

Részletes szakmai beszámoló

Uráli nyelvek névszói szerkezetei / Nominal Structures of Uralic Languages NKFI 125206

Az *Uráli nyelvek névszói szerkezetei / Nominal Structures of Uralic Languages* (röviden NSUL) című, 2017. szeptember 1. és 2021. augusztus 31. között támogatott, majd a futamidő meghosszabbításával 2023 május 31-i dátummal záródó, 125206 sz. FK típusú kutatási projekt szakmai beszámolója az alábbi szerint épül fel:

- Rövid bevezető
- A kutatócsoport működésének áttekintése
- . A kutatók eredményeinek összefoglaló ismertetése
- Az adatbázis részletes bemutatása

Bevezető

A kutatás leíró és elméleti nyelvészeti módszereket egyesített uráli szintaktikai jelenségek összehasonlító vizsgálata céljából. Névszói szerkezetekkel és nominális, tágabban nonverbális állítmányú mondatokkal foglalkoztunk veszélyeztetett uráli nyelvek egy kiválasztott csoportjában, illetve a magyar nyelv két korszakában. A projekt keretén belül született tudományos előadások és publikációk eredményei új szempontokkal bővítették az uralisztikai kutatásokat, finomították a tipológiai általánosításokat, és a szintaxiselmülethez is hozzájárultak, pl. a fejevégség, az ellipszis, vagy a szórendváltás kutatásának területén. Az összehasonlító, parametrikus megközelítés számos olyan kérdésre is ráirányította a figyelmet, amely az egyes nyelvek izolált vizsgálata során egyszerűen nem merült fel korábban. Reményeink szerint számos jövőbeni, akár egyéni, akár csoportos kutatásnak adunk így lendületet, mivel a projekt keretében összeállított és közzétett nyilvános adatbázis meghatározó empirikus alapokat teremtett ennek. (Tudomásom szerint több pályázat is elbírálás alatt áll kapcsolódó témákban.)

Az FK típusú, kutatócsoportos projekttel párhuzamosan indult a vezető kutató második posztdoktori kutatása *A magyar főnévi kifejezések történeti nyelvtana* címmel (NKFI 125070), amely 2021-ben zárult.

A kutatócsoport működése

A projekt munkaterve egy viszonylag nagy taglétszámmal működő, finnugrista és elméleti nyelvészekből álló kutatócsoport összehangolt munkájának elképzelésére épült.

A támogatási időszak első két évét lendületes, csoportos munka jellemezte, rendszeres szakmai találkozók, a közös olvasmányok és egymás kutatási eredményeinek megvitatásával. Az adatbázis szerkezeti előkészítése, az elvi alapok lefektetése is ebben az időszakban történt. Az intézményi átszervezésekkel is súlyosbított covid-időszak alatt az első évekre jellemző dinamika azonban megtört és ezt később már nem sikerült visszaállítani. A közös feladatok görgetése, a találkozók elmaradása azt eredményezte, hogy az egyes kutatók inkább saját, egyéni kutatásaik felé fordultak, illetve többen újabb projektek, újabb együttműködések felé mozdultak el (amit szakmai előmenetelük, illetve anyagi biztonságuk sok esetben meg is követelt). Ez érthető módon lekötötte azokat a humánerő kapacitásokat, amelyekre a NSUL-projekt ismételt fellendítéséhez is szükség lett volna. A csoport tehát bizonyos mértékig dezintegrálódott, ugyanakkor a projekt utolsó szakaszában (a finanszírozási kereten belüli szabadabb átcsoportosításnak köszönhetően) lehetőség nyílt az

tervezett adatbázis gyors és hatékony megvalósítására. Bár az adatbázis konkrét kivitelezésén már lényegében a vezető kutató dolgozott, a korábbi, közösen előállított anyagok mind beépítésre kerültek (részleteket lásd alább). Az adatbázisba bevitt nyersanyag túlnyomó többsége is a kutatócsoport tagjainak gyűjtéséből, feldolgozásaiból származik.

