

# “EFENDİM’E SÖYLEYİYİM..”

- Bilgisayarlar dünyasında sağlanan bunca gelişmeye karşın, bu süper-sistemler niçin hâlâ insan efendilerinin sözlerini doğrudan tam olarak anlayamıyorlar? Araştırmacılar, bu sorunun çözümünü bulmak için yoğun çaba harcıyorlar.

**Arthur FISHER**

**K**onumuza, "yapay zekâ" alanında araştırma yapan uzmanlar arasında anlatılan küçük bir fıkra ile girelim: **2020 yılında, düşmanla girilen atom savaşını bir yeraltı sığınagından yöneten mareşale, üstün zekâlı ve üstün hızlı bilgisayar ilk uyarıyı bildiriyor:**  
**Bilgisayar :** "--Düşman füzeleri geliyor!"  
**Mareşal :** "--Havadan mı, denizden mi?"  
**Bilgisayar :** "--Evet"  
**Mareşal :** (Sabırsızlık ve öfkeyle) "Evet, ne?"  
**Bilgisayar :** "Evet komutanım!..."

Aslında dinleyenlerin gülüp geçtikleri bu fıkra, yıllardır insan konuşmalarını 'anlayabilecek' bir bilgisayar geliştirme çalışmaları sadece acı acı gülümsetiyor... Bu alanda araştırma yapanlar, bir bilgisayara konuşmaları anlama yeteneği kazandırmanın, ona görme yeteneği vermekten çok daha zor olduğuna inanıyorlar.

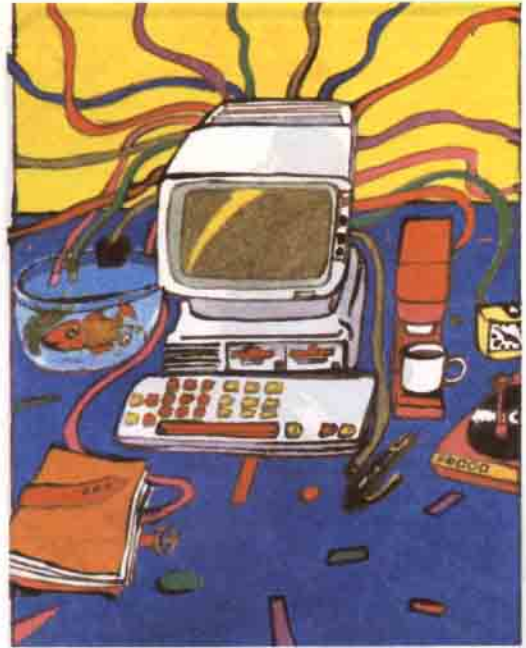
Bu zorluğa karşın araştırmalar sabırla sürdürülüyor. Nitekim, konuşmaları anlayabilecek bir sistemin yapımı, Japonların "Beşinci kuşak bilgisayarlar" araştırma programlarının başlıca hedefini oluşturuyor. ABD Savunma Bakanlığı da, 1971 yılından bu yana, konuşmaları ileri düzeyde anlayabilecek bir bilgisayar yapım projesi üzerinde çalışan M.I.T. ve Carnegie Mellon Üniversitelerine parasal destek sağlıyor. Ünlü M.I.T. Üniversitesi'nden Profesör Michael L. Dertezos, "Bu bilmeceyi kim çözerse, bilgi iletişim alanında devrim yaratacaktır." diyor.

İnsanların konuşmalarını anlayabilecek bilgisayarların yapılması, hiç kuşkusuz büyük yararlar sağlayacaktır.

Bir kere, bu yolla veriler/komutlar, bilgisayara (tuşlar aracılığı ile olduğundan) çok daha büyük bir hızla aktarılabilecektir. Konuşma yoluyla bir bilgisayara, dakikada 150-180 sözcük aktarılabılır. Bu, ortalama bir daktilografin hızından üç, üstün bir daktilografin hızından en az iki kat fazladır.

