

## BAŞKURT DİLİNİN MAKİNE FONU HAKKINDA

Sirazitdinov Z.A., Filoloji bilimi adayı

Şamsutdinova G.G., asistan araştırmacı

Tarih, Dil ve Edebiyat Enstitüsü

Ufa Bilim Merkezi Rusya Bilimler Akademisi

Ufa, Başkurdistan Cumhuriyeti, Rusya Federasyonu

sazin11@mail.ru

İnsan toplumunda bilgisayarların hızlı yaygınlaşması kendi yörüngesine dilsel faaliyetleri de çekmektedir. Bugün birçok Türk dilleri kendi kaynakların elektronik varyanta çevirmektedir: bilgisayar sözlükler, filolojik veritabanları ve arama sistemleri CD'ye veya İnternet ağına yerleştirilmiş bilgi kaynakları oluşturmaktadır [1]. Başkurt halkının dili de bu süreçten uzakta kalmamaktadır.

Ulusal bir dilin bilgi küresine sokulması arasında Başkurt dilinin de olması ile birlikte kardeş dillerin dilbilimsel gerçeklerin nispeten araştırma yapılması için geniş umutlar açmaktadır. Bu perspektifler büyük hacimli araştırma malzemelerine hızlı erişim açılması ile bağlıdır. Ama bu bilgi kaynakların kullanmanın önemli kısıtlamaları da bulunmaktadır. Bunlar aşağıdaki gibidir:

- 1) dilbilimsel verilerin güvenilirliği;
- 2) verileri derleyicinin yetkililiği;
- 3) bu verilerli sitenin İnternette işleyiş vakti.

Bu faktörler genelde vikipedinin bilgi verileri, ticari kuruluşlar ve özel bireyler ile ilgilidir. Ama itibarlı akademik yapılar veya otoriteli üniversitelerin bilimsel araştırmalar enstitüsünde yapılmış, giriş sayfasında ya da CD'nin astarında yapımcının adı yazılmış ve veri tabanları basılı birincil kaynaklara bağlantı ile donatılmış dilbilimsel kaynaklar için böyle problemler ortaya çıkmamaktadır.

Başkurt dilinin böyle bir ciddi kaynağı Başkurt Dilinin Makine Fonu olmaktadır. Bu kaynaklar Rusya Bilimler Akademisinin Ufa Bilim Merkezi Tarih, Dil ve Edebiyat Enstitüsünün dilbilimi ve bilgi teknolojileri laboratuvarında yapılmıştır [2; 3; 4; 5; 6].

IT teknolojilerinde bugün kullanılan terimlere göre Makine Fonun Bilgisayar Fonu olarak adlandırmak daha doğrudur. Ama bu adın tarihi kökenleri 1980'lere bağlıdır. O zaman SSCB Bilimler Akademisi Rus Makine Fonu (RMF) oluşturulması için program kabul etmiştir. Bu fonu oluşturma problemleri farklı konferanslarda tartışılmıştır [7; 8]. Bugün RMF İnternette bu adıyla kullanılmamaktadır [9].

Aynı yıllarda SSCB halkları dillerinin de makine fonların oluşturma fikri öne sürülmektedir [10]. Birçok bilim adamları Türk dilleri için birleşik bir makine fonu oluşturma hakkında bahsetmiştir [11]. Böyle bir fonu oluşturmak için hatta Çimkent şehrinde matematik ve uygulamalı

dilbilimi profesörü K.B.Bektaev başta olmak üzere koordinasyon kurul da düzenlenmiştir. Ama bu fonların hiç biri gerçekleştirilmemiştir.

Başkurt Makine Fonu (BMF) oluşturma fikri, onun temel konuları tarafımızdan 1990'larda öne çıkarılmıştı [12]. Nedir bu Başkurt Makine Fonu? Başkurt Makine Fonu - dilbilimciler, öğretmenler, öğrenciler için mahsus yapılmış filolojik bilgiler arama sistemidir. Bu bilgi sistemi mahsuslaştırılmış veritabanının, bu veritabanından bilgi isteme arayüzeyleri ve bu bilgi istemelerini işletmek için program sağlamaların içermektedir.

