

TİKA
I. ULUSLARARASI
TÜRKOLOJİ
SEMPOZYUMU

Kırım / UKRAYNA, Mayıs-2005



ÇAĞDAŞ TÜRK LEHÇELERİ ÜZERİNE BİLGİSAYAR MERKEZLİ ÇEVİRİ ÇALIŞMALARI

Doç. Dr. Mehmet KARA, Kırıkkale Üniversitesi

İçinde bulunduğumuz çağda teknolojiyi üreten veya bunu kendi işlerinde en verimli bir şekilde kullanan milletlerin dünyada önemli bir konum ve güce sahip olduğunu hepimiz biliyoruz.

Öte yandan, yeterli güce sahip olmayanların güçlü devletlerce ne hâle getirildiğini yine son yıllarda çok açık ve acı bir biçimde anlamış olduk.

Türk medeniyetini yükseltmek istiyorsak, artık bizim de bilgiyi ve teknolojiyi çağa uygun biçimiyle kullanmamız gerekmektedir.

Günümüzde bilgisayar, kimilerince itici bulunsa bile, artık hayatın vazgeçilmezleri arasındadır ve bu gidişle girmedigi ev kalmayacaktır.

Hayatımıza hızlı giren bu sihirli ya da itici makine, internetin de kendisine eşlik etmesiyle birlikte diğer birçok alanda olduğu gibi dille ilgili çalışmalarda da yoğun bir şekilde kullanılmaya başlanmıştır. Diğer dünya dillerinin yanı sıra Türkçe ve kollarıyla ilgili bilgisayar merkezli çalışmalar da son yıllarda gelişme göstermiştir. Bunların birçoğu çeviri çalışmalarıyla ilgilidir.

Çağdaş Türk lehçeleri üzerine yapılan bilgisayar merkezli çeviri çalışmalarını dörde ayır-mak mümkündür:

1. Lehçeden Lehçeye Makine Çevirisi Çalışmaları.
2. Sesi Yazıya, Yazıyı Sese Dönüştürme Çalışmaları.
3. Kesim Olarak Algılanmış Yazıyı Metne Çevirme ya da Yaygın Adıyla Optik Karakter Tanıma (Optical Character Recognition = OCR) Çalışmaları.
4. Alfabe Çevirisi Çalışmaları.

1. Lehçeden Lehçeye Makine Çevirisi Çalışmaları

Doğal dil işleme yöntemiyle geliştirilen makine çevirisi, iki dil veya lehçe arasında yapılıp ve karşılıklı tanımlamaları, eşleştirmeleri içerir. Doğal dil işleme, dört aşamada gerçekleştirilmektedir:

1. Biçimbirim Çözümlemesi (Morfolojik Analiz): Biçimbilim, kelimelerin yapısını analiz eder. Yazılan morfolojik analiz programları yoluyla bu işlem, otomatik olarak gerçekleştirilir. Bu tür programlar, dil işleme tekniklerinin en temel kesitini oluşturur. Makine çevirisinin diğer aşamaları, bu temel çözümleme üstüne kurulurlar. Karşılıklı iki dil veya lehçenin kelime yapısı eşleştirildiğinde makine çevirisinin birinci adımı gerçekleştirilmiş olur. Eklemeli bir dil olan Türkçe için yazılan morfolojik analiz programları, kök ve ekleri ayırıştırıran bir sistemden ibarettir.

2. Sözdizimi Çözümlemesi (Sentaktik Çözümleme): İkinci basamakta kelime ve kelime gruplarının birleşmesiyle oluşan cümlelerin analizi yapılır. Bu kısımda ana amaç, var olan bir cümlemin doğru olarak öğelerine ayrılması veya öğeleri sağlam bir cümlemin oluşturulmasıdır. Bu aşamada özellikle kelime gruplarının tespit edilmesi sağlanır. Bunların birleştirilmesiyle cümlemin çatısı kurulmuş olur.

3. Anlam Çözümlemesi (Semantik Çözümleme): Makine çevirisinin üçüncü basamağında; yapısı tespit edilen kelimelerin, kelime gruplarının ve cümlemin anlamlandırılması amaçlanır. Bu aşamada, X dilindeki bir ifadenin Y dilindeki anlamının sağlanmasına çalışılır.

4. Anlam Kargaşasının Giderilmesi (Pragmatik Çözümleme): Makine çevirisinin bu son basamağında metindeki cümlelerin birleşerek oluşturduğu ifadeler bütünü içerisinde birbiriyle ilgili olan kısımların ilişkilendirilmesi ve anlam bütünlüğünün sağlanması amaçlanır. Bu basamakta anlam bulanıklıklarının giderilmesi için özel bir çaba sarf edilir.

Bu aşamalardan sonra elde edilen bilgi istenildiği takdirde bir “ara dil”e çevrilir. Daha sonra da bu ara dilden istenilen anlamları ifade edecek cümlelerin üre-til-me-si sağlanır.

