

DİL ENGELİNİ KALDIRACAK BİLGİSAYARLAR

Dennis ROUVRAY — Gordon WILKINSON

Bugün dünyada konuşulan 5.000'den fazla dil ve lehçe vardır ve iletişim arttıkça, profesyonel tercümanlar için yeni ve yoğun iş imkânları ortaya çıkmaktadır. Bilgisayarların tercüme işlerinde kullanılması fikri ilk kez 30 yıl önce ortaya atıldı. Bu tür tercüme bazılarında kalıcı ve uzun vadede ilerlemeye hââ açık görünmüyor ise de, bazı özel uygulamalar için, bütün anadillerin 1990'ların ilk yıllarına kadar bilgisayarlar aracılığı ile tercüme edilebilecekleri hesaplanmaktadır. Halen Lüksemburg'daki Avrupa Topluluğu Komisyonlarında olduğu kadar, IIT, XEROX, GM gibi şirketlerce de milyonlarca kelime, her gün muntazaman bilgisayarlar aracılığı ile tercüme edilmektedir. Tercüme, şimdilik bilimsel, teknik, hukuksal ve idari konular ile sınırlandırılmıştır. Edebi tercüme, yazılımlar için oldukça güçtür. Yeni filizlenmekte olan makina ile tercüme yeni bir hizmet endüstrisi olup, bu endüstri dalında hizmet veren yazılım şirketleri için gelecek parlak görülmektedir.

Günümüzde artık her dilin dilbilgisi ve sözdizim kuralları analiz edilebilir olduğundan, bu kuralları kullanarak makina ile tercüme kolaylıkla gerçekleştirilmektedir. Dilin bilimsel olarak incelenmesi (Dilbilim) 1950'lerin ortalarından itibaren Zellig A. Harris ve talebesi Noam Chomsky tarafından yapılan çalışmalar ile büyük sıçrayış göstermiştir. Chomsky, önce belli bir dilin cümle yapım kurallarını, bilgisayarlarca anlaşılabilir şekilde belirledi, sonra ikinci dilin kurallarını aynı şekilde tespit ederek, bu iki dilin birbiri ile nasıl ilişkili olabileceğini düşündü ve formülendirdi. Bundan sonra, bilgisayarların hafızasını uygun sözlük ile yükleyip, bir dili diğere kolaylıkla tercüme edilebilir şekilde soktu.

Makina ile tercümede ilk ilerlemeler, Georgetown Üniversitesi'nde 1950 sonları ile 1960

Bilgisayarlar teknik dökümanları % 80 doğrulukla tercüme edebiliyorlar. Birkaç yıl içinde Arapça ve Çince dahil herhangi bir anadil den tercüme yapabilecekler.

başlarında ABD Savunma Bakanlığı'nın, Rusça teknik raporları İngilizce'ye çevirmek üzere bir program geliştirmesi için 20 milyon dolarlık bir fon ayırması ile başladı. Programlar 1960'ların başlarında kullanılabilir hale geldi; fakat çalışmalar 1966'da ABD Bilimler Akademisinin, makina ile tercüme pratiği bulmayan raporu üzerine durduruldu. Ancak bütün kısıtlara rağmen, çalışmalar, Georgetown Üniversitesi'nden Peter Toma gibi projenin güçlü savunucuları tarafından sürdürüldü. Halen birçok dil çifti birbirine klasik metot ile tercüme edilmesine rağmen, İngilizce, Fransızca, Almanca, İtalyanca gibi birçok dil birbirine bilgisayarlar aracılığı ile muntazaman tercüme edilmektedir. Tercüme programlarının yazımı pahalı olup, çok zaman almaktadır. İlk çalışmalar ABD'de gerçekleştirildiği için, dil çiftlerinden biri hep İngilizce olmaktadır. Geçen birkaç yıl içinde Çince, Japonca, ve Arapça dillerinin çiftleri de kullanılabilir hale gelmiştir. Toma tarafından yazılmış bulunan en gelişmiş yazılım paketlerinden (software package) biri SYSTRAN olarak anılmaktadır. İlk kez 1976 yılında tesis edilen bu sistem, IBM'in herhangi bir 370 veya 3400 serisi ile SIEMENS'in bilgisayarlarında 2000 sistemine uygun yazılımla birlikte kullanılabilir.

