

**GASPIRALI: KİRİL LATİN ALFABELERİ ARASINDA  
ÇEVİRİ AMAÇLI PROGRAM**

**Mehmet Kara**

**Türklük Araştırmaları Dergisi**  
**16**  
(Ayrıbasım)



Marmara Üniversitesi Fen-Edebiyat Fakültesi

İstanbul, 2004

**Değerlendirme- Tarihî Kaynaklar**

***GASPIRALI:*  
KİRİL VE LÂTİN ALFABELERİ  
ARASINDA ÇEVİRİ PROGRAMI**

**MEHMET KARA**

(Kırıkkale üniversitesi)

Türkiye’de artık diğer alanların yanı sıra genelde dilbilim ve özelde Türk lehçeleriyle ilgili olarak da bilgisayar merkezli çalışmalar giderek artmaktadır. Bu güne kadar çağdaş Türk lehçeleri üzerine yapılan bilgisayar merkezli çeviri çalışmalarını dörde ayırmak mümkündür: Alfabe çevirisi çalışmaları; sesi yazıya, yazıyı sese dönüştürme çalışmaları; resim olarak algılanmış yazıyı metne çevirme ya da yaygın adıyla optik karakter tanıma (*optical character recognition=OCR*) çalışmaları; lehçeden lehçeye makine çevirisi çalışmaları.

Biz, bu yazımızda Kiril alfabesinden Latin alfabesine çeviri yapan "Gaspirali Ver 1.0" adlı ("exe" uzantılı) alfabe çevirisi programı ve bu programın ASP tabanlı biçimi üzerinde duracağız. Ancak, bu bahse geçmeden önce, bugün Türkçe metinlerde kullanılan alfabeler ve alfabe çevirisi çalışmaları konularına kısaca değinmek istiyoruz.

Günümüz Türk dünyasında Latin, Kiril ve Arap alfabeleri kullanılmaktadır. Çin’deki Uygur Türkleri Arap, İran’daki Türkler Arap asıllı Fars, Bağımsız Türk cumhuriyetleri ve Türk topluluklarının bir kısmı Latin, bir kısmı ise Kiril alfabesi kullanmaktadırlar. Alfabede birlik, dilde birliğin en temel kesiti olduğu için 1990’lı yıllarda bu konuda çok çaba harcanmıştır. Gün-

müzde bu yöndeki gayretler biraz gevşemiş gözükse de, alfabe dönüşümleriyle ilgili bilgisayar merkezli çalışmaların artmaya başlamış olması sevindiricidir.

Türkçeyle ilgili sözkonu alfabe çevirisi çalışmaları, birkaç şekilde yürütülmektedir. Türkçe 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 Üniversitesi'nden Özlem Çetinoğlu, Türkçe karakterlerin yükü olmadığı bilgisayarlara Türkçe yazı göndermek zorunda kalındığında iletilerin okunmadığını 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 Üniversitesi'nin 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 halini 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 Latin harfelerine dönüştürmeye yöneliktir. Arap harfli metinleri Latin 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 Azerbaycan Türkçesi yayınlanan *Varlık* dergisi metinlerini Latin 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 Latin alfabesine paralel bir şekilde soldan sağa yazılması gibi avantajlar bakımından Kiril harflerini Latin'e dönüştürme çalışmaları daha başarılı bir şekilde yürütülebilmektedir. Kiril'den Latin'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 Latin alfabesine geçmesi konusunda büyük çaba harcanmış, zaman zaman Kiril'den Latin'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-Latin çevirisi konusunda kafa yoranlardan birisi de Kemal Altıntaş'tır. Altıntaş, kardeşiyle birlikte Kiril harfleriyle yazılmış Kırım Tatarcası metinlerini Latin 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 Latin harflerine çeviren bir programın Ahmad Salpagarov tarafından hazırlandığını bir internet sayfasından öğrendim.

Ben, bütün Türk lehçelerinin Kiril harfli metinlerini Latin'e çeviren bir program üzerinde beş yıl (1998 yılından beri) çalıştım. Daha önce şahsî 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 Bakanlığı'nın web sayfasında aktif sayfa olarak ücretsiz hizmete sunulmuş bulunan<sup>1</sup> bu programla ilgili bilgiler vermek istiyorum.

