

Vonzatok és szabad határozók szabályalapú kezelése

Gábor Kata¹, Héja Enikő¹

¹ MTA Nyelvtudományi Intézet, Korpusznyelvészeti osztály, Postafiók 701/518,
H-1399 Budapest, Magyarország

{gkata,eheja}@nytud.hu

Kivonat: A cikkben bemutatjuk egy szintaktikai szabályrendszer kidolgozásának módszertanát, melynek segítségével elkülöníthetők a vonzatok és a szabad határozók, valamint megfogalmazhatók a szabad határozók mondatba illesztését végző szabályok. Kitérünk a szabályok elméleti hátterére, implementálására, valamint az eredmények értékelésére.

1 Bevezetés

A cikkben tárgyalt munka célja egy olyan szabályrendszer kialakítása, mely az automatikus szintaktikai elemzés során az esetragos főnévi csoportokat funkciójuknak megfelelő annotációval látja el. A szabályrendszer az Intex/NooJ szövegfeldolgozó eszköz [7] magyar moduljába [8] épül. Az Intex, illetve legújabb változata, a NooJ magyar szintaktikai modulja részleges elemzést végez: felismeri a tagmondatok legfelsőbb szintű összetevőit, és meghatározza a köztük lévő szintaktikai függőségi viszonyokat. Ennek megvalósításához szükség van arra, hogy meg tudjuk különböztetni az igei vonzatkeret részeként előforduló főnévi csoportokat a szabad határozói (adjunktum) funkciót betöltő NPktől. Ebben a cikkben kizárólag az ige valamilyen típusú bővítményeként¹ előforduló legfelsőbb szintű NPK funkciójának meghatározásával foglalkozunk.

Ellentétben azzal az elterjedt vélekedéssel, mely az argumentumszerkezetet és az ige bővíthetőségét annak lexikai tulajdonságaként kezeli, mi a szintaktikai relációkat megjelenítő esetragokat választottuk kiindulópontul. Az esetragoknak saját funkciót tulajdonítunk, és az esetragos NPK szintaktikai szerepét az esetrag funkcióiból kívánjuk levezetni. A megközelítésünk azon a gondolaton alapul, hogy a vonzatszerep és a szabad határozói szerep közti különbség valójában a bővítményt tartalmazó szerkezet kompozicionalitásának és produktivitásának mértékében rejlik. Míg a vonzatok előfordulásai azokra a tagmondatokra korlátozódnak, melyek a vonzatot előíró igét tartalmazzák, az adjunktumok ugyanazt a szerepet igék egy tágabb csoportja, egyes esetekben valamennyi ige mellett betölthetik. Eszerint a gondolatmenet szerint a vonzatság egy az esetragok lehetséges funkciói közül, mely abból a szempontból speciális, hogy csak nagyon megszabott környezetben tölthetik be. Az esetragok funkcióit szabályokkal írjuk le, melyek a főnévi csoporthoz egy nyelvtani szerepet rendelnek. Az

¹A “bővítmény” szót összefoglaló névként használjuk a vonzatokra és a szabad határozókra.

esetragos főnevek azon előfordulásai, melyek leírhatók anélkül, hogy a szabály bemenetében hivatkoznánk a főnévi csoport tagmondatában előforduló ige lemmájára, adjunktum szerepűnek tekinthetők.

A szabályok által kiosztott címkék szemantikai tartalommal rendelkeznek, de magukat a szabályokat szintaktikai műveletként fogjuk fel, melyek lehetővé teszik az adott NP használatát az aktuális kontextusban.

Cikkünkben ismertetjük a módszert, melynek segítségével elkülönítjük az esetragok funkcióit, valamint leírjuk és az automatikus szintaktikai elemzésben megvalósítjuk az adjunktív szabályok rendszerét. Munkánk eredménye egy kritériumrendszer, melynek segítségével a vonzatok elkülöníthetőek a produktívan használt, kompozicionális szerkezeteket alkotó adjunktumoktól. A szabályrendszerünk a lexikon méretét is csökkentti, mivel a többszintű szabályaink az eddiginél több esetrag-jelentést tudnak leírni. Ez azért is hasznos, mert egy lépésben a szintaktikai függőségek mellett hasznos szemantikai információt is kódolnak, így a szöveg „megértéséhez” is hozzájárulnak. A szabályainkkal annotált szöveg XML formátumú, így indexálható és lekérdezhető például a Xaira korpuszlekérdező eszközzel. Ez az eszköz lehetővé teszi, hogy az NPk bármilyen attribútumának értékei szerint keressünk, például kigyűjthetjük az annotált korpuszból az *okot* kifejező főnévi csoportokat.

A következőkben bemutatjuk a magyar szintaxis néhány fontos jellemzőjét, valamint leírjuk, miért nehéz a magyar nyelvre alkalmazható vonzattesztet találni [2], majd bemutatjuk saját módszerünket [3]. Ezután egy konkrét példát ismertetünk [4], végül bemutatjuk a kiértékelés eredményét [5].

