.), *Türkçe'nin Ağızları: Çalıştay Bildirileri* (s. 35-46). İstanbul: Boğaziçi Üniversitesi Yayınları.
- Burnage, Gavin & Dunlop, Dominic. (1992). Encoding the British National Corpus. In Jan Aart, Peter de Haan and Nelleke Oostdijk (Eds.), *English Language Corpora: Design, Analysis and Exploitation: Papers from the 13th International Conference on English Language Research on Computerised Corpora*. Nijmegen: Editions Rodopi.
- Crowdy, Steve. (1993). Spoken corpus design. *Literary and Linguistic Computing*, 8 (4), 259-265.
- Delice, H. İbrahim. (1995). Anadolu ve Rumeli ağızları metinlerinde kullanılan seslerin transkripsiyonu meselesi. *Türklük Bilimi Araştırmaları*, 1, 119-131.
- Demir, Nurettin. (1996). *Einige Merkmale der yörükischer Dialekte, Symbolae Turkologicae, Studies of Honour Lars Johanson*. İstanbul.

- Demir, Nurettin & Yılmaz, Emine. (2006). *Türk Dili: El Kitabı (3. Baskı)*. Ankara: Grafiker Yayıncılık.
- Demiray, Erdiç. (2008). Ağız araştırmalarında transkripsiyon işaretlerinin farklı kullanılması sorunu. *Turkish Studies*, 3 (6), 199-211.
- Demircan, Ömer. (1977). *Türkiye Türkçesinde Kök-Ek Bileşmeleri*. Ankara: Türk Dil Kurumu Yayınları 436.
- Ercilasun, Ahmet B. (2000). Ağız çalışmalarına toplu bir bakış. A. Sumru Özsoy ve Eser E. Taylan (Haz.), *Türkçe'nin Ağızları: Çalıştay Bildirileri* (s. 1-4). İstanbul: Boğaziçi Üniversitesi Yayınları.
- Ergenç, İclal. (1995). *Konuşma Dili ve Türkçenin Söyleyiş Sözlüğü: Bir Deneme*. Ankara: Simurg.
- Ergenç, İclal. (2002). *Konuşma Dili ve Türkçenin Söyleyiş Sözlüğü*. İstanbul: Multilingual Yayınları.
- Ergin, Muharrem. (1993). *Türk Dil Bilgisi*. İstanbul: Bayrak Basın/Yayın/Tanıtım.
- Gülensoy, Tuncer. (1988). *Kütahya ve Yöresi Ağızları*. Ankara: Türk Dil Kurumu Yayınları 336.
- Günay, Turgut. (1978). *Rize İli Ağızları*. Ankara: Kültür Bakanlığı Yayınları.
- Günşen, Ahmet. (2006). Anadolu ağızlarında farklı bir gelecek zaman eki ve çekimi: -ıcı / -ici; -ucu / -ücü. *International Periodical for the Languages, Literature and History of Turkish or Turkic*, 1 (2), 54-84.
- Hatipoğlu, Çiler. (2006). Sesbilgisi ve Sesbilim. Ahmet Kocaman (Haz.), *Dilbilim: Temel Kavramlar Sorunlar Tartışmalar* (s. 15-42). Ankara: Dil Derneği Yayınları.
- İlhan, Nadir & Yavuz, Serdar. (2006). Karakeçililer ve Karakeçili ağızında ikinci uzun ünlüler. *Doğu Anadolu Bölgesi Araştırmaları*, 114-117.
- Karahan, Leyla. (2000). Ağız araştırmalarında sorunlar. A. Sumru Özsoy ve Eser E. Taylan (Haz.), *Türkçe'nin Ağızları: Çalıştay Bildirileri* (s. 21-26). İstanbul: Boğaziçi Üniversitesi Yayınları.
- Karpuz, H. Ömer & Uysal, İdris Nebi. (2010). Karaman İli ağızlarının özellikleri ve sınıflandırılması. *Türkiyat Araştırmaları Dergisi*, 27, 1-17.
- Kennedy, Graeme. (1998). *An Introduction to Corpus Linguistics*. London & New York: Longman.
- Korkmaz, Zeynep. (1976). Anadolu ağızları üzerindeki araştırmaların bugünkü durumu ve karşılaştığı sorunlar. S. Buluç, H. Eren ve A. Dilaçar (Haz.), *Türk Dili Araştırmaları Belleten 1975-1976* (s. 143-172). Ankara: TDK.
- Korkmaz, Zeynep. (1963). *Nevşehir ve Yöresi Ağızları*. Ankara: Ankara Üniversitesi Yayınları.
- Korkmaz, Zeynep. (2007). Nevşehir ve yöresi ağızlarının şekil bilgisini belirleyen temel özellikleri üzerine. *Turkish Studies/Türkoloji Araştırmaları*, 2 (2), 492-501.
- Özsoy, Sumru & Taylan, Eser E. (Haz.). (2000). *Türkçe'nin Ağızları: Çalıştay Bildirileri*. İstanbul: Boğaziçi Üniversitesi Yayınları.
- Ruhi, Ş., Hatipoğlu, Ç., Eröz-Tuğa, B., Işık-Güler, H. (2010). *A guideline for transcribing conversations for the construction of spoken Turkish corpora using EXMARaLDA and HIAT*. Ankara: Setmer Publishing.
- Sağır, Mukim. (1999). Ağız Araştırmalarında Çevriyazı. *Ağız Araştırmaları Bilgi Şöleni* (9 Mayıs 1997) (s. 126-138). Ankara: TDK Yayınları 697.
- Sinan, Ahmet Turan. (1992). *Akçadağ, Doğanşehir ve Darende Ağızları: Fonetik ve Morfolojik İnceleme*. MA Thesis, Fırat University, Elazığ.
- TDK. (1945). *Türk Dialektleri Çevriyazı Sistemi*. İstanbul: TDK.
- Tekin, Talat. (1988). *Orhon Yazıtları*. Ankara: Türk Dil Kurumu Yayınları.
- Tulu, Sultan. (2008). Anadolu ve Rumeli ağızlarında (-i verir) yapısı. *Karadeniz Araştırmaları*, 19, 129-136.
- Yavuz, Handan, Balcı, Ayla & Turan, Ümit Deniz. (2004). *Turkish Phonology, Morphology and Syntax*. Eskişehir: Anadolu Üniversitesi.

