

A BIRTOKOS ESETE AZ ABLAKKAL

Vadász Noémi – Indig Balázs

Bevezetés

Jelen cikk a jelöletlen és a *-nAk* ragos birtokos szerkezetek¹ kezeléséről szól. Közelebbről arról, hogy egy performanciaalapú, pszicholingvisztikai indíttatású elemzőmodell – az ANAGRAMMA – a balról jobbra és szavanként történő elemzés során hogyan birkózik meg a birtokos szerkezetekkel és az eset-többértelműséggel.

Az ANAGRAMMA elemzőrendszer működési elveinek rövid bemutatása után a magyar birtokos szerkezeteket ismertetjük. Ezután korpuszmérésekkel illusztráljuk a jelöletlen birtokos viselkedését. Végül az ANAGRAMMA keretein belül egy lehetséges megoldást kínálunk az eset-többértelműségek tisztázására és a birtokos szerkezetek kezelésére.

1. Az ANAGRAMMA

Az ANAGRAMMA performanciaalapú elemzőrendszer, l. Prószyński–Indig–Vadász (2016), egy pszicholingvisztikai indíttatású nyelvelemző modell, vö. Indig–Vadász–Kalivoda (2016), amely az emberi mondatmegértés mintájára balról jobbra és szavanként elemzi a szöveget. Az elemzőmodell többmondatos szövegegységeket is képes elemezni.

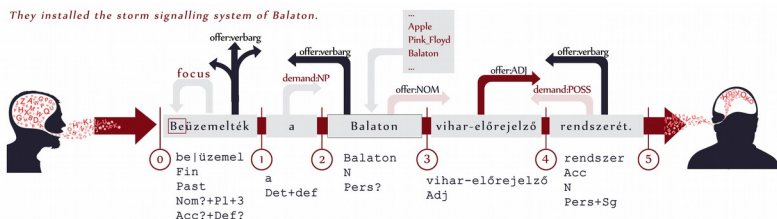
1.1 Az ANAGRAMMA működése

A rendszer működésének három alapvető eleme a **kereslet-kínálat** elvű keretrendszer, l. Prószyński–Indig (2015), a **tározó** és az **ablak** (a részleteket l. Indig–Vadász 2016a). Az elemző kimenete egy különböző típusú irányított éleket tartalmazó függőségi gráf.

A kereslet-kínálat elvű keretrendszer azt jelenti, hogy az elemző látóterébe kerülő tokenek más – korábbi vagy éppen később érkező – elemek számára kínálatként viselkedhetnek, ugyanakkor saját keresleteik is lehetnek (pl. az ige keresi vonzatait, a melléknévek kínálatként viselkednek a főnevek módosítójának szerepében). A keresleteket különböző **keresőeljárások** valósítják meg, amelyek megszorításokat is tartalmazhatnak például a keresés irányára (például a determinánsokat csak balra keresi) vagy a keresett elem különböző

¹ Cikkünkben Kiefer (1992) terminusaihoz tartjuk magunkat.

morfológiai jegyeire (például az ige ragozása meghatározhatja a keresett alany vagy tárgy számát és személyét).



1. ábra: Az elemzett mondaton jól megfigyelhetők a keresletek és kínálatok által létrejövő, különböző típusú függőségi élek.

Az elemzés során az adott pillanatig még be nem kötött éleket, tehát az eddigi kínálatokat – az éppen elemzett token megelőző tokeneket, azok morfológiai elemzésével együtt –, a keresleteket, valamint a kész részszerkezeteket az ún. tározó tartalmazza. A tározó tartalma az elemzés bármely pillanatában lekérdezhető a keresletek számára, így annak tartalma az elemzés közben folyamatosan változik.

Az elemzés során a függőségi élek bekötéséhez a tározón kívül rendelkezésre áll egy, az aktuálisan elemzett szótól jobbra tekintő, két token méretű elemzési ablak is. Ebben az ablakban megjelenhetnek az aktuális token elemzését befolyásoló elemek. A kínálatok az ablakban lévő tokenek morfológiai jegyei alapján jelennek meg, ám a tokenek egészen addig nem indítják el a saját keresőiket, amíg nem kerülnek az ablak bal szélére aktuálisan elemzett tokenként.

1.2 Kétfázisú mondatfeldolgozás

Az ANAGRAMMA az emberi szövegfeldolgozás mintájára működik. A modell kidolgozásához a *Sausage Machine* kétfázisú mondatfeldolgozását vettük alapul, l. Frazier–Fodor (1978). Frazier–Fodor (1978) modelljében az első fázis a PPP ('Preliminary Phrase Packager'), amely a szöveges bemenet szócsoportjaiból készít frázisokat. A második fázisban ('Sentence Structure Supervisor') a frázisok további nemterminális csomópontok hozzáadásával megkapják a szerepüket a mondatban².