2 Vonzatok és szabad határozók a magyar mondatban

A magyar mondatban az összetevők felszíni szórendje nem tükrözi szintaktikai szerepüket. A komplementumok és adjunktumok csaknem bármilyen felszíni sorrendje elfogadható, bár egyes szerkezeti pozíciók különböző diskurzus-funkcióknak feleltethetők meg [1]. Fókuszálás vagy topikalizáció által bármilyen funkciójú összetevő az ige elé mozgatható. Ezen kívül az igemódosítók (névelőtlen NPk, igekötők, adverbiumok) is megelőzhetik az igét. Magyar szövegek automatikus elemzésekor azzal kell tehát szembesülnünk, hogy a felszíni sorrendet legfeljebb a diskurzus-funkciók feltérképezésére használhatjuk, a szintaktikai függőségi viszonyokat azonban nem határozhatjuk meg a sorrend alapján. Mivel a felszíni sorrendet nem használhatjuk, egyéb konfigurációs információ pedig nem áll rendelkezésünkre, nem használhatjuk a transzformációs, konfigurációs nyelvtanok tesztjeit a vonzatok és adjunktumok elkülönítésére. Ezt támasztja alá az is, hogy Radford [6] (angol nyelvre vonatkozó) tesztjei közül egyik sem használható a magyarra:

a) Passzíválás: A vonzatszerépű PP-ből kiemelt NP passzíválható, az adjunktum PP főnévi csoportja nem:

[This job] needs to be worked at by an expert.

*[This office] is worked at by a lot of people.

b) Pronominalizálás: A 'do so' szerkezet, ami a V' kategóriát helyettesíti, tartalmazhat újabb V'-t képző adjunktumot (i), de el is hagyható (ii), míg a komplementum kötelezően benne foglaltatik a V'-ban (iii), nem hagyható el (iv).

- i) John will [buy the book on Tuesday] and Paul will *do so* as well.
- ii) John will [buy the book] on Tuesday and Paul will *do so* on Thursday.
- iii) John will [put the book on the table] and Paul will *do so* as well.
- iv) *John will [put the book] on the table and Paul will *do so* on the chair.

c) Felsőszíni sorrend: A vonzatok közelebb vannak az igehez, mint az adjunktumok, mert a szintaktikai fában előbb csatlakoznak hozzá, és a fa élei nem keresztezhetik egymást.

d) Ellipsis: Bármilyen frazális kategória elliptálható. V' kategóriájú összetevő akkor elliptálható, ha tartalmazza az igei fejet vonzataival és adjunktumaival (i), a fejet vonzataival, adjunktumok nélkül (ii), de a fej egyik vonzatával, de a másik nélkül nem alkot elliptálható összetevőt (iii):

- i) – Who might be going to the cinema on Tuesday?
– *John might be _____* .
- ii) – Who might be going to the cinema when?
– *John might be _____ on Tuesday.*
- iii) – Who will put the book where?
– **John will _____ on the table.*

Az a) és b) tesztek azért nem alkalmazhatók, mert a magyarban nincs hasonló passzivalás, illetve pronominalizálás. A c) teszt nem teljesül a magyar mondatok egy részében:

*A gyerekek nyírják a kertben a fűvet.
Bea tegnap megtalálta a kutyádat.*

A fenti mondatokban a szabad határozók az ige és vonzata (tárgya, illetve alanya) közé ékelődnek, tehát az ige nem közvetlenül szomszédos a vonzataival.

Ha feltételezzük, hogy a magyar 'tesz' ige az angolhoz hasonlóan három vonzattal (alany, tárgy, lokatívusz) rendelkezik, megmutathatjuk, hogy a d) teszt sem használható a magyarra:

- *Ki megy hová kedden?*
- – *János __ moziba __.*

- *Ki tette a könyvet hová?*
- – *János ___ az asztalra.*

A második mondatpár, melyben a *tesz* ige két vonzata szerepel, és a harmadik az igével együtt elliptálódik, megmutatja, hogy a vonzatok nem csak közösen hagyhatók el. Azaz egyik teszt sem alkalmazható a magyar nyelvre.

Komlósy [3] a magyar igei argumentumszerkezetről írott tanulmányában azt állítja, hogy a vonzatok és szabad határozók elkülönítése olyan művelet, melynek elvégzéséhez a nyelv egész nyelvtanának ismeretére szükség van. A vonzatot olyan összetevőként határozza meg, melynek szintaktikai és szemantikai tulajdonságait az őt kormányzó ige írja elő. Három tesztet javasol a szerkezetek elkülönítésére, melyek, amint a szerző is vállalja, nem elégségesek az összes szerkezet meghatározásához:

- e) ha egy bővítmény kötelező, akkor vonzat;
- f) ha egy opcionális bővítmény kitétele lehetővé teszi a szerkezet kibővítését egy másik bővítménnyel, mely mellől az első bővítmény már nem hagyható el, akkor az első bővítmény vonzat;
- g) ha X szónak bővítménye Y, és van olyan Z szó, ami szisztematikusan helyettesítheti X+Y szerkezetet, valamint helyettesítheti X -et, amikor Y nincs jelen, de nem helyettesítheti X -et, ha Y jelen van, akkor Y X opcionális vonzata.

A kötelezősége hivatkozó e) teszt használatát mindenképp mellőzni szeretnénk, mivel ellipsis vagy egyéb műveletek által gyakorlatilag bármit elhagyhatunk a mondatból, és nehézségekbe ütközhet annak eldöntése, hogy egy adott mondat tartalmaz-e kötelező, de elhagyott összetevőt. Az f) és g) tesztek megbízhatóságát nem vitatjuk, ám használhatóságuk korlátozott, így mindenképp szükségünk van más kritériumokra is.

Bár a GB vonzatesztjeit elvetettük, az összetevők szintaktikai szerepeinek azonosítására használt koordinációs teszt a mi munkánkban is fontos szerepet játszik. Ha feltételezzük, hogy csak azonos szerepű összetevők koordinálhatók, a vonzat – adjunktum ellentét segítségével kell számot adnunk az alábbi mondatról:

**János beszennyezte a szőnyeget sárral és a cipőjével.*

A *sár* és a *cipő* tehát különböző funkciót lát el ebben a mondatban, bár szemantikailag mindkettő a beszennyezés eszközének tekinthető.

