

# Névutók, előre!

## Korpuszvezérelt elemzés a névutószerű elemekről

Ligeti-Nagy Noémi<sup>1,2</sup>

<sup>1</sup>MTA-PPKE Magyar Nyelvtechnológiai Kutatócsoport,  
1083 Budapest, Práter u. 50/A

<sup>2</sup>Pázmány Péter Katolikus Egyetem Bölcsész- és Társadalomtudományi Kar  
2087 Piliscsaba, Egyetem u. 1.  
ligeti-nagy.noemi@itk.ppke.hu

**Kivonat** Jelen tanulmányban a magyar névutók csoportját vizsgálja. A korpuszvezérelt elemzés során hat – bináris – tulajdonságra fókuszálunk, a névutók kategóriájára zárt és merev osztály helyett kontinuumként tekintve. Eredményeink egy újszerű megközelítést jelentenek, amellyel hatékonyabban elemezhetők a névutószerű elemek a mondatban.

**Kulcsszavak:** névutók, korpusznyelvészet, mondatelemző

### 1. Bevezetés

Ebben a cikkben egy átfogó, korpuszvezérelt elemzést készítünk a magyar névutók folyamatosan változó csoportjáról. A cél az, hogy ezekre a szavakra ne egy merev és homogén kategória elemeiként tekintsünk, hanem inkább kontinuumként, amelyben több különböző tulajdonság alapján helyezhetők el az egyes szavak. Távollabbi célunk egyfajta függőségi nyelvtan keretein belül elemezni őket, lehetőség szerint mindvégig szem előtt tartva egy pszicholingvisztikailag motivált elemzőrendszer (ANAGRAMMA, [1,2]) igényeit.

A legfontosabb tisztázandó kérdés az elején, hogy mit értünk névutó alatt. Számítógépes nyelvészeti szempontból a névutók azok az – önálló – elemek, amelyek megkérdőjelezhetetlenül jelzik egy főnévi csoport végét a mondatban (ld. például az 1-es példában). Ugyanakkor elkerülhetetlen egy mélyreható irodalmi áttekintés arról, hogy pontosan mely elemek tartoznak ebbe a csoportba. A 2.1. fejezetben bemutatjuk, hogy a jelentősebb nyelvészeti iskolák hogyan vélekedtek a névutókról, noha a mi célunk ezeknek a szavaknak az elemzése függetlenül attól, hogy minden iskola szerint névutók-e vagy sem. A 2.2. fejezetben a kutatás háttérét jelentő elemzőrendszert mutatjuk be röviden. A 3. fejezetben bemutatjuk a saját kritériumainkat és osztályozási feltételeinket. A 4. fejezetben összefoglaljuk mindazt, ami vizsgálataink során a korpuszból nyilvánvaló vált, illetve bemutatjuk, hogyan jellemezhetők az egyes névutószerű elemek a szófaj jellemző tulajdonságait képviselő bináris jegyek segítségével. Cikkünket az 5. fejezetben összefoglalással zárjuk.

(1) *A főnév után jön a névutó.*

## 2. Bevezetés

### 2.1. Kapcsolódó irodalom

Aligha találunk még egy olyan szóosztályt, amely viszonylag kevés számú eleme ellenére olyan hosszan tartó vitát eredményezne, mint a magyar névutók. Szinte lehetetlen két különböző tanulmányt találni, amelyek ugyanolyan klasszifikáció mellett érvelnének. Ebben a fejezetben megpróbálunk áttekintést adni arról, amit az elméleti nyelvészek eddig mondtak a magyar névutókról. A hagyományos magyar nyelvtanokban [3] a névutók a funkciószavak egy csoportját alkotják, amelyek morfológiai jellegű egységet alkotnak az őket megelőző főnévvel (vagy főnévi kifejezéssel), így esetragként működnek. Ezek a nyelvtanok a következőképpen csoportosítják a névutókat:

- (2) a. **Valódi névutók**, a névutók, amelyek esetrag nélküli főnevet követnek: *által, alá, alatt, alól, elé, előtt, elől, ellen, felé, felől, fölé, fölött, helyett, iránt, köré, körül, közé, között, közül, mellé, mellett, mellől, miatt, mögé, mögött, mögül, nélkül, szerint, után.*
- b. **Ragvonzó névutók**: *fogva, kezdve, nézve, keresztül, túl, együtt* stb.
- c. **Birtokos személyjelet tartalmazó névutók**: *alapján, céljából, ellenére, érdekében, értelmében, esetében, folyamán, javára, következtében, nyomán, részéről, számára, táján* stb.

