
MAGYAR NYELVJÁRÁSOK 52 (2014): 31–42.

A DEBRECENI EGYETEM MAGYAR NYELVTUDOMÁNYI TANSZÉKÉNEK LEKTORÁLT FOLYÓIRATA

Az Uráli Etimológiai Adatbázis mint dokumentáció és mint az etimológiai kutatás eszköze

BÁTORI ISTVÁN

Fejtegetéseim az Uráli Etimológiai Adatbázis (Uralic Etymological Database, rövidítése: UEDb., Uralotéka) kifejlesztése során nyert tapasztalataimra támaszkodnak. Az adatbázis előkészületi munkáiról és az internetes közhasználatba vételéről beszámoltam a Piliscsabai Finnugor Kongresszuson (BÁTORI 2011) és legutóbb Debrecenben a *Folia Uralica Debreceniensia* című folyóiratban is megjelent (magyarul) róla egy beszámoló jellegű tanulmányom (BÁTORI 2013).

1. A szótár, az elektronikus szövegtárolás és az adatbázis

A szótárkészítésben a számítástechnikai eszközök iránti érdeklődés viszonylag korán, már a múlt század 60-as éveiben felébredt, amikor még fennállt a veszélye annak, hogy a korai kezdeményezés elvész az éretlen technika labirintusaiban.

A gépi szövegfeldolgozás még nem létezett. Központi probléma volt az adatfelvétel, a szövegbetáplálás. Ezen belül külön problémát jelentett az ékezetes betűk kezelése, amely kompromisszumok kötésére kényszerítette a szakembereket. A számítógép betápláló billentyűzete szegényebb volt, mint az egyszerű irodai írógépe. A hiányzó ékezetes betűk jelölésére és visszaadására természetesen ki lehetett fejleszteni átíró rendszereket, és egy időben szinte úgy tűnt, hogy a komputerlingvisztika lényege egy ilyen transliteráció kifejlesztése. A szövegfelvételezés és a mellékjeles betűk kezelése akkora erőbevetést igényelt, hogy a számítástechnikai nyelvészet közép-európai híveinek alig maradt idejük más, lényegesebb problémákkal foglalkozni. Gondolok itt például PAPP FERENC akkor szinte hősi kezdeményezésére, idézve 1967-ben megjelent cikke címét: *Die Bearbeitung des ungarischen Wortschatzes auf Lochkartenmaschinen* (A magyar szókinés feldolgozása lyukkártyás gépeken). Talán éppen ennek az éretlen technikának az átvétele vezetett a számítógépes nyelvészettel szembeni tartózkodáshoz.

A modern szövegfeldolgozás kialakulásához több tényező is hozzájárult. Ki kell emelni itt az Unicode-kódolási rendszer elterjedését (a 90-es években), amely túlszárította az ékezetes betűk idegölő, ide-oda kódolgatási szükségességén. Tudatosítani kell azt is, hogy a lexikográfusok érdeklődése nem (lexikális) adat-

bázisok építésére irányult, hanem a szótárkészítés modernizálására (CALZOLARI 1989). A szótárírók korán felismerték, és tisztán látták a számítástechnikai támogatás előnyeit és szükségszerűségét, de továbbra is nyomtatott szótárakat akartak készíteni (PAJZS 1990: 7).

2. Az Uráli Etimológiai Adatbázis keletkezéséről

Az MTA Nyelvtudományi Intézetének a munkatársai (Molnár Zoltán, Oszkó Beatrix, Salánki Zsuzsa, Sipos Mária, Falk Nóra és Fejes László) Csúcs Sándor kezdeményezésére, az UEW. (Uralisches Etymologisches Wörterbuch) háromkötetes szótári anyagát 1992–1996-ban géppel olvashatóvá tették, ami óriási munka volt. Az UEW.-ben használatos nyelvek ékezetes betűinek a betáplálása (kb. 600–800 betű) meghaladta a közkeletű irodai PC billentyűzetének a kapacitását. Nemcsak az etimológiai szótár poliglott, rövidítésekkel teli anyagát kellett betűhíven géppel olvashatóvá tenni, hanem ki kellett találni egy átírást is a szótárban használatos nyelvek ékezetes betűinek a betáplálására.