² Az elemzési ablak fontosságát jól mutatja, hogy nehezebb olvasni a hírcsatornákon a képernyő alján végigfutó hírszalagot, ahol nem tudunk előretekinteni (l. <http://users.itk.ppke.hu/~yanzigy/olvaso>).

Az angol nyelvre kidolgozott modellben az első fázis eredetileg egy kb. hat szó méretű ablakban dolgozik. Mi a modellt a magyar nyelvre alkalmazzuk, amelyben – a magyar nyelv agglutináló jellege miatt – célszerűbb egy kisebb ablak beállítása, így a PPP fázist az ANAGRAMMA keretében egy három token (aktuálisan elemzett token + két token) méretű elemzési ablak valósítja meg.

A Sausage Machine alkalmazása az ANAGRAMMA modellben azt jelenti, hogy a balról jobbra és szavanként történő feldolgozás egy lépésében az elemző a három token méretű elemzési ablak bal szélén lévő elemmel foglalkozik, elindítja az összes, az adott token morfológiai elemzése által indítandó folyamatot, megvizsgálja, hogy a token mint kínálat kielégíthet-e egy keresletet a tározóban. Az ablakban található további tokenek – az elsődleges morfológiai információjukkal együtt – módosíthatják az aktuális elemzési lépéseket (pl. jelentős szerepük van a szófaji egyértelműsítésben is). Modellünkben a kétfázisú mondatelemzés kimenete a fent említett, különböző típusú éleket tartalmazó függőségi gráf.

1.3 Performanciaalapúság

A performanciaalapú elemző létrehozásakor olyan szerkezetek kezelésére is gondolnunk kell, amelyek a hagyományos elméleti nyelvészeti ítéletek alapján agrammatikusak lehetnek, ám korpuszokban mégis találhatunk rájuk példát. Az elemzőnek minden olyan bemenetet kezelnie kell, amelyet az ember is megért – még ha nehezebben és lassabban is, mint a gyakori és grammatikus szerkezeteket. Ezzel szemben a csak elméletben létező, de valójában soha elő nem forduló, ritka szerkezetek kezelése nem cél.

A különböző birtokos szerkezetek előfordulásait a performanciaalapúság jegyében korpuszon vizsgáltuk és osztályoztuk a kezelésük módja szerint. Az eredményeket a 2.1 fejezet ismerteti.

2. A birtokos szerkezetek

Kiefer (1992) alapján a magyarban kétféle birtokos szerkezetet különböztünk meg: 1) jelöletlen birtokos és 2) *-nAk* ragos birtokos. A kétféle szerkezet közös tulajdonsága, hogy a birtokos csak olyan mondatban szerepelhet, amely egyéb birtokossal kapcsolatban nem álló, vele számban és személyben egyeztetett birtokszót tartalmaz.

A kétféle birtokos között két fontos – a számítógépes kezelés szempontjából kritikus – különbség van. Az egyik, hogy amíg a jelöletlen birtokos és

birtoka közé nem kerülhet a birtok névelője (1a), addig a *-nAk* ragos birtokost követheti a birtoka névelője (1b).

- (1) a. **a kapitány a hajója*
 Nom Pers
 b. *a kapitánynak a hajója*
 Dat Pers

A másik fontos különbség, hogy a jelöletlen birtokos és birtoka csak fix sorrendben jelenhet meg, szemben a *-nAk* ragos birtokossal, amely a birtokától függetlenül bárhol helyet foglalhat a mondatban. Vö. a (2) és a (3) alatti példákat a kétféle szerkezet közötti különbségekhez.

A (2) alatti példákon megfigyelhető, hogy a jelöletlen birtokos és birtoka egy frázisként kezelendő. A (2a) és (2c) mondat mutatja, hogy a jelöletlen birtokos és birtok sorrendje kötött, bárhol foglaljon helyet a mondatban a szerkezet. A (2b) mondat agrammatikusságát az okozza, hogy nem a megfelelő sorrendben áll a jelöletlen birtokos és birtoka. Az *elsüllyedt* token melléknévi igenévként a birtokot módosítva bekerülhet a jelöletlen birtokos és birtoka közé a birtokos szerkezet NP-jében (2d).

- (2) a. *A kapitány hajója elsüllyedt.*
 Nom Pers Fin
 b. **Hajója a kapitány elsüllyedt.*
 Pers Nom Fin
 c. *Elsüllyedt a kapitány hajója.*
 Fin Nom Pers
 d. *A kapitány elsüllyedt hajója.*
 Nom *Fin Pers

A (3) alatti példák mutatják, hogy a *-nAk* ragos birtokos szerkezet birtokosa és birtoka külön frázisként nagyobb szabadsággal jelennek meg a mondatban, egymáshoz képesti sorrendjük nem kötött, és mivel beékelődhet a birtokos és birtoka közé a finit ige, ezért nem egy frázisként kell kezelnünk.