Ezt a kategorizációt erősen vitatják a strukturalista vagy generatív háttérű tanulmányok. A *Strukturális magyar nyelvtan* 3. kötetében [4] a névutókat három (és fél) csoportba sorolják:

- (3) a. **Esetszerű névutók**: a csoport tagjai esetragként viselkednek. Nem adnak esetet az NP-nek, esetrag nélküli főnévi csoportot vonzanak. Például: *által, alá, alatt, alól, elé, előtt, elől, ellen, felé, felől, fölé, fölött, helyett, iránt, köré, körül, közé, között, közül, mellé, mellett, mellől, miatt, mögé, mögött, mögül, nélkül, szerint, után.*
- b. **Valódi névutók**: azok az elemek, amelyek lexikai esettel rendelkező névszót vonzanak: *alul, át, belül, kívül, együtt* stb.
- c. **Egyéb névutók**: azok a névutók, amelyek sosem fordulnak elő személyes névmással: *végett, közben, óta, gyanánt, hosszát.*
- d. (Névutószerű elemek): egy átmeneti osztály, birtokos személyjeles alakokat tartalmaz.

Felsorolásunkból még hiányzik egy tisztán generatív megközelítés. Ezt biztosítja például É. Kiss a *The Syntax of Hungarian*-ben [5]. Itt az eddig említett szavak a következő kategorizációban jelennek meg:

- (4) a. **Vonzatos határozószók** (lexikai esettel rendelkező főnévi csoportot vonzó „névutók”): *együtt, alul* stb.
- b. **Idiomatikus melléknévi vagy határozói igenevek**, lexikai esettel rendelkező főnévi csoportot vonzanak: *nézve, kezdve* stb.
- c. **Esetrag nélküli főnévi csoportot vonzó névutók**: *alá, alatt, alól, elé, előtt, elöl, felé, felől, fölé, fölött, fölül, köré, körül, közé, között, közül, által, ellen, ellenére, helyett, szerint, iránt, miatt, nélkül, után, végett és révén*.

Ez a három különböző kategorizáció jól példázza, milyen bizonytalan a névutók meghatározása a magyar nyelvészeti irodalomban. Ugyanakkor szükséges egy másik, erre a kérdésre vonatkozó irodalmi tételre is hivatkoznunk, nevezetesen Dékány Éva doktori disszertációjára, amely részletes listát tartalmaz az összes magyar névutóról [6]. Dékány Éva csoportosítását a *naked* és a *dressed* névutók fogalmára alapozta, amely meghatározások eredetileg Marác Lászlótól származnak [7]. *Naked* névutó alatt az esetragos, *dressed* névutó alatt pedig az esetrag nélküli névszóval társuló névutókat értik. Ez a kategorizáció is szerepel az 1. táblázatban, amely egy kimerítő, és összehasonlító jellegű felsorolása az irodalomban valaha névutóként említett szavaknak (a fent említett négy tanulmány alapján).

Az 1. táblázatban dőlttel szedett szavak a névutók. Az ezeket követő négy oszlop a jelen tanulmányban említett négy hivatkozást jelöli (MG mint Magyar grammatika [3], Str mint Strukturális magyar nyelvtan [4], SoH mint The Syntax of Hungarian [5] és D.É. mint Dékány Éva disszertációja [6]). Az oszlopokban található jelek az adott tanulmány névutókkal kapcsolatos véleményét jelölik: egy szám azt jelenti, hogy az adott szót az adott tanulmányban névutónak tekintik; egy betű azt jelenti, hogy az adott szót említik a tanulmányban, de mint nem-névutót. A különböző számok és betűk különböző alkategorizációra utalnak. Ha zárójel szerepel az adott jel körül, azt jelenti, hogy ez a szó nem szerepel az adott tanulmányban, és nem vagyunk biztosak a(z adott tanulmány szerinti) kategorizációjában. A - azt jelenti, hogy az adott irodalom nem állít semmit az adott szóról.

A névutók kategorizációjának fő problémája azonnal világossá válik a táblázat alapján: van a szavaknak egy csoportja (hívjuk őket „tisza névutóknak”), amelyek minden helyzetben névutóként viselkednek, a névutók minden fontos tulajdonságával rendelkeznek, és nem utolsósorban minden jelentős tanulmányban névutóként határozzák meg őket. Más szavakkal: a táblázat minden oszlopában egy szám tartozik hozzájuk. Ugyanakkor vannak olyan szavak is, amelyek – és pontosan erre kívánunk rávilágítani az 1. táblázattal, – bizonyos szempontból névutók, de más szempontból valami egészen mások. A célunk az, hogy ezt a folyamatosságot ragadjuk meg: meghatározzuk a prototipikus névutók csoportját és jellemzőit, illetve körülírjuk az ezt a magot körülvelő rétegeket, amelyek a nem prototipikus névutókat tartalmazzák.

1. táblázat. A névutóként tárgyalt szavak listája és csoportosításuk összefoglalása. Rövidítések: MG: Magyar grammatika [3], Str: Strukturális magyar nyelvtan [4], SoH: The Syntax of Hungarian [5], D.É.: Dékány Éva disszertációja [6]. Számok: a szó az adott helyen névutó; betűk: nem névutó (eltérő szám / betű: az adott tanulmány alkategóriái). –: nem meghatározható a besorolás. Zárójel: nem egyértelmű kategorizáció.

