

DİL ÇALIŞMALARININ MAKİNELEŞTİRİLMESİ

G. HAZAI

Sayın Meslektaşlarım,

Her şeyden önce, bana, büyük bir şeref saydığım bu dildiri olanağımı sağladığı için Türk Dil Kurumu'na teşekkür etmek istiyorum.

Bu konferansında daha sonraki tartışmaların fikir teatisini ve kongre sorunların ortaya çıkması olanağımı sağlayacağı umuduyla, yalnız birkaç sorun ve düşünce üzerinde duracağım.

Geçen çeyrek yüzyıl bilimde dev gelişmelerin ışığında geçti. En köklü gelişmeyi, şüphe yok ki doğal bilimler gösterdi. Bu gelişmeyi gözlerimizin önünde canlandırabilmemiz için, radarı, atom enerjisini, ya da insanın uzaya ulaşmasını anlamam yeter sanıyorum.

* 5.III.1966'da Ankara'da Türk Dil Kurumu'nda verilen konferansın metni.

1. Verilerin makinelerle değerlendirilmesi sürecini burada elbette basitleştirilmiş biçimde tanıtmamız gerekir. Amacımız, teknik ayrıntıları enine boyuna incelemek değil, tam aksine verilerin makinelerle değerlendirilmesi ile dilcilik çalışmaları arasındaki ilişkileri aydınlatmaktır.

2. Şüphesiz bu makineler, sadece sözlük çalışmaları için de kullanılabilir. Çünkü makinenin yalnız tek tek kelimeleri değil, aynı zamanda bu kelimelerden hareketle onların ortamını, yani bütün bir cümleyi "çıkarması" olanağı da vardır. Burada yalnızca cümle sınırını belirtecek bir işaret seçmek ve makineye iki işaret arasında yer alan bütün kelimeleri, yani pratik olarak bütün cümleyi çıkarması için kumanda vermek yeter.

3. Bizim çalışma alanımızda somut bilimsel denemeler henüz pek azdır. 1962 yılında Budapeşte'de Ural-I tipi hesap makinesinde, orta büyüklükte bir transkripsiyon metninin incelenmesinde çalıştım. Başarılı olan bu denemenin sonucunu, Berlin'de bulunmam dolayısıyla, ancak ileride yayımlayabileceğim. Konferansta örnek olarak verdiğim dil anıtının değerlendirilmesi işini ise, matematik-istatistik metodu uygulamakla beraber, elle yaptım. Olağanüstü çok zaman ve iş isteyen ve makine yardımı ile birkaç saat içinde tamamlanabilecek bu değerlendirme süreci, verilerin makine ile değerlendirilmesi işleminin araştırmalarımızda kullanılmasının kaçınılmaz bir zorunluk olduğunu bir kere daha gösterdi. Son zamanlarda ünlü altayist J.R.Krueger (Bloomington), olağanüstü büyük önem taşıyan ve parlak sonuçlar veren çalışmalarda bulundu. Konusunu Moğol dili sorunları arasından seçti. Basılmakta olan yazısı, verilerin makine ile değerlendirilmesi ile Altay dilleri araştırmaları arasındaki ilişkileri prensipler temelinde ele almaktadır ki, türkologlar için özel bir önem taşımaktadır.

Başka bir ilim dalında çalışan araştırmacıları ilgilendiren yalnızca parlak sonuçlar değil, aynı zamanda “nasıl” sorusudur, bilimsel metod sorunudur. Şüphe yok ki, bilimsel devrimde bu sorunun kesin rolü olmuştur.

Birçok metodik etken arasında, bilimsel çalışmaların belirli dönemlerinin makineleştirilmesi şüphesiz başta gelmektedir. Elektronik hesap makinelerinin yardımı olmadan, yukarıda andığımız bilimsel başarıları düşünebilmek bile güç olurdu. Herkesçe bilinir ki, değil bir uzay gemisini yola çıkartmak, hatta basit bir uçağın yapımına başlamak bile, ancak on binlerce matematik işlemin tamamlanmasından sonra gerçekleşebilir. Bu işlemleri, makinelerin yardımına başvurmadan, geleneksel metotlarla yapmayı denemiş olsalardı, uzay gemileri bugün daha yer yüzünden ayrılamazdı. Kısaca özetlersek, daha önceleri belki insanoğlunun birkaç kuşağının ancak başarabileceği işleri makineler bugün birkaç saat içinde, hatta birkaç dakika içinde tamamlıyorlar. Yukarıda da değindiğimiz bilim devriminin temel dayanaklarından biri işte budur.