- (3) a. *A kapitánynak a hajója elsüllyedt.*
Dat Pers Fin
b. *A hajója a kapitánynak elsüllyedt.*
Pers Dat Fin
c. *A kapitánynak elsüllyedt a hajója.*
Dat Fin Pers
d. *A hajója elsüllyedt a kapitánynak.*
Pers Fin Dat
e. *Elsüllyedt a kapitánynak a hajója.*
Fin Dat Pers
f. *Elsüllyedt a hajója a kapitánynak.*
Fin Pers Dat

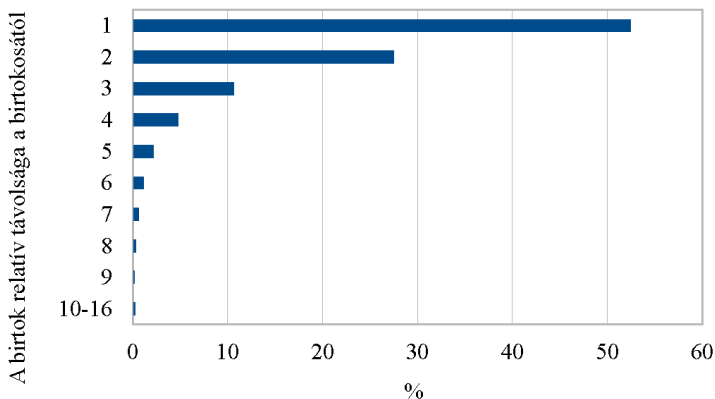
Ahhoz, hogy a kétféle birtokos szerkezetet megfelelően kezelje az elemző, először meg kell figyelni azok viselkedését valós nyelvi adatokon. A korpuszmérés eredményeit a következő fejezet ismerteti.

2.1 A jelöletlen birtokos szerkezet a korpuszban

Méréseinket a Pázmány Korpuszon végeztük, l. Endrédy (2016), amely 1,2 milliárd tokenből áll (amelyet több mint 30.000 weboldalról gyűjtöttek). Két részből tevődik össze, a főkorpusz (954.298.454 token) szerkesztett szövegeket tartalmaz, a kommentkorpusz (58.985.126 token) pedig szerkesztetlen szövegeket (kommenteket, fórumhozzászólásokat), a két alkorpuszt egyben kezeltük. A korpusz építéséhez felhasznált különböző típusú weboldalak biztosítják a korpusz heterogenitását.

A korpuszon megmértük³, hogy milyen messze kerülhet a jelöletlen birtokostól a birtoka. A méréskor a nominatívuszi esetű elem és az azt követő első birtokos ragozású elem távolságát mértük. Több mint 7.700.000 találatot kaptunk.

³ A mérést köszönjük Kalivoda Ágnesnek.



2. ábra: A birtok relatív távolsága a jelöletlen birtokosától. A függőleges tengelyen a tokenben mért relatív távolság, a vízszintes tengelyen a százalékos eloszlás látható.

Amint a 2. ábrán látható, a birtok az esetek 52,46%-ában közvetlenül követi a jelöletlen birtokosát. 27,45%-ban egy token ékelődik közéjük (általában egy melléknév vagy számnév), 10,66%-ban két token (általában egy határozószóval módosított melléknév vagy számnév) kerül a birtok és birtokosa közé.

Kevesebb, mint 10%-ban fordul elő, hogy a birtok ennél messzebbre kerül a jelöletlen birtokosától. Ezekben az esetekben általában a névszói frázist egy melléknévi igenév és vonzatai vagy egy felsorolás növeli szélesre, amint a (4) példában is látható. **Vastag** betűvel emeltük ki a jelöletlen birtokos szerkezet birtokát és birtokosát.

- (4) ***Krisztina** különleges, Swarovski kristályokból és minőségi japán gyöngyökből készült, egyedi tervezésű romantikus, nőies **nyaklánc***

Az eredményekből látszik, hogy a jelöletlen birtokos birtoka az esetek kb. 80%-ában beleesik a birtokos két token méretű elemzési ablakába. A cikkben ismertetett eljárás az ilyen eseteket képes kezelni. A maradék 20% megoldására – amikor a birtok két tokennél messzebb kerül a jelöletlen birtokosától – egyéb megoldást kell keresnünk. Feltételezzük, hogy az ember a szövegfeldolgozás során az elemzési ablak méretét meghaladó NP-k értelmezésekor

egyéb elemzési stratégiát használ, ám ezen elemzési stratégiák megértése és az elemzőbe való beépítése további kutatásokat igényel.

3. Az NP-k az ANAGRAMMA elemzőben

Az ANAGRAMMA elemzőrendszerben az elemzés során a tokenek a hagyományos szófaji címkék helyett különböző hierarchikus jegyszerkezetekkel rendelkeznek. A jegyszerkezetek egyes jegyei indítják el a különböző keresési vagy kínálati eljárásokat, így a kevert szófajok kezelése leegyszerűsödik⁴.