ÇEVİRİYAZIDA GERİBİLDİRİM, SESLİ DURAKLAMA, SESSİZLİKLER VE ÜNLEMLERİN ÖLÇÜNLEŞTİRİLMESİ

Hale IŞIK-GÜLER & Betül ERÖZ-TUĞA

Orta Doğu Teknik Üniversitesi

hisik@metu.edu.tr, beroz@metu.edu.tr

1. GİRİŞ

Alanyazında sözlü söylemin çeviri yazıya aktarılmasındaki farklı yaklaşımlar sebebiyle, genel amaçlı sözlü derlem oluşturulmasında, özellikle de ODTÜ Sözlü Türkçe Derlemi Projesinde (ODT-STD) olduğu gibi çoklu çeviri yazıcı sistemiyle çalışılan durumlarda, çeviri yazının ölçünleştirilmesi elzemdir. ODT-STD kapsamında iletişimi düzenleyici özelliği kabul edilen (Schegloff, 1981) geri-bildirim, sesli duraklamalar ve ünlemler için başlıca iki tür uygulama yapılmıştır: (1) uzunluk ve hece sayısı ölçünleştirmesi, (2) sesin yazıya dönüştürülmesi. Bu bildiri de söz konusu ölçünleştirmeler ele alınacaktır.

Proje ekibi düzenli olarak uygulamalı çeviri yazı ölçünleşme çalışmalarıyla çeviri yazılarda karşılaştıkları örnek ve problemleri paylaşarak ölçünleştirmeyi sağlayıp, “doğunluk” ve “yineleme” (İng. saturation and iteration) tekniklerini kullanarak çeviri yazı yönergesini oluşturmuştur. Ayrıca, edebi eserlerdeki geribildirim ve ünlemlerin yazıya yansıtılma şekilleri incelenmiş ve ölçünleştirme kuralları ve listeleri zenginleştirilmiştir.

Çeşitli dillerde karşılıklı konuşmalardaki sözcelerin yaklaşık beşte biri geribildirimlerden oluştuğu için (Çubukçu, 2005) sözlü derlem çalışmalarında bu dilsel öğelerin incelenmesi ve ölçünleştirilmesi önemlidir. Derlem verisinde bu öğelerin sözlü Türkçedeki kullanımının yazıya yansıtılmasındaki olasılıklar ve değişkenlikler gözlemlenmiş ve ölçünleştirilmiştir (örn. hı/hu/hı-hı/hı-hı/hı-hı, he ya/ha ya, vb.).

Konuşmada çeşitli işlevler gören sessizliklerin de iletişim (bkz. Ephratt, 2008), incelik ve kültürle bağlantılı anlamlarının örneklendirilebilmesi (Sifianou, 1992; 1997) veya ruhdilbilimsel işlevlerinin dilbilimciler tarafından araştırılabilmesi için derlem çalışmalarında hassas/kesin ölçüme olanak verecek şekilde işaretlenmeleri gerekmektedir. *Sözlü Türkçe Derlemi*'nin çeviri yazılarında kullanılan EXMARaLDA Partitur Editor'deki süre göstergesi ve ses dalgası panelinin buna olanak vermekte ve böylelikle ölçünleştirmede güvenilirlik artmaktadır. Sesli duraklamalar için ise derlemde duraklamanın uzunluğuna göre “e/em” ya da “ee/eem-emm” kullanılmıştır. Yapılan bir ön ölçünleştirme çalışmasıyla ancak bu ses/yazımlar kullanıldığında çeviri yazıcıların daha güvenilir aktarımlara ulaşılabilirdiği gözlenmiştir.

Sözcüksel ve sözcüksel olmayan ünlemlerin üzerinde de çeşitli çalışmalar yapılmış ve ünlem yazımları söyleyiş ve varsa anlamsal ayrışmalarını da yansıtabilecek bir şekilde çeviri yazı yönergesine yansıtılmıştır (örn. uyarı olarak “Aman!” ve önemsizlik belirten “Amaan!”). Bu çalışmada *Sözlü Türkçe Derlemi*'ndeki özellikle sözcüksel olmayan çoklu-şekilli (İng. multi-mode) geribildirim ve sesli duraklamaların işlevsel yapıları, kullanılan ODT-STD için HIAT sözcce sonu işaretlemeleriyle birlikte tartışılacaktır.

1.1 Neden Ölçünleştirme?

Özellikle çoklu çeviriyazıcı kullanılan derlem projelerinde çeviriyazıcılar arasındaki güvenilirliği (Eng. inter-transcriber reliability) en üst düzeye çıkarma gerekliliği bulunmaktadır. Bunu yaparken göz önünde bulundurulması gereken bir kaç önemli etmen mevcuttur. Aktarım için çeşitliliği göz ardı etmeyen bir sınırlama getirilmesi gereklidir. Bununla birlikte çoğu çeviriyazıcıların uzmanlık sahibi olmadığı da unutulmaması gereken bir diğer etmendir.

