

Javaslat az etimológiai minősítés egységesítésére

Sass Bálint

MTA Nyelvtudományi Intézet, Budapest
joker@nytud.hu

Kivonat Elektronikus nyelvi erőforrások esetén az egységes formátum mindig egyszerűbb feldolgozhatóságot és könnyebb felhasználhatóságot jelent. Dolgozatomban az összegyűjtött követelmények alapján kialakítom az etimológiai minősítés egyfajta egységes, a korszerű, XML alapú lexikai adatbázisokban alkalmazható modelljét. A rendszer működését példákon keresztül mutatom be.

Kulcsszavak: etimológia, egységes etimológiai minősítés, DTD

1. Előzmények

A Webster szótár 1828-as kiadása óta terjedt el, hogy a szótárak a szócikk alapvető részének tekintik az etimológiai minősítést [2], mely „azt van hivatva megjelölni, hogy a kérdéses szó eredete sommásan miként határozható meg.” [1]

Benkő Loránd a téma szempontjából nagy jelentőségű cikkében hangsúlyozza az egységes etimológiai szemlélet és ennek folyományaként az egységes etimológiai minősítési rendszer szükségességét. „Szabatos etimológiai minősítések hiányában egyáltalán nem lehetne statisztikát készíteni egy-egy nyelv szókészletének származásbeli kategóriáiról.” Fontos szerepet tulajdonít a különféle szóalkotási módok számbavételének. Elvként fogalmazza meg, hogy egy nyelvi elem történetében azokat a pontokat kell megragadni, amikor az elem újnak vehető minőségbe lép, illetve hogy a közvetlen etimológiai előzményhez kell viszonyítanunk, azaz a lépések közül számunkra mindig a legutolsó, az időben hozzánk legközelebbi a legfontosabb. Megjelenik az a sokak által fontosnak tartott követelmény is, hogy azonos jelenségeket konzekvensen azonos, különbözőeket különböző minősítéssel kell ellátni [1].

A különféle szótárak etimológiai minősítési rendszere meglehetősen eltérő. Tapasztalatok szerint az etimológiai leírás szinte bármit tartalmazhat egy-egy szavas meghatározástól egészen a szó történetéről szóló terjedelmes esszéig [2], ezenkívül a szótárakban az etimológiai információ jelentős része rejtetten jelenik meg.

Dolgozatom célja egy a fenti kívánalmaknak megfelelő és az említett hátrányokat kiküszöbölő egységes etimológiai minősítési rendszer kidolgozása absztrakttal modellt, illetve konkrét XML DTD formájában. Ezutóbbi teszi lehetővé, hogy a modellt lexikai adatbázisokban alkalmazzák.

Napjainkig jellemző, hogy a lexikai adatbázisokban az etimológiai minősítést egyszerű szöveggént kezelik. Lemondanak a nyelvi elemek közötti kapcsolatok ábrázolásáról [7]. Olyan modellt, ami minősítés a belső szerkezetét megragadná

tudomásom szerint még nem készült. Az általam ajánlott modell beilleszthető bármelyik formátumba az egyszerű <etym> tag helyére.

A kiinduló anyagot az ÉKSz-ben [6] található etimológiai minősítések adták. Jelen dolgozatban a modell ennek megfelelő részletességű kidolgozását tűztem ki célul, vállalva, hogy ezzel leegyszerűsíttem a problémát.

2. Modell

Nézzük azokat a tulajdonságokat, aminek a modell meg kell, hogy feleljen.

- Legyen *egyetemes*: képes legyen magába foglalni minden lehetséges etimológiai megoldást [5].
- Legyen *rugalmas*: majdnem minden részlet megadása opcionális legyen.
- Legyen *érzékeny*: ragadja meg a nyelvi elem összes minőségi változását.
- Legyen *többrétű*: engedjen meg alternatívákat és legyen alkalmas eszköze valószínűségi kategóriák kifejezésére.
- Legyen *explicit*: tegye könnyen hozzáférhetővé az információt.
- Legyen *konzekvens*: azonos jelenséget azonos módon, különbözőt különböző módon kezeljen.

Az egyetemesség kívánalmába beletartozik, hogy a nyelvi elemekhez időpontot is rögzíteni lehessen. A köznévi és a tulajdonnévi eredet jól elkülönül, ezt érdemes jelölni. A nyelvi elemekhez általában vagy a forrásnyelvet adják meg vagy a konkrét szót, amiből származik. Hasznos, ha opcionálisan mindkettőt meg lehet adni.

Alább a modell két formájának közös leírása következik, zárójelben mindig utalok a DTD vonatkozó kulcsszavára.