Birbirine yakın diller veya benzer Türk lehçeleri arasında anlaşmayı sağlayacak bir makine çevirisi programı yapmak aslında pek zor değildir. Türk lehçelerinin gramerleri ana hatlarıyla aynıdır. Kökler ve eklerin denk düştüğü durumlarda sadece bir kök kelimenin yerine bir diğerini koymak, ya da bir eki diğer lehçedeki ekle değiştirmek aktarmayı büyük ölçüde halletmektedir. Bazen basit fonetik eşleştirmeler

bile meseleyi çözebilir. Özellikle Oğuzca, Kıpçakça gibi bir ana lehçenin alt kolları arasında bunu yapmak çok daha kolaydır. Örnek olarak iki Oğuz grubu Türk lehçesi olan Türkiye Türkçesi ile Türkmence arasında makine çevirisi programı çok daha kolay gerçekleştirilecek; biri Oğuz grubuna, diğeri ise Kıpçak grubuna mensup olan Türkiye Türkçesi ile Kazakça arasında bu programı hazırlamak biraz daha zor olacaktır. Ancak İngilizce ile Türkiye Türkçesi arasında yapılacak makine çevirisi programına oranla da biri Oğuz, diğeri Kıpçak grubuna ait iki Türk lehçesinde çeviri programı hazırlamak daha kolay ve başarılı olacaktır. Kısacası diller veya lehçelerin birbirine benzerlikleri arttıkça, makine çevirisi programı yazmak kolaylaşacak ve başarı oranı yükselecek; fakat benzerlikler azaldıkça, makine çevirisi programı yazmak zorlaşacak ve başarı oranı da bir o kadar düşecektir.

Bu açıdan bakıldığında Türk lehçeleri arasında karşılıklı anlaşmayı sağlayacak makine çevirisi programlarını üretmek pek zor olmayacaktır. Ancak edebî metinlerin aktarımında veya başarı oranı yüksek metin elde etme işlemlerinde kullanılacak makine çevirileri için uzun zaman, güçlü uzman ekip, fazla masraf ve emek gerekir.

Çağdaş Türk lehçelerine ait makine çevirisi çalışmaları daha çok morfolojik analiz düzeyindedir. Yani doğal dil işleminin birinci basamağı üzerine epeyce çalışma yapılmıştır. Bunların birçoğu Türkiye Türkçesi üzerinedir. Yazdığı morfolojik analiz programıyla ismini en çok duyuran bilim adamı Kemal Oflazer'dir. Daha önce Bilkent Üniversitesinde çalışmış olan Oflazer, bu gün çalışmalarını mensubu olduğu Sabancı üniversitesi bünyesinde yürütmektedir. Hacettepe Üniversitesinden A. Köksal'ın doktora tezi de Türkçe otomatik morfolojik analiz üzerinedir. İstanbul Teknik Üniveristesinden Gülşen Cebiroğ-lu'nun "Sözlüksüz Köke Ulaşma Yöntemi" adlı yüksek lisans tezi de morfolojik analiz programları arasında değerlendirilmelidir. Bu yöntemle, tıpkı morfolojik analizde olduğu gibi sözlük kullanmadan bir kelimenin biçimbirim analizinin yapılması hedeflenmiştir.

Elbette bu konuda diğer Türk lehçelerine yönelik çalışmalar da bulunmaktadır. Bilkent Üniversitesinden Kemal Altıntaş, Kırım Tatarcası için bir morfolojik analiz programı yazmıştır. Ancak bu çalışmaların hemen hepsi daha yeni zamanların verimi olduğu için dünyada kendini ispatlamış ve yaygınlık kazanmış Türk lehçeleri makine çevirisi programı henüz yoktur. Yaptığımız araştırmalar bizi bu fikre götürmüştür. Ancak bizim haberdar olmadığımız çalışmalar da elbette vardır. Özellikle Azerbaycan'da buna yönelik çalışmalar olduğunu duyuyoruz. Eğer bu tür başarılı programların varlığını bilen var da bizi de bilgilendirirse seviniriz.

Bilkent Üniversitesi Bilgisayar Mühendisliği Bölümü'nden Kemal Altıntaş, yüksek lisans tezinde morfolojik analiz programının yanı sıra Türkiye Türkçesi ile Kırım Tatarcası arasında karşılıklı tercüme yapacak bir sistem geliştirmiştir. Boğaziçi Üniversitesinden Cem Say'ın hedeflediği çalışmalar içerisinde Türk lehçeleri arasında makine çevirisinin de bulunduğu bilinmektedir. Boğaziçi üniversitesinde yapılmış bir Azeri Türkçesi-Türkiye Türkçesi makine çevirisi çalışmasının varlığından haberdar oldum, ancak bu çalışmayla ilgili ayrıntılı bilgi edinemedim.

Çağdaş Türk lehçeleriyle ilgili başarılı makine çevirilerinin yapılabilmesi için kapsamlı gramer-lerin, genel sözlüklerin, deyimler sözlüklerinin, istem (valenz) sözlüklerinin önceden hazırlanmış olması gerekir. Bu gün çağdaş Türk lehçeleri ile ilgili temel çalışmalar hâlâ tamamlanamamıştır.