Dil araştırmacıları, doğal bilimlerdeki bu devrimi gördükten sonra, haklı olarak şu soruyu sormaktadır: Bu metotlar bizim anlamımızda da kullanılmaz mı? Makineler bizim çalışmalarımızı da kolaylaştırılmaz mı? Bu bildirim başlığı da, öyle sanıyorum ki, bu soruya olumlu cevap verebileceğimizi göstermektedir. Amacımız da, çalıştığımız bilim dalında bu sorunların ve bununla ilgili isteklerin aydınlatılması ve bunların çözümünün ana çizgilerinin belirtilmesidir.

Dilciliğin hemen bütün alanlarında verilerin büyük yığınlarıyla çalışmak gerektiğini dilciler yakından bilir. Örneğin, dil tarihi araştırmaları için, dil anıtlarının temelli olarak incelenmesi, büyük sayıda verilerin düzenlenmesi, değerlendirilmesi gereklidir. Çünkü ne kadar çok dil anıtı değerlendirilirse, adı geçen dönemde dilin durumu üzerinde o kadar ayrıntılı bir görüş edinebiliriz. Sözlük çalışmaları da olağanüstü “materyal” ve “zaman” gerektiren işlerdir. Çünkü bildiğimiz gibi, ideal bir dil tarihi sözlüğü, bütün dil anıtlarının bütün verilerini kapsamına alan sözlüktür. Bu gibi örneklerin sayısını daha da arttırabiliriz. Ama bizim amacımız, yeni yöntemin sağladığı olanakları belirli bir sorun çerçevesi içinde etrafıca göstermek olduğundan bu birkaç örnekle yetineceğim.

Sayın meslektaşlarımın, yıllardan beri Anadolu ve Rumeli Türkçesi tarihinin devirlere bölünüşü meseleleri üzerinde çalıştığını bildiklerini sanıyorum. Bu sorun, gerçekten de büyük sayıda materyalin, birçok dil anıtının dikkate alınmasını gerektiren bir alanı kapsamaktadır. Bu soruna bu açıdan yanaşırsak, -Ármin Vámbéry'nin yüzyılımızın başında yayımlanan monografisini

(Altosmanische Sprachstudien. Leiden 1901) çıkış noktası olarak aldığımızda en iyi durumda olmadığımızı söylememiz gerekir. Gerçi o zamandan beri gerçekten çok, değerli dil anıtları yayınlandı ama, -elimiz altında bulunan büyük sayıda materyalle karşılaştırma yaparsak durumumuzu yeterli bulmaktan bugün uzağız. Bu ödevin titizliği bakımından ise özellikle yeterli sayılacak bir durumda değiliz. Çünkü, herhangi bir kriteriyumu ele aldığımızda, -bu ister *İYOR*-şimdiki zamanının, ya da *ECEK*- gelecek zamanının doğuşu olsun, ister *GİL*-morfeminin kayboluşu, ya da seslilerin düz ve yuvarlak uyumu olsun- her devirden dakik verilere ihtiyacımız olacaktır. Karşılaştırmalarda ancak bu güvenli bir temel olabilir.

