

'CIBAKAYA' ADLI BİLGİSAYAR PROGRAMININ DİL ÇALIŞMALARINA YAPABİLECEĞİ KATKILAR

CEVAL KAYA

1. Giriş

- 1.1. Metin yayınında dizinin önemi,
- 1.2. Eski yöntemle dizin hazırlanışı,
- 1.3. Dil çalışmalarında bilgisayar kullanımı.

2. Cibakaya'nın özellikleri.

- 2.1. Teknik özellikler.
- 2.2. Programın kullanımı.
- 2.3. Programın yapabildiği işlevler

3. Cibakaya'nın, türdeşi programlarla karşılaştırılması

4. Sonuç

1. Giriş

Bu yazıda, dil çalışmalarına katkıda bulunmak üzere kendi yaptığım 'Cibakaya' adlı bilgisayar programını tanıtacağım.

1.1. Metin yayınında dizinin önemi

Dil araştırmasının malzemesi metindir. Araştırmanın doğru sonuçlara ulaşması için sağlam metinlere ihtiyaç vardır. Sağlam olmayan metinler üzerine kurulan araştırma, güvenilir olamayacağı gibi, kendisinin sonuçlarına dayanacak sonraki araştırmaları da yanlış yöne sevk eder. Sağlam bir metnin kurulması, yani metin yayını işi, bu sebepten, dilcilikte çok önemlidir.

İyi bir metin yayını, kendisinden yararlanmayı mümkün kılacak ayrıntılı dizinleri içermelidir. Metinde geçen bütün unsurları sistemli bir şekilde içine alan bir dizin, araştırmacıya zaman ve emek tasarrufu sağlar. Kullanıcıyı, metni baştan sona okumak ve gereken verileri fisleme zahmetinden kurtarır. Hazırlayıcı-

ya ise, metni sağlam bir şekilde kurmaya ve ikili biçimlerden kurtulmaya olanak verir. Metin, ancak dizini hazırlandıktan sonra, gerçekten sağlam olur.

1.2. Eski yöntemle dizin hazırlanışı

Bilgisayar öncesinde, eski yöntemle bir metnin dizini şöyle hazırlanırdı:

1. Fişleme: Metnin transkripsiyonu, ve gerekiyorsa çevirisi, yapıldıktan sonra, standart boyda fişler kestirilirdi. Standart fiş, A4 kâğıdının dörtte biri idi. Bir kelime bir fişe olmak üzere, metindeki bütün kelimeler fişe geçirilirdi. Fişin sol üst köşesine, kelime; sağ üst köşesine, o kelimenin çeviri metnindeki karşılığı; ortaya, o kelimenin içinde bulunduğu bağlam (cümle, cümlecik, beyit, satır vb.); sağ alt köşeye ise, kelimenin metindeki yeri (satır, beyit vb. numarası) yazılırdı.

2. Sıralama: Metnin fişlenmesi bittikten sonra, bütün fişler ilk harflerine göre, tasnif edilirdi. İlk harflerine göre tasnif sonunda, 28 öbek oluşurdu. Bu öbeklerin her birini ikinci harflerine göre yeniden tasnif etmek gerekirdi. İkinci harflerine göre tasnif sonunda oluşan yeni öbekleri de üçüncü harflerine göre ayırmak icap ederdi. Bu iş, madde başları belirinciye kadar sürerdi. Üçüncü harften sonra, genellikle madde başları ortaya çıkardı.

3. Düzenleme: Madde başları ortaya çıktıktan sonra, her bir madde başı alınır, ekli şekilleri alfabetik veya gramatikal olarak kendi içinde sıraya konurdu. Sonra, dizin metninin yazılması işine geçirilirdi. Maddeyi yazarken, fişler yardımıyla metnin bütünüünün transkripsiyonu ve çevirisi kontrol edilir, ikili şekiller ve diğer yanlışlıklar düzeltilirdi.

Bu eski yöntemle dizin hazırlarken, şu tip sorunlarla sıkça karşılaşılırdı: Fişlenirken kelimelerin atlanması veya iki kere yazılması, kelimenin metindeki yerini gösteren numaranın unutulması veya yanlış yazılması, fişlerin kaybolması gibi. Bunlarda sorun çıkmasa bile, binlerce fişin yazılması, tasnifi ve korunması ayrı bir dertti. Zekâ gerektirmeyen ve aylarca süren bu dizin hazırlama işi, zaman ve emek kaybına sebep olduğu gibi, bilim adamında usanç da oluştururdu. Bilim adamları, ancak genç ve dinç oldukları ve dışardan bir mecburiyetin bulunduğu doktora tezlerinde bu işe cesaret edebilirlerdi.