2. Sesi Yazıya, Yazıyı Sese Dönüştürme Çalışmaları

Sesi yazıya, yazıyı sese çevirme çabaları; bir dilin kendi içinde yapılan çalışmalar kapsamında değerlendirilebilir. Sesi yazıya, yazıyı da sese çeviren programlar; dilin fonolojik özelliklerinin çok ayrıntılı bir şekilde incelenmesiyle geliştirilmektedir.

Bilkent Üniversitesi Türkçe Dil ve Konuşma İşleme Merkezi, doğal dil ve konuşma işleme alanlarında araştırmalar yapmak, Türk lehçelerine yönelik dil ve konuşma işleme uygulamaları geliştirmek amacıyla kurulmuştur. Bu merkezde konuşmayı tanıma; yazıyı sese, sesi yazıya çevirme programları üzerine çalışmalar yürütülmektedir.

Orta Doğu Teknik Üniversitesi bünyesinde de Türkçe üzerine doğal dil işleme çalışmaları yürütülmekte, makine çevirisiyle ilgili çalışmalar ve OCR çalışmaları yanında konuşmayı tanımaya yönelik çalışmalar da yapılmaktadır. Ayrıca bu merkezde Osmanlıca dokümanların bilgisayar desteği ile hızlı ve verimli bir şekilde okunmasını sağlayacak bir Uzman Transkripsiyon Sistemi (UTS) üzerine çalışmalar yapılmaktadır. Şu anda Amerika'da bulunan Kemal Altıntaş da Osmanlı arşiv belgelerinin önce okunup ses olarak kaydedilmesi ve daha sonra da bu sesleri yazıya dönüştüren bir program yardımıyla bütün belgelerin daha hızlı bir şekilde Lâtin harfleriyle yayımlanarak kullanıma

sunulmasını düşünenlerdendir. Ona göre, Osmanlıca OCR programı kullanıp metni tanımaya çalışmak yerine, bir ses tanıma programı geliştirerek belgeleri okumak ve sesi yazıya dönüştürmek daha akılcı bir yöntemdir. Çünkü bu işlem, OCR'ın zorluklarına bir alternatif olabilir.

Cumhuriyet Arşivleri Genel Müdürü Doç. Dr. Yusuf Sarınoy, bir görüşmemizde aynı yöntemi bir başkasının da kendilerine teklif ettiğini belirtmişti.

Yaptığımız araştırmalara göre, bütün dünyada yazıyı sese dönüştüren çalışmalar; sesi yazıya dönüştürenlere oranla çok daha başarılı durumdadır. Bu konuda İngilizce için geliştirilen programlar arasında "ReadPlease" ve "Dragon NaturallySpeaking"ın önemli bir yeri vardır. Türkçe içinse, Arçe-lik'in ürettiği "GVZ" adlı bir programın varlığı bilinmektedir.

Türkçe ile ilgili yazıyı sese, sesi de yazıya dönüştürecek programlar yazılmaya çalışılmakla birlikte; bu gün için adını duyurmuş başarılı programlar yoktur.

3. Resim Olarak Algılanmış Yazıyı Metne Çevirme ya da Yaygın Adıyla Optik Karakter Tanıma (Optical Character Recognition = OCR) Çalışmaları

Resmî yazıya dönüştürmelerde kullanılan OCR (Optical Character Recognition = Optik Karakter Tanıma) programları, dünyadaki dillerin birçoğunu içermesi amaçlanarak hazırlanmaktadır. Çünkü bir OCR programının başarısı temelde iki şekilde ölçülmektedir: Başarı oranının yüksekliği ve kapsadığı dil sayısı.

OCR çalışmaları, dil içi veya karşılıklı dillerle ilgili çalışmalar olmaktan ziyade dillerin resim olarak taranan karakterlerinin yazıya çevirilmesiyle ilgili çalışmalardır. Elde edilen metinlerin üzerinde değişiklik yapılması büyük kolaylıklar sağlamak ve uzun emek harcanarak dizilebilecek metinler, başarılı OCR programları kullanıldığı takdirde çok az emek ve düzeltmeyle yeniden kullanılabilir ve basılabilir hâle getirilmektedir. Bu yöntem, daha hızlı üretmek için bütün dünyada sık başvurulan bir yöntemdir.

İlk OCR programlarının dil seçeneğinde Türkçe bulunmadığı için Lâtin temelli dillerden biri seçilerek yapılan metin taramalarında "ç, ş, ğ, ı, ö, ü" gibi Türkçe karakterler büyük oranda karışık karakterlerle karşılanmaktaydı ve OCR programları Türkçe için yetersiz kalmaktaydı. Ancak sonradan tarama dilleri arasına Türkiye Türkçesi de dahil edilmiş ve bu problem büyük ölçüde halledilmiştir. Daha sonraları ise, birçok dünya dilinin yanı sıra bütün Türk lehçelerini kapsayan bir OCR programı üretilmiştir. Bu programın adı Fine-Reader'dır. Ancak hâlâ bazı lehçelerde Türkçe karakterler ya tam tanınamayıp başka karakterle karşılanmakta ya da kare şeklinde çıkmaktadır.