	MG	Str	SoH	D.É.		MG	Str	SoH	D.É.
<i>alá</i>	1	1	1	1	<i>körül</i>	1	1	1	1
<i>alapján</i>	3	a	-	(b)	<i>következtében</i>	3	a	-	(b)
<i>alatt</i>	1	1	1	1	<i>közben</i>	1	3	-	-
<i>alól</i>	1	1	1	1	<i>közé</i>	1	1	1	1
<i>által</i>	1	1	1	1	<i>közel</i>	2	-	(a)	2
<i>alul</i>	2	2	a	2	<i>között</i>	1	1	1	1
<i>át</i>	2	2	a	2	<i>közül</i>	1	1	1	1
<i>belül</i>	2	2	a	2	<i>létére</i>	1	a	-	b
<i>céljából</i>	3	a	-	(b)	<i>mellé</i>	1	1	1	1
<i>együtt</i>	2	2	a	2	<i>mellett</i>	1	1	1	1
<i>elé</i>	1	1	1	1	<i>mellől</i>	1	1	1	1
<i>ellen</i>	1	1	1	1	<i>miatt</i>	1	1	1	1
<i>ellenére</i>	3	a	-	1	<i>módjára</i>	1	a	-	-
<i>elől</i>	1	1	1	1	<i>módra</i>	1	-	-	-
<i>előtt</i>	1	1	1	1	<i>mögé</i>	1	1	1	1
<i>esetén</i>	3	a	-	1	<i>mögött</i>	1	1	1	1
<i>felé</i>	1	1	1	1	<i>mögül</i>	1	1	1	1
<i>felől</i>	1	1	1	1	<i>múltán</i>	1	-	-	-
<i>felül</i>	2	-	(a)	2	<i>múlva</i>	1	-	-	1
<i>fogva</i>	2	2	b	a	<i>nélkül</i>	1	1	1	1
<i>főlé</i>	1	1	1	1	<i>nézve</i>	2	2	b	a
<i>fölött</i>	1	1	1	1	<i>nyomán</i>	3	a	-	1
<i>főül</i>	2	-	(a)	1	<i>óta</i>	1	3	(1)	1
<i>folytán</i>	3	-	-	1	<i>részére</i>	3	a	-	b
<i>gyanánt</i>	1	3	(1)	1	<i>révén</i>	3	a	-	b
<i>hasonlóan</i>	-	-	-	2	<i>során</i>	3	a	-	1
<i>helyett</i>	1	1	1	1	<i>számára</i>	3	a	-	b
<i>hosszat</i>	-	3	-	-	<i>szembe</i>	(2)	-	(a)	2
<i>innen</i>	2	-	-	2	<i>szemben</i>	2	-	(a)	2
<i>íránt</i>	1	1	1	1	<i>szemből</i>	(2)	-	(a)	2
<i>ízben</i>	1	-	-	-	<i>szemközt</i>	(2)	-	(a)	2
<i>javára</i>	3	a	-	b	<i>szerint</i>	1	1	1	1
<i>kedvéért</i>	3	a	-	b	<i>tájban/tájt</i>	1	-	-	1
<i>képest</i>	-	-	(a)	2	<i>túl</i>	2	2	a	2
<i>keresztül</i>	2	2	a	2	<i>túlra</i>	(2)	-	(a)	2
<i>kezdve</i>	2	2	b	a	<i>túlról</i>	(2)	-	(a)	2
<i>kivéve</i>	(2)	-	(b)	a	<i>után</i>	1	1	1	1
<i>kívül</i>	2	2	a	2	<i>útján</i>	3	a	-	1
<i>kívülre</i>	(2)	-	-	2	<i>véggett</i>	1	3	(3)	1
<i>kívülről</i>	(2)	-	-	2	<i>végig</i>	2	-	(a)	2
<i>köré</i>	1	1	1	1					

Az említett csoportosítások mind egyetértenek a névutók egy aspektusában: az esetragoktól való elkülönítésükben. Éppen ezért itt most nem kívánunk szólni az esetragok és a névutók kapcsolatáról. Ami igazán érdekel minket, azok az utóbbiak jellemző tulajdonságai: ezek okozzák ugyanis a kategorizációk eltéréseit. A korpuszt vizsgálva ezekre a jellemzőkre koncentrálnunk, próbálva egy nagyobb képet megragadni, a névutókra mint kontinuumra tekintve: a prototipikus névutók minden tulajdonsággal rendelkezni fognak, míg a perifériás névutók csak néhányal.