A kutatók kutatási és szakmai eredményei

Az egyéni kutatási eredményekről a projekt honlapján lehet részletesebben tájékozódni (<https://nytud.hu/palyazat/urali-nyelvek-nevszoi-szerkezetei>). Ezek az egyéni kutatások hol szorosabban, hol lazábban kötődtek a projekt központi kutatási kérdéseihez, de minden esetben jelentős publikációkat és hazai vagy nemzetközi konferencia-előadásokat eredményeztek.

Asztalos Erika fő kutatásai az udmurt szórend sajátosságaira és az információs szerkezetre irányultak, majd érdeklődése a tagadás és kvantorhatókörök összefüggéseire is kiterjedt. Dékány Éva átfogóan foglalkozott a magyar főnévi kifejezések grammatikájával, illetve tágabb perspektívában vizsgálja a nominalizáció és a nem-véges alárendelés kérdéskörét. Ekaterina Georgieva érdeklődése nagyon széleskörű: a projekthez kapcsolódó eredményei elsősorban az ellipszis és a nominalizáció témaköreiben születtek, miközben az udmurt mellett egyéb uráli és nem uráli nyelveket is vizsgált. F. Gulyás Nikolett külső munkatársként csatlakozott a projekthez, és a predikatív birtoklás vizsgálatával járult hozzá az ismereteinkhez. Hegedűs Veronika általánosságban foglalkozott a nem-igei állítmányok szerkezetével és típusaival különböző uráli nyelvekben. Mus Nikolett a tundrai nyenyec nyelv szakértője és a nyenyec kopulás mondatok számos sajátos vonását vizsgálta, mielőtt saját kutatócsoportos projektjét beindította. Sipos Mária északi hanti nyelvváltozatok nominális és lokatív mondataival, illetve kopulahasználatával foglalkozott. (Tánczos Orsolya a projekt korai szakaszában kilépett, Gugán Katalin pedig szülési és gyermeknevelési távollét miatt végül csak nagyon korlátozottan vett részt a projekt munkájában.)

A jelen beszámolóban foglalt FK típusú, kutatócsoportos projekttel párhuzamosan indult a vezető kutató - jelen sorok írója - második posztdoktori kutatása *A magyar főnévi kifejezések történeti nyelvtana* címmel (NKFI 125070). Ez lényegében egy szerkezettörténetről szóló monográfia összeállítását célozta, amely egyrészt részletesen leírja az ó- és középmagyar főnévi kifejezések nyelvtani rendszerét, másrészt dokumentálja és elemzi a szerkezeti felépítésekben megfigyelhető változásokat. A vállalkozás messze túllépte a támogatási időszak kereteit és a kézirat jelenleg is publikálatlan formában bővül és gazdagodik a folyamatos anyaggyűjtés eredményeképpen. Ugyanakkor a monográfia és a NSUL adatbázis szempontrendszerét és módszertanát párhuzamosan, egymást elősegítve dolgoztam ki a főnévi kifejezések tekintetében. A két projekt párhuzamosítása tehát egyrészt nagyon előnyös vállalkozás volt, másrészt a munka mennyisége és komplexitása miatt olyan mértékű túlvállalás, amit az elnyert pályázatok ambiciózus megkezdésekor még nem láttam át. Szerencsére jelenlegi munkaköröm a Nyelvtudományi Kutatóintézet Történeti Nyelvészeti és Uralisztikai Intézetében (a Történeti és összehasonlító mondattani kutatócsoportban) lehetővé teszi, hogy a befejezetlen feladataimat akadályoztatás nélkül folytassam, vagyis a kéziratot és az adatbázist az eredeti koncepciók alapján építsem tovább.

A vezető kutatónak a projekthez szorosan kapcsolódó publikációja nem jelent meg. A készülő monográfia mellett konferenciaelőadásokkal, rendezvényszervezéssel, illetve az adatbázis megtervezésével és létrehozásával járult hozzá a projekt tudományos eredményeihez. A projekthez szorosan nem kapcsolódó témában azonban fontos publikációi jelentek meg (pl. egy teljes tanulmánykötetet szerkesztett az Oxford University Pressnél generatív történeti szintaxis témában, önálló fejezettel járult hozzá a *Syntax of Hungarian*,

Comprehensive Grammar Resources című sorozat második kötetéhez, illetve egy óegyiptomi vallásos szöveg kritikái, kommentált magyar kiadását is elkészítette.)