A modell lényegi ötlete, hogy az etimológiai minősítést *elemek* (`elem`) és rajtuk értelmezett *műveletek* (`op`) rendszerének fogja fel. Az elemek egyszerűen fogalmazva szavak, a műveletek pedig a különféle szóképzési módok. Szavakból kiindulva a műveletek segítségével új szavak képződnek, míg végül elérkezünk a minősítendő szóhoz. A *lépés* (`step`) összefoglalja a műveletet az operandusként hozzátartozó elemekkel, az etimológiai minősítést így lépések sorozatának lehet tekinteni. Szokás szerint a lépések a jelenből haladnak a múlt felé, a minősítés legfontosabb részét az első lépés adja. Az etimológiai minősítés tehát egy fastruktúra alakját veszi fel, ahol a fa gyökerében a minősítendő szó foglal helyet, a csúcsok az elemeket, az élek, élcsoportok a műveleteket jelentik, adott csúcsba vezető élcsoport a kiinduló csúcsokkal együtt pedig egy lépést alkot.

Igyekeztem számbavenni az összes lehetséges műveletet [4]. Ezeket az alábbi táblázat foglalja össze, az ajánlott operandusszámmal, illetve a DTD-ben szereplő rövidítéssel együtt.

leszármazás	1	desc
átvétel	1	loan
nemzetközi szó és vándorszó	1	internat
tükörfordítás	1	calque
félig tükörfordítás	1	lc
származékszó	1	deriv
betűszó	2≤	acr
szóösszevonás, szóösszerántás	2≤	contr
szóösszetétel	2≤	comp
szóvegyülés	2≤	mix
hangutánzó, hangfestő szó	0	onom
gyermeknyelvi szó	0	child
mesterséges szóalkotás	0	artif
állathívogató szó, indulatszó	0	interj
egyéb kapcsolat	1≤	rel

Nulla az operandusszám, ahol nem adható meg elem, amiből a szó kialakult, egy ott, ahol egy szóból alakul ki egy másik, és kettő vagy több a „szóösszerakás” különféle fajtáinál. Szóösszetételnél legalább az egyik tagnak teljes szónak kell lennie, ha mindkét rész csonka, akkor szóösszevonás a művelet, a betűszóknál a tagokból csak egy-egy betű kerül át, szóvegyülésnél a betűk sorrendje is változhat.

Az elemeknél meg lehet adni a *nyelvet* (*lang*), a *szóalakot* (*orth*), a *jelentést* (*def*), a *magyarázatot* (*gloss*), az *átadott alakot* (*pass*), az *időpontot* (*time*) és hogy tulajdonnévről vagy köznévről van szó (*proper*). Közneveknél általában a jelentés, tulajdonneveknél a magyarázat használandó. Az átadott alak arra szolgál, hogy például szóösszevonás esetén megadja az elem azon részletét, mely a műveletben részt vesz.

Az alternatívák lehetőségét műveletként volt a legkényelmesebb bevezetni (*alt*). Operandusszáma legalább kettő, eredménye pedig mindig üres elem. A valószínűségeket elem és lépés esetén is megadhatjuk (*p*). Öt valószínűségi osztályt hoztam létre, a *kizárttól* (*impossible*) a *biztosig* (*sure*). Alapértelmezett az utóbbi, és a kérdéses részletek esetén általában a középső *kérdéses* (*quest*) használandó.

A modell DTD formában a következőképpen néz ki.

```
<!ELEMENT etym (step+) >
<!ELEMENT step (elem*) >
<!ATTLIST step
  op (desc|loan|internat|calque|lc|
    deriv|acr|contr|comp|mix|
    onom|child|artif|interj|rel|
    alt) #REQUIRED
  id ID #IMPLIED
  p (impossible|perhaps|quest|probable|sure) "sure" >
<!ELEMENT elem (lang?,orth?,def?,gloss?,pass?,time?) >
```

```

<!ATTLIST elem
    proper (y|n) "n"
    stepid IDREF #IMPLIED
    p (impossible|perhaps|quest|probable|sure) "sure" >
<!ELEMENT lang (#PCDATA) >
<!ELEMENT orth (#PCDATA) >
<!ELEMENT def (#PCDATA) >
<!ELEMENT gloss (#PCDATA) >
<!ELEMENT pass (#PCDATA) >
<!ELEMENT time (#PCDATA) >

```

3. Példák

Végül nézzük meg néhány példán, hogy hogyan működik a gyakorlatban a rendszer. Három példa az ÉKSz-ből [6], a negyedik pedig egy angol szótárból való. Először mindig a címszót és az etimológiai minősítést idézem, majd rövid magyarázattal és a modell szerinti XML alak következnek.