Lüksemburg'daki Avrupa Komisyonlarında makina ile tercüme yolu ile İngilizce'nin % 50'si İtalyanca'ya, Fransızca'nın % 20'si İngilizce'ye tercüme edilmektedir. Bu yolla tercüme edilen miktar, şimdilik yılda 500 sayfa civarındadır. Bütün dünyada ise yılda 500.000 sayfa bilgisayarlar ile tercüme edilmektedir.

Bu iş, 750 tercümanın klasik metot ile tüm zamanını aldığı bir iştir. Gelecek birkaç yılda bu rakamın artacağı kesindir.

Avrupa Komisyonları'na makina ile tercüme şu şekilde gerçekleştirilmektedir: Önce metin kodlanarak kelime işlemcisine (Word Processor) depolanır, buradan bir telefon hattı ile bilgisayarla gönderilen metin, tercüme edilir ve kelime işlemcisine geri döner. Diğer taraftan tercü-

me, ekranda yalnız veya kaynak metin ile birlik- te görünür. Bu şekle "Ham Tercüme" denir. Bu tercüme ehliyetli tercümanlar tarafından, hataların ve dil kalitesinin düzeltilmesi amacı ile kontrol edilir. Son şeklini almış olan tercüme, basılı veya elektronik olarak müşteriye ulaştırılır. Makina ile tercümede birçok sistem geliştirilmeye çalışılmakta olup, pek azı günlük kullanıma girmiştir.

En Başarılı Sistemler :

SYSTRAN : ABD Hava Kuvvetleri, Avrupa Komisyonu ve XEROX - GM şirketleri tarafından kullanılmaktadır.

WEIDNER : Bir kısım şirket tercüme bürosu ve ajans tarafından kullanılmaktadır.

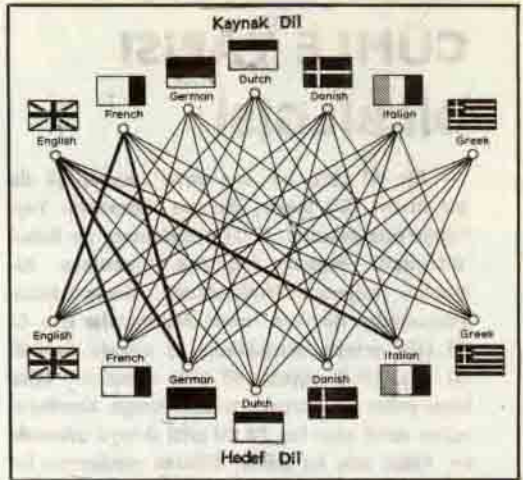
SMART : Kuzey Amerika'da Caterpillar ve Kanada Muhaceret Bürosu tarafından kullanılmaktadır.

ALPS : Tercüme büroları tarafından kullanılmaktadır.

Müşteriler makina ile tercüme sistemini satın aldıktan sonra, günün şartlarına göre yenilenmiş programları otomatikman alırlar.

Sözlükler, belli konular için ısmarlama olarak hazırlanır ve kaynak malzeme gerektirdikçe gerekli değişiklikler muntazaman yapılır.

Hızla gelişen teknolojilerde yenileme ve günün şartlarına uydurmanın önem kazandığı aşkırdır. Değişik dillerde sözlükler tercüme edilmek ve bilgisayarda depolamak üzere derlenir. Bu sözlükler içinde; kelimeler, ifadeler ve esas anlamlarından değişik anlamlara gelen kelimeler için farklı dosyalar vardır. Farklı konular için farklı sözlükler gerekir. Benzer şekilde, müşteri sözlükleri ısmarlama olarak hazırlandığı için, farklı konuların değişik sözlüklerde başarı şansı farklıdır. Avrupa Komisyonunun SYSTRAN sistemi idari konuların tercümesinde, teknik konulara göre hazırlanmış Kanada SYSTRAN'ına göre daha başarılıdır. Tercümelerde çoğunlukla, muğlak kelimeler, sözlükte olmayan ifadeler ve belirsizlikler hatalara sebep olmaktadır. Fakat bu konu genellikle uluslararası olarak ortak kavramlar üzerinde anlaşmaya varılmış bulunan teknik ve bilimsel metinler için geçerli değildir. Bir ham tercüme, tercüme edilen dilde % 80 doğruluk sağlayabilir. Geri kalanın % 15'i anlaşılabilir ve küçük düzeltmeleri gerektirici hatalardır. Kalen % 5 muğlak ve yanlış tercüme edilmiş kısımdır. LOGOS Şirketinden Chuck Staples'e göre bu rakamlar en iyi tercüme için aşağı yukarı doğrudur. Tamamıyla doğru cümleler %