Elimizdeki anıtları incelerken, belirli görüntülerle ilgili olarak, bunların dil düzeni içindeki ağırlığını, yaygınlığını ve seyrek ya da sıklığını dakik verilerle bilmek isteriz ki, bu dakik karşılaştırma temelini dayanarak, başka bir dil tarihi kesimini incelerken aynı soruları sorabilelim. Daha önceki değerlendirmelerde belirli görüntülerin yaygınlığı hakkında söylenenler, bilinen genellemeler içinde, sübjektif değerlendirmeler düzeyinde kalmaktaydı. "Seyrek", "az", "sık", "daha sık" ve benzeri kelimelerle dile getirilen değerlendirmeler, bu çalışmalar için aslında hiç bir dayanak vermezler. Kısaca diyebiliriz ki, dil anıtlarının değerlendirilmesinden bugün, görüntülerin tekrarlanmasını da yansıtan tam morfem göstergeleri bekliyoruz. Bunlar belirli morfem ve morfem-dizilerinin, bundan başka bileşik dil unsurlarının incelenmesi, dil düzenindeki yerlerinin ve ağırlıklarının tam olarak tesbit edilebilmesi için güvenilir bir temel olur. Başka bir deyimle, belirli dil anıtlarının dakik olarak, istatistik bakımdan incelenmesi gereklidir.

Ben, geniş bir transkripsiyon metni üzerinde bu metodu uygulamayı denedim. Vardığım sonuçların tanıtılması bu konuşmamın çevresini fazlasıyla aşar. Ama şunu söyleyebiliriz ki, dakik istatistik incelemeler bazen beklenmedik sonuçlar doğurdu. Örneğin, XVII'nci yüzyılda seslilerin yuvarlak ve düz uyumu sürecinin bittiğini daha söyleyemeyeceğimiz açık-seçik ortaya çıktı. Çünkü bazı morfemler için, örneğin *ÜN*, *ÜP*, ya da *Yİ/Nİ* ve *Mİ* morfemleri için daha bu sürecin hatta başlamadığını görüyoruz. Öyle ki, 409-118-43-18 sayıları eski durumu gösterirken, yeni durumu gösteren ancak bir tek sayı ile karşılaşıyoruz. Şimdiki ve gelecek zamanın hallerinde alınan sonuç da ilginç oldu. Adı geçen metinde *İYOR*- ve *ECEK*- şekillerine olağanüstü az sayıda karşılaştığımız, yani dil düzeninde bunların ağırlığının olağanüstü azımsı olduğu aydınlandı. Bu örnekler, istatistik incelemeler boyunca ortaya çıkan sonuçlar, daha da çoğaltılabilir.

Demek ki, belirli sorunların istatistik-karşılaştırma ile incelenmesi sonucunda, çeşitli görüntülerin ayrıntıları ve tümü daha dakik olarak ortaya çıkmaktadır. Dil tarihinin devirlere bölünüşü kriterlerinin ayrı ayrı dönemlerdeki özelliklerini araştırırken, bu görüntülerin zaman sınırlarını çizerek, birçok dönemle ilgili istatistik dil incelemeleri yapmamız gerektiği açıktır. Eğer bu görüntüleri kâğıda geçirerek birkaç genel özellikle yetinmeyip bu görüntülerin birbirinin yerini alan düzen tiplerindeki yeri ve ağırlığı ile ilgili sorulara dakik cevaplar vermek istiyorsak, yukardaki gibi inceleme yapmamız gereklidir.

Bu özellikte bir inceleme ise, -geleneksel metotlar uygulandığında- sonu gelmez el işi ve zaman ister. Örneğin, incelediğim anıtta 14.000 kelime ve 30.000 morfem gözden geçirdim. Vardığım sonuçlara, belki de 1.000 saatten çok bir çalışma ile erişebildim. Halbuki ödevi normal bir delikli kart makinesi birkaç saat içinde çözebilirdi.

Sayın meslektaşlarım, bütün bunlardan sonra haklı olarak, dil anıtlarının makineler yardımı ile nasıl değerlendirilebileceği sorusu aklınıza gelebilir. Bu yolla dil tarihi sözlüklerinin makineler yardımı ile düzenlenmesi sorununa da yaklaşmış olacağımız açıktır. O halde, bu çalışmalarda hangi makineler söz konusu olabilir ve bunlar yukardaki ödevleri nasıl çözebilir?