Hali hazırda bu gibi süreçlerde derlem projelerinin yöneticilerine yardımcı olacak, Sözlü Türkçe'nin yazıya aktarılması sırasında kullanılacak kapsamlı (geribildirimleri, ünlemleri vb. içeren) bir sözlük ve yazım kılavuzu bulunmaktadır. Bu çalışma bu amaca yönelik olarak ODT-STD kapsamında atılan adımları örneklemek adına yapılmıştır.

2. ALANYAZIN

Büyük ölçekli ulusal derlemlerde (örn., BNC, ANC) çoğunlukla ölçünleştirme amaçlı olarak standard yazım'a (Eng. standardized orthography) başvurulmuş ve bu bir ölçek olarak çeviriyazıcı yapanlar tarafından kullanılmıştır.

Bir önceki bölümde de dile getirildiği üzere sözlü Türkçe derlemleri henüz oluşum aşamasındadır. Bu sebeple alanyazında kapsamlı bir ölçünleştirme çalışmasına rastlanmamış, ancak Türkçe'nin bu çalışma kapsamında ilgilendiğimiz bazı özellikleri (örn. geribildirim ve ünlemler) ile ilgili derlem temelli olmasa da araştırma çalışmaları mevcuttur. Bu çalışmaların veri inceleme ve veri paylaşım yöntemleri, Türkçe ile ilgili geniş kapsamlı ölçünleştirme kitapçıkları hazırlanırken yardımcı olabilecek niteliktedir.

Türkçe ünlemler ile ilgili olarak daha önce sözdizimsel özelliklere (Akar, 1988) ve ünlemlerin edimbilimsel yapı ve işlevlerine (Büyükkantarçoğlu, 2000, 2006; Külebi, 1990) bakılmıştır. Türkçe'de geribildirimler için ise Çubukçu (2005) detekleyici geribildirimlerin özellikleri ve sınıflandırılmaları üzerine çalışmıştır. Bu çalışmalarda amaç sözlü dilin yazıya aktarılması olmadığı için ölçünleştirme amacına gidilmemiştir. Türkçe'de özellikle sessizlik ve sessiz duraklamalar'ı konu alan bir söylem analizi çalışması rastlanmamıştır.

3. YÖNTEM

3.1 STD Ölçünleştirme Süreci

ODTÜ Sözlü Türkçe Derlemi Projesinde ölçünleştirme sürecinin ilk aşamasında, Ekim 2008'den başlayarak alanyazında geribildirimlerin, ünlemlerin ve sesli duraklamaların ne şekilde yazıya aktarıldıkları ve kategorize edildikleri incelendi. Bunun için örneğin Aksan (1983), Bayraktaroğlu (1992), Büyükkantarçoğlu (2000, 2006) gibi ilgili çalışmalara bakıldı. Bunun yanında özellikle ünlemler için edebi eserlerin taranması yoluna gidildi. Bu amaçla Haldun Taner, Aziz Nesin, ve Muzaffer İzgü'nün eserlerinde ilgilendiğimiz Türkçe birimlerin yazıya ne şekilde yansıtıldığı incelendi.

Bir sonraki aşamada STD proje grubu, toplanan verilerin pilot çeviriyazısına başladı. Bazen aynı sözlü metinler her bir grup çalışmanı tarafından ayrı ayrı çözümlendi, bazen de farklı

sözlü metinlerden edinilen bilgiler bir araya getirildi. Bu paylaşımların yapıldığı çeviriyazı karşılaştırma toplantıları haftada birkaç kere, 6-8 kişilik bir grupla yapıldı.

Bu toplantılarda yineleme (eng. iteration) ve doygunluk (eng. saturation) yöntemleri izlenerek aşağıdaki iki çeşit ölçünleştirme uygulamaları yapılmasına karar verildi:

1. Uzunluk ve hece sayısı ölçünleştirilmesi
2. Sesin yazıya dönüştürülmesi ile ilgili ölçünleştirme

Ekim 2008 ile Mart 2010 arasında farklı gelişim süreçlerinden geçilmiş, yeni gelen çeviriyazı dosyaları ve karşılaşılan güçlükler göz önünde bulundurularak bazen bir önceki uygulamaya geri dönüş olmuştur.

3.2 Uzunluk ve Hece Sayısı Ölçünleştirilmesi

Bu tip ölçünleştirmede (a) geribildirimler ve (b) ünlemler ele alındı. Örneğin, uzunluk olarak "haa", "haaa" ve "haaaa" olarak duyulan geribildirim için iki yazım biçimi saptanmıştır. Kısa duyulan geribildirim için "ha", uzun duyulan geribildirim için, duyulan "a" sesinin uzunluğu ne olursa olsun, yazımda iki "a" ile sınırlandırılmıştır. Ünlemlerde ise, örneğin derlem verisinde "vaaay!" ve "vaaaaaay" ile karşılaşılmıştır. Burada da geribildirimler için uygulanan ölçünleştirme prensibi benimsenmiş ve ünlemlerde uzun duyulan "a" kısmı iki sınırlandırılarak "vaay" biçimi benimsenmiştir.

3.3 Sesin Yazıya Dönüştürülmesi

Sözlü Türkçe'de ve derlem verisinde sıkça karşılaşılan ama yazılı dilde benimsenmiş bir karşılığı olmayan bazı sesler için yazıya dönüştürme çalışması yapıldı. Bunlardan bir tanesi "cümle başı belirteci" (Eng. Sentence Initialiser) olarak isimlendirdiğimiz ve dudak/ağız şaplatma sesine benzer cümle başlangıç sesinin yazımıdır.

Örnek (1)

328 [10:03:8]	329 [10:06:6]
ISA [v]	((0.9)) tıh ben ölürüm ya öyle. _birine

Bu sesin sözlü Türkçe'nin yazımında yerleşmiş olan karşı olma/kabul etmeme için kullanılan (ve aynı şekilde kulağa gelen) "cık" ile karıştırılmaması için, ayrı bir şekilde yazıya "tıh" olarak geçirilmesine karar verilmiştir.