Avrupa Ekonomik Topluluğu'nun 7 resmi dili vardır. Siyah hatlar, belgelerin doğrudan çeviri yönlerini gösterir.

60 ile % 80 arasındadır. Böylece, kimsenin klasik metotlarda yüksek kalite beklemediği hal ve şartlarda makina ile tercüme birçok amaç için uygun sonuçlar vaat etmektedir.

Daha önemli sınırlı düzeltme tercüme profesyonel standart getirmektedir. Tipik bir ham tercüme, yalnız bir insan ile yüksek süratle yapılabilmektedir. Metin girişi, günde adam başına 30 sayfa civarında ele alınır ve tercüme 30 dakika sürer (gerçek zaman). Düzeltme dar boğaz teşkil eder ve bu safhada kelime PROCESOR'u saat başına 5 sayfadan yukarı çıkmaz. Bütün bunlara ilaveten, tercümanların normal hataları düzelttiği "basım sonrası" kontrol işleminde gözle görünür bir zaman kazanımı söz konusudur.

MINİ BİLGİSAYAR SİSTEMLERİ

Son senelerde, makina ile tercüme konusunda ihtisaslaşmış birçok büro kuruldu. Avrupa'da ilk kuruluş 1977'de ITT tarafından Harlow - Essex'de kuruldu. PTT dil ve dilbilim merkezinden Ulla Magnusson - Murray, ilk altı ayda makinalı tercümenin, klasik metottan iki kere daha hızlı olduğunu, bu hızın ham tercümede arttığını, bazı dil çiftlerinde 3 misli daha yüksek hıza ulaştığını belirtmektedir. Belli tip sözlüklerin büyüklüğü arttıkça, tercüme ve düzeltme zamanı kısalmaktadır. Bilgisayar cihazları için yapılan yatırım göz önüne alınmaz ise ham tercüme, klasik tercümenin yarı fiyatına mal olmaktadır.

Son zamanlara kadar, makinalı tercüme sistemi, uluslararası ajanslar ve devlet sektörleri

CÜMLE YAPISI

DİLBİLGİSİ

Bir dil çifti, bir ana veya kaynak dil ile bir hedef dil içerir. Avrupa Ekonomik Topluluğu'nda resmen 7 dil onaylanmış ve kabul edilmiştir. Bunlar: İngilizce, Fransızca, Almanca, İtalyanca, Hollandaca, Danimarkaca, Yunancadır. Böylece resmi dokümanlar için 42 dil çifti ortaya çıkmaktadır. N sayıda değişik dil için (N-1) sayıda dil çifti olacaktır. Eğer İspanyolca ve Portekizce de Avrupa Komisyonu'na dahil olur ise 72 dil çifti ortaya çıkacaktır. Fakat işin iyi tarafı, dillerin yapılarının bilinmesidir. Bir dilin tarif edilebilmesi için gerekli sentaks ve dilbilgisi kuralları bu konuda yardımcı olacaktırlar. İsimler, fiiller, sıfatlar ve benzerleri muhtelif şekillerde görev yaparlar. Eğer S cümle, NP İsim İfadesi, VP fiil ifadesi, D belirleyici, AUX, yardımcı (fiil), N isim, V fiil olarak kabul edilirse, kural şöyle olur.