Verilerin makinelerle değerlendirilmesi, ya da dilcilik verilerinin makineler yardımı ile değerlendirilmesi temel olarak iki çeşit makine yardımı ile oluyor: Bunlardan biri delikli kart makinesi; ötekisi ise elektronik hesap makinesidir. Birçok işlem ve bu arada dilcilik verilerini değerlendirme işlemi bakımından da, iki makine çeşidi arasındaki başkalık sonuçlar açısından önemli değildir. Tabii, elektronik hesap makineleri bu işi daha çabuk yapmaktadır. Bir benzetme yapmak istersek, örneğin delikli kart makinesine 1938 modeli bir otomobil, elektronik hesap makinesine de 1965 modeli bir yarış otomobili gözü ile bakabiliriz. Belirli bir yolu bugün her ikisi de alabilir. Ancak hedefe biri daha çabuk varır, biri daha geç. Bazı alanlar var ki, bu zaman farkı çok büyük önem taşımaktadır ve bunu gözden uzak tutmanın olanağı yoktur. Örneğin bir uzay gemisi aya inerken, elektronik hesap makinesinin yardımı kaçınılmaz bir zorunluktur. Şüphe yok ki, dilcilik çalışmaları bu alanlar arasında değildir. Delikli kart makineleri de bizim çalışmalarımız için yeterlidir.

Çalışmalarımız için önemli bir soru da şudur: Bu makine nasıl çalışır, makineden neler bekleyebiliriz?

Diyelim ki, -örneğin XVI'ncı yüzyıla ait- Arap harfleriyle yazılmış 10 formalık bir dil anıtını makine ile değerlendirmek istiyoruz. Her şeyden önce

bu metni Latin harflerine çevirmek gereklidir. İşte bu bizim işimizdir. Ama bunu yaptıktan sonra artık metinle işimiz kalmaz ve makine bunu bize hazır sözlük olarak düzenler. Arap harflerinin Latin harflerine çevrilmesi, metni makineye “vermek” için gereklidir. Diyelim ki, delikli kart makinesi kullanacağız. Metnin her kelimesinden bir delikli kart hazırlanır. Her bir kelimenin karşılığı belirli bir delik kombinasyonudur. Örneğin, körler için kabartma yazıdaki nokta kombinasyonu gibi. İşimizi kolaylaştırmak için, her bir kelimenin karşılığı olarak bir sayı kombinasyonu seçiyoruz. Diyelim ki, metinde şöyle bir cümle var: *Askerün sözi tesirlü olmuşdur.*

Önceden tesbit edilen sayı alfabesine göre, / $a = 1, b = 2, c = 3, \dots$ / bu cümleyi şu sayı dizileri biçiminde yazabiliriz:

01	23	14	6	22	27	18	
<i>a</i>	<i>s</i>	<i>k</i>	<i>e</i>	<i>r</i>	<i>ü</i>	<i>n</i>	
23	20	30	12				
<i>s</i>	<i>ö</i>	<i>z</i>	<i>i</i>				
25	7	23	12	22	15	27	
<i>t</i>	<i>e</i>	<i>s</i>	<i>i</i>	<i>r</i>	<i>l</i>	<i>ü</i>	
19	15	16	11	24	5	26	22
<i>o</i>	<i>l</i>	<i>m</i>	<i>ı</i>	<i>ş</i>	<i>d</i>	<i>u</i>	<i>r</i>

Demek ki, 23–20–30–12 sayı kombinasyonu, *sözi* kelimesinin karşılığıdır ve yalnız bu kelimeyi verir. Bu yolla hazırlanan delikli kartlar, bundan sonra makineye “verilir”, ve makine şu emri alır: Kartları sayı, yani alfabe sırasına göre düzenle! Böylece kartları örneğin şu sıra ile makineden almış oluruz:

*askerün
olmuşdur
sözi
tesirlü.*

İstersek, yalnızca ters yönde kartları dizmesi emrini veririz ve böylece ters dizide bir sözlük çıkar. O zaman sıra şöyle olacaktır:

*izös
rudşımlo
nüreksa
ülriset*

Bu biçimde ters bir düzenleme takı ve eklerin durumunu aydın bir şekilde görebilmemiz açısından önemlidir.