Buraya kadar üzerinde durduğumuz resmî metne dönüştürme çalışmaları matbu metinlerle ilgiliydi. Bu, Türkçede bir dereceye kadar başarılıdır.

İşin bir de el yazısı boyutu bulunmaktadır. El yazısını tanıma çalışmalarında ilk olarak üzerinde durulması gereken husus, harflerin ve kelimelerin nasıl bölümleneceğidir. Hangi nokta kümelerinin kelime, hangilerininse harfi temsil ettiğinin tespiti oldukça zor bir işlemdir. Zorluk doğuran bir diğer taraf da el yazısının kişiden kişiye değişmesidir. Bu özellik, belki de tanımlamayı en çok zorlaştıran ve her defasında değişken hâle getiren kritik bir faktördür.

Bu gün hem Türkçe Lâtin harfli el yazısı metinler hem de Türkçe Arap harfli el yazısı metinler üzerine çalışmalar yürütülmektedir. Ancak bunlarla ilgili OCR çalışmalarında henüz istenilen düzeyde başarı elde edilememiştir.

4. Alfabe Çevirisi Çalışmaları

Türkçeyle ilgili alfabe çevirisi çalışmaları, birkaç şekilde yürütülmektedir. Türk-çe karakterler ("ç, ş, ğ, ı, ö, ü") dikkate alınarak yazılan bir metnin İngilizce karakterlere dönüştürülmesi ya da İngilizce karakterler dikkate alınarak yazılmış Türkçe bir metnin Türkçe karakterlere dönüştürülmesine yönelik programlar, bu tür çalışmaların birinci ayağını oluşturmaktadır. İnternetin hızlı bir şekilde hayatımıza girmesi ve Türkçe karakter taşıyan bazı iletilerin okunamaması sebebiyle bu tür program çalışmalarına ihtiyaç duyulmuştur.

Sabancı Üniversitesinden Özlem Çetinoğlu, Türkçe karakterlerin yüklü olmadığı bilgisayarlara Türkçe yazı göndermek zorunda kalındığında iletilerin okunamadığını düşünüp Kasım 2003'te Türkçe metni otomatik olarak Türkçe karakterlerden arındıran ve söz konusu sıkıntıyı gideren bir program hazırlamıştır. Bu programın adı "Türkçe Karakter Temizleme Programı"dır

ve Sabancı Üniversitesinin web sayfasında aktif sayfa olarak hizmet vermektedir. Elinizdeki Türkçe metinleri bu şekilde değiştirmek için, metni kutucuğa yazarak veya yapıştırarak “Metni Gönder” tuşuna basmanız yeterlidir. Elde edilen metni kopyalayarak dilediğiniz programda kullanabilirsiniz. Bu şekilde değiştirilen yazıyı alan kişiler, istedikleri takdirde aynı yerde hizmete sunulan “Yazım Türkçeleştirme Programı”nı kullanarak gönderdiğiniz metni eski haline getirebilir, yani metnin Türkçe karakterli hâlini yeniden elde edebilirler. Dilerseniz, kendinize gelen metinde de aynı işlemi uygulayarak uzun müddet uğraşmak zorunda kalmaktan kurtulmuş olursunuz. “Yazım Türkçeleştirme Programı” da Gökhan Tür tarafından Ocak 2000’de yazılmıştır.

Alfabe çevirisiyle ilgili ikinci tip çalışma ise, Kiril veya Arap harfli Türkçe metinleri Lâtin harfelerine dönüştürmeye yöneliktir.

Bu gün Türk dünyasında Lâtin, Kiril ve Arap alfabesi kullanılmaktadır. Çin’deki Uygur Türkleri Araḫ, İran’daki Türkler Arap asıllı Fars, Bağımsız Türk cumhuriyetleri ve Türk topluluklarının bir kısmı Lâtin, bir kısmı ise Kiril alfabesi kullanılmaktadırlar. Alfabede birlik, dilde birliğin en temel kesiti olduğu için 1990’lı yıllarda bu konuda çok çaba harcanmıştır. Bu çabalar günümüzde biraz gevşemiş, ancak alfabe dönüşümleriyle ilgili bilgisayar merkezli çalışmalar artmaya başlamıştır.

Arap harfli metinleri Lâtin harflerine dönüştürme çalışmaları son yıllarda ağırlık kazanmakla birlikte; Arap yazı sisteminin bitişken olması, sağdan sola yazılması, harekesiz metinlerin daha çok ünsüzler sistemine dayanması ve okunma alternatiflerinin zaman zaman birden fazla olması gibi sebeplerden dolayı oldukça ağır ve başarı oranı düşük bir şekilde devam etmektedir. Bu konuda İran’da Fars alfabesiyle Azerice yayınlanan *Varlık* dergisi metinlerini Lâtin harflerine dönüştüren bir program Ali Rıza Sarrafî ve arkadaşları tarafından geliştirilmiştir. Ancak bu program, bilgisayarda dizilmiş şekli bulunan Fars alfabesiyle yazılmış metinleri çevirmektedir. Oysa OCR programı yardımıyla kâğıda basılı metinlerin çevrilmesi çok daha büyük önem arz etmektedir. Osmanlı metinleri söz konusu olduğunda bu konuyla ilgili çalışmaların önemi bir kat daha artmaktadır.

