

## Szófaji beosztás névszói csoportok elemzéséhez

Naszódi Mátyás: MorphoLogic, 1126 Budapest, Orbánhegyi út 5.  
naszodim@morphologic.hu

**Kivonat:** A klasszikus nyelvosztályozás – főnév, melléknév, számnév stb. – nem elégséges a gépi elemzéshez. A névszók pontosabb kategorizálása lehetőséget ad a többértelműségek számának csökkentésére, és pontosabb mondat-elemzésre. Jelen cikk egy finomabb, de jelentéstant nem használó felosztást javasol. A felosztás szerepet játszik a szóalaktanban és a mondattanban egyaránt. Segítségével pontosítható a névszói szerkezetek határa és definiáltsága, mely hasznosnak bizonyult a korábbi projektekben, de használata várhatóan a fordítási projektben elengedhetetlenné válik.

### 1. Szófaji osztályozás

A szófaji osztályozás, mint kategorizálás két célt szolgál. Egyrészt a szótani szabályok megadásánál játszik szerepet, másrészt a mondattani szabályok ezekre, mint alapelemekre épülnek. A szófaji osztályok nyelvenként eltérnek. Ennek ellenére nyelvészeink, mint más népek nyelvészei is igyekeznek hasonló osztályokat, kategóriákat létrehozni, hagyomány szerint a latin grammatikát alapul véve. Ez hasznos, ha össze akarunk vetni különböző nyelveket, vagy nyelvtanulóknak ad jó támpontot. A gépi feldolgozás szempontjából is hasznos. Sajnálatosan vagy szerencsére a nyelvek különbözőek. Egyesek szerint angol nyelvben kisebb a különbség az ige és a névszó között. A szó aktuális szófaját nem a szó, mint inkább a szó mondatbeli helye határozza meg. A *file* szó éppúgy lehet ige, mint köznévi. Jelentése ugyanaz, csak a szerepe egyik esetben igei, a másikban főnévi. Persze lehet találni tisztán igei, és teljesen köznévi szavakat is, de nem ez a jellemző. Magyarban az igeik és a névszók jobban elkülönülnek, de a névszók kategóriája szótani szempontból kevésbé különböznek. Különbségek vannak, de ezek elsősorban nem ragozási, szóalaktani, hanem mondattani különbségek.

A szokásos névszói kategóriák a következők: főnév, melléknév, számnév, és névmás. [EL], [PF] Ez utóbbi szerencsétlen osztály, mert a szintaxis szempontjából sokkal fontosabb, hogy a névmás főnévi, melléknévi, netán számnévi-e, mint az, hogy névmás. A főnév alosztályai a szokás szerint köznévi és tulajdonnévi, míg a számnevek három alosztálya tő-, sor- és törtszámnév. A melléknéveket nem szokták tovább osztani, legfeljebb megemlítik, hogy alap, közép vagy felsőfokú-e a szó, kifejezés.

## 2. A névszói szerkezetről

Az általánosan elfogadott kategóriák a szótani vizsgálatoknál használhatóak, de elégtelen a mondattanhoz. Én most elsősorban a névszói szerkezetek szempontjából vizsgálom a szófajokat.

Névszói szerkezetnek – esetünkben a magyar nyelvre vonatkozóan – nevezem azokat a nem igei szerkezeteket, melyeket kötött sorrenddel kell írni a magyarban. Ezek azok a szerkezetek, melyek részei az utolsót kivéve különböző alanyesetű névszóból, illetve alanyesetű jelzői szerkezetekből állnak, esetleg névelővel, kvantorral kezdve. (a melléknév jelentésének egy kiterjesztése a jelzői szerkezet több szóból álló szintagmákra).

Pl.: *mindhárom sárga csőrű pelyhes kiskacsával*

Az ilyen szerkezetekben a szórend kötött. A hagyományos szófaji beosztás alapján a szó végén egy köznév áll, ezt különböző jelző(s szerkezete)k előzik meg. Ezek előtt lehet egy mennyiségjelző, egy birtokos jelző, és legelől a névelő és/vagy kvantor. Minden része opcionális, tehát nem kötelező, de (a névelőn kívül) legalább egy elemet kell tartalmaznia.