Örnek olarak aldığımız 10 formalık metni makine birkaç saat içinde alfabe sırasına koymakta, ayrıca her kelimenin, takının, ekin metinde kaç kere geçtiğini sayısı ile, yani "izafî ihtimaliyet sayısı" ile vermektedir. Bunun yanında, istersek, bu kelimelerle nerede karşılaşacağımızı da sayfa sayısı ile sözlüğe geçirir. Yani böylece aslında, ihtimaliyet sayıları ile birlikte tam bir morfem ya da söz göstergesi, bir sözlük elimize geçiyor. Bundan sonra, ister göstergeye dayanarak, ister makineye emir vererek, bütün *sözi* ve *olmuşdur* tipindeki şekilleri elde edebiliriz. Böylece *sözi* ve *sözü* tipindeki şekillerin birbiriyle nöbetleşe geçip geçmediğini, böyle geçtiği takdirde de bunun ne oranda olduğunu çıkarabiliriz. Her görüntü ve her morfem için bu soruları sorabiliriz ve cevaplarını da alabiliriz.

Gördüğümüz gibi, elimizdeki metni ister dil tarihi incelemesi, ister dil tarihi sözlüğü için kullanalım, buna bağlı olmaksızın, el işi olağanüstü azalmakta, bir araştırmacının aylarca sürecektir böyle titiz bir işini makine birkaç saat içinde yapmaktadır.

Bunun ayrıntılarına daha çok girmek istemiyorum. Bunun yerine izin verirseniz, bu çalışmaların geleceği ile ilgili olarak birkaç söz söyleyeceğim.

İşimizin zorunlu gerekleri yararıma, verilerin değerlendirilmesinde geleneksel el işi metotlarından bir an önce kurtulmamız gerekiyor. Dilci kuşaklarının el işini makinelerin birkaç yüz saat içinde tamamlayabilmesi gerçeği açıkça gösteriyor ki, çalışmalarımızın gündemine çok daha büyük araştırma programları, daha fazla materyalin değerlendirilmesini gerektiren ödevleri almamız mümkün. Yoksa, benzer sonuçları -şimdiye kadarki tecrübelerimizin gösterdiği gibi ancak araştırmacıların sayısını arttırarak ve 20-30 yıllık çalışmalardan sonra bekleyebiliriz.

Materyalleri toplama ve düzenleme çalışmalarını ele alırsak, her yüzyıldan 5-10 ortalama 15-20 formalık dil anıtının ihtimaliyet göstergelerinin çıkarılmasını çalışma programlarımıza alabiliriz. Bu ise, dil tarihinin devirlere bölünüşü sorunun cevaplandırılması için sağlam bir dayanak sağlayacağı gibi, tarihsel sözlüklerle ilgili bilgilerimizi önemli ölçüde zenginleştirmiş olur. Çalışmaların aslan payını makine üzerine alır ve dilciye hazırlık ve değerlendirme ödevi kahr.

Bu saydığım örnekler, makineler yardımı ile değerlendirilebilecek konular dizisinin ancak çok küçük bir kesimidir. Makineler yardımı ile, el yazma varyantlarının karşılaştırılması; tarihsel ve sinkron istatistik incelemeleri; örneğin, dilde yenileşme etkisinin devirlere göre dakik olarak değerlendirilmesi; yazar-

lar sözlüklerinin düzenlenmesi vb. çalışmalar için olanaklar vardır. Başka bir deyişle, makinelerin çalışmalarımıza girmesi, dilciye ihtiyaç kalmayacağı, ya da daha az ihtiyaç olacağı anlamına gelmez. Makinelere yararlanmak demek, aynı dilcilerin, şimdiye kadar daha ağır koşullar altında belirli sonuçlar almış olmalarına karşılık, bundan sonra daha elverişli koşullar içinde daha çok ödevi çözebilme olanaklarına kavuşmaları demektir.

Verileri değerlendirme işleminin makineleştirilmesi gerçekten de dilcilik araştırmalarında “devrim” yapabilir. Bunun getireceği yenilik ve ilerleme ise, -materyallerimizi, sorunlarımızı ve mütevazı isteklerimizi gözönünde tutarsak- bizim için doğal ilimlerin parlak başarıları kadar değerli olur.