1.3. Dil çalışmalarında bilgisayar kullanımı

Bilgisayar her alanda kullanılabilir. Bu harika âlet, dılcılık alanına şu olanakları sunar: Metin yazımı, metne dayalı çeşitli dizinlerin hazırlanması, bilgi-

nin depolanması ve gerektiğinde süratle ulaşılarak kullanılması, depolanmış bilgi üzerinde her türlü istatistikî araştırmanın yapılması.

Bilgisayarlarda donanım, yani makine aksamı önemli olmakla birlikte asıl mühim olan yazılımdır. Yazılım, bilgisayarı yönlendiren programdır. Program ise, bilgisayara hangi durumda ne iş yapacağını söyleyen ayrıntılı buyruklar dizisidir. Bir bilgisayar ne kadar mükemmel olursa olsun, uygun programı yoksa, sonuç alınmaz. Bir taşıtta sürücü neyse, bilgisayarda da program odur.

Kullanım kolaylığı, dizgi olanaklarının sınırsızlığı, program kütüphanesinin zenginliği, programlarındaki standart, Türk klavye sisteminin benimsenmiş olması, birçok programın Türkçe versiyonunun bulunması, Türkçeye göre sıralama yapabilmesi gibi üstünlükleri dolayısıyla, Türkiye'deki dilciler ve öteki sosyal bilimciler arasında Macintosh marka bilgisayarlar yaygınlaştı. Basım sanayisinin de bu makinelerle çalışması, bu markanın yaygınlaşmasında etkili oldu. Türkiye dilcileri, şimdi de en çok bu markayı tercih etmektedir. 'Cibakaya' da bu marka bilgisayarlarda çalışmaktadır.

2. Cibakaya'nın özellikleri

Uygurca *Altun Yaruk* üzerinde doktora tezi hazırlarken, metni kurup eski yöntemle fişlemeye başladığım 1986 yılında bilgisayarla tanıştım. Bilgisayarın sunduğu olanakları görünce, 70.000'den fazla kelimeye sahip olduğunu tahmin ettiğim metnin dizimini hazırlayacak bir program arayışına yöneldim. Ancak, böyle bir program bulamadım, varlığını bilen kişiye de rastlamadım. Çevremde hem dilciliği, hem de bilgisayarı bilen kişi yoktu. Prof. Dr. Zikri Altun'dan programlamacılık mantığını öğrendim. Sonunda, program kitaplarından bakarak, büyük ölçüde de deneme yanılma yoluyla kendi programımı kendim yapmak durumunda kaldım. 1987 yılında programın algoritmasını çözmüş, iskeletini kurmuştum. 'Cibakaya' adını verdiğim program, 1989'da başkaları tarafından da kullanılabilir duruma geldi. Aynı yıl, dizimini bu programla hazırlayarak tezimi sundum. Sonraki yıllarda yeni işlevler ekleyerek programı biraz daha geliştirdim. Şimdiye kadar programın toplam 9 versiyonu ortaya çıktı. Burada 1 Mayıs 1995 tarihini taşıyan ve en son versiyon olan 5.2'yi tamtacağım.

2.1. Teknik özellikler

Program, yalnızca Macintosh marka bilgisayarlarda çalışmaktadır. Bu makinelerin standartlarına uygun olarak QuickBASIC dili ile yazıldı. Diskte 108K

yer kaplamaktadır. Programın kullanımı zor değildir. Ayrıca yardım dosyası da bulunmaktadır. Kopyalanması herhangi bir hukukî veya teknik sorun oluşturmaz. Her şeyi kendi içinde halleden paket bir program değildir. Metin girişi ile çıktının görüntülenmesi kelime işlem (word processor) programında, sıralama ise veri tabanı (data base) programında yapılır. Diğer bütün işlevler, Cibakaya'da gerçekleştirilir.