Az adatfelvitelről az 1995-ös jyväsckyläi 8. Nemzetközi Finnugor Kongresszuson MOLNÁR ZOLTÁN tartott bemutatót és rövid beszámolót, ahol elmondta, hogy az UEW. komputerezálásának a célja egy számítógépezített szótár, amihez később talán egy szófeldolgozó program fog kapcsolódni, nem pedig egy adatbázis (1996: 122). Az elhangzottak alapján úgy tűnt, hogy az UEW. adatfelvivőiben, de a szótár szerkesztőiben sem tudatosult, hogy a vállalkozásuk lényege valójában nem más, mint egy hagyományos és nagy etimológiai szótárat átalakítani elektronikus tárolásra, s ez világszerte az első ilyen horderejű kezdeményezés. Az UEW. anyagának géppel olvashatóvá tétele ugyanis megnyitotta az etimológiai anyag gépi kiaknázásának az útját még akkor is, ha a szótár akkori munkatársai ebben csak a szövegszerkesztők (elsősorban a Word) közvetlen előnyeit látták, és nem perspektívákat.

3. A konverzió és az egyenértékűség

A jyväsckyläi kongresszusi bemutató után Bátori érintkezésbe lépett az UEW. főszerkesztőjével, Rédei Károllyal, valamint az MTA Nyelvtudományi Intézetének akkori igazgatójával, Kiefer Ferencsel, és felajánlotta, hogy egy kooperációs program keretében a Koblenzi Egyetem Számítástechnikai Karának Számítástechnikai Nyelvészeti Intézetében transzformálja az UEW. géppel olvasható, szövegszerű anyagát egy adatbázissá, amelynek a keretében lehetőségessé válik az etimológiai anyag sokrétű kiaknázása.

1998. január 29-én lezajlott egy megbeszélés Rédei Károllyal Bécsben az egyetem finnugor szemináriumában, ahol megszületett a kooperáció az MTA Nyelvtudományi Intézete és a Koblenz-Landaui Egyetem Számítástechnikai Nyelvészeti Intézete között. A megbeszélésen Rédei Károly (vendéglátó) mellett Bakró-Nagy Marianne, Honti László és Bátori István vettek részt. A kooperáció

az UEW. etimológiai anyagának jobb hozzáférhetőségére, az RTF-formátumú anyag adatbázissá való átalakítására irányult. Rédei számára különösen fontos volt a copyright kérdése, a szótár anyagának változatlan átvétele, illetőleg az, hogy az etimológiai anyag ne kerüljön illetéktelenek kezébe. Bátori (és a koblenziek) számára központi kérdés volt az interdiszciplináris kihívás: közvetíteni számítástechnika és uralisztika között és felépíteni az első — a világon az első — etimológiai adatbázist. Az adatbázissá való átalakítás felhasználta az UEW.-ben már meglévő, de szigorúan még nem formalizált etimológiai és lexikológiai kategóriákat (szócikk, forrásnyelv, -nyelvjárás, kútfő/forrás stb.), és átalakította azokat egységes, koherens és az egész szótár anyagát átfogó szerkezetekké. Problematikus volt viszont, hogy az adatokat (szerkezeteket) a szótár szövegszerű utalásai alapján kellett rögzíteni, ami több volt, mint a mezők egyszerű leképzése (például meg kellett határozni a szótárban a „megelőző szó” jelentését).

A Budapesti Uráli Műhely második tanácskozásán 1999-ben BÁTORI így fogalmazta meg az Uráli Etimológiai Adatbázis célkitűzését: „Az adatbázisrendszer célja az, hogy a tárolt anyagot sokoldalúan, hatékonyan és a felhasználóhoz alkalmazkodva tegye hozzáférhetővé. Három sajátságáról van itt szó: a hozzáférhetőségről, a kombinálhatóságról és a felhasználó igényeinek messzemenő kielégítéséről.

1. Az adatok elérhetőségét a szerkezeti egységek szerinti tárolás (és a keresési technológia) biztosítja. Indextáblázatok segítségével minden mező és elemi adat közvetlenül megtalálható.

2. A kombinálhatóság természetesen összefügg az elérhetőséggel, de itt az a lényeg, hogy az adatbázis nemcsak az egyes mezők adatainak a felsorolását támogatja, hanem azok egybevetését és kiválasztását is. Idetartozik az adatoknak és a kombinálhatóságuknak az ellenőrzése is.

3. A felhasználóhoz való alkalmazkodás, a felhasználó igényeinek, adatszemeletének tekintetbevétele túlnyúl a tároláson.” (2001: 28).