Azt feltételezzük, hogy a fentihez hasonló mondatokban más szabály kapcsolja az igehez a nem koordinálható, azonos esetragot viselő NPket, és ezek a szabályok az eltérő szintaktikai szerep mellett különböző szemantikai címkét is társítanak a főnévi csoportokhoz. Az alábbi mondatban:

Párizsban még bíztam az apámban.

a két, azonos esetragú NP koordináció nélkül szerepel együtt, és nem is koordinálhatók. Ennek oka, hogy az egyik NP ('*az apámban*') vonzat, míg a másik szabad határozó. A mondat szerkezetileg kétértelmű, de teljesen kizárja az olyan értelmezéseket, melyben a két NP azonos szerepet tölthetne be. Egy esetrag ugyanis csak egyszer jelenhet egy funkciót egy tagmondatban. Kérdés azonban, hogy a vonzat-adjunktum

szembeállításal hogyan magyarázzuk meg az alábbi mondat helyességét?

2005-ben Párizsban még bíztam az apámban.

Ebben a mondatban két adjunktumszerepű NP-t találunk, melyek koordináció nélkül is jólformált szerkezetet alkotnak. Ezt a jelenséget úgy próbáljuk magyarázni, hogy a vonzat – szabad határozó szembeállítás helyett egy többfokozatú skálát alkalmazunk, melyben minden adjunkciós szabály különböző funkcióért felelős.

3 Kompozicionalitás és produktivitás

A NooJ magyar moduljában véghezvitt szintaktikai elemzés célja, hogy a szöveg valamennyi legfelsőbb szintű esetragos főnévi csoportját nyelvtani szerepe szerint annotálja. Ehhez a tagmondatok konfigurációs szerkezete helyett az esetragok szintaktikai szerepjelölő funkcióját kívánjuk használni. A dependencia-nyelvtan terminológiájával élve az esetragos NP szerepét a predikátumhoz való viszonyában határozzuk meg. Mindazonáltal a szerepek leírásakor kerülni akarjuk a predikátum tulajdonságaira való hivatkozást, és minél több esetrag-funkciót szeretnénk általános szabályokkal megragadni. A predikátum-vonzat relációt is az esetragok egyik funkciójának tekintjük.

Esetragnak azt a toldalékot tekintjük, ami a magyar főnév jobb szélén jelenik meg, másik toldalék nem követheti, és egy főnévnek csak egy esetragja lehet. Ezek alapján a magyarban 19 esetragot sorolhatunk fel. Feladatunk, hogy valamennyi esetrag lehetséges funkcióit szabályokkal leírjuk, szabályokkal nem kezelhető szerkezeteket pedig vonzatként felsoroljuk.

Elsőként megállapíthatjuk, hogy az alany- és a tárgyeset nem rendelkezik default jelentéssel: minden előfordulásukban az igei argumentumszerkezetet részét képezik². A többi esetragról azt feltételezzük, hogy rendelkeznek saját szintaktikai és szemantikai tulajdonságokkal, melyek szabályokkal leírhatók. Ezeket az általános szabályokat, melyek az esetragok alapértelmezett funkcióját/funkcióit definiálják, default szabályoknak nevezzük. A default szabályok bemenete utalhat az őt tartalmazó főnévi csoport fejének szemantikai vagy morfoszintaktikai tulajdonságaira, de soha nem utalhat annak az igének a lemmájára, amelyik az esetragos főnévi csoport tagmondatának állítmánya. Ennek értelmében egy esetragnak egynél több default funkciója is lehet, bár a funkciókat leíró szabályok közül szigorúan véve csak egy szabály valódi „default”, ami az esetrag összes olyan előfordulását lefedi, melyre a többi szabály nem illeszkedik. A szabályok kimenete a főnévi csoport szerepét leíró címke. Annyiféle nem vonzat szerepet különböztetünk meg esetragonként, ahány szabályt használunk az egyes esetrag funkcióinak leírásához (különböző esetragok funkciói viszont egybeeshetnek). Mivel a szerepeket az NPhez társító szabályokat szintaktikai (adjunkciós) szabálynak fogjuk fel, a szabályok kimenetében megjelenő szerep-címkék is szintaktikainak tekinthetők. Itt azonban fontos megjegyezni, hogy a szerepek erős szemantikai

²Természetesen ez nem vonatkozik a névutós frázisokban, illetve a főnevek vagy melléknevek vonzataként előforduló, nem legfelsőbb szintű NPkre.

tartalommal bírnak, valamint a szabályok jellegéből is kiderül, hogy egyes adjunkciós műveletek szemantikailag megszorított bemeneten működnek. Ezek alapján úgy tekintettük, hogy a szintaxist és a szemantikát nem kezelhetjük külön modulban.