## 2.2. Az elemzőrendszer

Az ANAGRAMMA nyelvi elemzőrendszer az emberi szövegfeldolgozást igyekszik modellálni, ezért az elemzendő szöveget balról jobbra, szavanként elemzi. Az aktuálisan vizsgált szó mellett azonban az azt követő néhány tokenre is „kitekint”: ezt egy két szó méretű elemzési ablakkal modelláljuk. Az ANAGRAMMA elemzésének kimenete egy függőségi fa. Az elemző egyik fontos működési alapelve a *kereslet-kínálat* keretrendszer, amelyben az épülő függőségi fa kapcsolatai a mondatban szereplő szavak keresletei és kínálatai alapján valósulnak meg. Minden egyes token több tulajdonsággal rendelkezik, amely tulajdonságok keresletként vagy kínálatként jelennek meg a mondatban. Ha egy kereslet összetalálkozik egy megfelelő kínálattal, akkor megszületik a függőségi kapcsolat közöttük.

## 3. Módszer

Az 1. fejezetben azt állítottuk, hogy a névutók jelentősége – számítógépes nyelvészeti szempontból – abban rejlik, hogy az esetragokhoz hasonlóan minden kétséget kizáróan jelzik egy főnévi csoport végét. Ugyanakkor az 1. táblázat számos eleme képes megjelenni a főnévi frázis előtt és után is. Jelen tanulmányunkban – tekintve, hogy távolabbi célunk minden névutószzerű elem viselkedésének algoritmikus leírása – egyet tartunk csak meg a kritériumok közül: a névutó-jelöltnek minden esetben együtt kell járnia egy főnévi csoporttal, mindegy, hogy maga előtt vagy maga után veszi, esetraggal vagy esetrag nélkül. Ezt a különbséget szemlélteti az 5a és az 5b alatti példa. Az 5a-ban szereplő *kívülről* beletartozik a vizsgált szavak körébe, de az 5b-ben szereplő nem.

- (5) a. *A hang a szobán kívülről jött.*  
b. *A hang kívülről jött.*

A vizsgálat során használt korpusz az MNSz2 2.0.4-es verziója [8]. A vizsgálat egyik szakasza motiválta a másikat; kereséseinket az 1. táblázat elemeiből indítottuk, de a keresések során előbukkanó újabb névutó-jelöltekkel bővítettük azt.

Az irodalom, valamint korpuszlekérdezéseink eredményei alapján a következők tulajdonságokat határoztuk meg:

- **pozíció:** a névutó a névszói csoport előtt vagy után szerepel? Ennek értéke független attól, hogy azok szigorúan követik-e egymást, vagy kerülhet közéjük más elem. A pozíció tulajdonságának értéke 1, ha a névutó minden esetben a névszó után következik, 0 egyébként.
- a vonzat **esetragja:** első lépésként, a tulajdonságok binaritását megtartandó, az esetrag nélküli és az esetragos főneveket vonzó névutók elkülönítését célozzuk meg. Ugyanakkor a későbbiekben szükséges lehet a névutók megkülönböztetése az általuk vonzott esetragok alapján is. Ennek a tulajdonságnak az értéke 1, ha a névutó mindig esetrag nélküli főnévhez járul, 0 egyébként.
- **szomszédosság:** értéke 1, ha a névutó és a névszó minden esetben közvetlen szomszédai egymásnak, 0 egyébként.
- pozíciójuk a **kérdőszavas mondatokban:** minden esetben követi a névutó a kérdőszót, vagy hátul maradhat? Az előbbi esetben az érték 1, egyébként 0.

Ezek jellemzően szintaktikai attribútumok, főként a – jelen tanulmányban alkalmazott – számítógépes nyelvészeti nézőpontunk motiválja őket. Ezen kívül azonban szükséges lehet néhány morfológiai jellemző számbavétele is, amelyek ugyan nem befolyásolják jelentősen ezeknek a szavaknak a számítógépes elemzését, de kellene a komplex képhez.

- **mutató névmások:** a névutó is mindig másolódik a mutató névmásra, vagy csak az esetrag? Az első esetben az érték 1, egyébként 0 (lásd 6a és 6b).
- személy-szám **egyeztetés:** ha a névutó egy személyes névmáshoz járul, hol jelenik meg az egyeztetés? Értéke 1, ha az egyeztetés a névutón jelenik meg, 0 egyébként (ld. 7a és 7b).

(6) a. *a rét alól*  
       *az **alól** a rét **alól***

b. *a réten át*  
       *azon a réten **át***

(7) a. *alólam*

b. *rajtam keresztül*

#### 4. Eredmények

A 2. táblázatban látható a névutók bináris értékű tulajdonságokra lebontott elemzése. A „poz” oszlop a névutónak a névszóhoz képest vett pozícióját értékeli: ha a névutó mindig követi a névszót, akkor ennek értéke 1, különben 0. A „NOM” oszlop a névutó által megkívánt esetraggal kapcsolatos: ha a névszó mindig esetrag nélküli, akkor ennek értéke 1, különben 0. A „szom” oszlop a szomszédosságot vizsgálja: ha a névszó és a névutó mindig közvetlenül egymás

mellett találhatóak, akkor ennek értéke 1, különben 0. Fontos kitétel, hogy a névszók előtt is megjelenni képes névutóknál a névszó előtti (és a névutó utáni) névelőt nem számítottuk, csak egyéb módosító elemek jelenlétét. A „wh” oszlop a kérdőszavas kifejezésekben vizsgálja a névutókat: ha azok a kérdőszó névutóiként jelennek meg, akkor ennek értéke 1, különben 0. Az „az\_a” oszlop a mutató névmási determinánsú főnévi csoportokkal kapcsolatos: ha a névutó a mutató névmáson is ismétlődik, akkor ennek értéke 1, különben 0. Végül a „pers” oszlop azt értékeli, hogy ha személyes névmáshoz járul a névutó, az egyeztetés a névutón (1) vagy egy külön elemen jelenik-e meg (0).