4. ÖRNEKLEM

Bu bölümde ölçünleştirme çalışmaları derlem verisi kullanılarak sesli duraklamalar, sözcüksel olmayan geribildirimler, sessizlikler ve ünlemler için örneklenecektir.

4.1 Sesli Duraklamalar

Sesli duraklamalar için ilk aşamada toplanan veride ağızlar ve yöresel konuşma biçimlerinde sesli duraklamalardaki sesli harf kullanımında çeşitli farklılıklar gözlemlenmiştir. Bu farklılıkları çeviriyazıya geçirirken, çeviriyazıcı sayısının fazla olması nedeniyle sorunlar yaşandığı ve değişik çeviriyazıcılar arasındaki güvenilirliği (Eng. intertranscriber reliability) sağlamak için ölçünleştirme yoluna gidilmiştir. Duraklar yazıya ya tek başına “e” ünlüsü tek başına kullanılarak veya “m” ünsüzü ile beraber kullanılarak geçilmiştir. Ancak bilgi kaybını önlemek ve farklılıkları vurgulayabilmek için de aşağıdaki örnekte görüldüğü gibi EXMARaLDA programındaki *comment tier*, yani yorum satırına duyulduğu düşünülen ses yazılmıştır (bkz. Örnek 2).

Örnek (2)

66 [01:40.5]
CEV000041 [v] ee

Sesli duraklamalardan sonra ölçülebilir bir sessizlik bulunmaktaysa arkasına satır üstü işaretli konulmuştur.

Örnek (3)

142 [02:48.5]
RUK000029 [v] ((0.6)) ee' ((1.0)) ben ne gün geldim? Pazartesi
144 [02:53.1]
MUS000031 [v] geldi. _e' ((0.5)) Cuma mı ne gelmiş de İstanbul'dan
186 [04:11.0]
BUR [v] ((0.3)) eem' ((0.2)) Nisa. _ne o İsa gibi.

Sesli duraksamadan sonra belirgin sessizlik olmadığı durumlarda herhangi bir satır sonu işaretlemesi yapılmamıştır.

4.2 Sözcüksel Olmayan Geribildirimler

Sözcüksel olmayan geribildirimleri ölçünleştirirken “tam kelime olmayan sesler” (Carter & McCarthy, 1997) tanımı kullanılmıştır. Konuşucular arasında *destekleyici/kabullenici* (İng. supportive/ agreement) ve *düzenleyici* (İng. regulative) işlev gören sözcükler ile *reddetme ve karşı çıkma* (İng. rejection/ disagreement) işlevlerini sağlayan sözcükler geribildirim sayılmıştır. Aşağıdaki tabloda derlem verisinde rastladığımız ve ölçünleştirdiğimiz geribildirim sözcükleri bulunmaktadır. Çeviriyazı yaparken bu geribildirim sözcüklerinin sonuna satır sonu sembolü olarak EXMARaLDA ODT-STD klavyesinde bulunan satır üstü nokta işareti (') kullanılmıştır.

Tablo 1
Derlemde kullanılan geribildirim sözcükleri

Destekleyici/kabullenici, düzenleyici	Reddetme ve karşı çıkma
hı/hıı/hı-hı/hı-hıı/hıı-hıı	i-ih
hm/hmm/hm-hm/hm-hmm/hmm-hmm	i-ım
ha/haa/ha-ha/ha-haa/haa-haa	
he/hee/he-he/he-hee/hee-hee	

Bu aşamada çeşitli sorunlarla karşılaşmış ve bunlara çeşitli çözümler bulunmuştur. Mesela aynı formdan türeyen pek çok versiyon olması bir sorun teşkil etmiş ve ölçünleşmeyi elzem hale getirmiştir.

Birden fazla heceli geribildirimlerde (hı-hı/hı-hıı/hm-hm/hm-hmm/ha-ha/ha-haa/haa-haa/i-ih/i-ım) iki hecenin arasında, önünde veya ardında boşluk bırakmadan, satır üstü işareti ile birlikte tire işareti kullanılmıştır.

Örnek (4)

117 [03:17:8]	118 [03:22:9]
ISA	var. _hı-hı'

Şayet geribildirim sözcük dizilerinde sözcükler arasında ölçülebilir bir duraklama yoksa (0.1 saniyeden az), Örnek 5'te görüldüğü gibi tek bir (') kullanılmıştır. Ancak, eğer duraklama ölçülebiliyorsa her bir sözcükten sonra Örnek 6'daki gibi (') konulmuştur.

Örnek (5)

124 [02:37.5]
BUR [v] hı hı hı'

Örnek (6)

134 [02:37:0]	134 [02:37:8]	134 [02:38:6]
RUK [v] yaptık hep beraber	ha •	onlar. • ha-ha'