Például a *-ban* esetrag alapértelmezett jelentése attól függ, hogy milyen szemantikai jegyekkel rendelkező NPn jelenik meg: az időt kifejező főnévi csoportnak időhatározói szerepet ad ('*januárban találkozunk*'), míg egyéb esetekben szabályos helyhatározói funkciót ad az NPnek ('*a hordóban találtam*'). A szabályok, melyek az alapértelmezett szerepeket osztják ki, természetesen a vonzathatározók megállapítása után futnak le, mivel bemenetük kevésbé specifikus: a kontextustól függetlenül működnek. Azok az [ige + NP + esetrag] szerkezetek, melyek nem írhatók le általános szabályokkal, [ige + vonzat] szerkezetként elemzendők. Azért nem rendelhető hozzájuk default szabály, mert ezek a szerkezetek nem kompozicionálisak: az NP igéhez képesti szerepét nem lehet olyan szemantikai címkével ellátni, mely nem utal az ige jelentésére. Például:

A közönség elhalmozta az előadót kérdésekkel.

Ha a fenti mondatban az [ige + NP + -val] szerkezet kompozicionális lenne, az NPhez tudnánk olyan absztrakt címkét társítani (pl. hely, idő, mód stb.), ami leírja az igehez való viszonyát anélkül, hogy az ige jelentésére bármilyen módon hivatkozna. Hogy ez nem lehetséges, az abból is látszik, hogy természetes nyelven sem találunk hozzá olyan parafrázist, amely kifejezi az ige és az NP viszonyát, de nem tartalmazza sem az igét, sem annak szinonimáját.

Ezek alapján a 4) mondat az ige + vonzat szerkezet példájának tekinthető – tehát az *elhalmoz* ige lexikai tételébe fel kell vennünk.

Vannak azonban olyan esetragos szerkezetek is, melyek köztes kategóriát képviselnek az adjunkció teljes produktivitása és a vonzatság teljes lexikalitása között. Az esetragok ezen használatai csak egyes *szemantikai* igeosztályok mellett mondhatók produktívnak. Például a *-tól* esetragnak ilyen módon megkülönböztethetjük két funkcióját: a mozgást jelentő igék mellett a mozgás kiindulópontját jelentő NPn jelenik meg, míg állapotváltozást jelentő igék mellett az állapotváltozás közvetlen okát jelentő NPt azonosítja. Az esetragnak ezt a két funkcióját két szabállyal tudjuk leírni, melyek közül mindkettő utal a tagmondat állítmányának szemantikai osztályára. Azt állítottuk, hogy az adjunkció teljesen produktív művelet, mely a tagmondat állítmányától függetlenül alkalmazható (az egyetlen követelmény, hogy a tagmondatnak *legyen* állítmánya), míg a vonzatság az egyedi igei lemmák lexikális tulajdonságától függ. A szemantikai igeosztályokon működő műveleteket leíró *nem-default* szabályok kevésbé produktívak, mint a default-szabályok, így besorolásuk nem egyértelmű. Mindazonáltal érdekünkben áll, hogy ne tekintsük vonzatnak a nem-default szabályok által létrehozott szerkezeteket, mert így a bennük szereplő NPk szerepéről több információt tudunk adni, mintha csak vonzat-státuszukra hivatkoznánk. Emellett egy szintaktikai teszt is alátámasztja, hogy a főnévi csoportok funkciói szélesebb körűek a vonzat-adjunktum kettősnél. Ha feltesszük, hogy egy jólformált magyar tagmondat nem tartalmazhat kettő vagy több olyan NPt, melyek ugyanazt az esetragot viselik, ugyanazt a szerepet töltik be, és nincsenek koordinálva, akkor problémát okoz a 3) mondat helyessége, amelyben két *-ban* esetragos adjunktum szerepű NP van (az ugyanolyan esetragos vonzat mellett). Mivel mi azt feltételezzük, hogy az esetragos NP annyiféle szerepet tölthet be, ahányféle szabály alkalmazható az esetragra, vagyis minden sza-

bály kimenete különböző címkével látja el az NPt, egyszerűen megfogalmazhatjuk a jólformáltsági feltételt: minden szabályunknak csak egy találat lehet tagmondatonként (a találat azonban koordinált NPt is tartalmazhat), így minden NP különböző szerepet kap.

4 A szemantikai igeosztályok meghatározása

Miután végigvettük azokat az általános vezérelveket, amelyek kutatásunk alapjául szolgálnak, egy konkrét példa részletesebb tárgyalásával folytatjuk. Az alábbiakban a *–val* esetrag előfordulásait vizsgáljuk meg. Azt feltételeztük, hogy a szóban forgó esetraghoz két default szabály tartozik, vagyis két olyan szabály, amely anélkül határozza meg a megfelelő főnévi csoportok mondatban betöltött szemantikai szerepét, hogy bármilyen formában is hivatkozna a predikátumra. (Mint később látni fogjuk ez a feltételezésünk nem igazolódott.)

Ezek közül az egyik a default társhatározói szabály, amely ASSOCIATE nevű címkével látja el a releváns főnévi csoportokat. A szóban forgó főnévi csoportok közös jellemzője, hogy a predikátum által jelölt eseményben betöltött szemantikai szerepük az alany szerepével egyezik meg (*'ül'*). A szabály akkor alkalmazódik, ha a megfelelő főnévi csoport rendelkezik a +HUMAN szemantikai jeggyel.

János Marival ül a kertben.

A másik szabályunk a default eszköz szabály. Ez azokat az NPket jelöli meg, amelyek az ige által jelölt eseményt végrehajtására szolgáló eszközre referálnak..

János kocsijával hazavitt mindenkit.

Fontos kiemelni, hogy a fenti szabályok esetében nem hivatkoztunk a predikátumokra, legfeljebb a kérdéses NPK szemantikai vagy szintaktikai jegyeire. Ez összhangban van azzal a hagyományos nézettel, hogy az adjunktumok szinte bármilyen ige mellett megjelenhetnek, vagyis az adjunktumok jelentése független az igétől.