A módszertannal kapcsolatban megjegyzendő, hogy 5 darab ellenpélda szükségeltetett ahhoz, hogy egy adott névutó ne kapjon 1-es értéket egy tulajdonságra. Ha például 5 esetben szerepel a névszó előtt, akkor már 0 látható a „poz” oszlop adott sorában.

A „wh” tulajdonság kiértékelése különösen nehéz a korpusz alapján, ezért a legtöbb esetben itt az értéket saját nyelvi intuíciónk alapján határoztuk meg. A ?-et tartalmazó cellák azt jelzik, hogy az adott tulajdonság tesztelésére kiötlött példáink elfogadhatóságában nem vagyunk teljesen biztosak.

Az első, ami szembetűnik a 2. táblázatban, hogy a 4-6. oszlopban gyakori a -. Ez azt jelzi, hogy az adott szerkezetben az adott névutó nem fordul elő. Például a *gyanánt* névutót nem lehet személyes névmáshoz illeszteni, ezért szerepel - a *gyanánt* vektorának utolsó pozícióján.

Ha a kizárólag 1-es értéket tartalmazó vektorú névutókat tekintjük, akkor kirajzolódik a prototipikus névutók köre: ezek szinte teljesen megfelelnek a szakirodalmi kategorizációk alapján is „tisza névutóknak” nevezett csoport tagjaival. A kivételt egyrészt azoknak a névutóknak egy része jelenti, amelyeknek az egyes szám harmadik személyű névmáshoz illesztett alakja megegyezik az alap alakjukkal (*alá, elé*). Ezeknél megfigyelhető, hogy a névszó előtt is megjelenhetnek *elé a ...-nak* szerkezetben (ld. 8). Kivételt jelent még az *által*, amely főleg a szépírodalmi alkorpuszban, de kisebb számban a személyes alkorpuszban is előfordul esetragos névszóval (9).

(8) - *Mestrében meg sem állunk - szólt elé a kocsisnak.*

(9) *Természetesen szintén szigorúan tévén által.*

Ha eltekintünk a 4-6. tulajdonságok értékétől, kirajzolódik még egy jelentős csoport, amelyek az első három, szintaktikai szempontra 1 értéket kaptak. Mivel a 4-6. tulajdonság mint feltétel nem alkalmazható teljes körűen ezekre a szavakra, kihagyásuk egy algoritmikus feldolgozásnál nem okozhat gondot. Az 1-3. tulajdonságra kapott 1 érték azonban azt jelzi, hogy ezeket a szavakat érdemes lenne névutónak annotálni a korpuszban, mivel mindig főnévi csoportot lezáró pozícióban szerepelnek, közvetlenül a magán testes esetragot nem viselő névszó után.

A táblázatban a kakukktójas a *szemből*: ezt, mivel a fő szintaktikai tulajdonságokra még értéket sem kaphat, lévén hogy nem fordul elő olyan szerkezetekben, nem tekintjük semmilyen szempontból sem névutónak.

2. táblázat. A magyar nyelvészeti szakirodalomban valaha névutóként említett szavak tulajdonságvektorként történő megjelenítése.

	poz	NOM	szom	wh	az	_a	pers
<i>alá</i>	0	0	1	1	1	1	
<i>alapján</i>	1	0	1	1	0/1	0	
<i>alatt</i>	1	1	1	1	1	1	
<i>alól</i>	1	1	1	1	1	1	
<i>által</i>	1	0	1	1	1	1	
<i>alul</i>	1	0	0	1	0	0	
<i>át</i>	0	0	0	0	0	0	
<i>belül</i>	?	0	1	1	0	0	
<i>céljából</i>	1	0	1	1	0	-	
<i>együtt</i>	0	0	0	0	0	0	
<i>elé</i>	0	0	1	1	1	1	
<i>ellen</i>	1	1	1	1	1	1	
<i>ellenére</i>	1	0	1	1	0	-	
<i>elől</i>	1	1	1	1	1	1	
<i>előtt</i>	1	1	1	1	1	1	
<i>esetén</i>	1	0	1	1	0	-	
<i>felé</i>	1	1	1	1	1	1	
<i>felől</i>	1	1	1	1	1	1	
<i>felül</i>	1	0	0	0	0	0	
<i>fogva</i>	1	0	1	1	0	0	
<i>főlé</i>	1	1	1	1	1	1	
<i>főlott</i>	1	1	1	1	1	1	
<i>főül</i>	1	0	1	1	0	1	
<i>folytán</i>	1	0	1	?	0	-	
<i>gyanánt</i>	1	1	1	-	-	-	
<i>hasonlóan</i>	0	0	0	1	0	0	
<i>helyett</i>	1	1	1	1	1	1	
<i>hosszat</i>	1	1	1	-	-	-	
<i>innen</i>	0	0	0	?	0	0	
<i>íránt</i>	1	1	1	1	1	1	
<i>ízben</i>	1	1	1	-	-	-	
<i>javára</i>	1	0	0	1	0	1	
<i>kedvéért</i>	1	0	0	1	0	1	
<i>képest</i>	1	0	0	1	0	0	
<i>keresztül</i>	0	0	0	0	0	0	
<i>kezdve</i>	0	0	1	1	0	0	
<i>kivéve</i>	0	0	0	1	0	0	
<i>kívül</i>	0	0	0	1	0	0	
<i>kívülre</i>	1	0	0	1	0	0	
<i>kívülről</i>	1	0	1	1	0	-	
<i>köré</i>	1	1	1	1	1	1	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	