Bugün bu sistemin 7 büyük veritabanı vardır, onlar birleşik fonun bölümlerin içeriyordur:

- Genel dosya bölümü;
- Leksikografik bölüm;
- Gramatik bölüm;
- El yazmalı kitaplar katalogu bölümü;
- Siyah harflı kitaplar katalogu bölümü;
- Deneysel fonetik bölün;
- Diyalekt bölümü.

Şunu belirtmek gerekiyor ki, fon da veritabanlarına verilen bağlantılar geçicidir. Yakında onlar ayrı bir bilgi sistemi olarak yapılacaktır.

1. Genel dosya bölümü dil sisteminin temel bilgilerini içermektedir. Bu taban yapıları şeklinde 100 000'den yukarı kök sözcüklerin ve dilin her tabakasından türevleri içermektedir.

Bu bölüm sözün kökeni, stili, diyalekt/edebiyat, tarihselcilik/ yenilikçilik/ artık kullanılmayan deyim gibi özellikleri hakkında bilgi vermektedir. Her leksik birim için 50'den fazla sözcük işareti tasarlanmıştır. Tabanda her kelime dilde kullanma paradigmaları ile sunulmuştur, ve her paradigma için edebiyat eserlerinden, gazetelerden ve ilmi eserlerden, sözlü halk sanatı eserlerinden basılı malzemenin kaynağı (ismi, yazarı ve basılma tarihi) gösterilerek kullanma örnekleri verilmektedir. Kullanma örnekleri Rusya Bilimler Akademisi Ufa Bilim Merkezi Tarih, Dil ve Edebiyat Enstitüsü dosyalarından alınmıştır.

Bu bölümün veritabanı başka bölümlerin tabanlarıyla bağlıdır, özellikle genel dosyadaki herhangi bir sözün hangi leksikografik çalışmalarda kaydedilmesini belirlemek mümkündür (bu leksikografik bölüm).

2. Leksikografik bölümde söz hakkında bilgi Başkurt dilinde mevcut olan sözlüklerden kaynaklanarak verilmiştir. Bugün bu bölümün veritabanı 58 yapıları sözlükten ibarettir, ve 500 000'ne yakın sözlük maddeleri içermektedir.

Bu tabanda akademik sözlükler ve öğrenciler için sözlükler mevcuttur: tek dilli sözlükler, iki dilli sözlükler, sıklık sözlükler, terim sözlükleri, deyimler sözlükleri, eşanlamlı sözlükler, sözlük-kılavuzlar (semt isimleri, su nesnelere isimleri, şehir sokakları isimleri, BC dağ isimleri) vb.

Bu bölümde kullanıcı aramayı sözcüğün ilk parçasına veya onun herhangi bir parçasına göre yapabilmektedir. Bulunan sözlük maddesinin sonraki çalışma için almak mümkündür.

3. Deneysel-fonetik bölümde Başkurt dilinin sesli ünlülerin ve ünsüzlerinin artikülasyon ve deneysel özelliği verilmiştir. Şimdi deneysel özellikler laboratuvar görevlileri elde etmiş genlik-sıklık grafikler ve sonogramlar olarak gösterilmiştir. Fonemler dinlenebilirler. Aynı zamanda bu bölümde 8000 söz içeren fonetik sözlük de vardır. İlmi ilgiden başka bu bölüm Başkurt dilinin bireysel öğrenenlerin de büyük ilgisini çekmektedir.

4. Halkı ve bilim dünyasını Başkurt dilinde mevcut olan Başkurt yazılı edebiyat dili hakkında bilgilendirmek amacıyla Başkurt Makine Fonunda iki katalog vardır: el yazmalı kitaplar ve siyah harfli kitaplar katalogu. Bu katalogların toplam taban hacimi 2000 birimden de fazladır. El yazmalı kitaplar katalogu Rusya Bilimler Akademisi Ufa Bilim Merkezi Tarih, Dil ve Edebiyat Enstitüsünün el yazmalı kitaplar fonundaki kitapların kısaca açıklamasını içermektedir.