### **Kiril-Latin Çeviri Programı: *Gaspirali***

"Gaspirali", yirmi Türk lehçesinin Kiril harfleriyle yazılmış metinlerini Latin harflerine çeviren hem ASP tabanlı biçimi hem de "exe" uzantılı biçimi bulunan bir programdır. Bu program, Kiril ile yazılmış web sayfalarının birçoğunu Latin'e çevirebilme özelliğine de sahiptir.

1996'da bitirdiğim doktora tezimde Kiril harfleriyle yazılmış 704 sayfalık bir metni Latin'e çevirmek için yaklaşık bir yılımı harcamıştım. Bunun bir bilgisayar programı yardımıyla çok kısa sürede yapılabileceğini düşündüm ve 1998 yılında Kiril-Latin çevirisiyle ilgili çalışmalara aktif olarak başladım. Birkaç yazılımcıyla görüştüğümüzden 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. Bilgisayarına bu programı yükledim. Artık Kiril harfli bir kitap sayfasını metin formatında tarayıp *Word*'de kullanmayı başarabiliyordum. Önce kritik ikili okunuşları (Kiril "e" => Latin "e" veya "ye" gibi) tespit edip bunları "bul-değiştir" komutu aracılığıyla değiştirmeye başladım. 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 Türk lehçelerindeki birçok karakteri çevirmeyip sadece Rus alfabesinde bulunan

<sup>1</sup> <http://www.kultur.gov.tr/gaspirali>

karakterleri çevirmektedir ve bazı Türkçe karakterleri de ikili olarak (meselâ "ş"yi "sh" şeklinde) dönüştürüyordu. Benzer bir programın Azerbaycan Türkçesi için yapılan sürümü de Latin karakterleri epeyce yanlış çeviriyordu. Dilci meslektaşarımdan 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 serî çalışabilen bir programın yazılması gerektiği kanaatine vardım.

Uzun görüşmeler ve tartışmalardan sonra 2001 yılı Ekim ayı içerisinde özel bir firmanın yazılımcısıyla hafta sonları buluşup *Delphi*'de "exe" uzantılı bir program yazmayı kararlaştırdık. Yaklaşık bir aylık hazırlık devresinden sonra Türkmence'den ilk denemeyi yapmayı plânladık. Programcı, yazdığı programı ilk kez uygulamaya koyduğunda büyük bir hayal kırıklığı yaşadım ve çok moralim bozuldu. Latin'e çevrilen bütün karakterler, soru işareti şeklinde çıkmıştı. Yarım saat sancılandıktan sonra yanlışıımı 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. Nihâyet gizlenmiş karakterlerin kodlarını çözmeyi de başardık. Sonuçta yazılan programın Türkmence'deki başarısı % 95'in üzerine çıkmıştı. Artık diğer lehçelere geçebilirdik. Gün geçtikçe, düşlediğim program yavaş yavaş ferdî maksatlı için yazılmış bir program olmaktan çıkıyor, bütün Türk dünyasının kullanımına sunulacak bir yazılıma dönüşüyordu. Yaklaşık bir yıllık çalışmanın ardından 20 lehçenin taranmış metinlerinde alfabe çevirisi yapabilecek başarılı bir program tamamlanmış oldu. Bu lehçeler şunlardır: Altay, Azerbaycan, Başkurt, Çuvaş, Gagavuz, Hakas, Karaçay, Karkalpak, Kazak, Kazan Tatar, Kırgız, Kırım Tatar, Kumuk, Malkar, Nogay, Özbek, Tuva, Türkmen, Yakut, Yeni Uygur.

Azerbaycan ve Türkmenistan cumhuriyetleri Latin harflerine geçtiği için programın, bu lehçelere ait metinleri hem Türkiye'deki bilimsel çalışmalarda kullanılan Latin'e hem de Yeni Azerbaycan ve Türkmen Latini'ne çeviri yapan "exe"leri bulunmaktadır. Programın ASP tabanlı biçimi de aynı

özelliklere sahiptir. Üstelik ASP tabanlı biçim, aynı zamanda Kiril ile yazılmış web sayfalarının birçoğunu Latin'e çevirebilme özelliğine de sahiptir.