A kétfázisú mondatelemzés első fázisában történik a névszói frázisok előkészítése. Az NP-t a bal szélén egy determináns nyithatja. A determinánsok az elemzés során kínálatként bekerülnek a tározóba, ahol majd a fej Det-keresőjét elégítik ki.

Az NP-n belül a determináns után több különböző szófajú módosító elem szerepelhet, amelyeket az **NPMod** bennfoglaló kategória alá sorolunk. A főnévi fejet módosító elemek lehetséges címkéi a következők:

melléknév:	CAS/Nom: tő+NPMod+Adj+Sg/Pl(+PersSg/Pl1-3)
számnév:	CAS/Nom: tő+NPMod+Num+Sg/Pl(+PersSg/Pl1-3)
melléknévi igenév	
folyamatos:	CAS/Nom: tő+NPMod+PartPres+Sg/Pl(+PersSg/Pl1-3)
befejezett:	CAS/Nom: tő+NPMod+PartPast+Sg/Pl(+PersSg/Pl1-3)
beálló:	CAS/Nom: tő+NPMod+PartFut+Sg/Pl(+PersSg/Pl1-3)

Az **NPMod** jegyű elemek kínálatként viselkednek, ami azt jelenti, hogy a saját jegyeikkel együtt bekerülnek a tározóba készen arra, hogy egy később jövő elem (az NP feje) megfelelő keresői megtalálják őket és függőségi éleket hozzanak létre velük. Az **NPMod** jegy ugyanakkor saját keresőket is indít, hiszen ezeknek az elemeknek saját módosítóik is lehetnek. Ezeket a módosító elemeket balra, tehát a tározóban keresi.

A főnévi fejet módosító elemek jegyei között szerepel egy szófaji kategóriáját jelölő jegy (**Adj**, **Num**, **PartPres**, **PartPast**, **PartFut**), egy egyes illetve többes számot jelölő jegy (**Sg** vagy **Pl**), valamint egy opcionális jegy, amely a birtokos ragozás meglétét (**Pers**), illetve a birtokos ragozás által kijelölt birtokos számát (**Sg** vagy **Pl**) és személyét (**1**, **2** vagy **3**) jelöli.

⁴ Például a melléknévi igenévnek vannak melléknevekre és igékre jellemző vonásai. A melléknevekre jellemző viselkedést a melléknévi jegye által indított eljárások biztosítják (kínálatként viselkedik egy főnév számára), az igei viselkedését pedig az igei jegye (saját vonzatkereső eljárást indít), l. Indig–Vadász (2016b).

Ha az NP feje egy főnév – és nem záródott le az NP korábban egy esetragos **NPMod** kategóriájú elemmel –, annak a jegyei a következőképpen alakulnak:

főnév: **CAS/Nom: tő+N(+PropN)+Sg/Pl(+PersSg/Pl1-3)**

A főnevek jegyei között is megtalálható a szófaj jegye (**N**). Emellett egy opcionális jegy fejezi ki, ha tulajdonnévről van szó (**PropN**). A főnevek jegyei között is szerepel az egyes, illetve többes számot kifejező jegy, valamint a birtokos ragozás jegye a birtokosnak megfelelő számmal és személynél.

A főnevek és az **NPMod** jegyű elemek közös tulajdonsága, hogy magukon viselhetnek testes esetragot. Ha magukon viselik a testes esetrag jelenlétét jelölő **CAS** jegyet, akkor egy NP lezáró határt tesznek ki maguk mögé. Ez azt jelenti, hogy a frázis készen áll arra, hogy egy vonzatigényt kielégítsen, például az ige vonzataként létrejöjjön közte és az ige között egy függőségi él. A nominatívusz testetlen esetrag – amelyet a **Nom** főkategória jelöl – és a névutók azonban problémát okozhatnak, ekkor az elem mindaddig nem kapja meg a **CAS** jegyet, amíg nem történik meg az eset-egyértelműsítés. Az elemző előretételező elemzési ablaka segít abban, hogy a testes esetraggal nem rendelkező elemek mondatbeli szerepét mielőbb tisztázzuk.

A névutók kezelése további kutatásokat igényel, ám kijelenthető, hogy a névutók mint esetet kifejező elemek Indig–Vadász (2016b) alapján az esetragokhoz hasonlóan kezelendők, hiszen a névutók is a névszói frázist lezáró elemek. A névszói fejjel történő összekapcsolódásukkal válik a névszói frázis kínálattá az ige vonzatkeresői számára. Az elemző a névutók névszóhoz kapcsolását az elemzési ablak segítségével oldja meg csakúgy, mint a testes esetrag nélküli elemek eset-egyértelműsítését, amelyet a következő fejezet részletez.