A kooperáció során ismételtén kitént, hogy a Nyelvtudományi Intézet Uráli Csoportjának mérvadó munkatársai számára továbbra is a nyomtatott szöveg és az elektronikusan tárolt anyag egyenértékűsége volt a legfontosabb. (A tárolt szöveg és az adatbázis közti alapvető különbség — úgy tűnt — elsikkadt. A szótárszerű használaton túlmenően csak a statisztikai kiértékelés, táblázatok összeállítása váltott ki érdeklődést.). 2010-ben Bátori átadta a konvertált és egy OTKA-projekt keretében certifikált adatbázist az Uráli Csoportnak (BÁTORI 2011). A Koblenzben átalakított adatbázishoz a Nyelvtudományi Intézetben készült egy SQL-en alapuló szoftvercsomag, amelynek segítségével az UEW./UEDb.-anyaga 2010 óta Budapestről az interneten is elérhető uralonet néven. Ugyancsak elérhető (azonos adatokkal az eredeti XML-formátumú) Uralotéka Koblenzből (<http://uralothek.uni-koblenz.de:8080/Uralothek/pdom/basis.html>) és Budapestről (uraloteka.nytud.hu) is.

Az uráli etimológiai anyag Csúcs Sándor által kezdeményezett gépi feldolgozása világszerte az első volt, amit végül az Uralotéka koronázott be. Ez az első működő, az egész nyelvcsaládot felölelő etimológiai adatbázis. Ilyen méretű és szintű mű jelenleg nincs még egy a nyelvészetben, sem a turkológiában, sem az indogermanisztikában.

Szükségtelen volt ugyanakkor a két álláspont, az egyenértékűséghez való ragaszkodás, illetve a technikai lehetőségek kihasználására való törekvés argumentatív szembeállítás, hiszen tulajdonképpen mind a két szemlélet tartalmaz pozitív momentumokat. Az Uráli Csoport megkövetelte az egyenértékűséget, amelynek a certifikálása nélkül az adatbázis használhatatlan. Az adatbázis tudományos szintje nem süllyedhet, veszhet el a komputerezálás érdekében. Bátor álláspontja ugyanakkor az adatbázis innovatív jellegének a hangsúlyozásával az ő informatikával való közelebbi kapcsolatán nyugodott, és felkészített a kommunikációs paradigmaváltásra. Az UEDb. nem új nyelvészeti kategóriákat akar bevezetni, nem akarja az uralisztikát forradalmasítani, csupán a meglévő, hallgatólagosan elismert keretek között akarja az etimológiai kutatómunka határfokát javítani.

4. Dokumentáció az Uralotékában

A dokumentációt két különböző szinten lehet vonatkoztatni az etimológiára. Először úgy, mint általában a tudományos publikációkban: a dokumentáció a tézisek felállításához felhasznált (szak)irodalom. Másodsor az etimológia sajátosságát tekintve: az etimológia önmagában is dokumentáció, amely rögzíti a kutatás aktuális állását.

4.1. Dokumentáció az etimológiában mint forrásmegadás

A dokumentáció az etimológia bizonyító erejű alátámasztása, tipikus módon a források és a szakirodalom. A dokumentáció ebben az értelemben a tulajdonképpeni etimológián kívülálló csatolmány, amely ugyan formális szempontból nélkülözhetetlen, de a problémamegoldás szempontjából mellékes.

Mivel a forrásmegadás/kútfő minden egyes adat esetében elvileg szükséges, fenyeget a redundancia, mert lépten-nyomon fel kellene sorolni ugyanazokat a szerzőket. Ezt elkerülendő az UEW. (mint az etimológiai szótárak általában) tartalmaz egy bevezető fejezetet, amelyben a szerkesztők megnevezik, hogy a cikkekben felhasznált egyedi nyelvi adatok honnan származnak. Magukban a cikkekben rendes körülmények között mellőzhető a forrás ismétlődő és feleslegesnek tűnő megnevezése. Az UEW. elsősorban egyedi nyelvi szótárakra támaszkodik, amennyiben ilyen szótárak léteznek. A szegény írásbeliségű nyelvek esetében nyelvjárási szótárakra vagy szójegyzékekre van szükség, amelyek anyaga etimológiai szempontból érdekesebb és fontosabb lehet, mint a standardizált köznyelv. Kézenfekvő feltételezés, hogy minden nyelvhez (minden nyelvjáráshoz) tartozik egy kútfő. Ez az egyszerű kép azonban fellazul: vannak olyan nyelvek, amelyekre vonatkozóan az UEW. több szótárt is használ (így a ma-

gyar, a finn, az észt, a zürjén esetében), ugyanakkor viszont vannak nyelvek, amelyeknek az adatai ugyanattól a szerzőtől (szótárból) származnak (például a karél, az olonyec, a vepsze és a vót a SKES.-ből, a koibál, a motor és a taigi JANHUNENTÓL, 1977, stb.). Emellett adott esetben származhatnak a szótárban az adatok egyedi, nem standard szerzőktől is. Ilyen esetben a forrást a szócikkben adják meg. Így előáll a default-paradox, azaz hogy a legfontosabb és leggyakoribb forrásművek a szótár szövegében nem (vagy alig) fordulnak elő és fordítva: a szócikkben idézett források éppen a kivételesek.