Kiril harfli metinlere gelince; bu metinlerin ilk bakışta tek görüldüğü hâlde iki harfle karşılanan karakterler bulundurması, duruma göre farklı şekillerde okunup yazılabilmesi, her Türk lehçesinin kendine göre farklı karakterlerinin olması gibi dezavantajlarının yanında bitişken olmaması ve Lâtin alfabesine paralel bir şekilde soldan sağa yazılması gibi avantajlar bakımından Kiril harflerini Lâtin’e dönüştürme çalışmaları daha başarılı bir şekilde yürütülebilmektedir.

Kiril’den Lâtin’e çeviri yapan programlar, yakın zamana kadar genellikle Rusça metinler ve Rusça karakterler esas alınarak hazırlanmış; bütün Türk lehçelerini kapsayacak benzer bir program yazılmamıştır.

Sovyetler Birliği’nin dağılmasından sonra yapılan Türkiye merkezli toplantılarda Türk cumhuriyet ve topluluklarının Lâtin alfabesine geçmesi konusunda büyük çaba harcanmıştır. İşte bu çabalar, zaman zaman Kiril’den Lâtin’e çeviri yapan programları da gündeme getirmiştir.

Bildiğimiz kadarıyla bu konudaki ilk ciddi çalışma, TÜBİTAK Marmara Araştırma Merkezi’nde Yapay Zeka Bölüm Başkanlığı yapmış olan Şakir Kocabaş ve ekibi tarafından yapılmıştır. Kocabaş ve ekibinin hazırladığı program, taranan Kiril harfli metinlerin TIF formatındaki görüntülerini karakterlere ayıran, bu karakterleri karşılık gelen harflerle birlikte veri tabanında saklayan, yeni taranan harfleri ‘öğrenilmiş’ harflerle karşılaştıran bir programdır. Ancak bu program, nedense kullanıma sunulmamış ve hayata geçmemiştir.

Kiril-Lâtin çevirisi konusunda kafa yoranlardan birisi de Kemal Altıntaş’tır. Altıntaş, kardeşiyle birlikte Kiril harfleriyle yazılmış Kırım Tatarcası metinlerini Lâtin harflerine çeviren bir “makro” hazırlamıştır. Söz konusu program tek bir Türk lehçesi için yazılmıştır.

Benzer bir programın Hollanda’da yazılmaya çalışıldığını, ancak Türkçe karakterler tanımlanamadığı için bundan vazgeçildiğini Kemal Altıntaş bir sohbet sırasında bana aktarmıştı.

Yakın zamanda, Karaçay-Malkarca metinleri Lâtin harflerine çeviren bir programın Ahmad Salpagarov tarafından hazırlandığını bir internet sayfasından öğrendim.

Ben kendim, bütün Türk lehçelerinin Kiril harfli metinlerini Lâtin’e çeviren bir prog-ram üzerinde beş yıl (1998 yılından beri) çalıştım. Daha önce kişisel bilgisayarlarda kullanılacak “exe” uzantılı bir versiyonunu yazdığımız programın geçtiğimiz birkaç ay içerisinde ASP tabanlı biçimi de yazılıp bitirildi.

Şimdi sizlere Türkiye Cumhuriyeti Kültür ve Turizm Bakanlığının web sayfasında aktif sayfa olarak ücretsiz hizmete sunulan bu programla ilgili bilgiler sunmak istiyorum.

KİRİL-LÂTİN ÇEVİRİ PROGRAMI: *GASPIRALI*

“Gaspıralı”, yirmi Türk lehçesinin Kiril harfleriyle yazılmış metinlerini Lâtin harflerine çeviren hem ASP tabanlı biçimi hem de “exe” uzantılı biçimi bulunan bir programdır. Bu program, “Kiril”le yazılmış web sayfalarının birçoğunu “Lâtin”e çevirebilme özelliğine de sahiptir.

Bu Program Nasıl Ortaya Çıktı?

1996’da bitirdiğim doktora tezimde Kiril harfleriyle yazılmış yedi yüz dört (704) sayfalık bir metni Lâtin’e çevirmek için yaklaşık bir yılımı harcamıştım. Bunun bir bilgisayar programı yardımıyla çok kısa sürede yapılabileceğini düşündüm ve 1998 yılında Kiril-Lâtin çevirisiyle ilgili çalışmalara aktif olarak başladım. Birkaç yazılımcıyla görüşükten sonra, böyle bir işlemi yapabilmek için tarayıcı yanında Kiril harfli metinleri resim olarak taradıktan sonra metne dönüştürebilecek OCR programına ihtiyaç olduğunu ve bunun Türkiye’de bulunmadığını öğrendim.