2.2. Programın kullanımı

Program ile dizin hazırlığına geçmeden önce, metin, kelime işlem programında yazılmış olmalıdır. Metni girerken kodlu olarak yazmak gerekmektedir. Bilgisayar, kelimenin tabanının nerede bittiğini, çekim eklerinin nerede başladığını bilemeyeceği için, bunu ona göstermek, yani kelimeyi kodlamak lâzımdır. Kodlama, isim tabanından sonra nokta (.), fiil tabanından sonra eksi (-) koymaktan ibarettir. Bır de, her kaydın (satır, beyit, paragraf vb.) başına referans numarasının yazılması gereklidir.

Örnek:

626/01 ötrü ilig.lerim örü kötür-üp ulı-yu

626/02 sıgta-yu inçe ti-p takşur-tılar

Bu şekilde bütün metnin yazımı bittikten sonra, dizinin hazırlanması işine geçilir. Dizini hazırlama sırası, eski yöntemdeki dizin yapma sırası ile aynıdır. Program, önce metni fişler. Fişleme, örnekli veya örneksiz olabileceği gibi, bütün kelimeleri veya sadece istenen harfle başlayanları kapsayabilir. Fişleme işinden sonra, veri tabanı programında, fişlenen kelimelerin sıralanması işi gerçekleştirilir. Daha sonra, sıralanan kelimeler, Cibakaya'da düzenleme işlemine tâbi tutulur. Bu işleminden sonra dizinimiz hazır. Son işlemin çıktı dosyası, kelime işlem programıyla açılarak dizin görüntülenir ve üzerinde istenildiği gibi sayfa mizanpajı yapılır.

Kelime işlem programı olarak MS Word'ü kullanıyoruz. Veri tabanı programı olarak da File Maker Pro'yu tercih ediyoruz. Bu iki programı, türlerinin şimdilik en iyileridir.

2.3. Programın yapabildiği işlevler

Program, temelde bir dizin programıdır ve şu dizinleri hazırlar:

1. Genel dizin: Bir metindeki bütün kelimelerin metnin hangi sayfa ve satırında geçtiğini, aldığı ekleri de göz önünde tutarak ayrıntılı olarak gösteren

dizin. Bu dizinde, madde başlarının çekimli şekilleri, yani alt maddeler kendi içinde gramatikal değil, alfabetik olarak dizilir. İstenirse, her madde başından sonra kaç defa geçtiği verdirilebilir. Geçtiği kaydın da dizinde gösterilmesi, yani örnekli dizin (concordance) alınması mümkündür. Dizinin sonunda, toplam kelime sayısı, alt maddelerin sayısı ve madde başlarının sayısı çıkarılır.

2. Çekim ekleri dizini: Metindeki çekim eklerini, bu ekleri alan madde başlarıyla birlikte gösteren dizin. Bu dizinde, çekim eki birden fazla ekten oluşsa da, bir bütün olarak tek ek gibi değerlendirilir ve bu eklerin dizini verilir. Dizinde referans olarak, söz konusu eki alan madde başları gösterilir. Böylece, o ekin veya ek öbeğinin hangi madde başlarının alt maddeleri olduğu, dolayısıyla hangi madde başlarına bakılması gerektiği ortaya çıkar.

3. Sıklık dizini: Metindeki madde başlarını, en çok geçenden en az geçene doğru, geçme sıklığıyla birlikte gösteren dizin. Bu dizinde, önce sıklık belirten sayı, ardından referans olarak da o sayıda geçen madde başları verilir. Böylece, en çok geçen madde başları ve kaç defa geçtikleri ortaya çıkmış olur.

4. Sondan dizin: Madde başlarının tersinden yapılmış dizini.

Herhangi bir dilin, lehçenin veya eserin madde başları bir dosyada toplanırsa, program, buradaki kelimelerin sondan dizinini de hazırlayabilir.

Bu dört dizinin örnekleri, Türk Dil Kurumu yayınları arasında çıkmış olan şu eserde görülebilir. C. Kaya: *Uygurca Altun Yaruk*. Giriş, Metin ve Dizin. Ankara, 1994.

Programın başka işlevleri de vardır:

5. Metni girerken her kaydın başına konulan referans numaraları otomatik olarak da verdirilebilir. Bu numaralar birer birer olabileceği gibi üçerli, beşerli, onarlı vb. atlanarak da yazdırılabilir.