Az UEDb. egységesítette a kútfőkre (forrásokra) való utalást, és kihelyezte a szócikk testéből egy külön ablakba. A főablakban a Kútfő oszlopban egységesen ott áll a szótár szerkesztőjének a neve, mellette az oldalszám (amennyiben az UEW.-ben megvan). A használó kérésére a kérdéses forrásmű megjelenik az erre a célra rendszeresített információs ablakban. Itt az elérhető bibliográfiai adatok egyöntetűen jelennek meg függetlenül attól, hogy az UEW. előszavából, a rövidítésjegyzékből vagy pedig az aktuális szócikkből származnak-e. Amennyiben egyes nyelvek esetében az Előszó több szótárt is felsorol, a tájékoztató ablakban egy lista jelenik meg. Az UEW.-ben ugyanakkor, amennyiben a kútfő rövidítése nem állt az adat előtt, a használónak kellett kiderítenie, hogy az idézett adat honnan (kitől) származik. Az UEDb.-ben (pontosabban a megjelenítő szoftverben, az Uralotékában) az adathoz tartozó forrás automatikusan kinyomozódik, és ezzel tehermentesül a használó. Az UEW. fellistázza minden szócikk végén a felhasznált (a kútfőkön túlnyúló) irodalmat is.

Az UEDb. támaszkodik a már az UEW.-ben is meglevő irodalomjegyzék rövidítéseire. Az adatbázisban ellenőrizve van, hogy minden irodalomra utaló rövidítés megtalálható. (Fordítva viszont nem, tehát lehetséges, hogy van a felsorolt 356 cím között olyan is, amelyet a szerzők, szerkesztők nem idéznek.) Az adatbázis integritása szempontjából fontos volt továbbá az irodalomjegyzék és a nyelvjárási jegyzék egyeztetése is.

4.2. Dokumentáció mint az aktuális státus rögzítése

A másik lehetséges szemlélet szerint az etimológia már önmagában is dokumentáció. Az adatbázisban elektronikusan tárolt anyag a kutatás eredményének a letéteményese, amely rögzíti a kutatás aktuális állását.

1. Ez összefügg a bevezetőben tárgyalt egyenértékűséggel is, amely elveszti a nyomatékát. Az etimológiai anyag megbízhatóságát, hitelességét nem a vele egyenértékű nyomtatott szótár igazolja, hanem az elsődleges források. Az adatbázis természetesen támaszkodhat másodlagos (vagy x-edleges) forrásokra is, mint például az UEW., de a hitelesség axiómája éppúgy érvényes az UEDb.-re, mint az UEW.-re. Az etimológiai szótár, illetve adatbázis lényege éppen az, hogy a szerző a rokonított szavak összetartozását a helyszínen dokumentálja is.

2. Ezen nyugszik a lexikográfiában ismételten hangoztatott és általánosan uralkodó nézet, mely szerint az elektronikus szövegtárolás radikálisan felgyor-

síthatja és olcsóbbá teheti a szótárak felújítását (updating) (PAJZS 1990: 9, MOLNÁR 1996: 121). Ez vonatkozik az egyszerű sajtóhibák felszámolására és érdemi felújítás(ok)ra is. Amennyiben arra törekszünk, hogy az etimológiai adatbázis anyaga ismereteink aktuális szintjét mutassa, akkor etimológiai ismereteink gazdagodásának (változásának) is tükröződnie kell benne. Ez azt jelenti, hogy a változást be kell vezetni az adatbázisba. A felújítás technikailag viszonylag egyszerű: elég a kérdéses cikket (adatot) megújítani, az adatállomány egésze különben változatlan marad, és máris kész az új, garantáltan hibamentes adatbázis. Az adatbázis üzemeltetése (karbantartása) átvezet az adatállomány soha meg nem szűnő felújításába.

3. A felújítás egyszerűsége viszont egyúttal problémát is jelent. Amennyiben ugyanis a felújítás olcsó és egyszerű, akárki megteheti, így viszont kétségessé válhat az adatbázis megbízhatósága, integritása.

Ezért „Az elektronikus tárolás esetében [...] biztosítani kell, hogy éppen a felújítás egyszerűsége következtében csak az erre jogosult szerkesztő változtathassa meg az anyagot. Az adatfelújítás önmaga viszonylag egyszerű feladat összevetve a jogosultság megállapításával és biztosításával. Itt kell megjegyezni az elektronikus tárolás hitelességének veszélyeztettségével kapcsolatban, hogy a számítástechnika szükség esetén eredményesen lép fel a hamisításokkal és az adatok eltulajdonításával szemben. Általában érvényes, hogy drágább szoftverrel és jobb kódolási eljárással eredményesebben lehet a visszaélések ellen küzdeni. De nem szabad szem elől téveszteni az arányokat: egy pénzügyi anyagának a védelme és egy etimológiai szótár anyagával való visszaélés nem hasonlítható össze.” (BÁTORI 2013: 18).