*mindhárom pelyhes kiskacsával*  
*mindhárom sárga csőrűvel*

*mindhárom sárga csőrű pelyhessel*  
*mindhárommal*

A *sárga csőrű* egy jelzői szerkezet az adott esetben. Kérdés, lehet-e a fenti kifejezésnek vele egyenértékű sorrendezése. Mi az, ami szintaktikus szabályokkal biztosítja a fenti sorrendet? Talán egyetlen más sorrend fogadható el, de azt sem érzem teljesen azonosnak: *mindhárom pelyhes sárga csőrű kiskacsával*

Másik példában jobban látszik a kötöttség.

*Kovács Katalin sikeres magyar sportolónő*  
tulajdonnév melléknév melléknév köznév  
tulajdonnév melléknév köznév köznév

A hagyományos szófaji felosztás szerint a két melléknév közt nincs különbség szófaji felosztás szerint, ezért semmi sem indokolja, hogy ne cserélhetnénk fel. Amennyiben a szót főnévnek tekintjük, akkor viszont az utolsó két szó sorrendjét lehetne megváltoztatni. Mindkét megoldás elvetendő.

## 3. A névszók finomabb osztályozása – pontosabb névszói szerkezet

Ennek egy megoldását mutatja az 1988-ban készült kísérleti magyar mondatelemző [FN]. Ebben a munkában a következő új névszói szófaj szerepel: népcsoport, foglalkozás, mértékegység. A szófajok szemantikainak tűnnek, de nem tisztán azok. A lényeg, hogy a jelzők osztályozásánál a köznév és a különböző minősítésű jelzők közt sorrendi precedenciát lehet felállítani. Nem lehet tisztán főnév és melléknevekről beszélni. Ahelyett, hogy kétarcú névszókról beszéljünk, melyek a hagyományos beosztás szerint lehetnek pl. főnevek és melléknevek is, olyan kategóriát lehet nekik tulajdonítani, amelyek nem köznevek, nem is melléknevek, hanem a (prioritás alapján) a kettő közti szófaj. A jelenleg kísérleti beosztásom és azok közti prioritás a következő:

...<szelektor<mennyiség<tulajdonság<fajta<köznév

**Magyarázat:**

**köznév** a szokásos szófaj, ha nem teszem másik osztályba.

**fajta:** ezek többsége a köznév gyakori jelzői lehetnek. Személy esetén három alkategóriája van, a következők prioritással

népnév<vallás/pártállás<foglalkozás

Pl. *magyar református ács*

Állatoknál az **alfajt** jelentő jelző: *kardfogú tigris*

Élettelennél ilyen az **anyagnév**: *vörösréz névtábla*

Ez utóbbi két eset persze csak akkor érdekes, ha nem írják egybe az öt követő köznévvél

**Tulajdonság:** ezek az általános melléknevek, illetve jelzős szerkezetek. Jelzők lehetnek a szelektorban is.

**Mennyiség:** olyan részszerkezet, mely egy **számnévből** és opcionális **mértékegységből** áll esetleg egyszerűbb jelzővel díszítve. A számnéven itt csak tö- illetve tö- és törtszámnévből álló kifejezést értek, melyet most nem részletezek. A mértékegység tulajdonképpen köznév a hagyományos szófaji kategória szerint. Szerepe persze jól elkülöníthető az egyéb köznévtől.

A **szelektor** is több részre osztható: **birtokos<pozíció=kiválasztó**

**Birtokos jelző** a szokásos birtokost megjelölő részkifejezés, amelynek száma, személye az egész kifejezés birtokragjának számával és személyével megegyezik.

**Pozíció** vagy egy **sorszám**, vagy egy úgynevezett **pozicionáló melléknév**. A sorszám a szokásos, a pozicionáló melléknév pedig a melléknevek azon alosztálya, amely a középfok *bb*-je nélkül is kaphat *leg*- felsőfokot: *utolsó, szélső, hátsó*...

A **kiválasztó jelző** első pillanatra általános jelzőnek tűnik. Nem is lehet sokszor megkülönböztetni őket. A humán ellenőrzés, vajon egy jelző a tulajdonság, avagy a szelektor kategóriába tartozik a rákérdezés módszere. Ha a *milyen* kérdésre válaszol, akkor tulajdonságjelző, ha *melyik* kérdésre, akkor szelektor. Egyes melléknevek egyértelműen szelektorok. Ezek az *-ik* végű melléknevek, mint a sorszámnevek, vagy pl. *jobbik, másik*. Általában az összetett szerkezetű jelzők is ide tartoznak, de ennek eldöntéséhez további vizsgálatra van szükség: *tegnap vásárolt, első helyet elért*... A mai magyar nyelvben a szelektort tartalmazó névszói kifejezés névelővel vagy kvantorral kezdődik. Részletesebben a névszói szerkezetről lásd [FN].