Az adatbázisban felhasznált eredeti nyelvi anyag, és annak nyelvészeti értelmezése is az egyes kutatók munkájához köthető:

- mai magyar és ómagyar: Dékány Éva, Egedi Barbara, Hegedűs Veronika
- udmurt: Asztalos Erika, Ekaterina Georgieva
- északi hanti: kazimi - Sipos Mária, színjai - F. Gulyás Nikolett, Sipos Mária
- keleti hanti: szurguti – Egedi Barbara (Csepregi Márta útmutatása és anyaggyűjtése alapján), illetve Gugán Katalin
- tundrai nyenyec: Mus Nikolett

Ugyanakkor meglévő korpuszok (pl. OUIDB - München, Typological Database of the Ugric Languages - ELTE) annotált adataiból is bőségesen merítettünk. A szurguti hanti nyersanyag alapvetően Csepregi Máriával történő, több éven át folyó konzultációk eredménye, aki engedélyezte ennek az anyag felhasználását. Az adatbázis végső kialakításában elsősorban Asztalos Erika, Ekaterina Georgieva és Hegedűs Veronika volt a segítségemre.

Az adatbázis

A projekt egyik legfontosabb eredménye egy szabadon hozzáférhető, on-line adatbázis (<https://nsul.nytud.hu>), amely a projekt címében is szereplő, különböző uráli nyelvekből gyűjtött névszói szerkezeteket mutat be alaposan kidolgozott nyelvészeti szempontok szerint. Ezek a szempontok tulajdonképpen paraméterek és változók, amelyek segítségével nagyon hatékonyan összehasonlíthatók a vizsgált nyelvek szerkezetei közötti egyezések és eltérések.

Az adatbázis mérföldkőnek számít abban a tekintetben is, hogy bár a Nyelvtudományi Kutatóközpont oldalán számos korpusz és elemzett nyelvi adatbázis is megtalálható, ezek többsége magyar nyelvű, míg a NSUL adatbázis teljes egészében angol nyelvű, így közvetlenül tud kapcsolódni a nemzetközi kutatásokhoz, és gyakorlatilag a világon bárkinek hozzáférhető.

i) Az adatbázis felépítése

Az adatbázis a Nyelvtudományi Kutatóközpont szerverén található és <https://nsul.nytud.hu> linken érhető el. A nyitófelület azonnal feltárja az adatbázis szerkezetét, amely végig áttekinthető marad a bal oldali menüsornak köszönhetően. Alapvetően három típusba sorolhatók a funkciók:

- A bal oldali menüben legfelül az adatbázis és a mögötte húzódó koncepció bemutatása kapott helyett, illetve rövid összefoglalók az adatbázisban szereplő nyelvekről. Az "About data display and glossing" menüpont részletesen bemutatja az egyes adatpontok felépítését, és hogy felhasználóként hogyan kell értelmezni ezeket.
- A bal oldali menüben legalul kapott helyett a "Resources", amely egyrészt az adatbázis egészében használt szakirodalmi és forráshivatkozásokat oldja fel, illetve egy glosszáriumot is tartalmaz az adatbázis szempontjából legfontosabb nyelvészeti szakterminusokkal és magyarázatukkal ("Linguistic terms and definitions"). Ezekre a terminusokra az adatpontokból is linkek mutatnak, mindig értelemszerűen csak a relevánsakra.
- Az adatbázis lényegi része a "Topics" menüsor-csomagban található, amely 12 címre osztja a jelenségeket. Ezek a "topikok" tovább mélyülnek, három, vagy négy szintet, vagyis különböző kategóriák és feltételek mentén választhat továbbhaladási utat a felhasználó. A hierarchia megfelelő mélységében pedig maga az adatpont található, amely egyrészt tartalmaz egy leírást a nyelvi jelenségről, másrészt magát az eredeti nyelven gyűjtött példát (gyakran tudományos átírással kiegészítve), angol nyelvű glosszákat, angol nyelvű fordítást, és végül forrásmegjelölést. Adott jelenségre gyakran egynél több példa szolgál illusztrációként,

ugyanakkor vannak olyan jelenségek, amelyek adott nyelvre nem értelmezhetők. Ilyenkor a "This parameter is not applicable" kifejezés tűnik fel.