A szoftvertechnológia és különösen a szoftver hálózati használata oda vezethet, hogy a könyv alakú szótár helyébe egy szoftvercsomag/hálózati licens lép, amely mint szolgáltatás rendszeresen felfrissítődik. Az üzemeltető időnként (vagy rendszeresen) közzé tesz új verziókat, amelyben a felfedezett hibák ki vannak javítva, az anyag bővül, aktualizálódik. Az üzemeltető gondoskodik arról, hogy a változtatásokat csak a felelős ügyvezető hajthassa végre, és a szótár anyaga megőrizze integritását. A szoftvertechnológia ilyen irányú fejlődése figyelmet érdemel akkor is, ha a módszerek nem alkalmazhatók mechanikusan az UEDb.-re.

5. Az Uralotéka a kutatás eszközeként

Az Uralotéka mint a kutatás eszköze feltételezi az etimológiai kutatás tudományos megalapozottságát, amin konkrétan itt hitelességet és módszerességet értünk. A tudományos kutatás bizonyos mértékben leszűkíti az Uráli Etimológiai Adatbázis lehetséges kisugárzását. Kissé általánosítva úgy is fel lehetne vetni a kérdést, hogy hogy kinek a kezébe kerüljön az UEDb.

Három felhasználói csoport (típus) jön számításba:

1. szakemberek, elsősorban az etimológusok, a nyelvtörténészek és az összehasonlító nyelvészek; ez egy nagyon szűk kör,

2. a felsőoktatás, elsősorban a tanárképzés szakemberei; ez egy tágabb kör, ahol a határok mind a két irányban elmosódtak,

3. az érdeklődő (művelt?) nagyközönség; nagyságrendben ez lehet a legnépesebb használói kör, amelynek belső struktúrája is érdekes.

Mind a három felhasználói típusnak vannak saját inherens problémái. Közlelbbi vizsgálat után az első két csoport igénye olyan közel áll egymáshoz, hogy nehéz és célszerűtlen lenne őket megkülönböztetni. Az etimológus kezében az UEW. a kutatás eszköze, amelyből a megbízható és aktuális adatokat meríti. Funkcióját tekintve ugyanilyen eszköz az adatbázis is, melynek a szótárral való tartalmi egyenértékűségéről fentebb már volt szó. A kérdés konkrétan az, hogy az adatbázis hol nyújthat segítséget az etimológus kezében.

1. A keresés gyorsabb, kényelmesebb. Ez nyilván a legszembetűnőbb előny. A keresés eredménye még bonyolult műveletek esetében is szinte azonnali. Vegyünk egy példát az „intelligens” keresésre!

Szeretnénk kikeresni azokat a szócikkeket, amelyekben a magyar adatban a szó belsejében előfordul egy palatális *n* (azaz *ny*). Az Uralotéka először is az adatbázisban meglévő magyar adatszavak alapján összeállítja és táblázatba rendezi a hangok jegyzékét (felismerve a kettős és hármas betűkkel jelölt hangokat is). A felhasználónak csupán le kell szűkíteni a keresést a szó belsejére (két + jelölés között megadva a keresni kívánt hangot), le kell kattintani a palatális sort, és a keresendő hangosztályt be kell vezetni (az enter lenyomásával) a betápláló mezőbe.

The screenshot shows the Uralotéka web application interface. At the top, there is a search bar with the text "Ungarisch" and "belég" selected, and a search query ".+ [ny ny] . +". Below the search bar are buttons for "enter", "reset", "billentyűzet (2X leütni)", "tanács", "fon.jegyek", "tovább", and "restart". To the right of the search bar are checkboxes for "teljes szavak", "nyelvek nyjársok", "német angol magyar".

Below the search bar is a section titled "Virtuális billentyűzet az uráli fonológiához Ungarisch (DI-Nr: 16000)". It features a text input field containing "nyput" and a dropdown menu for "szókezdő", "szóbelső", "szóvégi", and "minden uráli nyelv".

The main part of the interface is a large grid of phonetic categories. The categories are: mssh, zárhang, nazális, réshang, affrikáta, liquida, labiális, dentális, alveoláris, palatális, velár-/glotál-is, jegyek, zöngés +, lágy, lat. +, hosszú, labiális. The "palatális" category is highlighted, and the "ny" sound is selected within it.