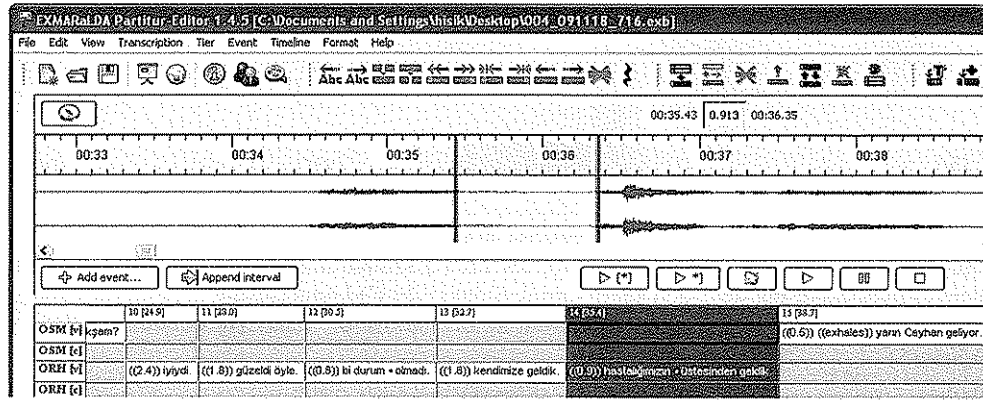
Aşağıda, Örnek 7'de görüldüğü gibi çoklu şekilli geribildirimlerin satır sonu işaretlemesi yapılırken, geribildirim işlevini doğru yansıtabilmek adına verinin çoğu için satır üstü nokta işareti (') kullanılırken, çok belirgin bir şekilde soru işlevi gören geribildirimler için soru işareti (?) kullanılmasına karar verilmiştir.

Örnek (7)

64 [01:19.1]	65 [01:20.0]
RUK [v]	uyuyor mu?
BUR [v]	hı-hı'

4.3 Sessizlikler

Derlemdeki 0.1 saniyeden kısa sessizlikler için EXMARaLDA Partitur Editor'daki ODTU-STD klavyesi kullanılarak büyük siyah yuvarlak (bkz. Şekil 1, saniye 35.4) kullanılır. Çeviriyazıda HIAT ilkesi takip edilerek 0.1 saniyeden uzun olan sessizlikler ölçülür ve bir sonraki konuşmacıya atanır. Sessizliklerin uzunlukları ((,)) olarak, yani tek ondalık belirtilerek, yazılır.



Şekil 1. Sessizlik ölçümü

Sözlü Türkçe Derlem dosyalarında 2009'da yapılan EXMARaLDA programı güncellemesi ile çok hassas sessizlik ölçümü sağlanmıştır.

4.4 Ünlemler

Ünlemler için Haldun Taner ve Aziz Nesin'in eserleri ve Türkçe üzerine yapılmış çeşitli çalışmalar taranmıştır. Bu çalışmalardan bazıları: Atabay, Kutluk ve Özel (1983), Büyükkantarcioglu (2000), Külebi (1990) ve Akar (1988).

Çoklu şekilli ünlemlerin belirlenmesinde yükselen ezgiyi diğer sesletimlerden ayırt etmek için Örnek 8'deki gibi ünlem işareti ("!") kullanılmıştır.

Örnek (8)

182 [03:48.5]	183 [03:49.0]	184 [03:50.0]
RUK [v]	tüh!	
BURC [v]	((0.2)) daha yeni.	((0.5)) uyanır ama. Uyukusu fazla şey değil.

Örnek 9'daki "ay" ve Örnek 10'daki "yav" için olduğu gibi ünlem yüksek sesle ya da yükselen ezgiyle söylenmemişse herhangi bir sözce sonu işareti kullanılmamış ve aynı sözcenin parçası kabul edilmiştir.

Örnek (9)

71 [01:20.5]	
RUK [v]	((0.8)) ay ne numaracı bu da!

Örnek (10)

..	72 [02:52.0]	
OSM [v]	((laughs)) ((6.0)) şu makinanın sigortasız olması durumu kötü oldu ya.	
ORH [v]	((laughs))	
73 [03:00.3]	74 [03:03.2]	75 [03:05.1]
OSM [v]	((1.4)) hı-hı'	
ORH [v]	((2.3)) Colin'in?	((1.7)) yav ((0.4)) benim anlamadığım böyle
..	76 [03:10.2]	
ORH [v]	şey ((0.7)) Colin((0.3)) üç kuruştan kaçan bir firma da değildir. nasıl oldu	

STD'de temsil edilen ya da edilmesi beklenen ünlemlerden bir kısmına ve diğer ölçünleştirme çalışmaları sonucu oluşturulan çeviriyazı kurallarının ilk baskısına Ruhi, Hatipoğlu, Eröz-Tuğa, Işık-Güler (2010)'dan ulaşmak mümkündür.

5. SONUÇ

Sözlü derlem çalışmalarında ölçünleştirme yabancı derlemlerde çoğunlukla standart yazım kuralları gözetilerek (Eng. standard orthography) ve sözlü dil sözlükleri kullanılarak yapılmıştır. Türkçe için bu şekilde çok kapsamlı bir kaynak olmadığından STD için proje ekibi tarafından ölçünleştirme çalışmaları yapılmıştır. STD'deki geribildirim, ünlem ve sesli duraklamalar ile ilgili ölçünleştirme çalışmalarında ağırlıklı olarak iki amaç gözetilmiştir. Birincisi, derlem kullanıcıları için olanakların genişletilmesi ve aramaların güvenilirliğinin artırılmasıdır. İkincisi ise çeviriyazıcılar için çeviriyazı sürecinde işlevsellik ve kolaylık sağlanmasıdır.