KURAL

- 1 S → NP + VP
- 2 VP → Verb + NP

- 3 NP → Det + N
- 4 Verb → Aux + V
- 5 Det → the, a
- 6 N → boy, race
- 7 Aux → will, can
- 8 V → win, see

Bu basit formülendirmeye "cümle dilbilgisi yapısı" denir ve "The boy will win the race" şeklindeki cümlenin ortaya çıkmasını sağlar. Aşağıda kuralın nasıl çalıştığını görelim:

i	S	
ii	NP + VP	1
iii	NP + Verb + NP	2
iv	Det + N + Verb + NP	3
v	Det + N + Verb + Det + N	3
vi	Det + N + Aux + V + Det + N	4
vii	the + N + Aux + V + Det + N	5
viii	the + N + Aux + V + the + N	5
ix	the + boy + Aux + V + the + N	6
x	the + boy + Aux + V + the + race	6
xi	the + boy + will + V + the + race	7
xii	the boy will win the race	8

Chomsky aktif ve pasif cümlelerin temelini teşkil eden yapı ile ilgili dilbilgisi geçişlerini kurallandırmıştır. Bu karmaşık kurallar ile "Çocuk yarışı kazanacak" cümlesi "Yarış çocuk tarafından kazanılacak" sekline sokulabilir.

gibi hızlı ve çok tercümenin istendiği organizasyonlar ile sınırlı idi. Bu durum değişmektedir. Örneğin bugün İTT tercüme büroları imkânları sunmakta, ABD'deki serbest yazar ve sanatkarlar tercüme ekipmanlarını ortaklaşa almak üzere konsorsiyum oluşturmaktadır. Bugün dünya üzerinde var olan iletişim sistemlerinin önerdiği imkânlar, makina ile tercümeden faydalanabilmek için muhakkak komple bilgisayar sistemlerine sahip olunması gerektiğini ortaya koymaktadır. Telefon hatları ile kilometrelerce uzaktan bilgisayara bağlanılabiliyor. Alım gücü yüksek müşteriler şüphesiz programları bilgisayarlı veya bilgisayarsız olarak alabilirler. Bilgisayar program paketleri tekli dil çiftleri için aşağı yukarı 11.000 US Dolar civarında başlar. Fakat birkaç dil çifti ve birden fazla istasyon içeren sistemler için fiyatlar hemen 10 misline çıkmaktadır. Bilgisayar programları lisanslı olan ve başlangıçta Almanca'dan İngilizce'ye tercüme konusunda ihtisaslaşmış olan LOGOS şirketi Almanya'da şimdilik 20 müşterisi olduğunu, ekipmana bağlı olarak hardware fiyatının 30.000 ile 60.000 dolar arasında olabileceğini belirtmiştir. WEIDNER haber şirketinin Bob Adams

firmasının iki sistem önerdiğini, bunlardan birincisinin mini bilgisayar esaslı ve aynı anda birden fazla tercüman terminallerini içeren sistem olduğunu, ikincisinin ise gene bilgisayar esaslı ve tek dil içeren, tek tercümanlı sistem olduğunu belirtmiştir. Bu paket 256 K hafızalı bilgisayar ile çalıştırılabilmektedir. Makina ile tercümenin çok sayıda klasik tercümanın yerini alacağı ihtimal dahilinde değildir. Bu teknoloji denince duyanlar, bu konuda çok az şey bilirlerdir. Bilgisayar program paketleri muhtemelen hiçbir zaman tam anlamı ile bir düzeltme veya editörün yardımı olmaksızın çalışamayacaklardır. Çünkü dil, makinalar ile mükemmelen tercüme edilebilmek için karmaşık ve zor bir iştir. Uzun vadede tercümanların bilgisayarlar karşısında avantajlarını kaybedeceklerini düşünmek imkânsızdır. Böylece tercümanlar yalnızca zaman alıcı ve sıkıcı monotonluktan kurtulmayacaklar, aynı zamanda daha iyi tercümeleme, daha kısa zamanda ve daha iyi fiyatla yaparlarken daha karmaşık boyutlar üzerinde düşünmek imkânına sahip olacaklardır.

New Scientist'den Çev: H. Selçuk BATUALP