Below the main grid is a smaller grid for vowel classification: mgnh, monoftong, elülső, középső, hátulsó, felső, középső, alsó. The "ny" sound is also highlighted in this grid.

At the bottom left, there is a sidebar with the text "Unicode Information Etimológiák Suche" and "Rekonstruktumok Címzójegyzék Származtatás Hasonlítás DTD Kereső". At the bottom right, there is a "Restart" button.

A betáplálás után előttünk van a keresett lista:

Mit dem Suchmuster . + [ny my] . + sind in der UEDb 10 Etymologien gefunden worden.

16000						
Rekonstruktionsebenen						
Nr.	UENr.	Ur	FU	Ug	Lexeme	Bedeutung
1	1726			āðəms, āðmz	enyv	Leim
2	101		čavz, čapa	čavz, čapa	savanyú	sauer, Molken, Käsewasser, Sauerampfer
3	1775			kenz	könyörög, könyörül	flehen, inständig bitten, beten, sich erbarmen, Mitleid haben
4	311		künā, künā, kinā, kinā		könyök	Ellbogen
5	349		konz + [Sfz]		hanyatt	rüchlings, auf den Rücken
6	419		kurnā		hornyol	kerben, falzen, nuten, kannelieren, riefen, rillen, rippen
7	1790			kšnz	könyű	leicht
8	813		puia		bonyolód-	sich verwickeln sich einwickeln, sich verwirren
9	844		pšnz		fenyő	Tanne, Fichte, pinus, abies
10	1875			šnz	enyelég	liebkosen,

Ha nem magyar szavak származását, hanem például a zürjénét nyomozzuk, akkor a zürjén fonémákat rendező a háttérbeli program táblázattá, és bocsátja a felhasználó rendelkezésére (virtuális billentyűzetként). A finn esetében pedig, ahol nincsenek palatális hangok, üresen marad a palatális sor.


A kérdés lehet bonyolultabb is. Kíváncsiak lehetünk például arra, hogy mi felel meg a finn diftongusoknak a magyarban. A keresési beállításokat és a találatokat az alábbi két ábra szemlélteti (a második ábra csak a lista elejét tartalmazza).

The screenshot displays a linguistic software interface. At the top, there are search filters for 'Finn --/Ung Lit' and 'RKN: / Beleg'. Below this is a 'Virtuális billentyűzet az uráli fonológiához' (Virtual keyboard for Uralic phonology) section. It features a grid for consonant classification (mssh, zárhang, nazális, réshang, liquida) and a detailed diftongus table. The diftongus table shows combinations of front, middle, and back vowels in various positions (elöltag, utótag).

In der UEDb sind mit dem obigen Suchmuster in den Sprachen/Dialekten: Finn , Ung , 56 Etymologien gefunden worden.

LfNr	UENr	Rekonstruktionen		Belege	
		Form	Bedeutung	Finn	Ung
1	<u>5</u>	<i>ajje</i>	Schläfe	<i>atvo(t), atvena</i>	<i>agy</i>
2	<u>25</u>	<i>arpa</i>	irgendein Wahrsagungs-, Zaubermittel	<i>arpa-mies, arpajaiset</i>	<i>orvos</i>
3	<u>49</u>	<i>čada, šada</i>	das Laufen, das Rennen; die Brunst der weiblichen Tiere; laufen, rennen; brunsten	<i>suota, suoti, olla_suodilla</i>	<i>szalad, szilaj</i>
4	<u>130</u>	<i>eje, úje</i>	Nacht	<i>yö</i>	<i>éj</i>
5	<u>142</u>	<i>íce, íše</i>	Schatten, Schattenseele	<i>juoppo meni itsettömäksi</i>	<i>és ~ s</i>
6	<u>187</u>	<i>jokse</i>	laufen, ? läufig od. brünstig sein, sich paaren	<i>juokse, juoksu-atka</i>	<i>ív</i>
7	<u>195</u>	<i>juve, juke</i>	trinken	<i>juo</i>	<i>i</i>

A kérdés irányulhat több nyelvre is, az (uráli/finnugor) rekonstruált alakokra vagy egy egész alcsaládra. A példában azokat az etimológiákat keressük, amelyekben van finn, lappN, zürjén, magyar és szelkup(Tas) megfelelés. A betápláló mezőben lesz meghatározva a keresendő hang, illetve hangsor: ami a példában szó belseji réshang — a rekonstrukciós szinten. (Az üresen maradt betápláló mező úgy értelmeződik, hogy minden megfelelést meg kell mutatni.)