  

	poz	NOM	szom	wh	az	_a	pers
⋮	⋮	⋮	⋮	⋮	⋮	⋮	
<i>körül</i>	1	1	1	1	1	1	
<i>következtében</i>	1	0	1	1	0	-	
<i>közben</i>	1	1	1	1	1	-	
<i>közé</i>	1	1	1	1	1	1	
<i>közel</i>	0	0	0	0	0	0	
<i>között</i>	1	1	1	1	1	1	
<i>közül</i>	1	1	1	1	1	1	
<i>létére</i>	1	1	1	-	-	-	
<i>mellé</i>	1	1	1	1	1	1	
<i>mellett</i>	1	1	1	1	1	1	
<i>mellől</i>	1	1	1	1	1	1	
<i>miatt</i>	1	1	1	1	1	1	
<i>módjára</i>	1	1	1	-	0?	-	
<i>módra</i>	1	1	1	-	-	-	
<i>mögé</i>	1	1	1	1	1	1	
<i>mögött</i>	1	1	1	1	1	1	
<i>mögül</i>	1	1	1	1	1	1	
<i>múltán</i>	1	1	1	1	0	-	
<i>múlva</i>	1	1	1	-	-	-	
<i>nélkül</i>	1	1	1	1	1	1	
<i>nézve</i>	1	0	1	1	0	0	
<i>nyomán</i>	1	0	1	1	0	-	
<i>óta</i>	1	1	1	1	1	-	
<i>részére</i>	1	0	1	1	0	1	
<i>révén</i>	1	0	1	1	0	-	
<i>során</i>	1	0	1	1	0	1	
<i>számára</i>	1	0	1	1	0	1	
<i>szembe</i>	0	0	0	1	0	0	
<i>szemben</i>	0	0	0	1	0	0	
<i>szemből</i>	-	-	-	-	0	0	
<i>szemközt</i>	0	0	0	1	0	0	
<i>szerint</i>	1	1	1	1	1	1	
<i>tájban/tájt</i>	1	1	1	-?	-?	-	
<i>túl</i>	0	0	0	0	0	0	
<i>túlra</i>	1	0	0	1	0	0	
<i>túlról</i>	1	0	0	1	0	0	
<i>után</i>	1	1	1	1	1	1	
<i>útján</i>	1	0	1	1	0	-	
<i>végett</i>	1	1	1	1	-	1	
<i>végig</i>	0	0	0	0	0	0	



Határozottan kirajzolódik a periférián lévő szavak csoportja is: ezek azok, melyeknek minden értéke 0. Az *együtt, közel, túl, végig* esetében közös, hogy minden szempontból megvizsgálhatóak, tehát minden olyan szerkezetben előfordulnak, amiben a tipikus névutók (zöme), azonban nem úgy viselkednek ezekben, mint a tipikus névutók. Ezeknek az annotációjában érdemes a határozószói jellegükhöz ragaszkodni, viselkedésük a „vonzatos határozószó” terminussal ragadható meg a legjobban: egy esetragos névszót vonzanak, maguk előtt vagy maguk mögött is elfogadják, és távolra is kerülhetnek a vonzatuktól.