A default-szabályok alkalmazásának megvan az az előnye, hogy ezáltal a szövegben szereplő minden megfelelő esetraggal rendelkező főnévi csoporthoz rendelünk szemantikai szerepet, így a lefedettség az ige felismerésétől függetlenül 100% lesz.

A tesztelés során azonban kiderült, hogy kezdeti feltevésünk nem volt helyes, amennyiben nemcsak default eszköz és default társhatározó szabályokat kell létrehoznunk. Ennek az az oka, hogy a *–val* esetragos főnevek egy további meglehetősen produktív használata a határozói használat. Az alábbiakban erre láthatunk példát:

Mari csökönyös és áhítatos erőszakkal ragaszkodik Bélához.

Bizonyos esetekben azonban meglehetősen problémás az ige által jelölt cselekvés módjára vonatkozó adverbialis és az eszközhatározói szerepet betöltő főnévi csoportok elkülönítése.

Szemem sarkával Liliant kerestem.

A gyermekem már késsel és villával eszik.

A probléma megoldására első lépésként létrehoztunk egy szabályt, amely azon az előfeltevésen alapul, hogy a melléknevekből és igékből képzett főnevek képesek betölteni az igemódosító pozíciót és gyakrabban is kerülnek ebbe a pozícióba, mint főnévibe. Így a harmadik default szabályunk bemenetét az –Ás illetve –sÁg végű főnevek alkotják. További szabályokat is létrehoztunk az adverbialis és eszköz típusú NP-k elkülönítésére. A szabályok mögött az a megfigyelés húzódott meg, hogy a cselekvés módjának és eszközének szétválasztása akkor igazán problematikus, amikor az ige nem egy konkrét eseményre, hanem egy esemény típusra referál. Ezt illusztrálja a fenti mondatpár második mondata. Ilyenkor a mondatban megnevezett eszköz nem egy konkrét eszköz lesz: inkább azt a módot jelöli, ahogyan az eseményt végre szoktuk hajtani, vagy ahogyan az esemény általában végbemegy. Ezért a fent említett default eszköz szabály esetében figyelembe vettük, hogy az NP rendelkezik-e névelővel. A névelős, vagyis határozott főnévi csoportokat eszközként jelölték meg szabályaink, míg a névelőtleneket módként.

A –val esetragos főnévi csoportok vizsgálata során még egy default szabályt alkalmaztunk, amely a bemeneti szöveget a MEASURE címkével láthatja el. Ez a szabály szintén támaszkodik szemantikai jegyekre, a szóban forgó főnévi csoportok a szabály alkalmazásakor már rendelkeznek a TIME és MEASURE szemantikai jegyekkel. Az ilyen jegyű –val esetragos NPk feltételezésünk szerint az ige által jelölt esemény (változás) mértékét vagy két esemény között eltelt időt fejezik ki. '[MEASURE Húsz évvel] ezelőtt' vagy '[MEASURE Három százalékkal] nőtt.'

A default-szabályok kialakítása során nem hivatkozhatunk az igéhez kapcsolódó szisztematikus morfológiai és szintaktikai változásokra, mivel ezzel ellentmondásba kerülnénk a default szabályok definíciójával, amely semmilyen formában nem engedi meg az igére való hivatkozást.

Az alábbiakban a nem default szabályokat tekintjük át. Ebbe a szabályosztályba is felvettünk egy INSTRUMENTUM szabályt. Emlékeztetőül: ezek a szabályok azért nem default szabályok, mert a főnévi csoport egyes tulajdonságain kívül az igére is hivatkoznak. Ilyen predikátum például a már fent is említett 'beszennyez', ahol az igének van egy eszköz típusú argumentumhelye. A koordinációs teszt segítségével megmutattuk, hogy a 'beszennyez sárral' és a 'beszennyez a cipővel' két különböző argumentum. Nem-default eszköz szabályra példa az alábbi, teszt-szövegből származó mondat is:

1) „...azok előtt, akik kezüket [INS saját kommunista elvtársaik vérével] mocskolták be.”

Kérdésként merülhet fel azonban, hogy miért van szükség a default és nem-default eszköz szabályok megkülönböztetésére. Egyfelől láttuk, hogy szintaktikailag

motivált az elkülönítésük. Másfelől egy nem-default szabály illeszkedése egy sztringre megakadályozza a default szabályok alkalmazását. Ez nyilvánvalóan akkor fontos, ha olyan default szabályok illeszkednének rá, amelyek más szemantikai szerepet tulajdonítanak az NPnek. Esetünkben pontosan ez lenne a helyzet, hiszen az ilyen NPK a fent leírt MODE szabály bemenetét képeznék.

Áttérve a nem-default társhatározói szabályra, szintén meg kell válaszolnunk a fenti kérdést. Míg a default szabály hivatkozott a releváns NP +HUMAN szemantikai jegyére, a default szabály nem használja fel ezt az információt. Ez a megkülönböztetés azt a tényt tükrözi, hogy létezik egy olyan igeosztály, amely esetében a *-val* esetragos főnév mindig társként viselkedik, vagyis szemantikai szerepe mindig megegyezik a mondat alanyának szemantikai szerepével. Ezt mutatja az alábbi mondat is.

2) *János veszekszik az autóval.*

A fenti mondat – 3)-mal szemben – csak úgy értelmezhető, hogy az *autó* is részese –és nem eszköze – volt a veszekedési eseménynek., annak ellenére, hogy az *autó* nem rendelkezik a +HUMAN jeggyel.