Türkmenistan’a giderek Kiril harflerini metin hâlinde taramayı sağlayan programlar üzerine araştırma yaptım. Orada Kiril harflerini tarayıp metne dönüştüren “FineReader” adlı bir OCR programının varlığından haberdar oldum ve bu programın iki farklı versiyonunu Türkiye’ye getirdim. Tarayıcıma bu prog-ra-mı yükledim. Artık Kiril harfli bir kitap sayfasını metin formatında tarayıp “word”de kullanmayı başarabiliyordum. Önce kritik ikili okunuşları (Kiril “e” => Lâtin “e” veya “ye-” gibi) tespit edip bunları *bul-değiştir* komutu aracılığıyla değiştiriyordum. Ancak bu metot belirli bir başarı elde etmemi sağlasa da oldukça uzun ve yorucu bir işlem gerektiriyordu.

Daha sonra bunun bir adım ilerisini düşünmeye başladım. Benim yaptıklarım otomatik olarak bir bilgisayar yazılımıyla yapılabilirdi ve bizi çok daha verimli sonuçlara götürebilirdi. Bu konuyla ilgili araştırmalarımı ciddi bir şekilde tartışacak kimse bulamayınca konuyu internet üzerinden araştırmaya karar verdim. Araştırmalarım esnasında Cyrillic-Latin Converter (Kiril-Lâtin Çevirici) adlı programların internet ortamında var olduğunu fark ettim. Ancak bunların hemen hepsi Rus alfabesinde bulunan karakterleri çevirmekteydi ve birçoğu da tek karakterli olan harfleri çift (meselâ Kiril harfindeki “ıı” “yi” “sh” şeklinde) dönüştürüyordu. Benzer bir programın Azerice için yapılan sürümü de Lâtin karakterleri epeyce yanlış çeviriyordu.

Bu arada dilci meslektaşlarımdan Doç. Dr. Çetin Pekacar ile bu konuyu internet üzerinden bir müddet tartıştık ve gerekirse bilgisayar programı yazmayı öğrenip bunu yapmak gerektiği sonucuna vardık. Bu aşamada Kemal Altıntaş ve kardeşi tarafından Kırım Tatarcası için hazırlanan “makro”yu Çetin Bey ile birlikte bir süre kullandık ve verimliliği üzerine tartıştık. Bu makro uzun sayfalarda çok fazla zaman harcıyordu. Bu yüzden ben uzantısı “exe” olan ve seri çalışabilen bir programın yazılması gerektiği kanaatine vardım.

Uzun görüşmeler ve tartışmalardan sonra 2001 yılı Ekim ayı içerisinde özel bir firmanın yazılımcısıyla birlikte Delphi’de “exe” uzantılı bir program yazmayı kararlaştırdık. Yaklaşık bir aylık hazırlık devresinden sonra Türkmenceden ilk denemeyi yapmayı plânladık. Programcı, yazdığı programı ilk kez uygulamaya koyduğunda büyük bir hayal kırıklığı yaşadım ve çok moralim bozuldu. Lâtin’e çevrilen bütün karakterler, soru işareti şeklinde çıkmıştı. Yarım saat sancılandıktan sonra yanlışımları buldum. Biz Türkmen Kirili’nin ASCII kodlarını dizilmiş metinlere göre oluşturmuştuk. Oysa tarayıcıdan geçirdiğimiz metinlerdeki karakterlerin ASCII kodlarını kullanmamız gerekiyordu. Fakat ne yazık ki taranmış metinlerde Türkçe karakterlerin hemen hepsi kendini gizliyor ve kare olarak görünüyordu. Nihayet gizlenmiş karakterlerin kodlarını çözmeyi de başardık. Sonuçta yazılan programın Türkmencedeki başarısı % 95’in üzerine çıkmıştı. Artık diğer lehçelere geçebildik.

Gün geçtikçe, düşlediğim program yavaş yavaş bireysel amaçlar için yazılmış bir program olmaktan çıkıyor, bütün Türk dünyasının kullanımına sunulacak bir yazılıma dönüşüyordu. Yaklaşık bir yıllık çalışmanın ardından yirmi (20) lehçenin taranmış metinlerinde alfabe çevirisi yapabilecek başarılı bir prog-ra-m tamamlanmış oldu. Bu lehçeler şunlardır: **Altayca, Azerice, Başkurtça, Çuvaşça, Gagavuzca, Hakaşça, Karaçayca, Karakalpakça, Kazakça, Kazan Tatarcası, Kırgızca, Kırım Tatarcası, Kumukça, Malkarca, Nogayca, Özbekçe, Tuvaca, Türkmençe, Yakutça, Yeni Uy-gur-ca.**

Azerbaycan ve Türkmenistan cumhuriyetleri Lâtin harflerine geçtiği için programın, bu lehçelere ait metinleri hem Türkiye’deki bilimsel çalışmalarda kullanılan Lâtin’e hem de Yeni Azeri ve Türkmen Lâtini’ne çeviri yapan “exe”-le-ri bulunmaktadır. ASP tabanlı program da aynı özelliklere sahiptir.

Türk dünyasında dil ve düşünce birliği konusunda neredeyse bütün ömrünü harcayan Gaspıralı İsmail Bey'i bir daha Türk dünyasında hatırlatmak ve idealini diri tutmak amacıyla, hazırlanan programa tarafımdan "Gaspıralı" adı verilmiştir.