İkinci olarak, bazı durumlarda bilgisayarlı sistemlere sözle komut verme tek çözümü olabilir. Konuşabilen ama hiç hareket edemeyen sakatların durumları, acil bir olayla karşı karşıya kalan pilotlar, bu ikinci gruba örnek oluştururlar.

Ancak, bilgisayar sistemlerinin insan konuşmalarını anlayabilmeleri için yenilmesi gereken zorluklar çok büyük boyutlarda görünüyor. Gerçi bu alanda bazı sınırlı başarılar elde edilmemiş değildir: Bu çalışmalar, aşağı yukarı 40 yıl kadar önce başlamış; 1960'lı yıllarda İngilizce olarak bir kaç sözcüğü anlayan sistemler geliştirilmiş; 1970'li yıllarda ise



ticari amaçlar için kullanılabilen ve birkaç düzine sözcüğü anlayabilen bilgisayarlar üretilmiştir. Bugün ise ses tanıyan sistemler, belli komutları uygulayan video oyunlarından; telefonla banka işlemleri yapılması gibi oldukça geniş kapsamda etkinlik gösterebiliyorlar. Hatta bunların arasında, sesli komutlarla havayolu bagajlarının varış yerlerine göre ayrılmaları, helikopterlerin pilot sesiyle yönetilmesi gibi, çok ilginç ve yararlı olanları da bulunuyor. Ama bu sistemlerin hepsi de, henüz 'konuşmaları gerçekten anlayabilmek' özelliğinden çok uzak...

Bu neden böyle? Çünkü bu bilgisayar sistemlerinin anlayabildikleri sözcük sayısı hâlâ çok sınırlı; sözcükleri bir arada anlayabilmeleri yeterince gelişmiş değil; ayrıca bu bilgisayarlar, her sözcüğü duyduktan sonra bir süre ara verilmesine ihtiyaç duyuyorlar; en küçük bir gürültüden 'rahatsız' oluyorlar, kendi 'efendilerinin' sesinden başka sesi anlayamıyorlar. Oysa bu alanda araştırma yapanların istedikleri; doğal tempoda ve kesintisiz bir biçimde konuşan kişinin sesini doğru olarak anlayabilecek bir bilgisayar yapmaktır. Ayrıca bu bilgisayarın sözcük hazinesinin zengin olması, kişisel ve bölgesel telaffuz farklılıklarından ve lehçelerden etkilenmemesi, gürültüsüz ortama gerek duymaması da isteniyor.

Bu temel amaca ulaşmakta, akustik (ses bilimi) ve lügistik (dilbilim) olarak karşımıza iki ayrı problem alanı çıkıyor. Konuşulan dil, kulağımıza 'fonem' denilen hece-sesler aracılığı ile gelir. Öyleyse bilgisayarımızı geliştirmek için yapılacak şey, bu 'akustik' sinyalleri 'elektronik' olarak analiz etmek; sonra da seslerin akustik 'imzalarını' aletin belleğindeki verilerle karşılaştırmaktır. Gerçekten de bugüne değin yapılan bilgisayarlarda kullanılan yöntem budur. Ne var ki, bu yöntemin başansı çok sınırlıdır; çünkü konuşma dediğimiz olgu, sandığımızdan daha karmaşıktır.

İlk olarak sözcük sayısına bakalım: Örneğin iyi düzenlenmiş bir İngilizce sözülükte 400.000 kadar sözcük bulunur.



“İşte, keman çalan genç buydu.”

Özel ve teknik terimleri de katarsak bu sayı çok daha fazla artar. Aslında iyi eğitim görmüş bir kimse, günlük konuşması sırasında 10-15 bin kadar sözcüktен yararlanır; ancak bildiği 100.000 dolayındaki sözcüklerden herhangi birisiyle, konuşmasını her an zenginleştirebilir. Oysa bu sayıda sözcük için yukarıda değinilen “tanıma” yönteminden yararlanmak, dev bir bilgisayar kapasitesi gerektirir.