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

### **Programla İlgili Kısa ve Özlü Bilgiler**

**Amaç:** Sayıları 20'yi bulan Kiril harfleriyle yazılmış çağdaş Türk lehçelerine ait metinleri OCR programı yardımıyla metin halinde taradıktan sonra her lehçenin özel karakterlerini dikkate alarak elektronik ortamda Latin harflerine çevirmek. Aynı zamanda Kiril ile yazılmış web sayfalarını da Latin'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 yazılımın Azerbaycan ve Türkmenistan gibi Latin harflerini kabul etmiş olan Türk cumhuriyetlerinin istenilen Kiril harfli metinlerinin yeni harflere aktarımında ve kültür erozyonunun önlenmesinde önemli rolü olacağını düşünüyoruz. Bundan sonra Latin harflerini kabul edecek olan bütün Türk cumhuriyet ve toplulukları için de aynı durum söz konusudur. Ortaya konulan programın Latin 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 Özbekçe'de başarı oranı yaklaşık %70-80 civarındayken, diğer lehçelerde bu oran % 100'lere yaklaşmaktadır. Özbekçe'de "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. Başarı oranına tesir eden bir diğer faktör de basılı metinlerin kalitesidir.

"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-Latin çeviri programı arıyorum. Kim söyleyebilir, haberdar edebilir?).

Türkiye başta olmak üzere Türk cumhuriyet ve topluluklarında; Türk lehçeleriyle ilgili çalışmaların yoğun olarak yapıldığı bütün ülkelerde Kiril harfli metinleri Latin'e çevirmek isteyen her şahıs ve kurum bu programdan faydalanabilecektir. Gaspıralı programının ASP tabanlı biçimi, Kültür ve Turizm Bakanlığı'nın web sayfasında, şimdilik lehçe uzmanları ve adresten haberdar olanlar tarafından deneme amaçlı olarak kullanılmaktadır. Deneme yapmaları için beş on kişiye verilen adrese bir ay içinde iki milyon kişinin girmesi ve programı kullanması, bizi çok şaşırtmıştır. Çok yakın bir zamanda düzenlenecek basın toplantısından sonra söz konusu adres açıklanacak ve bu program ücretsiz olarak bütün Türk dünyasının hizmetine sunulacaktır. Aynı programın "exe" uzantılı biçimi de tarafımızda mevcuttur. Talep olduğu takdirde o da piyasaya verilecektir.

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



## KAYNAKLAR

### Basılı Kaynaklar:

Atalay, V., M. Ozcilingir, N. Yalabik, "Computer Recognition of Ottoman Text", *Proc. International Symposium on Computer and Information Sciences-V (ISCIS)*, Capadoccia, Turkey, November 1990

Öztürk, A., *Osmanlıca Karakterlerin Bilgisayar Destekli Tanınması*, Gebze Yüksek Teknoloji Enstitüsü, Yayımlanmamış Yüksek Lisans Tezi, Gebze 1998.

Öztürk, A., S. Güneş, Y. Özbay, 2000, "Multifont Ottoman Character Recognition", *Proceeding of the 7<sup>th</sup> IEEE International Conference on Electronics Circuits and Systems, ICECS*, December 17-20, ss. 945-949, Portemilo Hotel, Jounieh, Lübnan.

Sabuncu, T., *PC World*, 25 (Mart 1993), s. 95-96.

Tika, *Türk Dünyası Bülteni*, 5 (Aralık 2002), s. 1.

### İnternet Kaynakları:

<http://salpagarov.narod.ru>

<http://www.mehkara.netteyim.net>

<http://www.hlst.sabanciuniv.edu/TL/>

<http://fens.sabanciuniv.edu/TL/ascii.html>

<http://fens.sabanciuniv.edu/TL/deascii.html>

<http://fens.sabanciuniv.edu/TL/deascii.html>

<http://www.gyte.edu.tr/jeodezi/kurultay.doc>

<http://www.dtm.gov.tr/turkcumhuriyetleri/toplanti3.htm>

<http://www.dtm.gov.tr/turkcumhuriyetleri/toplanti25.htm>

<http://salpagarov.narod.ru/alfabet/alantrans/alantranslit.htm>

<http://www.eurasianet.org/resource/cenasia/hypermail/200102/0126.html>