Program Hakkındaki Bazı Görüşler

"Çalışmanız hem bizlere kolaylık getirecek, hem de Türk dünyasında yeni Türk yazısının yaygınlaşması mümkün olacak".

Prof. Dr. Şükrü Halûk AKALIN (Türk Dil Kurumu Başkanı).

"Böyle bir çeviri programının yüzde yüz başarı ile çalışması, yalnızca bilimsel çalışmaların kolaylığını sağlamayacak, Kiril harfleri ile yazılmış Genel Türklüğün seçkin eserlerini hem Türkiye Türklerinin ilgisine sunmada hem de Kiril'den Lâtin'e geçen ve/ve-ya geçecek Türk cumhuriyetlerinin Sovyet dönemine ait iyisiyle-kötüsüyle bütün yazılı ürünlerini gelecek kuşaklara aktarmalarında hizmet edecektir".

Prof. Dr. F. Sema Barutçu ÖZÖNDER (Ankara Üniversitesi Dil ve Tarih-Coğrafya Fa-kül-tesi Çağdaş Türk Lehçeleri ve Edebiyatı Bölümü Başkanı).

Programla ilgili kısa ve özlü Bilgiler

Amaç: Sayıları 20'yi bulan Kiril harfleriyle yazılmış Çağdaş Türk lehçelerine ait metinleri OCR (Optical Character Recognition = Optik Karakter Tanıma) programı yardımıyla metin hâlinde taradıktan sonra her lehçenin özel karakterlerini dikkate alarak elektronik ortamda Lâtin harflerine çevirmek. Aynı zamanda Kiril ile yazılmış web sayfalarını da Lâtin'e dönüştürmek.

Kapsamı: Yapılan işlem bir dil çevirisi (lehçe aktarması) değil, alfabe çevirisidir.

Getirdiği Kolaylıklar: Bu güne kadar Kiril harfli metinler üzerine master ve doktora yapanlar, alfabe çevirisine en az bir yıl kadar zaman harcıyorlardı. Bu durum, bir süre sonra bıkkınlığa ve enerji kaybına sebep oluyordu. Artık bu yazılımla bir yıllık iş, bir veya iki hafta içerisinde bitirilmiş olacak.

Öte yandan hem şahsî hem de kurum bazında bu konularda çalışmalar yapılmaktaydı ve bunlar da uzun zamanda, yorucu bir şekilde bitirilebiliyordu.

Bundan sonra söz konusu yazılımla bu çalışmalarda adeta hamallık niteliği taşıyan alfabe çevirimi, çok kısa zamanda yapılabilecektir.

Aynı yazılımla Azerbaycan ve Türkmenistan gibi Lâtin harflerini kabul etmiş olan Türk cumhuriyetlerinin istenilen Kiril harfli metinlerinin yeni harflere aktarımında ve kültür erozyonunun önlenmesinde önemli rolü olacağını düşünüyoruz. Bundan sonra Lâtin harflerini kabul edecek olan bütün Türk cumhuriyet ve toplulukları için de aynı durum söz konusudur.

Ortaya konulan programın Lâtin harflerini kabul etmeyi düşünüp de eski mirasın yeni nesillere aktarılması konusunda derin sıkıntılar duyan Türk cumhuriyet ve topluluklarının bu tereddütlerinin giderilmesinde de büyük bir hizmet yapacağını ve kolaylıklar getireceğini düşünmekteyiz.

Başarı Oranı: Ünlüleri çakışan Özbekçede başarı oranı yaklaşık % 70-80 civarındayken, diğer lehçelerde bu oran % 100'lere yaklaşmaktadır.

Özbekçede "a" ile "ä", "ı" ile "i", "o" ile "ö", "u" ile "ü" aynı karakterle gösterilmektedir. Bu durumda daha çok geçen karakter dikkate alınarak programın Özbekçesi yazılmış ve hata payı en aza indirilmeye çalışılmıştır.

Duyulan İhtiyaç ve Önemi: "Türk Cumhuriyetleri Bilgi Teknolojileri Çalışma Grubu" nun yapmış olduğu bir toplantıda "Türkiye Bilişim Derneği" temsilcisi Ümit Karakaş, Türk cumhuriyetleriyle ülkemiz arasında alfabe farkının üzerinde önemle durulması gerektiğini vurgulamış; alfabe farklılığı giderilmez ve bu sorun bilgisayar ortamında çözülmezse bir yere varamayacağımızı ve köklü projelerin ortaya konulamayacağını belirterek böyle bir programa duyulan ihtiyacı vurgulamıştır.

Konuya duyulan ihtiyacı belirtmesi bakımından Karaçay Türkü olduğunu öğrendiğim Ahmad Salpagarov adlı bir internet gezgininin mesajlarından birini aşağıya alıyorum:

"Men Kiril Latin köçürüw (buruw) transliterator programma izleyme. Kim aytır, üretir?" (Kiril-Lâtin çeviri programı arıyorum. Kim söyleyebilir, haberdar edebilir?).

Kullanım Alanları: Türkiye başta olmak üzere Türk cumhuriyet ve topluluklarında; Türk lehçeleriyle ilgili çalışmaların yoğun olarak yapıldığı bütün ülkelerde Kiril harfli metinleri Lâtin'e çevirmek isteyen her şahıs ve kurum bu programdan faydalanabilecektir.