3) *János Marival ment moziba.*

A fenti mondatban *Mari* csak abban az esetben jelölhet társat, ha ember. Egyébként eszköz lenne.

A következő szabály az állapotváltozást kiváltó közvetlen okokat jelölő főnévi csoportokat látja el címkével. Akárcsak a többi nem-default szabály, ez is hivatkozik arra, hogy a főneves kifejezés környezetében található ige melyik igeosztályba tartozik. Idetartoznak például a '*megdöbbsent*', '*felidegesít*', '*megrémít*' igeik.

4.a) *János megdöbbsentette Marit a hírrel.*

A fenti mondatban szereplő ige nézőpontunkból lényeges szemantikai tulajdonságait az alábbi kifejezés szemlélteti:

5) CAUSE(János, E), ahol E<hír, CHANGE(S, S')> és CAUSE(hír, S')

Eszerint János létrehozott (CAUSE) egy szituációt (E), ahol a szituációt egy olyan kétargumentumú predikátummal írhatjuk le, amelynek az első argumentuma (*hír*) állapotváltozást okoz (CAUSE) *Mari* mentális állapotában, vagyis átmenetet idéz elő S-ből S'-be. A következő felmerülő kérdés, hogy hogyan igazolhatnánk szintaktikailag a három metapredikátum jogosságát (i.e. CAUSE, MENTAL, CHANGE)?

Az alábbi tesztet használtuk annak eldöntésére, hogy egy adott ige tagja-e ennek az osztálynak:

4.b) *A hír megdöbbsentette Marit.*

4.c) *Mari megdöbbsent a hírtől.*

Feltételeztük, hogy egy ige akkor és csak akkor tartozik ebbe az osztályba, ha a 4.a.), 4.b) és 4.c) példamondatok szerkezetével egyaránt jól formált mondatot alkot. 4.a) és 4.b) alapján azt állíthatjuk, hogy az ilyen típusú igéknek rendelkezniük kell legalább egy olyan olvasattal, ahol az alany nem ágens. Ha ez nem teljesülne 4.b) agrammatikus lenne, hiszen az alany jelölete ebben az esetben nem képes egy cselekedet szándékos végrehajtására. Ebből következik, hogy a csoportba tartozó legtöbb ige – bár korántsem az összes – mentális állapotváltozásra vonatkozik. Azt látjuk tehát, hogy a MENTÁLIS metapredikátumunk ebben az esetben ekvivalens azzal a követelménnyel, hogy az alany legyen legalább egy nem ágenses olvasata az ige mellett. 4.c) illusztrálja a CAUSE és CHANGE metapredikátumok szükségességét. Elfogadtuk Komlósy [4] azon nézetét, mely szerint bizonyos igék Okozó szerepű argumentumai megjelenhetnek *-tól* esetragos főnévként. Ezenfelül azok az igék, amelyek mindhárom szerkezetben megjelenhetnek, feltételeznek két állapot közötti átmenetet is, ahol a hangsúly nem magán az átmeneten van, hanem a második állapot elérésén. Ez a feltételezés párhuzamba állítható azzal a jelenséggel, hogy a szóban forgó igeosztály elemeit általában perfektív igealakokkal fordítjuk angolra. Ennek a jelenségnek az lehet az oka, hogy míg a perfektív igealakok az ige által jelölt esemény bekövetkezése utáni állapotot hangsúlyozzák, az imperfektív igealakok magát a folyamatot. Egy másik érv a CHANGE metapredikátum szükségessége mellett pedig azon alapul, hogy vannak olyan igék, amelyek mellett ugyan megjelenhet *-val* esetragos Okozó szerepű főnévi csoport, így ezek az igék az Okozó metapredikátum alá tartoznak, de nincsen két meghatározott állapot közötti átmenet, így a CHANGE metapredikátum nem alkalmazható. Vegyük példaképp az alábbi mondatokat:

6.a) *Az igazgató Jánost terhelte a feladattal.*

6.b) *A feladat Jánost terhelte.*

6.c) *János terhelve van.*

6.d) **János terhelve van a feladattól.*

A CHANGE metapredikátum szükségességére vonatkozó szemantikai intuíciónkat explicit módon támasztja alá 6.d) helytelensége. Ha a CHANGE metapredikátum által jelölt jelentéskomponens is jelen van az igében, mind a három szerkezet jól formált. A fenti példában szereplő igéből azonban csak ez hiányzik. Ez azt támasztja alá, hogy a CHANGE metapredikátum is disztingtív és éppen ezért a CAUSE-től függetlenül fel kell vennünk, ha meg szeretnénk adni az adott igeosztályba való tartozás szükséges feltételeit.

Egy másik igeosztályt alkotnak a faktitív igék. Az erre az igeosztályra hivatkozó szabályok egy AGENS2 nevű szemantikai címkét rendelnek hozzá a megfelelő főnévi

csoportokhoz. Azért így neveztük el ezt a csoportot, mert a faktitív műveltetés alapigéje mindig ágenses, így a *-val* esetragos főnévi csoport az alapige ágensét fogja jelölni. Mivel a kauzatív igék is képezhetők a *-(t)At* műveltető képzővel és ezek mellett szintén megjelenhetnek *-val* esetragos főnevek (amelyek azonban ebben az esetben nem lehetnek ágensek), hivatkoznunk kell a releváns főnévi csoport szemantikai jegyeire is, azaz ki kell kötnünk, hogy a szabály csak akkor alkalmazódjon, ha az rendelkezik a +HUMAN jeggyel.