A legérdekesebbek azok az elemek, melyekhez az 1 0 1 vektor tartozik (a 4-6. tulajdonságot figyelmen kívül hagyva): ezek a közvetlenül a főnév után szereplő, de esetragos főnevet vonzó névutók. Mint látható, ezek viselkedésükben sokkal közelebb állnak a prototipikus névutókhoz, mint a vonzatos határozószók. A problémát az általuk megkövetelt esetrag okozza: egy névszó és egy prototipikus névutó kapcsolatának elemzése nem sokban különbözik egy esetragos névszóétól; ha az ANAGRAMMA elemzőrendszer balról jobbra történő feldolgozását vesszük alapul, akkor az algoritmus lényege, hogy az esetrag nélküli névszó mellett, ha megtaláljuk a névutót, a kettőt összekapcsoljuk, és együtt lehetnek majd a mondatban egy másik elemnek a vonzatai. A vonzatos határozószók esetében a kívánt esetrag keresletként jelenik meg a mondatban, amit egy, az adott esetragot magán viselő főnévi csoport kínálata fog kielégíteni. Az 1 0 1 vektorú elemek esetében azonban a névutót megelőző főnévi csoport nem esetrag nélküli, az azt lezáró esetrag azt kommunikálja, hogy itt véget ért egy főnévi csoport. A probléma megoldására alább javasolunk egy lehetséges algoritmust.

A 2. táblázat néhány cellája további magyarázatot igényel. Az *alapján* 0/1 értéket kapott az „az\_” tulajdonságra. Ezt azért így jeleztük, mert minden más névutó vagy az egyik, vagy a másik módon viselkedik a mutató névmási determinánssal, az *alapján* azonban mindkét formában előfordul a korpuszban (10a és 10b). Ez csak erre a névutóra jellemző.

- (10) a. *az alapján az idézet alapján*  
 b. *annak az élettani tapasztalatnak alapján*

A 3. táblázatban szerepelnek a korpuszlekérdezéseink során gyakran előkerülő, névutószerű viselkedést produkáló szavak (melyeknek egy része már máshol is említésre került névutó-jelöltként [9]). Mint látható, a közvetlen szomszédosság mindegyikük jellemző tulajdonsága; ezen kívül egy-két kivételtől eltekintve mindig a főnévi csoport után található. Emiatt az 1 0 1 vektorú csoport elemeihez közeli szavak, azzal a különbséggel, hogy morfológiailag összetettebbek, vagy birtokos személyjeles, vagy esszívuszi módhatározó ragos szóalakok. Mondatbeli elemzésük nem sokban különbözik attól, amit a részletes morfológiai elemzésük aktiválna, azonban a jelentésük és a kizárólag ebben a tipikus névutó-pozícióban történő előfordulásuk azt indokolja, hogy felvegyük őket a névutók közé.

Az eddig elmondottak alapján a névutószerű elemek három nagy csoportja körvonalazódik. A gépi mondatelemzés során – az ANAGRAMMA balról jobbra haladó, *kereslet-kínálat* keretrendszerben működő eljárását szem előtt tartva – a következő algoritmusokat gondoljuk megvalósíthatónak:

3. táblázat. A korpuszlekérdezések során megfigyelt névutószerű elemek listája és a tulajdonságokhoz rendelt értékeik.

	poz	NOM	szom	wh	az_	a	pers
<i>címén</i>	1	0	1	1	0	-	
<i>dacára</i>	0	0	1	1	0	-	
<i>eltérően</i>	0	0	1	1	0	0	
<i>esetében</i>	1	0	1	1	0	1	
<i>érdekében</i>	1	0	1	1	0	1	
<i>értelmében</i>	1	0	1	1	0	-	
<i>fényében</i>	1	0	1	1	0	-	
<i>függően</i>	1	0	1	1	0	0	
<i>hiányában</i>	1	0	1	1	0	-	
<i>idején</i>	1	0	1	1	0	-	
<i>jegyében</i>	1	0	1	1	0	-	
<i>keretében</i>	1	0	1	1	0	-	
<i>kezdődően</i>	1	0	1	1	0	0	
<i>köszönhetően</i>	0	0	1	1	0	0	
<i>követően</i>	1	0	1	1	0	0	
<i>megegyezően</i>	1	0	1	1	0	0	
<i>megelőzően</i>	1	0	1	1	0	0	
<i>megfelelően</i>	0	0	1	1	0	0	
<i>terén</i>	1	0	1	1	0	-	
<i>ürügyén</i>	1	0	1	1	0	-	
<i>vonatkozóan</i>	1	0	1	1	0	0	

- A prototipikus névutók esetén: ezek a szavak mindig közvetlenül egy testes esetrag nélküli – jelenleg tehát alapértelmezetten NOM-nak címkézett – névszó után találhatóak. Feldolgozásuk lényege, hogy az esetrag nélküli névszóhoz érkeve, az ablak elemeire kitekintve látja őket az elemző, ezért további analízis nélkül összekapcsolja a névszót a névutóval, és onnantól kezdve együtt, az esetragos névszókkal parallel módon vesznek részt az elemzésben. A prototipikus névutók POS-tagjének a NU kategóriát javasoljuk – ahogy az az ide tartozó elemek nagy részénél jelenleg is van. Ezek a szavak a 2. táblázat azon elemei, melyeknél az első három cellában 1-es érték található.
- A prototipikus névutókhöz képest a periférián elhelyezkedő elemek esetén: ezek a szavak mindig esetragos névszót vonzanak, a névszó előtt is, más elemek közébekelelődésével is megjelenhetnek. A 2. táblázatban minden cellában 0 értéket kaptak. Ezeknek az elemzése más, régens szerepű elemekével megegyezően történhet. POS-tagjuknak a HA kategóriát javasoljuk, de emellett tulajdonságaik között fontos feltüntetni, hogy milyen esetragú névszót vonzanak – ez lesz ugyanis a keresletük a mondat elemzése során.
- Az esetragos névszót vonzó névutók esetén: ezek a szavak, bár mindig közvetlenül a névszó után találhatóak, különböznek a prototipikus névutóktól abban, hogy a névszó előttük testes esetragot visel(het). A 2. táblázatban 1 0 1 a hozzájuk rendelt vektor kezdete. Elemzésük nem egyezhet meg a vonzatos határozószónak tekintett, perifériás névutószerű elemekével, hiszen

