

## Személyragos főnévi igeneves bővítményt megengedő predikátumok kinyerése a Magyar Nemzeti Szövegtárból

Bottyán Gergely, Sass Bálint

MTA Nyelvtudományi Intézet, Korpusznyelvészeti Osztály  
1068 Budapest, Benczúr u. 33.  
{bottyang, joker}@nytud.hu

### Absztrakt

A szláv nyelvekben, valamint a legtöbb germán és újlatin nyelvben csak személyrag nélküli főnévi igenevek fordulnak elő. A magyarban azonban – néhány egymással különben nem rokon nyelvhez, például a walesihez és a portugálhoz hasonlóan – találhatunk személyrag nélküli és személyraggal ellátott főnévi igeneveket is [4].

A magyar főnévi igenév két típusát mutatja be az 1. táblázat.

#### 1. táblázat: A magyar főnévi igenév két típusa

A kulcsot meg kell keresni.	(1)
Sikerült mindent időben befejeznie.	(2)

A főnévi igenév személyragos változata (1. táblázat, 2) az utóbbi időben több, a magyar mondat szerkezet generatív szemléletű leírásával foglalkozó nyelvész figyelmét is felkeltette. A vonatkozó kutatások igyekeztek azokat a mondatkörnyezeteket azonosítani, amelyekben személyragos főnévi igenevek megjelenhetnek. Ezen kívül célul tűzték ki a személyragos főnévi igeneveket tartalmazó kifejezések szerkezetének ábrázolását is [1,5,6]. A mondattan chomskyánus iskolája művelőinek többségéhez hasonlóan a kutatás során a témáról szóló dolgozatok szerzői saját nyelvi intuícióikra hagyatkoztak, és nem alkalmaztak szisztematikus adatgyűjtési eljárást. Ennek ellenére [5] például azt állítja, hogy kimerítő felsorolását adja a személyragos főnévi igenevek használatát megengedő mondatkörnyezeteknek, valamint, hogy empirikus anyagon alapul.

Jelen dolgozat annak a kutatásnak az eredményeiről számol be, amelyet a 153,7 millió szavas szótövesített, morfológiai annotációval ellátott és egyértelműsített Magyar Nemzeti Szövegtár [7] alapján végeztünk. Elsődleges célunk az volt, hogy megvizsgáljuk annak az állításnak az érvényességét, miszerint [5]-ben minden, személyragos főnévi igeneves bővítményt megengedő predikátum (ezeket ezentúl licenzornak fogjuk nevezni) szerepel. Ehhez a korpuszadatokból kívántunk licenzorlistát készíteni. Egy további cél volt annak megállapítása, hogy a megfelelő korpuszmondatokban mely licenzorok mely személyragos igenevekkel fordulnak elő. A nyelvtechnológiában általános módon félig automatikusan, félig manuálisan oldottuk meg a feladatot, iteratív módszerrel.

Először a morfoszintaktikai kód alapján automatikusan kigyűjtöttük a Magyar Nemzeti Szövegtárból a személyragos főnévi igenevet tartalmazó mondatokat. Munkahipotézisünk az volt, hogy a licenzorok a személyragos főnévi igenévvel azonos tagmondatban fordulnak elő. Mivel a korpuszban tagmondat-annotáció nincs, a tagmondatjelölteket közelítő módszerrel próbáltuk megfogni. A közelítő módszerhez a tagmondathatároló írásjelek és tagmondatkezdő kötőszavak [3] listáját használtuk fel. A kapott tagmondatjelöltekből automatikusan kiszűrtük azokat, amelyek [2]-ben azonosított licenzort tartalmaztak, az előforduló személyragos főnévi igenevet pedig feljegyeztük a megfelelő licenzorhoz. Ezután következett az eljárás manuális része: a megmaradt tagmondatjelöltekben új licenzorokat kerestünk, majd az új, kiegészített licenzorlistával újra alkalmaztuk az eljárást előlről.

Három iteráció után a következő eredményekre jutottunk. A szótövesített licenzorlista 197 tagú. Az ezeket tartalmazó tagmondatok a Magyar Nemzeti Szövegtárban szereplő 228367 személyragos főnévi igenév tokenből 223140-et (98 %) fednek le. Szótövesítve 17874 licenzor – személyragos főnévi igenév párt kaptunk.

Vizsgálatunk eredményei alapján az alábbi következtetéseket vonhatjuk le. (i) Bőven akad olyan mondatkörnyezet, amely [5]-ben nem került azonosításra, tehát az ott található licenzorlista nem kimerítő. (ii) Ha egy nyelvtani konstrukciónak legalább egy jegye (esetünkben: a személyragos főnévi igenév személyragja) megfogható a Magyar Nemzeti Szövegtárban vagy bármely más, hasonló méretű gazdagon annotált korpuszban, megéri a fáradságot félig automatikus, félig manuális módszerrel kinyerni a megfelelő adatokat a nyelvi erőforrásból, mert ezzel elejét vehetjük sietve levont téves konklúzióknak.

## Bibliográfia

1. É. Kiss, K.: Agreeing infinitives with a case-marked subject. In *The syntax of Hungarian*. Cambridge: Cambridge University Press. (2002) 210-221
2. É. Kiss, K.: A személyragos alaptagú főnévi igeneves kifejezés. In É. Kiss, K., Kiefer, F. and Siptár, P., *Új Magyar nyelvtan*. Budapest: Osiris. (1999) 118-121
3. Gábor, K., Héja, E. and Mészáros, Á.: Kötőszók korpusz-alapú vizsgálata. In Alexin, Z. and Csendes, D. (eds.), *MSZNY 2003 – I. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged: Szegedi Tudományegyetem. (2003) 305-306
4. Miller, D. G.: Where do conjugated infinitives come from? *Diachronica*, 20, 1. (2003) 45-81
5. Tóth, I.: Inflected infinitives in Hungarian. Ph. D. dissertation. Tilburg: University of Tilburg. (2000)
6. Tóth, I.: Can the Hungarian infinitive be possessed? In Kenesei, I. and Siptár, P. (eds.), *Proceedings of the conference "Approaches to Hungarian"*. Budapest: Akadémiai Kiadó. (2002) 135-160
7. Várad, T.: On Developing the Hungarian National Corpus. In Vintar, Š. (ed.), *Proceedings of the Workshop "Language Technologies – Multilingual Aspects"*. Ljubljana: University of Ljubljana. (1999)