Az adatpontokat tartalmazó helyeken bármikor válthatunk a nyelvek között a vízszintesen található fülek segítségével. Ezáltal egyszerűen és közvetlenül hasonlíthatjuk össze bármely két nyelv azonos funkcióra használt mondattani szerkezeteit.

Fontos megjegyezni, hogy minden egyes adatpont exportálható és egyszerűen beilleszthető tetszőleges szövegszerkesztő programban megnyitott dokumentumba. Hasonlóképpen, az adott oldalra való URL hivatkozás is kimásolható és beilleszthető hivatkozásként tetszőleges platformra.

ii) Az adatbázis feltöltésének menete és a források

Az adatbázis építéséhez számos forrást használtuk, ezek természetesen nagyon különbözőek. A szakirodalomból átvett forrásoknál többnyire megőriztük a formátumot, de szükség szerint átalakítottuk az adatok egy részét (tipikusan a glosszázást), vagy hozzáadtuk tartalmat (pl. angol nyelvű fordítást) az uniformitás jegyében. Igyekeztünk a lehető legtöbb esetben felhasználni a kutatócsoport tagjainak már megjelent, vagy megjelenés előtt álló munkáit a nyelvi adatok prezentálásakor.

Az uniformításra való törekvés persze le is lassítja az évek során összegyűjtött nyelvi anyag feltöltését, hiszen az adatbázis informatikai és webes kialakítása tényleg a projekt legutolsó szakaszában történt, amikor végre megfelelő szakembert találtam a munkára, akivel kölcsönösen kielégítő munkakapcsolatot tudtam kialakítani. Az adatfeltöltés ennek megfelelően nem fejeződött be, vagyis még bőven folyamatban van. (Legszerényebb becslésem alapján is kb. 5-6000 adatpont feltöltéséről beszélünk.) Szintén lassítja a folyamatot, hogy minden adatponthoz tartozik egy leíró mező, vagyis nemcsak a puszta adatot jelenítjük meg, hanem egy rövid leírásban értelmezzük is a jelenséget, természetesen angolul. Az adatbázist informatikai szempontból megtervező nyelvtechnológus kolléga kérésemre egy nyelvészek által is könnyen használható adatbeviteli interface-t fejlesztett, ezen keresztül szerkeszttem és publikálom folyamatosan az off-line tárolt anyagokat.

A jelenleg megtekinthető adatbázisban a különböző nyelvek eltérő arányban képviselik magukat. A magyar és az udmurt adatok viszonylag problémamentesen konvertálhatók, ezek feltöltése sokat haladt az utóbbi időben. A mai magyar nyelvi anyagnál alapvetően a *Syntax of Hungarian, Comprehensive Grammar Resources* című, a magyar nyelv modern leírását célzó vállalkozás eredményeire (Alberti – Laczkó 2018), illetve a kutatócsoportunk tagjainak munkáira, míg az ómagyarnál leginkább saját kutatásaimra támaszkodtam. A hanti adatok feldolgozása ennél változatosabb képet mutat. A Csepregi Mártával, több éven át folytatott konzultáció és szövegfeldolgozások eredményeképpen óriás mennyiségű elemzett szurguti hanti anyag áll rendelkezésemre. Ennek a már általam leválogatott, felcímkézett nyelvi anyagnak a feltöltését az adatbázisba két technikai probléma némileg hátráltatja: (i) a szakszerű és többször ellenőrzött glosszák, illetve a fordítások eredetileg magyarul készültek el, ezeket angolra kell fordítani, (ii) a példaanyag konkrét közzétételében figyelembe kell venni, hogy a források egy része publikálatlan, így a közzététel jogi vonatkozásai nem magától értetődőek. Kérésemre, illetve É. Kiss Katalin kérésére, Csepregi Márta megkezdte az anyag ilyen jellegű áttekintését és előkészítését az open-access közzétételre. Terveink szerint a szövegek teljes terjedelmükben bekerülnek majd "Az uráli nyelvek változása aszimmetrikus kontaktushelyzetben" című korábbi projekt keretei között létrehozott adatbázisba (<http://archive.nytud.hu/oszt/elmnyelv/urali/adatbazisok.html>), míg a NSUL adatbázisba csak a céljainkat szolgáló, leválogatott és lefordított, illetve megfelelő módon kommentált adatokat töltjük fel. Az északi hanti adatok több forrásból és korpuszból érkeztek, ezek feltöltése szintén folyamatos és viszonylag problémamentes. A korábban legyűjtött tundrai nyenyec adatok helyzete azonban egyelőre rendezetlen. Mus Nikolett anyanyelvi adatközlője végül