3.1 *Nom: alany vagy jelöletlen birtokos?*

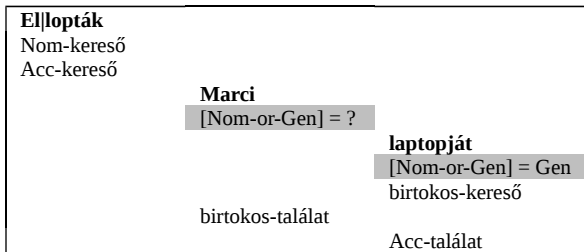
A jelöletlen birtokos kezelésekor az elsődleges feladat eldönteni azt, hogy a testes esetragot nem viselő NP kínálat lesz-e az ige vonzatkeresője számára, vagy egy birtokos szerkezet birtokosa. Az ilyen elemek mondatbeli szerepének eldöntésére kidolgoztuk az ún. [*Nom-or-Gen*] keresőeljárást, amely az elemzési ablak tartalmának megfelelően az első fázisban dönt.

A [*Nom-or-Gen*] keresőeljárás három lehetséges értéket vehet fel:

- | | |
|--------------------------------|---------------------------|
| 1. kezdetben, és ha nem ítélik | [Nom-or-Gen] = ? |
| 2. ha nominatívusznak ítélik | [Nom-or-Gen] = Nom |
| 3. ha birtokosnak ítélik | [Nom-or-Gen] = Gen |

A [Nom-or-Gen] keresőeljárás a testes esetragot magán nem viselő névszói elem esetében indul el [**Nom-or-Gen**] = ? értékkel. A keresőeljárás célja, hogy [**Nom-or-Gen**] = **Nom** vagy [**Nom-or-Gen**] = **Gen** értéket vegyen fel. Az elemzési ablakban található elemek segíthetnek a döntésben, azonban előfordulhat, hogy az elemzési ablakban nincsen olyan jegy, amely tisztázza a névszó testetlen esetét, így a keresőeljárás [**Nom-or-Gen**] = ? marad.

A keresőeljárás [**Nom-or-Gen**] = **Gen** értéket vesz fel akkor, ha az elemzési ablakában található egy olyan elem, amely birtokos ragozást visel magán. A jelöletlen birtokos szerkezet tulajdonságait és a 2.1 fejezetben kimért jelöletlen birtokos–birtok távolságot figyelembe véve elmondható, hogy a jelöletlen birtokostól birtoka az esetek legnagyobb részében elegendően kis távolságban lesz, tehát belefér annak elemzési ablakába.



3. ábra: A jelöletlen birtokost tartalmazó mondat elemzése

A 3. ábrán a jelöletlen birtokos szerkezetet tartalmazó mondat sematikus elemzése figyelhető meg. Az elemzés legelső lépésében az ige vonzatkeret-lehívása történik meg. Az *ellopták* ige egy nominatívusz és egy akkuzatívusz esetű vonzatigénnyel él, ezért egy Nom és egy Acc beállítású vonzatkeresőt helyez el a tározóban. A következő lépésben a *Marci* testes esetragot nem viselő főnév következik. A *Marci* elem [Nom-or-Gen] keresőjét szürkével emeltük ki. Ennek értéke a kereső indulásakor [**Nom-or-Gen**] = ?. A kérdéses elem ablakában azonban látható egy olyan elem, amely birtokos ragozást visel magán. Ennek megfelelően a *Marci* token [Nom-or-Gen] keresőjének értéke beállítódik [**Nom-or-Gen**] = **Gen**-re.

Fontos megjegyezni, hogy a [**Nom-or-Gen**] = **Gen** érték nem jelenti azt, hogy a testetlen esetragot magán nem viselő névszó biztosan birtokosa a birtokos ragozású elemnek. Előfordulhat, hogy a birtokos ragozású elem birtokosa a mondat hátralévő részében kapott helyet. A [**Nom-or-Gen**] = **Gen** érték tehát azt jelzi, hogy a testetlen esetrag nélküli elem egy megfelelő jelölt a

birtok birtokosának. Az (5) példában *Marci* egészen addig megfelelő jelölt a *kapitány* birtokosának, amíg nem érkezik egy jobb jelölt, a *kalózhajó*.

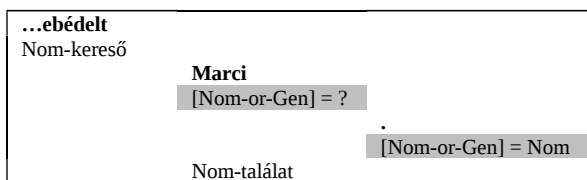
(5) *Marci kapitánya az elátkozott kalózhajónak.*

Az (5) mondat elemzésekor a következő alapelvekre támaszkodunk: 1) a birtoknak mindig van birtokosa, 2) az ígének mindig van alánya, 3) minden esetragos kínálat összekapcsolódik egy kereslettel. Fontos kiemelni, hogy mind a birtokos, mind az alany lehetnek néma elemek is.