Kitapların aşağıdaki gibi özellikleri tasvirlenmektedir: başlık (Rusçaya tercümesi), çevriyazılı başlık, yazar (eserin yazarı), çevriyazılı yazar, kopya eden (elden yazılı kitap kim tarafından kopya edilmiş), hacim (sayfalarda), ebat (metnin ebatı, sayfada satırlar sayısı), yıl (kitabın kopya edilme tarihi), özellik, özet (kitabın kısaca içeriği), koleksiyon (kitap kim tarafından, nerede, ne zaman bulunmuş, kitabın son sahibi), dil (metnin dili: Arapça, eski Türk dili, Osman dili vb.), paleografik malumat (kağıt, mürekkep, yazı, yorum varlığı), bibliografi (şu elden yazılı kitaba bağlantı), sigl (elden yazılı kitap nerede saklanıyor), şifre (kitabın muhafaza edilen yerdeki şifresi).

Kullanıcı istediği kitabı 6 alan yardımıyla arayabilmektedir: başlığı, dili, yazarı, yılı, sigl ve şifresi. Siyah harfli kitaplar katalogu Başkurdistan Cumhuriyeti Milli Müzesinde saklanan Arapça kitapların tasvirini içermektedir. O kitaplar 15 alana göre tasvir edilmektedir: yazarı, kitabın başlığı, kitabı yayınlayan, kitabın yayınlanma yeri, basımevi (basımevinin ismi), yayınlanma tarihi, cilt (eğer kitabın birkaç cildi varsa), hacimi (sayfalarda), kitabın ebatı, kitabın kapağı (kapağın varlığı veya yokluğu), özelliği (baş sayfa, imza, sayfaların numarası, yapraklar), özeti (kitabın kısaca içeriği), dili (kitabın metin dili), sigl (kitabın muhafaza edilen yeri), şifre (kitabın muhafaza edilen yerdeki şifresi). Kullanıcıya istediği kitabı 9 alan yardımıyla arama mümkünlüğü sunulmaktadır: başlığı, dili, yazarı, yılı, sigli, şifresi, yayınlanma yeri, yayınlayanın adı, basımevi.

Bu katalogların toplam taban hacimi 3000'den fazladır.

5. Diyalekt bölümü üç ayrı tabandan oluşmaktadır. Onlar leksik, kartografik, tekstolojik tabanlardır.

Leksik bölüm diyalekt taban hakkında bilgi içermektedir. Veri tabanı 52 000'den fazla diyalekt birimlerini içermektedir, 6 bilgi alanı vardır: diyalekt kelime, sözcük türü, diyalekt, söyleş, edebiyat normu, Rusçaya tercümesi. Bu alanların her birine göre bilgi araması yapılabilmektedir.

Kartografik bölüm Başkurt dilinin diyalekt atlası hakkında bilgi vermektedir. Diyalekt atlasının veritabanı dil olaylarının türün seçme mümkünlüğünü sunmaktadır (örneğin: fonetik, morfolojik, sintaksik, leksik), her tür için sarı isoglosslar tanımlanmıştır. Bu isoglosslar Tarih, Dil ve Edebiyat Enstitüsü görevlileri katkısıyla 1973'ten 1980'e kadar yapılan bilimsel araştırma gezileri sürecinde toplanan malzemeler üzerinde Başkurdistan Cumhuriyetinin ve komşu semtlerin 250 yerleşim merkezlerine göre belirlenmektedir.

Tekstolojik veritabanında Tarih, Dil ve Edebiyat Enstitüsü görevlileri türlü zamanlarda yapan bilimsel araştırma gezileri sürecinde toplanan sözlü dil örnekleri takdim edilmiştir. Veritabandaki metinler şu özelliklere göre seçilebilmektedir: diyalekt, söyleş, kayıt tarihi, habercinin cinsiyeti, habercinin yaşı, milleti.