Kaynakça

- Akar, D. (1988). Some Syntactic Properties of Turkish Interjections. *Studies on Turkish Linguistics* (s. 265-274.) Ankara : METU Yayınları
- Atabay, N., Kutluk, İ. ve Özel, S. (1983). *Sözcük Türleri*. Ankara: Türk Dil Kurumu Yayınları.
- Büyükkantarcioglu, N. (2000). A Pragmatic Analysis of Interjections in Turkish in the Context of Informal Reactive Responses (Resmi-Olmayan Sözel Tepkiler Bağlamında Türkçe Ünlemler Üzerine Edimbilimsel Bir İnceleme), 10th. International Conference on Turkish Linguistics (ICTL 2000), Bosphorus University 16-18 August 2000.
- Büyükkantarcioglu, N. (2006). An analysis of Turkish Interjections in the context of Reactive Idea Framing. *H.Ü. Edebiyat Fakültesi Dergisi*, 1(23), 19-32.
- Çubukçu, H. (2005). Karşılıklı konuşmada destekleyici geri bildirim. Ergenc, İ., İşsever, S., Gökmen, S. ve Özgür Adin (Haz.), *Dilbilim İncelemeleri* (s. 289-304). Ankara: Doğan Yayıncılık.
- Ephratt, M. (2008). The functions of silence. *Journal of Pragmatics* 40(11), 1909-1938
- Külebi, O. (1990). Türkçe ünlemlerin kullanımbilim (Pragmatics) yönünden incelenmesi. *Dilbilim Araştırmaları*, 1, 10-35.
- Ruhi, Ş., Hatipoğlu, Ç., Eröz-Tuğa, B., Işık-Güler, H. (2010). A guideline for transcribing conversations for the construction of spoken Turkish corpora using EXMARaLDA and HIAT. Ankara: Setmer Publishing.
- Schegloff, E. (1981). Discourse as an interactional achievement: Some uses of uh-huh and other things that come between sentences. In D. Tannen (ed.), *Analyzing discourse: Text and talk*. Washington, DC: Georgetown University Press, 71-93.
- Sifianou, M. (1992). *Politeness phenomena in England and Greece*. Oxford: Oxford University Press.
- Sifianou, M. (1997). Silence and politeness. In A. Jaworski (ed.), *Silence: Interdisciplinary perspectives*, Berlin: Mouton de Gruyter, 63-84.

TÜRKÇE İÇİN GENEL AMAÇLI SÖZLÜ DERLEM OLUŞTURMDA METAVERİ, ÇEVİRİYAZI ÖLÇÜNLEŞTİRMESİ VE DERLEM YÖNETİMİ

Şükriye RUHİ

Orta Doğu Teknik Üniversitesi

sukruh@metu.edu.tr; sukriyeruhi@gmail.com

1. GİRİŞ

Büyük veritabanlı sözlü derlemlerin geliştirilmesi birçok bakımdan hem kuramsal ve uygulamalı dilbilim kavramlarını gözden geçirmeyi (Čermák 2009; Weisser 2006) hem de geniş oylumlu derlem dilbilimi (İng. macro corpus linguistics) çalışmalarında dilbilim alanının ölçütlerine duyarlı çalışma yöntemlerinin geliştirilmesini gerekli kılmaktadır. Bu yuvarlak masa toplantısının amacı hazırlanmakta olan *Sözlü Türkçe Derlemi*'nin (STD) tasarımı ve ilk bulguları ışığında, Türkçe için genel amaçlı sözlü derlem geliştirmeye yönelik olarak bazı temel dilbilimsel öğeleri, çeviriyazı ilkeleri ile geniş oylumlu derlem dilbilimi gereklerine uyumlu bir derlem yönetim sistemini tartışmaktır. Oturumda, STD'nin amacı ve kullandığı derlem yazılım sisteminin (Schmidt 2004) özellikleri kısaca açıklandıktan sonra, aşağıdaki konular ele alınmıştır:

1. Yeni kuşak sözlü derlemlerin özellikleriyle uyumlu ve söylem çözümlemesine yönelik işaretleme içerecek derlemler için metaveri (İng. metadata) tasarımı; bağlam ve söylem türü kavramlarının derlem kayıtları çerçevesinde gözden geçirme; veribilgisi kodlaması ve uygulamaları (Ş. Ruhi);¹
2. Türkiye Türkçesi için oluşturulmuş bir sözlü dil derleminin bulunmaması nedeniyle, standart Türkçe ile ağız çalışmaları alanında yapılmış ve halen yapılmakta olan araştırmalar gözden geçirilerek hem standart Türkçe hem de çeşitli ağızların söyleyişlerini derlem boyutunda yansıtabilecek çeviriyazı ilkelerinin belirlenmesi; bazı çekim eklerinin temsil edilmesi çerçevesinde karşılaşılan başlıca sorunlar ve çözümler (Ç. Hatipoğlu ve Özlem Karakaş);
3. Alanyazında, söylemin çeviriyazıya aktarılmasındaki farklı yaklaşımlar sebebiyle, özellikle STD gibi çoklu çeviriyazıcı bulunan derlem çalışmalarında, çeviriyazıların tutarlılığı açısından geribildirim, durak, duraksamalar ve ünlemler için ölçünleştirme süreçleri ve çözümler (H. Işık-Güler ve B. Eröz-Tuğa);
4. Derlem kayıtlarını, veri bilgisini ve çeviriyazılarını web üzerinden yönetmeyi sağlayan özgün bir yazılımın geliştirilmesinde izlenen tasarım hedefleri; standart iş akışları, EXMARaLDA sistemi ile entegrasyon; TEI, XCES standartlarına uyum (M.G.C. Acar ve K. Eryılmaz).

Bu yazıda, XXIV. *Dilbilim Kurultayı*'nda düzenlenen yuvarlak masanın arka planı ve STD'nin hedefi ile bazı temel çeviriyazı özellikleri özetlenmiştir.

2. STD'NİN HEDEFİ VE ANA ÖZELLİKLERİ

STD, "ODTÜ Sözlü Türkçe Derlemi" başlıklı TÜBİTAK 108K283 numaralı proje tarafından desteklenerek oluşturulmakta olan, genel amaçlı bir çoklu ortam derlemidir. Proje, Ekim 2008 tarihinde başlamış olup Nisan 2011 tarihinde tamamlanacaktır. Derlem, yüz yüze ve iletişim araçlarıyla (televizyon ve telefon) Türkiye'nin çeşitli bölgelerinde gerçekleştirilen konuşmalar içermektedir.