A keresőeljárás [**Nom-or-Gen**] = **Nom** értéket vesz fel akkor, ha az elemzési ablakban olyan elemek következnek, amelyek már nem képezik az adott NP részét, például a finit ige (6a), a determináns (6b) és a mondatot vagy tagmondatot lezáró központozás (6c).

- (6) a. *Engem Marci hívott tegnap.*
 b. *Marci a második emeleten lakik.*
 c. *Tegnap otthon ebédelt Marci.*

A 4. ábrán a testetlen esetragos névszó sematikus elemzése figyelhető meg. Az elemzés legelső lépésében az ige vonzatkeretének lehívása történik meg. Az *ebédelt* ige egy nominatívusz esetű vonzatigénnyel él, ezért egy **Nom** beállítású vonzatkeresőt helyez el a tározóban. A következő lépésben a *Marci* testes esetragot nem viselő főnév következik. A *Marci* elem [**Nom-or-Gen**] keresőjét szürkével emeltük ki. Ennek értéke a kereső indulásakor [**Nom-or-Gen**] = ?. A kérdéses elem elemzési ablakában azonban látható egy olyan elem, amely tisztázza annak nominatívusz esetét. Ennek megfelelően a *Marci* token [**Nom-or-Gen**] keresőjének értéke [**Nom-or-Gen**] = **Nom** lesz.



4. ábra: A testes esetrag nélküli elemet tartalmazó mondat elemzése

Az elemző először a tisztázandó esetű névszót közvetlenül követő elemet vizsgálja meg. Amennyiben az nem képes a keresőeljárást [**Nom-or-Gen**] = **Nom** vagy [**Nom-or-Gen**] = **Gen** értékre állítani, akkor veszi a következőt. Ha az elemzési ablak jobb széléig nem talál olyan elemet, amely tisztázná

képes a névszó esetét, akkor [Nom-or-Gen] = ? értékű marad. Ekkor az elemző egy nehezebben elemezhető szerkezettel szembesül, aminek feloldásához – hasonlóan az emberi szövegfeldolgozáshoz – egy másik eljárást kell kidolgozni. Azokban az esetekben van szükség ilyen egyéb eljárásra, amikor a testes esetragot nem viselő elem elemzési ablakába nem fér bele olyan elem, amely tisztázhatná annak esetét. Itt megismételjük a 2.1 fejezetben bemutatott extrém nagy jelöletlen birtokos szerkezetet tartalmazó (4) példát.

- (4) **Krisztina** különleges, Swarovski kristályokból és minőségi japán gyöngyökből készült, egyedi tervezésű romantikus, nőies **nyaklánc**

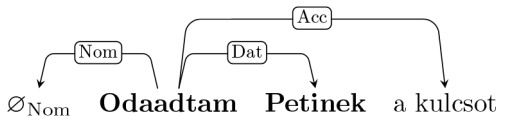
Az ilyen esetekben a [Nom-or-Gen] keresőeljárás értéke [Nom-or-Gen] = ? marad. Ez azt jelenti, hogy a kétfázisú mondatelemzés fázisában nem sikerült tisztázni a testes esetrag nélküli elem mondatbeli szerepét. Ekkor ennek a kérdésnek az eldöntését az elemző egyszerűen „későbbre halasztja”. A mondat végére várhatóan minden kínálat megtalálja a keresőjét, és nem marad találat nélkül egy kötelező kereső sem (amelyet nem tudjuk kiváltani egy zero csomópont bevezetésével), tehát nem marad „bekötetlen” függőségi él.

3.2 Dat: vonzat vagy -nAk ragos birtokos?

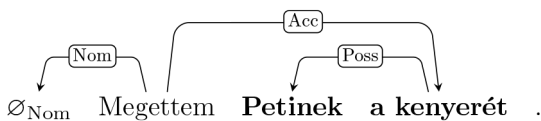
A datívuszi esetragot magukon viselő névszók esetében azt kell megvizsgálunk, hogy egy -nAk ragos birtokos szerkezet birtokosával, vagy egy igei elem részes esetű bővítményével állunk-e szemben. A -nAk ragos elemek minden esetben kínálatként kerülnek be a tározóba. Kínálatként a mondat elemzésének folyamatában egy igei elem vonzatkeresőjét vagy egy birtokos ragozású elem birtokoskeresőjét elégíthetik ki.

A (7) alatti példapár szemlélteti, hogy hol az ige, hol egy birtokos tart igényt a -nAk ragos elemre. A függőségi kapcsolatban álló elemeket vastag betűvel jelöltük. Az 5. és 6. ábrán a két mondat függőségi elemzése látható.

