

Az utónevek eredetleírásának formalizálása az Utónévportálon

Sass Bálint, Raátz Judit

MTA Nyelvtudományi Intézet
sass.balint@nytud.mta.hu, raatz.judit@nytud.mta.hu

Az MTA Nyelvtudományi Intézet feladata, hogy az anyakönyvi bejegyzésre alkalmasnak minősített női és férfi utónevek jegyzékét kezelje illetve bővítse. Folyamatos igény mutatkozik a mind újabb utónevek felvétele iránt, a lista évente 60-70 női és 40-50 férfi utónévvel egészül ki.

Az Intézet útjára indítja az Utónévportált, melynek célja, hogy a klasszikus Ladó-Bíró féle utónévkönyv [1] utódjaként szabadon hozzáférhetővé tegye az utónevekkel kapcsolatos közérdeklődésre számot tartó információkat a folyamatosan karbantartott utónévlistának megfelelően. A portál háttérét adó naprakész, hiteles utónév-adatbázis tartalmazza az egyes nevek összes fontos adatát, beleértve az aktuális gyakorisági értékeket is. A portál a már korábban is elérhető havi frissítésű jegyzéken kívül kiegészítő szolgáltatásokat nyújt.

Az adatbázisra épülő részletes keresési lehetőség elsősorban a névadáshoz kíván segítséget adni, magában foglal számos olyan szempontot, melyek a névadáskor szóba kerülhetnek. Ilyenek: a nevek hosszúsága, szótagszáma, hangrendje, névnapja mellett a részletesen kidolgozott eredet- és jelentés-információ, adott hangzású, ritmusú vagy magánhangzókombinációval bíró nevek keresésének lehetősége, illetve a konkrét aktuális adatokra támaszkodó gyakorisági keresés.

A cél az, hogy a utónevekhez tartozó adatok értelmük, szemantikus tartalmuk szerint legyenek kereshetők. A nevekhez jópár egyszerű, explicit, atomi adat tartozik (pl. szótagszám, névnap), ami könnyen kezelhető. Az eredetleírás, vagyis az utónevek etimológiája azonban nem ilyen. Itt a szabad szöveggént megfogalmazott leírás feldolgozására, az információ kivonatolására és explicit megadására van szükség ahhoz, hogy a célt megvalósítsuk, és a leírásban meglévő akár rejtett aspektusokra is külön-külön rákereshessünk.

Azt a megoldást választottuk, hogy a több mint 3500 eredetleírás mindegyikét egy szigorúan meghatározott formális alakra hozzuk, amely minden fontos tartalmi információt magában foglal, és melynek automatikus kezelése már kézenfekvő. Az ötlet egy korábbi cikkből [2] származik, mely az etimológiák formalizálására ad általános javaslatot.

A formalizálási eljárás két szakaszra bomlik. Először egységesítettük a leírásokat, azaz a különféle megfogalmazású, de azonos jelentésű fordulatokat egy szabályalapú megoldással egységes alakra hoztuk (pl.: a *becéző formája*, *becéző rövidülése*, *becéző alakja* egyaránt *becézője*-ként jelenik meg).

A második szakasz a tényleges formalizálás, ami valójában az eredetleírásban szereplő nyelvi elemek, illetve az egyik elemet a másikba alakító műveletek felfedését jelenti. Egy példa az 1. ábrán látható.

| | |
|------------------|---|
| eredetleírás | <i>A latin Dominicus név magyar formája</i> |
| formalizált alak | latin:Dominicus [megfelelője] magyar:~ |

1. ábra. A Domonkos név leírása és formalizált alakja. A leírásban két elemet (nevet) kapcsol össze a [megfelelője] művelet.

A tényleges formalizálás félautomatikus úton történt. Az egyszerűbb nevek (a nevek 55%-ának) formalizált alakját automatikusan állítottuk elő az egységesített alakból, a gyakori mintázatok alapján kialakított szabályrendszer segítségével. Az összetettebb nevekhez (pl.: *Lionella a francia 'kis oroszlán' jelentésű Leon férfinév angol Lionel változatának női párja*) manuális úton – részletes útmutató szerint dolgozó két párhuzamos annotátorral – rendeltük hozzá a formális alakot.

A formalizált leírásnak köszönhetően az eredetleírásokon belül szemantikus keresés válik lehetővé az Utónévportálon. Ennek segítségével olyan kérdésekre is választ kaphatunk, hogy közvetlenül mely nyelvből került át adott név a magyarba, vagy hogy melyek a névalkotással létrehozott neveink.

Hivatkozások

1. Ladó, J., Bíró, Á.: Magyar utónévkönyv. Vince Kiadó, Budapest (1998)
2. Sass, B.: Javaslat az etimológiai minősítés egységesítésére. In: II. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY2004), SZTE, Szeged (2004) 203–208