Uralothenk * 13.12.2014

RKN: / Beleg

Sprachen	Deutsch
Dialekte	Englisch
genet. Gruppe	Ungarisch

Restart ^

Unicode Informat Etymolog Suche

Reko Such Ablei Vergl DTD

In der UEDb sind mit dem obigen Suchmuster in den Sprachen/Dialekten: Finn , Lapp N, Syrj S, Ung , Selk Ta, 3 Etymologien gefunden worden.

LfNr	UENr	Rekonstruktionen		Belege				
		Form	Bedeutung	Finn	Lapp N	Syrj S	Ung	Selk Ta
1	<u>753</u>	<i>pesä</i>	Nest	<i>pesä</i>	<i>basse</i>	<i>poz</i>	<i>fészék</i>	<i>pitz</i>
2	<u>1088</u>	<i>uje, oje</i>	schwimmen	<i>ui</i>	<i>voggjá</i>	<i>uj-, ujan</i>	<i>úsz</i>	<i>uarn</i>
3	<u>1123</u>	<i>waske</i>	irgendein Metall, ? Kupfer	<i>vaski</i>	<i>vaŋke</i>	<i>ez-lš</i>	<i>vas, vaskos, vastag</i>	<i>kezi</i>

2. További előnyök: az ellenőrizhetőség, a teljesség, a szótári anyag következetes normalizálása, ami az adatkeresés megbízhatóságát növeli. A keresés kezdetén már előre ismert, hogy a jegyzékben hány címszó van, hogy az illető nyelv ábécéjében milyen betűk vannak, stb.

3. Az adatok átvétele kézi másolás nélkül: a szótárból ki kellett írni a szót, az adatbázisból viszont lehet közvetlenül nyomtatni, és ha kell, e-mailként elküldeni a szükséges adatokat. Ezzel eggyel kevesebb lesz a lehetséges íráshibák száma.

4. A kizárás technikája: az Uralotéka segítségével feleletet kaphatunk olyan kérdésekre, amelyek feltevése számítógép nélkül megengedhetetlenül költséges lett volna, példa erre egy hangsor keresése szavak belsejében. Az Uralotéka segítségével megengedheti magának a kutató, hogy olyan kérdéseket tegyen fel, amelyekre a válasz (esetleg) eredménytelen marad.

5. Az utalások közvetlenül nyomon követése: a szótárbeli szöveges vö. rövidítést a felszínen fel lehet oldani, és utalások segítségével az összevetendő szócikket rögtön meg lehet nyitni. Az ugrás úgy van programozva, hogy mindig vissza is lehet térni az eredeti helyre.

Azzal szeretném ezt az illusztrációt befejezni, hogy visszaemlékszem HELIMSKY (1984: 242–243) 30 évvel ezelőtt megjelent cikkére, amelyben az uráli fonológiai rekonstrukciók kérdését tárgyalta, és hiányolta az uráli rekonstrukciók tanulmányozását, beleértve a közbeeső szintek vizsgálatát is. Minden bizonnyal javult a helyzet időközben (példaként csak Honti László és Csúcs Sándor munkásságára utalok), de az uráli hangtörténeti kutatás növekvő igényessége és a bizonyító erejű hangmegfelelések lokalizálásának a költségessége mellett a tartózkodás érthető is. Ugyanakkor éppen az Uráli Etimológia Adatbázis ad a kutatók kezébe egy olyan eszközt, amellyel a rekonstrukciók felállításához szükséges hangmegfeleléseket hatékonyan és rendszeresen meg lehet találni.

6. Az Uráli Etimológiai Adatbázis a dokumentáción és a kutatáson túl

Befejezésül visszatérek a potenciális címzettek harmadik csoportjára, az ún. művelt nagyközönségre, mert úgy tűnik, hogy ez a csoport lenne talán a legfontosabb.

Az utóbbi években olyan írások jelennek meg uralisták tollából (RÉDEI 1998, KERESZTES 2003: 115, CSÚCS 2009, csak néhányat említve), amelyekben arról számolnak be, hogy Magyarországon az uráli nyelvrokonság és az uralisztika iránt kritikus és ellenséges hangulat kezd elterjedni. Az „alternatív” elméleteknek olyan népszerűsége van, hogy képviselőik egyre gyakrabban a sajtóba, a közigazgatásba és az iskolákba is beépülnek.

A finnugor nyelvrokonság tagadása heterogén eredetű. A propagálóit csak a finnugor nyelvészet elutasítása tartja össze. Közös viszont a nyelvészeti tájékozatlanságuk és felszínességük. Veszélyesek azért, mert propagálóik a közéletben, különösképpen a közművelődésbeli kapcsolataik révén a felnövekvő nemzedéket is félrevezetik.