- (7) a. **Odaadtam Petinek a kulcsot.**
 b. **Megettem Petinek a kenyerét.**



5. ábra: A datívusz mint vonzat.

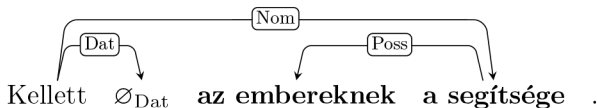


6. ábra A datívusz mint birtokos

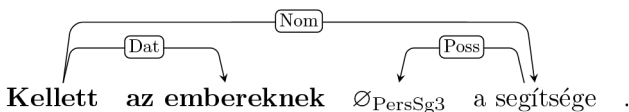
Sajnos ütközhetünk szerkezeti többértelműségbe, amely az emberi mondatmegértés számára is csak a kontextus vagy a világismeret alapján megoldható probléma. A (8a) mondat esetében kontextus nélkül kétféle jelentés is előhívható.

- (8) a. *Kellett az embereknek a segítsége.*
- b. *Kellett \emptyset_{Dat} az embereknek a segítsége.*
- c. ***Kellett az embereknek a segítsége $\emptyset_{PersSg3}$.***

A (8b) jelentés esetén az *az embereknek a segítsége* egy *-nAk* ragos birtokost tartalmazó birtokos szerkezet, a *kell* ige datívuszi vonzata pedig nem hangzik el. A (8c) jelentés esetén az *az embereknek a kell* ige datívuszi vonzata, az a *segítsége* birtokosa pedig egy zéró névmás, amelynek koreferenciáját a kontextusból kellene kiszámítani. A 7. és 8. ábrán a (8a) mondat kétféle jelentésének függőségi ábrája látható.



7. ábra: Néma datívuszi vonzat



8. ábra: Néma birtokos

3.3 A birtokos élek létrejötte

A kereslet-kínálat keretben fontos kérdés, hogy a birtokos szerkezetek esetében „ki keres kit”, és hogy a birtok vagy a birtokos viselkedik-e kínálatként. A birtok magán viseli a birtokos ragozást, amely egyértelműen jelzi, hogy

van birtokosa. Ennek megfelelően a birtokos viselkedik kínálatként, amelyet a birtok által elindított **birtokoskereső** keres.

A [Nom-or-Gen] keresőeljárás találat esetén sem eredményezi függőségi él létrejöttét. A keresőeljárást indító elem esetének tisztázására szolgál, így találat esetén (legyen az [Nom-or-Gen] = Nom vagy [Nom-or-Gen] = Gen) az NP egy CAS jegyet kap. Ez azt jelenti, hogy az NP az eset-egyértelműsítés után kész elfoglalni a mondatbeli helyét.

Ennek megfelelően a birtok (a birtokos ragozást magán viselő elem) két-féle lehetséges birtokost kereshet magának a mondatban: egy **Dat** főkéategóriájú elemet vagy egy [Nom-or-Gen] = Gen jegyű elemet. A birtokoskereső működése a következő:

1. a birtokos számának és személyének kiszámítása a ragozásból
2. [Nom-or-Gen] = Gen jegyű, egyező elem keresése balra a tározóban
HA talált → birtokos él
KÜLÖNBEN
3. **Dat** főkéategóriájú, egyező elem keresése balra a tározóban
HA talált → birtokos él
KÜLÖNBEN
4. **Dat** főkéategóriájú, egyező elem keresése jobbra az ablak jobb szélétől

Az 1. lépésben megtörténik a birtokos számának és személyének kiszámítása a ragozásból. Az elemző létrehoz egy zéró csomópontot a megfelelő számmal és személlyel, amely egészen addig betölti a birtokos funkcióját, ameddig nem érkezik egy megfelelő testes jelölt.

A 2. és 3. lépésben a megfelelő jegyű és főkéategóriájú testes birtokos keresése történik. Amennyiben bármelyik kereső talál, a birtokos és birtoka között létrejön az ideiglenes birtokos él. A mondat hátralévő részének elemzésekor ez az ideiglenes él még felbontható, csak a tagmondathatáron válik véglegessé. Amennyiben balra nem talált megfelelő jelöltet, a tározóba bekerül a **Dat** főkéategóriájú elemet kereső birtokoskereső.⁵

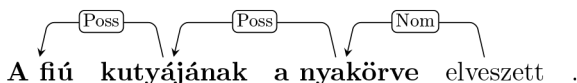
3.4 A többszörös birtokos

Az elemző képes kezelni a többszörös birtokosokat is. Az, hogy a többszörös birtokosok milyen típusú (jelöletlen vagy *-nAk* ragos) birtokos szerkezetek milyen szórendvariánsaiból állhatnak, nem témája a tanulmánynak. Az olyan, különböző szórendvariánsú, különböző egymásba ágyazott birtokos szerkeze-