7) *János levágatja a haját a fodrásszal.*

Az előbbi kvázi formalizmust használva a fenti példamondatot az alábbiak szerint írhatjuk le:

8) CAUSE(János, E), ahol E<fodrász, haj, ...> és AGENS2(fodrász, E)

Vagyis János létrehoz egy eseményt (E), amelynek legalább két szereplője van – hiszen csak a tranzitív igékből képzett műveltető igék mellett jelenhet meg az eredeti alany *-val* esetragos főnévként – és a fodrász az ágense annak az igének, amely E-t leírja. Következésképpen ezekben az esetekben a predikátum mellett megjelenő *-val* esetragos főnév az alapige ágense.

A fentiekben megmutattuk, hogy a default és nem default szabályok megkülönböztetése empirikus és elméleti nézőpontból egyaránt védhető. Munkánk jelenlegi állapotában ezekkel az igeosztályokkal rendelkezünk. A többi igét, amely mellett megjelenhet *-val* esetragos főnévi csoport, vonzatként kezeljük. Értelemszerűen az ilyen környezetben megjelenő *-val* esetragos főnévi csoportokat nem tudjuk szemantikai címkével ellátni. A megfelelő igéknek ezt a tulajdonságát kódolni kell az igei szótárban.

5 Implementáció

A munkafolyamat első lépése az igei szókincs kiválasztása volt. A Magyar Nemzeti Szövegtár (MNSz) [9] 2,800 leggyakoribb igéjét választottuk. Ezen az igeosztályon definiáltuk az esetragok default jelentéseit, a szemantikai igeosztályokat, valamint a vonzatkeret kódolását. Az esetragokat aszerint vizsgáltuk, hogy milyen gyakran fordulnak elő ezekkel az igékkel. Négy gyakori esetragot tanulmányoztunk részletesen: *-val* (instrumentális), *-nak* (datívusz), *-tól* (ablatívusz) és *-ra* (szublatívusz). Először az esetragok default jelentéseit határoztuk meg, mivel a nem-default szabályok kidolgozása előfeltételezi a default jelentések ismeretét. A nem-default szabályok kidolgozása úgy zajlott, hogy csoportosítottuk a (nem default jelentésében szereplő) esetraggal előforduló igéket aszerint, hogy az esetragot viselő NP milyen szerepet tölt be mellettük. Ezek a csoportok megadták a szemantikai osztályokat, és várakozásunknak megfelelően nemcsak egy-egy esetrag jelentéseinek elkülönítésében játszottak szerepet. Utolsó lépésként azokat az igéket, amik mellett az esetragos NP szerepét egyik szabály sem fedí le, megjelöltük, mint az adott esetragot vonzatként előíró igét.

A szabályrendszer a NooJ magyar szintaktikai moduljában végrehajtott részleges elemzésre épül. Az elemzés bemenete nem egyértelműsített, text formátumú magyar szöveg, sebessége 240K/perc. A kimenet szintaktikai jellegű annotációt tartalmaz. Az általunk használt morfológiai elemzés szóalakokat felsoroló szótárakra épül: a szótárak az MNSz 900,000 leggyakoribb szóalakját fedik le. A szótári bejegyzés a szóalakot, a hozzá tartozó lemmát és morfológiai kódot, valamint a lemma esetleges szemantikai jegyeit tartalmazza. A morfológiai kód a HUMOR [5] elemzésén alapul.

A nyers szöveg nyelvtani elemzése magában foglalja a tokenizálást, a mondatsegmentálást, lexikai és morfológiai elemzést, többszavas kifejezések és tulajdonnevek felismerését (a tokenizálás részeként), valamint a részleges szintaktikai elemzést. Egyértelműsítést egyáltalán nem használunk, mivel azt a nyelvtani szabályok, elsősorban az NP-nyelvtan nagyrészt elvégzi³. A szintaktikai elemzést egymásra épülő, többlépcsős nyelvtanok végzik, melyek a korábbi nyelvtanok kimenetére hivatkoznak. A tagmondat központi elemének a finit igét tartjuk, melyről azt feltételezzük, hogy a legfelsőbb szintű frázisokkal, illetve ezek fejével lép dependencia-viszonyba, vagyis az elemzés alapja a frázisok megtalálása. A frázisok felismerése után a tagmondat állítmányának megtalálása következik. Ezután a hétlépcsős tagmondathatárnyelvtanunk [2] bejelöli a határokat, amin belül az egyes igék vonzatait és szabad határozóit kereshetjük. Ettől kezdve a szintaktikai elemzés minden lépését a tagmondathatáron belül hajtjuk végre.

A szabályrendszer implementálásához szükséges további előfeldolgozás során a főnévi csoportokat és a finit igéket annotáljuk a szabályokban használt releváns tulajdonságaik szerint. Ehhez a NooJ-ban használt szótárakat kibővítettük a főnevek általunk használt szemantikai jegyeivel (pl. *time*, *human*, *measure*). A fej szótári jegyei alapján

³Az NP-nyelvtan fejlesztésekor [10] mért adatok alapján a magyar szövegek tokenjeinek kb. 69%-a valamelyik legfelsőbb szintű NP-be tartozik.

a felismert NPK annotációját kibővítettük ezekkel a jegyekkel, így a szintaktikai modul kimenet már tartalmazza őket. Az igei predikátumok pedig szintén szótári jegyként kapták meg a szemantikai csoportokat azonosító jegyeket.