Ezt a változást a nyelvi kultúra elmélyítése és a finnugor nyelvrokonság kézzelfoghatóvá tétele tartóztathatja fel a legeredményesebben. Az Uráli Etimológia

Adatbázis ezért nemcsak az etimológusok és nyelvtörténészek kezében munkaeszköz, hanem a nyelvi kultúra terjesztője is, amely módszeresen, meggyőzően és átfogóan indokolja a finnugor rokonságot. Lényeges ennek kapcsán, hogy az Uralotéka — az uralisztika és az idetartozó etimológiai kutatás aktuális szintjén állva — a finnugor (nyelvi) rokonság népszerűsítésében, meggyőzővé tételében ugyanazokra az adatokra támaszkodik, mint amit a kutatásban a finnugor (uráli) nyelvrokonság bizonyítására az interneten korszerű formában elér a művelt nagyközönség és ezen belül a felnövekvő nemzedék is.

Irodalom

- BÁTORI S. ISTVÁN 2001. Az etimológiai szótár mint adatbázis. In: OSZKÓ BEATRIX–SIPOS MÁRIA szerk., *Szófejtvő műhely. 1999. szeptember 8–10.* Budapesti Uráli Műhely 2. Budapest, Magyar Tudományos Akadémia Nyelvtudományi Intézete. 25–37.
- BÁTORI S. ISTVÁN 2011. Uralische Etymologische Datenbasis. Daten und Rekonstruktion. In: CSÚCS SÁNDOR–FALK NÓRA–TÓTH VIKTÓRIA–ZAICZ GÁBOR szerk., *Congressus XI Internationalis Fenno-Ugristarum Piliscsaba, 9–14. VIII. 2010. 6. Dissertationes symposiorum ad linguisticam.* Piliscsaba, Reguly Társaság. 9–15.
- BÁTORI S. ISTVÁN 2013. Az Uralotéka digitális adatbázis. *Folia Uralica Debreceniensia* 20: 3–22.
- CALZOLARI, NICOLETTA 1989. Computer-Aided Lexicography. Dictionaries and Word Data Bases. In: BÁTORI, S. ISTVÁN–LENDERS, WINFRIED–PUTSCHKE, WOLFGANG szerk., *Computational Linguistics. An International Handbook on Computer Oriented Language Research and Applications.* Berlin–New York, Walter de Gruyter. 510–519.
- CSÚCS SÁNDOR 2009. Gondolatok az alternatív nyelvrokonságról. *Finnugor Világ. A Reguly Társaság Értesítője* 14/2: 3–16.
- HELMISKY, A. EUGEN 1984. Problems of phonological reconstruction in modern Uralic Linguistics. *Советское финно-угроведение* 20: 241–257.
- JANHUNEN, JUHA 1977. *Samojedischer Wortschatz. Gemeinsamojedische Etymologien.* Helsinki, Suomalais-Ugrilainen Seura.
- KERESZTES LÁSZLÓ 2003. A magyar nyelv eredetkérdései. Eredmények és kihívások a finnugrisztikában. *Folia Uralica Debreceniensia* 10: 107–126.
- MOLNÁR ZOLTÁN 1996. Uralic Etymological Archive. In: LESKINEN, HEIKKI–MATICSAK, SÁNDOR–SEILENTHAL, TONU szerk., *Congressus Octavus Internationalis Fenno-Ugristarum* 10. 15. 8. 1995. 5. Jyväskylä, Moderatores. 121–123.
- PAJZS JÚLIA 1990. *Számítógép és lexikográfia.* Linguistica. Studia et dissertationes 4. Budapest, Magyar Tudományos Akadémia Nyelvtudományi Intézete.
- PAPP FERENC 1967. Die Bearbeitung des ungarischen Wortschatzes auf Lochkartenmaschinen. *Acta Linguistica Academiae Scientiarum Hungaricae* 17: 142–172.
- RÉDEI KÁROLY 1998. *Őstörténetünk kérdései. A nyelvészeti dilettantizmus kritikája.* Budapest, Balassi Kiadó.

SKES. = TOIVONEN, YJRÖ HENRIK–ITKONEN, ERKKI–JOKI, AULIS J.–PELTOLA, REINO
1955–1981. *Suomen kielen etymologinen sanakirja 1–7*. Lexica Societatis Fenno-
Ugricae 12: 1–7. Helsinki, Suomalais-Ugrilainen Seura.

UEW. = RÉDEI KÁROLY szerk. 1986–1991. *Uralisches Etymologisches Wörterbuch 1–3*.
Budapest, Akadémiai Kiadó.

uralonet (Budapestről): <http://www.uralonet.nytud.hu>.

Uralotéka (Koblenzből): <http://uralothek.uni-koblenz.de:8080/Uralothek/pdom/basis.html>.

Uralotéka (Budapestről): uraloteka.nytud.hu.