6. Metindeki istenmeyen karakterler temizlenebilir. Yahut da sadece istenilenler belirlenir, bunun dışındaki her şey sildirilebilir.

7. Dizin hazırlandıktan sonra, bazı madde başlarının birleştirilmesi söz konusu olabilir. Bu durumda, birleştirilen madde başlarındaki referans numaralarının birbirinin içine sıralı olarak geçme sorunu ortaya çıkar. Program, karışık durumdaki bu referans numaralarını yeniden sıralayıp düzenleyebilir.

8. Kodlu olarak yazılmış metni, paragraflanmış normal metin hâline dön-

dürmek de programın yapabildikleri arasındadır.

Ayrıntılar ve diğer işlevler için yardım dosyasında açıklama bulunmaktadır.

3. Cibakaya'nın, türdeşi programlarla karşılaştırılması

Şimdiye kadar, Cibakaya türünde yalnızca bir programa rastladım. 1994 yılında tesadüfen karşıma çıkan program. Amerika Birleşik Devletleri'nde, Dallas'ta, Summer Institute of Linguistics'te geliştirilmiş olup, 'Concordance Generating Program' (kısaca: Conc.) adını taşımaktadır. Ben, bu programın Şubat 1992 tarihli 1.70 Beta versiyonunu gördüm. Program, tıpkı Cibakaya gibi, metnin örnekli dizinini (concordance), hem de normal dizinini hazırlamaktadır. İki program karşılaştırıldığında ortaya çıkan sonuç şudur:

1. Conc.'ta fişleme, sıralama ve düzenleme işlemleri programın içinde gerçekleşmektedir. Cibakaya'da ise sıralama için veri tabanı programından yararlanıldığı için, programdan programa geçmek, dosyaları diğer programa taşımak zaman kaybına sebep olmaktadır. Ayrıca, Conc.'ın sıralaması daha hızlıdır. Yaklaşık 70.000 kelimelik bir metnin dizini Conc.'ta yarım saatte hazırlanabilmektedir. Cibakaya'da ise bu iş, iki üç saat almaktadır.

2. Conc.'ta sıralama düzeni ve font, kullanıcı tarafından belirlenmektedir. Dolayısıyla transkripsiyon alfabesi kullanma olanağı vardır. Cibakaya, sistem fontlarını kullandığı için, transkripsiyonlu olarak yazılan metinlerin dizini sorun olmaktadır.

3. Metin dosyası çok büyük olursa, Conc., dizini çıkaramamaktadır. Cibakaya'da metnin büyüklüğü için bir sınır yoktur. Kelime sayısı, veri tabanı programının sıralayamayacağı kadar çok olursa, dizini, harf harf veya birkaç harfi birlikte çıkartmak mümkündür.

4. Conc., kodlandıktan sonra eklerini de göz önünde bulundurarak kelimeleri doğru sıralamaktadır. Ancak, sıralanmış dosyanın düzenlenmesinde her kelimeyi madde başı kabul etmekte, alt maddeleri dikkate almamaktadır. Bunun sebebi, Conc.'ın, İngilizce gibi eklemeli olmayan iller için düşünülmüş olmasıdır. Bu sorun Cibakaya'da yoktur.

5. Cibakaya'nın yukarıda belirtilen (2.3) diğer özellikleri, Conc.'ta yoktur.

4- Sonuç

Daha önce, bir metnin dizinini hazırlamak için, kelimeler tek tek elle fişleni-

yordu. Aylarca süren ve kaçınılmaz bazı hataları da beraberinde getiren bu iş, şimdi bilgisayarlar sayesinde birkaç saat içinde hatasız ve eksiksiz olarak yapılabilmektedir.

Türk dili araştırmacılarının ihtiyaçları ve Türkçenin yapısı göz önünde tutularak hazırlanan 'Cıbakaya' adlı dizin programı, Türkiye üniversitelerinde metin yayımı biçiminde tez hazırlayan genç araştırmacılarca bir müddettir kullanılmaktadır. Türkçenin sistemi hakkında biraz bilgisi olan gerçek uzmanlar, dilcilerden de yardım alarak, daha iyi dizin programları hazırlayana kadar, adı geçen programın bir ihtiyacı karşılayacağına ve Türkçe araştırmalarına kendi çapında katkıda bulunacağına inanıyorum.