A kidolgozott szabályrendszerek közül a -val rag szabályait implementáltuk és teszteltük. A szabályok a szöveg valamennyi főnévi csoportját annotálják szerep szerint. A szabályok alkalmazásának sorrendjét specifikusságuk foka szabja meg. Három szintet különböztettünk meg: 1) először a legspecifikusabb szabályok, vagyis a vonzatók azonosítását végző lexikális szabályok futnak le, 2) őket követik a predikátumosztályokra alkalmazott nem-default szabályok 3) végül az igeire egyáltalán nem referáló default szabályokat alkalmazzuk. A csoportokon belüli alkalmazási sorrend tetszőleges, az egyetlen “minden más esetben” alkalmazandó default szabály kivételével, melynek az utolsónak kell lennie.

6 Értékelés

Az implementált szabályokat egy két regényből álló szövegen teszteltük: Szabó Magda *Abigél* (1978) és Nádas Péter *Leírás* (1979) című regényein. Mindkét szöveg része az MNSz szépirodalmi alkorpuszának. A kiértékeléshez Méray Tibor *Nagy Imre élete és halála* című művét használtuk. Ez a regény 12,545 mondatból áll, 130,027 szóalakat tartalmaz. A morfológiai elemzés során 1561 féle szóalak maradt ismeretlen.

A szövegben 29855 főnévi csoportot találtunk. Mivel a szabályaink pontosságát szeretnénk ellenőrizni, és a szabályok minden NPt annotálnak szerep szerint, csak a pontossági értékeket számoltuk ki, hiszen az egyes szabályok lefedettségének hiányai valamelyik másik szabály pontosságában is megmutatkoznak. Az értékelés során feltárt hibák – melyek nem az elemzés valamelyik korábbi lépésének hibás kimenetéből erednek – a szabályrendszer javításának módját is kijelölik.

Az eredmények manuális ellenőrzéséhez a Xaira (XML Aware Indexing and Retrieval Architecture) korpuszlekérdező eszközt [11] használtuk. Amint neve is mutatja, a Xaira jólformált XML dokumentumokból álló korpuszok indexálását és komplex lekérdezések megfogalmazását teszi lehetővé. A lekérdezések eredményéhez stíluslapok rendelhetők, így a szabályaink kimenetét emberek számára is könnyen olvashatóvá tudtuk tenni, ami megkönnyítette az eredmények manuális ellenőrzését.

Szabályok	1. Táblázat: Értékelés Pontosság	Találatok száma
Vonzat	71.50%	179
Műveltető	100.00%	1
Okozás	36.00%	11
Társh. szabály	65.00%	76
Társh. default	61.60%	129
Eszköz (szabály és default)	42.37%	573
Mód	54.76%	168
Default idő/mérték	88.09%	42
Lexikalizált	100.00%	59
Összesen	59.57%	1238

A kiértékelés során azt találtuk, hogy a hibás szemantikai címkék nagy része két forrásból származik. Az egyik, hogy az ige, amelynek lemmájára vagy szemantikai osztályára hivatkozni kéne nem szerepel a megfelelő listán. A hibás szemantikai szerepek másik fő oka a Mód szabály alkalmazásából adódó alacsony lefedettség. A Mód szemantikai szerepű összetevők ugyanis gyakorlatilag bármilyen ige mellett megjelenhetnek. Szerencsére munkánk jelenlegi szakaszában úgy tűnik, hogy a fent említett okok kiküszöbölésében még jelentős javulást lehet elérni, és ezáltal a pontosság nagy mértékben javítható.

Bibliográfia

1. É. Kiss, K.: The syntax of Hungarian. Cambridge University Press, 2002
2. Gábor K., Héja E., Mészáros Á.: Kötőszók korpuszalapú vizsgálata. In: Alexin Z., Csendes D. (szerk.): A Második Magyar Számítógépes Nyelvészeti Konferencia Előadásainak kötet, Szeged Egyetemi nyomda, 2004. Szeged, pp. 305-306.
3. Komlósy, A.: Régensek és vonzatok. In: Kiefer F. (szerk.): Strukturális Magyar Nyelvtan I. Mondattan. Akadémiai Kiadó, Budapest, 1992. pp.: 299-528
4. Komlósy, A.: A műveltetés. In: Kiefer F. (szerk.): Strukturális Magyar Nyelvtan III. Morfológia. Akadémiai Kiadó, Budapest, 2000. pp.: 215-291
5. Prószéky G., Tihanyi L.: Humor -- a Morphological System for Corpus Analysis. Proceedings of the first TELRI Seminar in Tihany. 1996. Budapest, pp. 149-158.
6. Radford, A.: Transformational Grammar. Cambridge University Press, 1988. Cambridge
7. Silberstein, M.: Dictionnaires électroniques et analyse automatique de textes: Le systeme Intex. Masson, 1993. Paris
8. Váradi, T., Gábor, K.: A magyar INTEX fejlesztéséről. In: Alexin Z., Csendes D. (szerk): A Második Magyar Számítógépes Nyelvészeti Konferencia Előadásainak kötet, Szeged Egyetemi nyomda, 2004. Szeged, pp. 3-11.
9. Váradi, T.: The Hungarian National Corpus. Proceedings of the Third International Conference on. Language Resources and Evaluation, 2002. Las Palmas pp.385-389
10. Váradi, T.: Főnévi csoportok annotálása Clark rendszerben. In: Az Első Második Magyar Számítógépes Nyelvészeti Konferencia Előadásainak kötet, Szeged Egyetemi nyomda, 2003. Szeged, pp. 65-71.
11. www.xaira.org