STD'nin ilk aşamadaki amacı, Türkiye Türkçesi için bir milyon sözcüklük, temel çeviriyazı ve açıklamalar (İng. annotation) ile yönlendirme sözeyleminin (ricalar ve emirler) işaretlemesini yapmaktır. STD'nin önümüzdeki yıllarda 10 milyon sözcük büyüklüğündeki bir derleme genişletilmesi hedeflenmektedir. Ağustos 2010 tarihi itibarıyla derlemede, 60 saat ses/görüntü kaydına karşılık gelen yaklaşık 500,000 sözcük büyüklüğündeki bir veritabanı çeviriyazının çeşitli aşamalarında bulunmaktadır. Halihazırda bu kayıtlarda 2008-2010 yılları arasını kapsayan süre içinde hem ölçünlü Türkçe hem de yerel ağızlardan örnekler bulunmaktadır. Konuşma türü ve bağlamları bakımından ise, sohbet, hizmet alımı karşılaşmaları, ders anlatımları ve televizyon ile radyo programları bulunmaktadır (STD'nin Demo ve Örnekçe Sürümleri için bkz. <http://std.metu.edu.tr>). STD, Türkçe dilbilim ve ilgili alanlarındaki araştırmacılara kar amacı gütmemek koşulu ile ücretsiz bir kaynak olarak sunulmaktadır.

STD'de hem kayıt alma hem de işleme aşamalarında kayıtların teknik ve toplumdilbilimsel özellikleri metaveri (İng. metadata) olarak web tabanlı bir Derlem Yönetim Sistemi'ne aktarılmaktadır (Ayrıntılı bilgi için bkz. Acar ve Eryılmaz, bu yayın). Çeviriyazı için kullanılan EXMARaLDA derlem oluşturma aracı (bkz. <http://exmaralda.org/>) farklı platformlara veri aktarımı olanağı sağlayan nitelikte olduğu gibi (örn. Praat ve TEI ile uyumlu uygulamalar), ses/görüntü ile çeviriyazıları birlikte izleme olanağı tanımaktadır. Bu bakımlardan, STD yeni kuşak derlemlere örnek teşkil etmekte ve veri aktarımı nedeniyle uzun ömürlü olması beklenmektedir. STD'de halihazırda açıklamalar İngilizce yapılmaktadır; ancak derlemin son halinde Türkçe ve İngilizce olarak sunulacaktır (bkz. aşağıda Şekil 1).

3. STD'DEKİ ÇEVİRİYAZILARIN TEMEL ÖZELLİKLERİ

Bu bölümde, STD'nin çeviriyazı metinlerinde izlenen genel kurallardan söz edilecektir.

STD'de çeviriyazılar HIAT'ın (Schmidt 2008) uyarlanmış ve STD projesinde geliştirilmiş kurallarına göre yapılmaktadır (bkz. Ruhi ve diğ. 2010b). Buna göre, sözlü dile özgü sözcük yapılarını belirleme ve açıklamada tümce yapısı, edimbilimsel özellikler ve konuşma diline özgü sesli/sessiz duraklamalar, tamamlanmamış veya değiştirilmiş sözcük/sözcükler ve konuşma örtüşmeleri işaretlenmektedir.

STD'de özellikle vurgulanan bir temel çeviriyazı kuralı, mümkün olduğunca çeviriyazıcı yorumuna yer vermeyecek şekilde konuşmaları yazıya aktarmaktır. Konuşma Çözümlemesi'nin çeviriyazı yöntemi ile uyumlu olarak, örneğin, bir konuşanın sözcüğü yapısal bakımdan tamamlanmamışsa çeviriyazıda bu sadece tamamlanmamış yapı olarak işaretlenmekte, söz kesme veya sözcük örtüşmesi mi olduğu araştırmacının incelemesine bırakılmaktadır. Örnek (1) ile açıklarsak, EMI000246 kimlik no.lu olan konuşanın, 'Sanayi girişinde bekleye' sözleri ile başlayan sözcüğü tamamlanmamış bir yapıdır ve bu 'birleşik üç nokta işareti' ile belirtilmektedir. Çeviriyazıda, söz konusu yapının XMA000366 kimlik no.lu olan kişinin sözcüğünün 00:00:21.6 saniyede EMI000246'nın sözcüğü ile örtüşecek biçimde seslendirmeye başlamış olmasından mı kaynaklandığı, yoksa EMI000246'nın kendisinin sözcüğünü tamamlamaktan vazgeçmesi nedeniyle mi oluştuğu belirtilmemiş, sadece örtüşme yeri mümkün olduğunca kesin bir biçimde çözümlenmiştir.

(1)

14 [21.6]	15 [22.1]	16 [23..17 [23.:
EMI000246 [v]	Sanayi girişinde bek ley...	
XMA000366 [v]	o zaman • şu köprü'nün üstünden direkt devam	

Öte yandan, derlemin mümkün olduğunca geniş bir araştırmacı kitlesi tarafından kullanımını sağlamak için ağırlıklı olarak ölçünlü Türkçenin yazım kuralları izlenmekte ve sadece sınırlı sayıdaki sözcük yazımı için seçenekli yazım kuralları kullanılmaktadır. Örneğin, sayı sıfatı *bir* söylenişe göre *bi* veya *bir* biçiminde yazıya aktarılmaktadır (Yazımda ağız farklılıklarını yansıtılması için bkz. Hatipoğlu ve Karakaş, bu yayın).