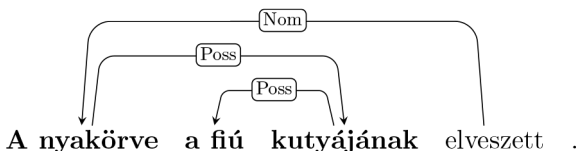
⁵ Már a birtok elemzési ablakában lehet egy **Dat** főkéategóriájú elem, ám mivel a *-nAk* ragos birtokos szerkezet birtokosát és birtokát két külön frázisként kezeljük, ezért ez a birtokos kapcsolat a kétfázisú mondatelemzés második fázisában jön létre.

tek kezelése, mint amilyenek a (9) alatt láthatók, az ANAGRAMMA számára nem jelentenek problémát. A 9. és 10. ábrán a két mondat függőségi elemzése látható.

- (9) a. *A fiú kutyájának a nyakörve elveszett.*
 b. *A nyakörve a fiú kutyájának elveszett.*



9. ábra: Egy beágyazott birtokos szerkezet elemzése



10. ábra: Egy másik beágyazott birtokos szerkezet elemzése

Az elemző ezeknél a mondatoknál is ugyanazokat a lépéseket végzi el, mint a beágyazás nélküli birtokos szerkezetet tartalmazó mondatoknál. Az NP-k előkészítésekor a testes esetrag nélküli elem eset-egyértelműsítése után a birtokok megkeresik a saját birtokosukat, és a mondat végére érve minden függőségi él létrejön.

4. Összegzés

Ismertettük a jelöletlen és a *-nAk* ragos birtokos szerkezetek kezelését az ANAGRAMMA elemzőben. A kétfázisú mondatelemzés első fázisában az előretekintő elemzési ablak segítségével megtörténik a testes esetrag nélküli elemek eset-egyértelműsítése, melynek eredményeképpen kiderül, hogy az adott elem nominatívusz esetű elemként a főige lehetséges alánya, vagy egy jelöletlen birtokos szerkezet lehetséges birtokosa. A kereslet-kínálat keretrendszerben a birtok lesz az, amely birtokos-kereslettel él. Ennek a keresletnek a kielégülésekor birtokos él jön létre a birtok és birtokosa között.

Hivatkozások

- Endrédi István 2016. *Nyelvtechnológiai algoritmusok korpuszok automatikus építéséhez és pontosabb feldolgozásukhoz*. PhD disszertáció, Budapest, PPKE-ITK.
- Frazier, Lyn – Fodor, Janet Dean 1978. The Sausage Machine: A New Two-Stage Parsing Model. *Cognition* 6(4): 291–325.
- Indig Balázs – Vadász Noémi 2016a. Window in Human Parsing – How Far can a Preverb Go? *Tenth International Conference on Natural Language Proceedings (HrTAL2016)*. Dubrovnik, Horvátország, 2016. szeptember, 29-30. Elfogadva, szerkesztés alatt.
- Indig Balázs – Vadász Noémi 2016b. POS Comes with Parsing: A Refined Word Categorisation Method. Poszter. *4th International Conference on Statistical Language and Speech Processing (SLSP2016)*. Pilzen, Cseh Köztársaság, 2016. október 11-12. (<http://grammars.grlmc.com/SLSP2016/Download/slides/pos-comes-with-parsing-abstract.pdf> utolsó letöltés: 2017.02.08.).
- Indig Balázs – Vadász Noémi – Kalivoda Ágnes 2016. Decreasing Entropy: How Wide to Open the Window? In Carlos Martin-Vide – Takaaki Mizuki (eds.) *Theory and Practice of Natural Computing: 5th International Conference, TPNC 2016, Sendai, Japan, December 12-13 2016, Proceedings*. Cham, Springer, 137–148.
- Kiefer Ferenc (szerk.) 1992. *Strukturális magyar nyelvtan I. Mondattan*. Budapest, Akadémiai Kiadó.
- Prószéky Gábor – Indig Balázs 2015. Magyar szövegek pszicholingvisztikai indíttatású elemzése számítógéppel. *Alkalmazott Nyelvtudomány* 15(1-2): 29–44.
- Prószéky Gábor – Indig Balázs – Vadász Noémi 2016. Performanciaalapú elemző magyar szövegek számítógépes megértéséhez. In Kas Bence (szerk.) *"Szavad ne feledd!" : Tanulmányok Bánréti Zoltán tiszteletére*. Budapest, MTA NYTI, 223–232.
- Vadász Noémi – Kalivoda Ágnes – Indig Balázs 2017. Ablak által világosan – Vonzatkeret-egyértelműsítés az igekötők és az infinitívuszi vonzatok segítségével. *XIII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY2017)*. Szeged, 2017. január 26-27. Szeged, Szegedi Tudományegyetem Informatikai Tanszékcsoport, 3–12.