#### 4. A szavak osztályozásának módszere

A jelenlegi nyelvi adatbázisok hagyományos szófajokat rendelnek a szavakhoz. A felhasználás céljából elengedhetetlen az új kategóriába való sorolás. Az előző fejezet ötleteket is ad egyes szófajok kimerítő feltérképezésére. A fajta szófaj többsége főnévként és melléknévként is szerepel az adatbázisokban. Ezek kigyűjtése után emberi erővel elvégezhető a népnév, foglalkozás stb. beosztás.

**Mértékegység** erősen véges halmaznak tűnik: *méter, liter, kilogramm, hüvelyk*... Ez nem egészen van így. A gyakoribbak természetesen egy fizika-, kémiakönyvből összeszedhetők. Vannak azonban „alkalmi” mértékegységek is. Ezek azok a fizikai objektumot jelentő köznevek, melyek mennyiségek összevetésére alkalmasak: *három szekér szalma, egy csipet só*. Ezeknek a szavaknak az a szótani tulajdonságuk, hogy megkaphatják a *-nyi* melléknévképzőt: *szekérnyi, csipetnyi, maroknyi*. Ha nagy korpuszokból kigyűjtjük az ilyen szavakat, akkor a gyakoribbakat mindenképpen megtaláljuk.

A pozicionáló mellékneveket szintén könnyű kigyűjteni, hisz szintén jellegzetes a morfológiai tulajdonságuk.

## 5. Néhány szerkezeti példa

*A kórházban sok ápolónő román.*

Ez a mondat nem tekinthető egy névszói kifejezésnek, mert a népnév nem követheti a foglalkozásnevet, ezért nem hiányos mondat, hanem szabályos nominális mondat két független névszói kifejezéssel.

*A szülő tegnap érkezett.*

Ez sem lehet egy névszói kifejezés, mert a *tegnap érkezett* jelző összetett, ezért a *szülő* (mint foglalkozásszerű jelentéssel bíró szó) nem előzheti meg a *tegnap érkezett* jelzőt, de mint melléknév sem, mert mint halmozott szelektor, a két rész közé vesszőt kéne tenni. Természetesen szemantika nélkül így is marad „mellélemzés”: az *érkezett* lehet ige és befejezett melléknévi igenév, sőt a *szülő* lehet a *tegnap* jelzője, vagy foglalkozása. Ezen a módszer nem segít, de szemmel láthatóan csökkenti a többértelműségek számát.

## 6. Összefoglalás

A szófajok nyelvfüggő kategóriák. Jó osztályozás elősegíti a pontosabb szóalaktani és mondattani elemzést. A szófaji kategóriák kiválasztása modellalkotási probléma. A szófajokat úgy kell létrehozni, hogy egyrészt azok szóalaktani és mondattani szempontból relevánsak legyenek, másrészt algoritmikusan meghatározhatók legyenek. Segítségükkel a mondat szerkezete világosabb lesz, értsd ezen, kevesebb félreértelmezés jön ki az elemzés során. A jelenleg lefutott projektekben nem volt szükség alapos magyar mondatelemzésre, de a hamarosan beinduló magyarról való fordításnál elengedhetetlen pontosabb szerkezeti felbontás. Ez nem zárja ki más módszerek használatát, de mivel kellően általános, bizonyára segíteni fog.

## Hivatkozások

1. [EL] Elekfi László: Magyar ragozási szótár. MTA Nyelvtudományi Intézete, Bp., 1994.
2. [FN] Farkas Ernő, Naszódi Mátyás: Magyar nyelvű mondatok elemzése természetes nyelvű interfész céljából. SzTAKI kiadvány, 1990 május.
3. [KF] Kiefer Ferenc (szerk.): Strukturális magyar nyelvtan 3. Morfológia. Bp., Akadémiai. Kiadó 1994.
4. [PF] Papp Ferenc: A magyar főnév paradigmikus rendszere Bp., 1975.