Çeviriyazıda EXMARaLDA'nın Partitur Editor'u kullanılmaktadır. Birçok sözlü derlemden farklı olarak Partitur Editor, konuşmaları alt alta gelen konuşan satırları gibi yansıtmak yerine, müzik partiyonu gibi, ses kaydı ile eşzamanlı olarak, soldan sağa ilerleyen zaman dilimleri gibi yazıda göstermektedir. Şekil 1'de görüleceği üzere, Partitur'da standart olarak bulunan konuşan sözcükleri ve sözcük açıklamaları (v ile c satırları) ile arka plan eylemlerinin yazıldığı *nn* satırlarının yanı sıra, STD'ye özgü olarak, 'tanımlanamayan konuşan' satırları (IND) ve toplu olarak söylenen sözcükler için ALL satırları bulunmaktadır. Bunun yanı sıra, çeviriyazı çalışmaları sırasında çeviriyazı yapan ve kontrol eden kişiler arasındaki iletişimi kolaylaştırmak için çeviriyazı biçim sayfasına *corr* (düzeltme) satır eklenmiştir. Söz konusu satırın çeviriyazı yapan kişilerin çok olduğu derlem çalışmaları sırasında ölçünleştirmeyi hızlandırdığını özellikle vurgulamak isteriz (Çoklu çeviriyazıcı projelerindeki ölçünleştirmeye konusu için bkz. Işık-Güler ve Eröz-Tuğa, bu yayın).

	3 [03.2]	4 [04.0]	5 [05.6]	6 [06.9]	7 [11.8]	8 [13.5]	9 [
SUK000057 [v]	verdi.	((0.5)) e soralım burda.					
SUK000057 [c]							
EMI000246 [v]	((0.3)) ((XXX))			affedersiniz.	((0.3)) Organize Sanayi girişi burası mı?		
EMI000246 [c]					((loudly))		
XMA000366 [v]						((0.5)) ((inhales)) burdan devam edin.	((
XMA000366 [c]							
IND000002 [v]							
IND000002 [c]							
ALL000001 [v]							
ALL000001 [c]							
[nn]			((rustling))			((traffic noise))	
[corr]							

Şekil 1: STD Partitur Editor'ünden bir çeviriyazı kesiti

4. SON SÖZ

STD'yi oluşturma deneyiminin Türkçe dilbilim alanına birçok yönden yeni girdiler sağlayacağı beklenmektedir. Günümüze değin, Türkiye Türkçesinin konuşma dili dilbilgisi üzerine toplu bir çalışma bulunmamaktadır. STD'nin bu tür, geniş oylumlu bir çalışma için önemli bir kaynak olacağını umuyoruz. Burada sadece bir örnek vermekle yetineceğiz. Sözlü Türkçe derlemleri araştırmacılara sunuldukça, günümüz derlem dilbilimi araştırmalarında önemli bir yer tutan kalıplı söz veya 'çok sözcüklü yapılar'ın (İng. multi-word expressions) Türkçede neler olduğuna derleme dayalı (İng. corpus-driven) bir biçimde tartışmak mümkün olacaktır (Ruhi, hazırlanmakta).

Notlar

1. Kurultay'da Ş. Ruhi tarafından sunulup bu yayında yer almayan "Sözlü Derlem için Veribilgisi Geliştirme: Bağlam ve Tür Kavramlarına Derlem Dilbilimi Açısından bir Bakış" bildirisindeki metaveri konusu için bkz. Ruhi ve diğ. (2010a).

2. Her derlem çalışmasında olduğu gibi, STD çok geniş katılımlı bir ekip çalışması sayesinde oluşturulmaktadır. Bu bağlamda, çalışmalarımıza özellikle kayıt ve çevriyazı aşamalarında büyük katkıda bulunan ODTÜ Yabancı Diller Eğitimi Bölümü 2009-2011 yılları arasındaki lisans ve lisansüstü öğrencilerimize teşekkürü bir borç bilirim. Diğer kayıt başışlayan kişi ve kurumlar için bkz. <http://std.metu.edu.tr/tesekkurler/>

Kaynakça

- Čermák, F. (2009). Spoken corpora design: Their constitutive parameters. *International Journal of Corpus Linguistics* 14 (1), 113-123.
- ODTÜ Sözlü Türkçe Derlemi Projesi. <http://std.metu.edu.tr/>
- Ruhi, Ş. (hazırlanmakta). 'LOOK' as an expression of procedural meaning in directives in Turkish: Evidence from the *Spoken Turkish Corpus*. 6th *International Symposium on Politeness*, 11-13 Temmuz 2011'e sunulmuş bildiri.
- Ruhi, Ş., Işık-Güler, H., Hatipoğlu, C., Eröz-Tuğa, B. & Çokal Karadaş, D. (2010a) Achieving representativeness through the parameters of spoken language and discursive features: the case of the Spoken Turkish Corpus. I. Moskowich-Spiegel Fandiño ve diğ. (Haz.), *Language Windowing through Corpora. Visualización del lenguaje a través de corpus. Part II* (s. 789-799). A Coruña: Universidade da Coruña.
- Ruhi, Ş., Hatipoğlu, Ç., Eröz-Tuğa, B., Işık-Güler, H. (2010b). *A Guideline for Transcribing Conversations for the Construction of Spoken Turkish Corpora Using EXMARaLDA and HIAT*. ODTÜ-STD: Setmer Basımevi.
- Schmidt, T. (2008). Overview of HIAT transcription conventions. http://www1.uni-hamburg.de/exmaralda/files/HIAT_EN.pdf
- Weisser, M. (2003). SPAACy – A semi-automated tool for annotating dialogue acts. *International Journal of Corpus Linguistics* 8 (1), 64-74.

YUVARLAK MASA 4

YÖNETEN:	Deniz ZEYREK
ARAŞTIRMA KONUSU:	Söylemde Yapı