6. Gramer bölümü akademik gramerin hipermetin sunmasını Başkurt dilinin söz değişim sisteminin algoritmik sunmasını ve kelimenin isim ve fiil değişimine göre veritabanının içermektedir. Bu istatistik tabanlar dilin üç fonksiyonel stiline göre sunulmaktadır: gazetecilik, nesir ve bilim. Fonksiyonel stiller aynı 1/3 oranda verilmiştir. Örnekleme aralıksız 1000 kelime formu seçme yöntemiyle yapılmaktadır. Toplam veritaban hacimi 600 000 kelimedir.

Veritabanı arayüzü bu kategorilerin ayrı morfolojik kategoriler ve sarı sesbirimciklerin aramasını yapmayı sağlayabilmektedir. Kategorilerin ve eklerin uyumluluğuna göre de arama yapılabilmektedir.

Morfolojik veritabanının sayısal verileri teorik araştırma için hem de pratik uygulama için de önemli. Özellikle veritabanların analizi Tarih, Dil ve Edebiyat Enstitüsü dilbilimi ve bilgi teknolojileri laboratuvarı görevlileri hazırlayan otomatik morfolojik analizörün işini kolaylaştırmaktadır.

Başkurt Dilinin Makine Fonu büyük pratik ve teorik öneme sahiptir.

1. İlk olarak, Fon, dilsel araştırma için yeni bir araçtır. Başından beri bu şekilde düşünülmüştür.

Dilin en önemli konusu kelimedir, başka hiç bir türlü dili öğrenmek, onu sözlüklerde aramak, metinde gramatik kullanışın bulmak mümkün değildir. Geleneksel yaklaşımda (Fondan önce) dilbilimciler kaynaklar (sözlükler, rehberler, kitapların metinleri) aramak için ve gerekli malzemeleri kopya etmek için çok zaman harcamaktadır. Fon bu tür meseleyi saniyeler içerisinde çözmektedir. Örneğin, dilbilimci -ҒАЗЫ/-ҒӘЗЕ formantının kökenin Başkurdistan toponimlerinde aramaktadır (Auyrgazy, Kuyugazy). Eskiden bunun için sözlükler bulup, sayfa-sayfa arayıp kopya etmek gerekirdi, fon ise hemen 30'dan fazla toponim temin etmektedir.

Yeni sözlükler yapıldığı zaman leksik birimin edebiyatta kullanma örneklerin aramak en zor iştir. İstatistiklere göre, 30 000-40 000 kelimeli bir sözlük düzenlemek için 10 milyon deyim içeren kartotek gerekmektedir. Onu hazırlamak çok yavaş ve zor meseledir. Ama birincil ürün (metin) ve son ürün (kartotek) makine içindeyse, işlem hızı yüz kere artmaktadır.

2. İkincisi, Makine Fonu dil öğrenmede, özellikle ders kitaplarında morfolojik veya semantik örnek resimlerin hazırlamakta önemli bir araçtır. Bu açıdan, Başkurdistan Cumhuriyeti Milli Eğitim Bakanlığı Başkurt Dili Makine Fonun Başkurt dili öğretmenlerine önermektedir.

3. Makine Fonu Başkurt dilin ikinci devlet dili olarak uygulamak için bir araçtır. Dili öğrenmek için birçok sayıda gramer, fonetik, sözlük gibi rehber malzemelerine ihtiyaç vardır. Evet, böyle malzemeler yayınlanmaktadır, ama her zaman el altında olmayabilirler. Fonun İnternette olması bu problemi çözmektedir. Bu açıdan Fon ile tüm cumhuriyetin terim komisyonları çalışmaktadır, çünkü onlara türlü tabela, levha, formların tercümesi için başvurular gelmektedir. 60 sözlükten aramaktansa veritabanından tercüme aramak çok daha kolaydır. Leksikografik bölüme okul ve lise öğrencilerinden de müracaatlar vardır.