İş daha da zorlaştıran, değişik kimselerin aynı sözcükleri farklı şekillerde telaffuz etmeleridir. Bunun en iyi bilinen örneği, ‘My Fair Lady’ müzikalindeki çiçekçi kız Eliza Doolittle’dir. Oysa bugün, bu amaçla kullanılan bilgisayarlar, genellikle bir tek kişinin konuşmasından anlayabiliyorlar. Oyle ki, bilgisayarla konuşan ‘efendisi’ soğuk algınlığı olsa, bilgisayar onun konuşmasını anlayamıyor. Ayrıca bilgisayarlar bugünkü gelişme düzeylerinde, kâğıt dışırtısı, adım sesi, dışarıdan duyulan trafik gürültüsü gibi seslerden de etkileniyorlar (Bilgisayarların sözcük hazineleri genişledikçe, gürültüleri ses sanma eğilimleri de artıyor).

Bu belirtilenler, akustik ile ilgili güçlüklerdir. Tek tek ses-heceler ve sözcüklerin anlaşılmasının zorlukları yanında, sürekli bir konuşmayı anlamamanın güçlüğü üzerinde de durulmalıdır. Her şeyden önce şunu belirtelim: Ses-heceler, bir konuşmanın akışı içinde az-çok farklı özellikler gösterirler. Örneğin biz insanlar “taş”, “dört”, “at”, “üstün”, “tartaqlama sözcüklerindeki ‘t’ sesini rahatça algılayabildiğimiz halde, bilgisayar bu sözcüklerdeki ‘t’ sesini farklı kabul ediyor. İşte bu yüzden, insan konuşmasını gerçekten algılaya-

cak bir bilgisayar sistemi, hem konuşulan dili bilmek; hem de bütünden anlam çıkarmak özelliklerine aynı anda sahip olmak zorundadır.

Bu konu ile ilgili olarak, 1963 yılında insanlar bile zor duruma düşüren bir deney yapıldı: Kişilere önce karşılardaki konuşmacıların ağızlarından kısa konuşmalar dinletildi. Dinleyenler, bu konuşmaların tümünü kolayca anladılar. Sonra konuşmalar banda alındı ve banttaki sözcüklerden rastlantıyla seçilen bazıları, sırasız olarak aynı dinleyicilere dinletildi. Bu kez dinleyiciler, konuşmaların % 20-30’unu anlayamadılar. Bu durum bize, bütünden anlam çıkarmanın önemini göstermektedir.

Son on yıldır “yapay zekâ” alanında ve bilgisayarlar üzerinde önemli araştırmalar yürüten ve Stanford Üniversitesi’nde profesör olarak görev yapan Terry Winograd, dildeki len-güistik zorlukların bir sıralamasını geliştirmiş ve her basamak için ayrı ayrı örnekler vermiş bulunuyor. Aşağıda belirtilen bu basamaklara sizler de kendi örneklerinizi getirebilirsiniz:

**Birinci basamak:** Bu basamakla ilgili zorluk, deyim yerinde ise “en basit” zorluktur. Sözcüklerin birden çok anlam taşımaları ile ilgili olan bu zorluğa şu örneği verebiliriz: “İşte kemanı çalan genç buydu” cümlesi, karakolda söylenirse başka; bir konser salonunun kulisinde söylenirse başka anlam taşır.

**İkinci basamak:** Cümle kuruluşunda farklı seçenekler izlenmesi ve sözcükleri değişik anlam yüklenmelerinden do-

“Sürahiyi bardağa çarpıp kırdı.”



Ön bu anlama zorluğu için şu örnek verilebilir: “Benzinin patlayıcı olduğunu en sonunda gördü.” Şimdi bu cümlede kişi acaba benzinin patlayıcı olduğunu bizzat “yaşadı mı”, yoksa “anladı mı”?