1. hordár [hord]

Itt egyetlen lépésből áll a minősítés: egy művelet szükséges egy elemmel. A művelet a származtatás, az elem pedig a *hord* szó, melynek jelen esetben csak a külső alakja van megadva. Az XML forma így a következő lesz:

```

<etym>
  <step op="deriv">
    <elem>
      <orth>hord</orth>
    </elem>
  </step>
</etym>

```

2. tegnap [teged 'minap' (té)+nap]

Két lépést, műveletet különíthetünk el: egyrészt egy szóösszetétel szerepel, mely két elemet kapcsol össze, másrészt az előtag kapcsolatban áll egy további elemmel. Látjuk, hogy a *teged* szónak csak egy része kerül át a címszóba, ezt a *pass* tagban adhatjuk meg pontosan. Ha egy elem egy lépés (művelet) eredménye, akkor szükséges azonosítani, hogy melyik lépésé. Erre szolgál az *elem* tag *stepid* attribútuma, mely egy *step* tag *id* attribútumára hivatkozik. Az első lépés eredménye maga a címszó, az összes többi lépés eredményét viszont elemként használjuk fel, így ezeknek mindig meg kell adni az *id* attribútumát.

```

<etym>
  <step op="comp">
    <elem stepid="id0-es1">
      <orth>teged</orth>
      <def>minap</def>
      <pass>teg</pass>
    </elem>
    <elem>
      <orth>nap</orth>
    </elem>
  </step>
  <step op="rel" id="id0-es1">
    <elem>
      <orth>té</orth>
    </elem>
  </step>
</etym>

```

```
</elem>
</step>
</etym>
```

3. fennhájáz [fenn+ ?háj 'szemháj' v. hája]

Ebben a példában megjelenik a kérdésség és a vaglyagosság. A kérdésség itt az `elem` tag `p` attribútumában adható meg. A példában a kérdőjel vélhetően a *háj* és a *hája* szavakra is vonatkozik, az XML formából ez egyértelműen kiderül. A vaglyagosság (`alt`) műveletének eredménye üres elem. Az XML formában a származékszó (jele itt: `quest`) művelet kapcsán még egy lépés megjelenik, ennek az eredménye is egy üres elem lesz, ami az eredeti formában az összeadásjel és a nyíl között kapna helyet.

```
<etym>
<step op="comp">
  <elem>
    <orth>fenn</orth>
  </elem>
  <elem stepid="id1-es1"/>
</step>
<step op="deriv" id="id1-es1">
  <elem stepid="id1-es2"/>
</step>
<step op="alt" id="id1-es2">
  <elem p="quest">
    <orth>háj</orth>
    <def>szemháj</def>
  </elem>
  <elem p="quest">
    <orth>hája</orth>
  </elem>
</step>
</etym>
```

4. polyp [F f. L f. Gk (pous foot)]

A modell egyetemességének (ld. 2. o.) illusztrálására az utolsó példát a láthatóan más formátumot követő Pocket English Dictionaryből vettem [3], Ebben a minősítésben három művelet rejlik: a szót az angol a franciából, az a latinból, az a görögből vette át (az átvétel jele itt az 'f.'). Zárójelben az eredeti görög szó van megadva jelentéssel együtt.

```
<etym>
<step op="loan">
  <elem stepid="id2-es1">
    <lang>F</lang>
  </elem>
</step>
<step op="loan" id="id2-es1">
  <elem stepid="id2-es2">
    <lang>L</lang>
  </elem>
</step>
<step op="loan" id="id2-es2">
  <elem>
    <lang>Gk</lang>
    <orth>pous</orth>
    <def>foot</def>
  </elem>
</step>
</etym>
```

Hivatkozások

1. Benkő, L.: Az etimológiai minősítés a szótárszerkesztésben. *Magyar Nyelv*, XC. évf (1994), 4. szám, pp. 385-392.
2. Drysdale, P. D.: Etymological Information in the General Monolingual Dictionary. In *Dictionaries – An International Encyclopedia of Lexicography* (szerk.: Hausman, F. J., Reichman, O., Wiegand, H. E., Zgusta, L.) pp. 525-530.
3. Fowler, F. H.: *The Pocket Oxford Dictionary of English*. 7th ed. Clarendon Press, Oxford, 1984.
4. Keszler, B. (szerk.): *Magyar grammatika*. Nemzeti Tankönyvkiadó, Budapest, 2000.
5. Kiss, L.: Kísérletek etimológiai képletek felállítására. *Magyar Nyelv*, LX. évf (1964), 4. szám, pp. 314-321.
6. Pusztai, F. (szerk.): *Magyar Értelmező Kéziszótár*. Akadémiai Kiadó, 2003.
7. Sperberg-McQueen, C., Burnard, L. (szerk.): *The XML Version of the TEI Guidelines. Print Dictionaries*. The TEI Consortium, 2002.
<http://www.tei-c.org/P4X/DI.html>