4. Bu bilgi sisteminin Başkurdistan Cumhuriyetinin dışında yaşayan başkurtların eğitimine de katkısı büyüktür, çünkü onların basılı malzemeye ihtiyaçları çoktur.

Makine Fonuna Orenburg, Perm, Çelyabinsk, Sverdlovsk, Saratov ve başka bölgelerden, Kazan, Sankt-Peterburg, Vladivostok şehirlerinden başvuru yapılmaktadır. Demek ki, Başkurt Dili Makine Fonu sadece bilim ve eğitim aracı değil, belki halkların ruh birleşimi için bir meydan da oluyordur.

Yabancı (mesela Norveç, Almanya, Polonya, Türkiye, Yemen, Ukrayna, Kazakistan, Özbekistan vb.) ülkelerden olan başvurular Fonu dünya medeniyatının Başkurt dili ile tanışması ve yayılması için araç olarak öne çıkmasını göstermektedir.

## EDEBİYAT

1. <http://sozdik.kz>; <http://solver.uz>; <http://www.muhanoff.travel.ru/turslov.htm>; <http://tili.kg/dict>; [maturtel.ru](http://maturtel.ru); <http://abc.marlamuter.ru/index.php/term/rus-tat,2844-bol-shoij.xhtml>; Электронный словарь АББYY Lingvo x5; <http://www.medialingua.ru>; <http://til.gov.kz/>; <http://mfbl.ru>; <http://fodor.ii.metu.edu.tr/content/metu-turkish-corporus>; <http://mfbl.ru/bashkorp/korpus>; [http://web-corpora.net/TatarCorpus/search/?interface\\_language=ru](http://web-corpora.net/TatarCorpus/search/?interface_language=ru).

2. Maşinnıy fond başkirkского yazıka. URL:<http://mfbl2.ru> (data obraşçeniya: 17.08.2016).

3. Prozaičeskiy korpus başkirkского yazık URL:<http://212.193.132.98/bashkorp/> bashkorp (data obraşçeniya: 17.08.2016).

4. Publitsističeskiy korpus başkirkского yazıka URL:<http://212.193.132.98/bashkorp/korpuspub> (data obraşçeniya: 17.08.2016).

5. Folklorniy korpus başkirskogo yazıka URL:<http://212.193.132.98/bashkorp/> korpusfol (data obraşçeniya: 17.08.2016).

6. Terminologičeskaya baza başkirskogo yazıka URL:<http://bashterm.ru> (data obraşçeniya: 17.08.2016).

7. Vtoraya Vsesoyuznaya konferentsiya po sozdaniyu MF RYa / Otv.red. çlen-korr. AN SSSR Yu.N. Karaulov. (Materialı konferentsii). 1988. – 230 s.

8. Materialı III Vsesoyuznoy konferentsii po sozdaniyu MF RYa. / Pod red. S.F.Gilyazova i Yu.N.Karaulova.M.: MGU, 1990. – 148 s.

9. Adres Maşinnogo fonda russkogo yazıka: <http://www.cfri.ru>.

10. Maşinnie fondı yazıkov narodov SSSR: Materialı raboçego soveşçaniya (Tbilisi, 15–22 sent. 1987 g.). –Tbilisi, 1988.

11. Maşinnie fondı yazıkov narodov SSSR: materialı raboçego soveşçaniya (Tallinn, 19–22 dekabrya 1988). – Tallinn: Inst. yaz. i literaturı AN EstSSR, 1988. – 21 s.

12. Sirazitdinov Z.A., Nadergulov I.U. O sozdanii Maşinnogo fonda başkirskogo yazıka/Vostokovedeniye v Başkortostane. Istoriya. Kultura. Vıp.3. – Ufa, 1992.



[www.atlasdergisi.org](http://www.atlasdergisi.org)