nem járult hozzá az adatok közzétételéhez, így a célirányosan lekérdezett nyelvi mondatok nem publikálhatók. Az adatpontok másodlagos szakirodalom felhasználásával persze - ha hiányosan is -, de feltölthetők bizonyos szintig.

iii) Az adatbázisban rejlő lehetőségek

Az adatbázis jelenleg hat nyelvi modullal, és 12 ún. nagy témakörön belül 446 paraméter-változóval rendelkezik. A bővítés lehetőségét horizontális és vertikális irányban is ígéretesnek találok. Kérésemre a html-fejlesztés kifejezetten támogatja a nyelvi modulok jövőbeli hozzáadását, vagyis a már meglévő struktúra bármikor bővíthető horizontálisan. Ha a jövőben anyagi és humán-erőforrást nyerhetünk erre a célra, akkor további uráli nyelvek és nyelvváltozatok adhatók hozzá az adatbázishoz. (A párhuzamos, folyamatban lévő kutatási projektek ismeretében, elsősorban manyisi, komi, középmagyar és moldvai magyar modul hozzáadását látom életszerűnek). Ha új, elméleti kutatási jellegű projekt indulna a jövőben a most záruló projekt teremtette alapokon (akár a jelenlegi kutatócsoport egyes tagjainak közreműködésével), akkor a vertikális fejlesztés is kivitelezhető. Ez egy bonyolultabb feladat lenne, mivel a nyelvészeti szempontok mélyítését, bővítését jelentené. Ebből a szempontból egyértelmű iránynak látom a negációt, azaz a tagadás kutatását, illetve a nominalizáció és a nem-véges alárendelés a jelenleginél sokkal részletesebb kibontását.

Maga az adatbázis jelenlegi formájában is felszínre hozott számos olyan kutatási kérdést, amire nem találtunk választ. Az adatbázis egyes pontjain az értelmezhető adatok hiánya pontosan azt mutatja, hogy egyelőre hiányos a tudásunk ezekben a kérdésekben. Ilyen, további kutatásra érdemes problémakörök például az információs szerkezet és a szórendi variációk összefüggése, kevésbé értett szerkezetek azonosítása és elemzése (ún. *belong-constructions*), a kvantifikáció, vagy a referencialitás jelölése. Mivel az egyes nyelvek független leírása és vizsgálata általában zárt nézőpont-rendszerek mentén folyik, ezeknek a problémáknak a felszínre kerülését én önmagában is óriási eredménynek találok.,

iv) Az adatbázis közzététele és népszerűsítése

Habár az adatbázis már elérhető a fenti linken, szélesebb körben még nem publikáltam, illetve népszerűsítettem. Ennek ugyanis az a reális feltétele, hogy az adatok feltöltése már kb. 70-80%-ban megtörténjen. Amint ez teljesül, több nyilvános fórumon tervezem az adatbázis bemutatását, illetve szakmai közleményben történő ismertetését, mindez nyilvánvalóan előremozdítja majd a kapcsolódó publikációknak és magának az adatbázisnak is a tudományos hasznosítását és az idézettségét.