Üçüncü basamak: Bu basamakla ilgili anlama zorluğu daha da gizlidir: Aynı cümleyi değişik tonlama ile söylemek, anlamı değiştirir, farklı yapar. “Can babasıyla birlikte İstanbul'a gitti” cümlesi ile Can babasıyla birlikte İstanbul'a gitti” cümlesindeki -koyu yazılan- iki farklı vurgu, iki değişik anlamlı cümle yaratır.

Dördüncü basamak: Bu basamakla ilgili zorluk, cümlelerin anlamlarının, konuşmanın bütününe göre kesinlik kazanmasından doğar. “Ahmet, Erzurum'lu ile evleniyor” dediğimizde, söz konusu bayanın herhangi bir Erzurumlu mu, yoksa belli bir kimse mi olduğu; daha önce söylenen ve daha sonra söylenecek cümlelerden çıkar.

Beşinci basamak: Eski edebiyatbilimcilerinin “Zaaf-ı telif” dedikleri bu anlam bulanıklığı, sık sık görülen bir karışıklıktır. “Sürahiyi bardağa çarpıp kırdı” cümlesinde, kırılan sürahi mi, yoksa bardak mıdır?

Profesör Winograd'ın açıklamalarına göre, böyle değişik anlamlara gelen cümleler, günlük yaşamda insanlara hemen hemen hiç güçlük çıkarmazlar. Çünkü biz insanlar, anlamı bir bütün olarak algılayabiliriz. Oysa bu alanda bugüne kadar geliştirilen bilgisayar sistemlerinin hiçbirisi henüz bu niteliğe kavuşamamışlardır.

Bütün bu güçlükler karşın, bu alandaki araştırmalar sabırla sürdürülüyor ve adım adım da olsa ilerlemeler kaydediliyor. ABD'de Kurzweil Yapay Zekâ Araştırmaları Merkezi-

nin kurucusu Raymond Kurzweil, kısa bir süre önce sınırlı olanaklı da olsa konuşmaları anlayabilen bir yazı makinesi geliştirilmiş bulunuyor. Geçtiğimiz baharda bu konu ile ilgilenenlerin bulunduğu bir toplantıda ilk kez sergilenen ve şimdi piyasaya verilmiş olan bu aleti, Boston'da çalışan radyoloji uzmanı bir hekim, bakınız nasıl kullanıyor: Boston'lu uzman hekim, önce aletin üst kapasite sınırı olan 1.000 sözcükten oluşan bir tıbbi sözlük geliştirmiş bulunuyor. Şimdi ise bir yandan çektiği röntgen filmlerini gözle incelerken saptadığı bulguları ve koyduğu tanıları yüksek sesle ve tane tane makineye söylüyor; makine ise o anda raporu yazıyor. Kurzweil, şu sırada sözcük hazinesi 15.000 adet olan yeni bir tipin üzerinde çalışmaktadır. Ancak bu yeni tip de yine belli bir kişinin sözlere uyacak ve her sözcükten sonra kısa bir duruşu gerektirecektir.

Konuşmaları anlayabilecek bir bilgisayar geliştirme çabaları, ünlü IBM şirketinin araştırma merkezinde de sürdürülüyor. Tabii, belirtmeye gerek yok ki burada çalışan uzmanlar da aynı sorunlarla karşılaşılıyorlar: Örneğin “Bir yata ihtiyacım var” cümlesi, sistemde “yat” sözcüğü bulunmadığından “Bir ata ihtiyacım var” biçimine dönüşüyor. Bu eksiklik ve aksaklıkları sabırla gidermeye çalışan, IBM araştırma merkezi görevlilerinden Frederick Jellinek, makinesine verdiği “Tangora” sözcüğünün ne anlama geldiği sorulunca gülümseyerek, “Bu, Guinness Rekorlar Kitabında belirtilen ve dünyanın en hızlı daktilografinin adidir” diyor.

Popular Science'den derleyerek çeviren: Melih ÖLÇER