nem klasszikus értelemben vett régensek; nem keresik a vonzatukat a mondatban, hanem egyértelműen tudják, hogy közvetlenül maguk előtt található. A problémát az jelenti, hogy ha az esetragos névszó elemzésekor azt nem kapcsoljuk össze a névutójával, az esetragja miatt az elemzés – rövid ideig ugyan, de – tévútra juthat. A 11-es példában a már feldolgozott *lát-tam* keres egy tárgyesetben lévő névszót – az *előadást* tokenhez érve meg is találhatója, és vonzatául veheti. Ezeknek az eseteknek a kiküszöbölésére azt javasoljuk, hogy ne csupán az esetrag nélküli, de a testes esetragot viselő névszók elemzésekor is vizsgáljuk meg az ablak első elemét, hogy nem névutó-e. Természetesen ez az eljárás is több problémát felvet: gazdaságosságának, pontosságának vizsgálata az itt felvázolt algoritmus implementációja után lehetséges.

(11) *Nem láttam sehol az előadást követően a főnököt.*

## 5. Összegzés

Tanulmányunkban a névutószerű elemek részletes, korpuszalapú elemzését mutattuk be. Megvizsgáltuk, hogy a nyelvészeti szakirodalomban is említett, a névutók egyes csoportjait ezekben a tanulmányokban egymástól élesen elválasztani hivatott tulajdonságok a korpuszon vizsgálva miként jellemzik a névutókat. A 2. táblázat eredményeit vizsgálva látható, hogy jól rendszerezhető a számtalan névutó-jelölt ezeknek a tulajdonságoknak a mentén. Találunk prototipikus és perifériás elemeket is, melyeknek az elemzése egy gépi mondatelemzés során eltérő kell, hogy legyen. Három nagy csoport körvonalazódik, melyeknek feldolgozására három különböző algoritmust javasoltunk. A kutatás következő fázisában ezeknek az implementációját kell megvalósítani, hogy kiértékelhessük a teljesítményüket. Emellett fontos konkrét számokat rendelni a 2. táblázat 0 értékű celláihoz, hogy lássuk azok szignifikanciáját, ezáltal tovább pontosíthassuk a névutók besorolását.

## Hivatkozások

1. Prószték, G., Indig, B., Vadász, N.: Performanciaalapú elemző magyar szövegek számítógépes megértéséhez. In Kas, B., ed.: „Szavad ne feledd!”: Tanulmányok Bánréti Zoltán tiszteletére. MTA NYTI, Budapest (2016) 223–232
2. Prószték, G., Indig, B.: Magyar szövegek pszicholingvisztikai indíttatású elemzése számítógéppel. *Alkalmazott nyelvtudomány* **15**(1-2) (2015) 29–44
3. Keszler, B., ed.: Magyar grammatika. Nemzeti Tankönyvkiadó (2000)
4. Kiefer, F., Bánréti, Z.: Strukturális magyar nyelvtan: Mondattan. Strukturális magyar nyelvtan. Akadémiai Kiadó (1992)
5. É. Kiss, K.: The Syntax of Hungarian. Cambridge Syntax Guides. Cambridge University Press (2002)

6. Dékány, É.K.: A profile of the Hungarian DP: the interaction of lexicalization, agreement and linearization with the functional sequence (2012) PhD dissertation. University of Tromsø.
7. Marác, L.: Dressed or naked: the case of the PP in Hungarian. In Werner, Abraham és de Meij, S., ed.: *Topic, Focus and Configurationality*. John Benjamins, Amsterdam and Philadelphia (1986) 223–252
8. Oravecz, Cs., Váradi, T., Sass, B.: The Hungarian Gigaword Corpus. In Calzolari, N., Choukri, K., Declerck, T., Loftsson, H., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., Piperidis, S., eds.: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, European Language Resources Association (ELRA) (2014)
9. Ligeti-Nagy, N.: Szövegtörzsek pontosabb annotációja gépi elemzéshez. In: *Többszempontú és kommunikáció Kelet-Közép-Európában. XXIV. Magyar Alkalmazott Nyelvészeti Kongresszus, Erdélyi Múzeum-Egyesület* (2015) 421–429