Gaspiralı programı, Kültür ve Turizm Bakanlığının aşağıda verilen web sayfasında ücretsiz olarak hizmete sunulmuştur: <http://www.kultur.gov.tr/portal/gaspirali>

Aynı programın, benim düşlerimden biri olan “Türk Dünyası Metin Bankası”nın kurulmasına zihni hazırlık yapmada ve metinlerin oluşturulmasında büyük bir hizmeti olacağını düşünmekteyim.

Beni sabırla dinlediğiniz için hepinize teşekkür ediyorum.

KAYNAKLAR

Basılı Kaynaklar:

- Atalay, V., M. Ozçifingir, N. Yalabik, 1990. “Computer Recognition of Ottoman Text”, Proc. International Symposium on Computer and Information Sciences-V (ISCIS). Capadocia, Turkey, November.
- Cebiroğlu, G., 2002, Sözlüksüz Köke Ulaşma Yöntemi, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü. Yayınlanmamış Yüksek Lisans Tezi, İstanbul.
- Cebiroğlu G., C. Tantug, E. Adalı, Y. Erenler, 2003, “Sentetik Türkçe Kelime Kökleri Üretimi”, Tainn 2003. Çanakkale.
- Cebiroğlu G., E. Adalı, 2002, “Sözlüksüz Köke Ulaşma Yöntemi”, TBD 19. Bilişim Kurultayı, İstanbul.
- Karşlıoğlu, M., 2000, “OCR Büyüsü”, Chip Dergisi, Kasım.
- Köksal, A., 1975. Automatic Morphological Analysis of Turkish, PhD Thesis, Hacettepe University, Ankara.
- Oflazer, K., 1994, “Two-level Description of Turkish Morphology”, Literary and Linguistic Computing, Vol. 9. Number 2.
- Oflazer, K., 1998, “Morphological Analysis”, Chapter in Syntactic Wordclass Tagging Hans van Halteren. Editor. Kluwer Academic Publishers.
- Öztürk, A., 1998, Osmanlıca Karakterlerin Bilgisayar Destekli Tanınması, Gebze Yüksek Teknoloji Enstitüsü. Yayınlanmamış Yüksek Lisans Tezi, Gebze.
- Öztürk, A., S. Güneş, Y. Özbay, 2000. “Multifont Ottoman Character Recognition” Proceeding of the 7th IEEE International Conference on Electronics Circuits and Systems, ICECS, December 17-20, pp. 945-949. Portemilo Hotel, Jounieh, Lebanon.
- Sabuncu, T., 1993, PC World, Mart, Sayı: 25, s. 95-96.
- Say, A. C. Cem, 2001, “Understanding Arithmetic Problems in Turkish”, International Journal of Pattern Recognition and Artificial Intelligence, Vol. 15, pp. 359-374.
- Tika, Türk Dünyası Bülteni, 2002, Aralık, Sayı: 5. s. 1.

İnternet Kaynakları:

- <http://salpagarov.narod.ru>
- www.cs.itu.edu.tr/~gulsen
- <http://www.freetranslation.com>
- <http://www.mehkara.netteyim.net>
- http://gb.lpt.ru/gb_arc.php?bid=1211
- http://gb.lpt.ru/gb_arc.php?bid=1211
- <http://www.kets.com/tr/sss/ocr.faq.htm>
- <http://www.ru.edu.tr/turktili/turko.html>
- <http://fens.sabanciuniv.edu/TL/ascii.html>
- <http://fens.sabanciuniv.edu/TL/deascii.html>
- <http://www.cs.bilkent.edu.tr/~ko/pubs.html>
- <http://www.gyte.edu.tr/jeodezi/kurultay.doc>
- <http://www.cs.itu.edu.tr/~gulsen/nlp/nlp.html>
- <http://www.cs.itu.edu.tr/~gulsen/nlp/nlp.html>
- <http://www.ceng.metu.edu.tr/~volkan/pub.html>
- <http://www.elektrik.selcuk.edu.tr/ogrgor/ozbay.htm>
- <http://www.dtm.gov.tr/turkcumhuriyetleri/toplantı3.htm>
- <http://www.dtm.gov.tr/turkcumhuriyetleri/toplantı25.htm>
- <http://www.teknoturk.org/docking/yazilar/tt000044-yazi.htm>
- <http://www.teknoturk.org/docking/yazilar/tt000038-yazi.htm>
- <http://salpagarov.narod.ru/alfabet/alantrans/alantranslit.htm>
- tdb.org.tr/listeler/kamu-int-cg/2003/May/att-0009/Kal_p.doc
- <http://www.geocities.com/nurayhaliloglu/yapayzeka/yapay3.htm>
- <http://www.fbe.ktu.edu.tr/tezler/bilgisayar/yukseklisans/99-t1178.htm>
- <http://www.eurasianet.org/resource/cenasia/hypermail/200102/0126.html>
- <http://research.po.metu.edu.tr/TR/EAK/MuhF/bilgisayarmuhendisligibolumu.htm>