

Nem mind VP, ami állít

A névszói állítmány azonosítása számítógépes elemzőben

Dömötör Andrea

PPKE BTK Nyelvtudományi Doktori Iskola
domotor.andrea@itk.ppke.hu

Kivonat: A kutatás annak lehetőségeit vizsgálja, hogyan azonosítható az esetrag nélküli névszók állítmányi szerepe egy pszicholingvisztikai indíttatású számítógépes mondatelemző számára. Ehhez egy konkrét szabályalapú modellt mutat be, amely 8 egyszerű, morfoszintaktikai címkékre alapozott szabály alapján, csak a kérdéses névszót megelőző mondatrészt figyelembe véve állapítja meg, hogy a névszó predikatív funkciót tölt-e be. A teszteredmények szerint amennyiben a kisebb, lokális elemzési feladatok (NP-sítés, a ragtalan névszók további eseteinek egyértelműsítése) jól vannak elvégezve, akkor a nagyobb egységekre alapozó második fázisú elemzésben a predikatív névszók azonosítása kevés erőforrással is nagy hatékonysággal megoldható.

1 Bevezetés

A tanulmány a nominális mondatok számítógépes elemzési lehetőségeit vizsgálja egy pszicholingvisztikai indíttatású elemzőrendszer keretében. Ezek a mondatok azért jelentenek a nyelvtechnológia számára különös kihívást, mert az elemzés középpontjában általában az ige és annak vonzatkerete áll, névszói állítmány esetén viszont legfeljebb a kopulával számolhatunk. A kopula viszont, mint ismeretes, nem mindig jelenik meg a mondatban hangzó formában, és a *van* szótó valamely alakjának jelenléte sem utal feltétlenül arra, hogy nominális mondatról van szó, hiszen a lokatív, egzisztenciális és birtokos mondatoknak is ez az igéje.

Éppen ezért a kutatás kiinduló elképzelése szerint a nominális mondatok elemzésének kulcsa nem a (zéró) kopula megtalálásában, hanem a predikatív névszók állítmányi szerepének felismerésében rejlik. Ezt erősítik meg a Universal Dependencies névszói predikátumokra vonatkozó irányelvei is. Eszerint, ha a kopula bármely számban vagy személyben elhagyható az adott nyelvben, akkor a függőségi elemzésben a predikatív névszó tekintendő fejnek, és ehhez kapcsolódik a kopula (ha van) egy 'cop' éllel (Nivre 2014). Azaz az elemzés során nem a kopula keresi a predikatív névszót, hanem fordítva.

Az elemzés szempontjából tehát kulcsfontosságú annak felismerése, hogy egy névszó predikatív szerepet tölt be. A magyarban azonban a morfológia alapján nem ismerhető fel az állítmányi szerep, hiszen a predikatív névszók a nominatívusziakhoz hasonlóan nem viselnek testes esetragot. A szintaktikai szerepre tehát csak a mondatbeli pozíció, azaz a kontextus mintázatai alapján lehet következtetni. Erre tesz kísérletet jelen kutatás.

Ludányi Zsófia, Krepsz Valéria, Gráczki Tekla Etelka (szerk.): *Doktoranduszok tanulmányai az alkalmazott nyelvészet köréből 2018*. Budapest: MTA Nyelvtudományi Intézet. 2018.

1.1 Az AnaGramma elemző

A célul tűzött elemzés az AnaGramma nyelvi elemzőrendszer (Prószéky–Indig 2015; Prószéky–Indig–Vadász 2016) működési elveit követi. Ennek fő tulajdonsága, hogy pszicholingvisztikai indíttatású, azaz az emberi szövegfeldolgozást próbálja modellálni. Ennek megfelelően az elemzés balról jobbra, szavanként halad. Az aktuális szó elemzése során így elsősorban az öt megelőző szavakra lehet támaszkodni, de rendelkezésre áll egy +2 token méretű ablak is, amely segítségével az éppen elemzett szót közvetlenül követő két szóra előretekintve valósul meg a kétfázisú elemzés első szakasza (Frazier–Fodor 1978). Az elemzés úgynevezett kereslet-kínálat elven működik, ami azt jelenti, hogy az éppen elemzett szó elindíthat keresleti szálakat (például: az ige keresi a vonzatait), vagy kielégítheti valamelyik aktív keresleti szál igényét (például: az alany kapcsolódik az igehez). Az elemző kimenete egy függőségi nyelvtanokra emlékeztető gráfrepresentáció.

Ezek alapján a névszók állítmányi szerepének e tanulmányban ismertetett vizsgálata sem a teljes mondat ismeretében, hanem a bal kontextus, illetve esetleg az ablak alapján történik. Ha a névszó felismerhetően predikatív, akkor a továbbiakban alanyt és esetlegesen kopolát fog maga mellé keresni.

1.2 Előzmény: nom-or-what

A testes esetrag nélküli névszók kezelésére az AnaGramma kereteiben már vannak jelentős eredmények. A nom-or-what elnevezésű algoritmus (Ligeti-Nagy et al. 2018) ezen névszók négy különböző funkcióját (esetét) tudja megkülönböztetni:

- (i) Nominativus (alany)
- (ii) Genitivus (jelöletlen birtokos)
- (iii) NP-t módosító elem (jelző)
- (iv) Névtós szerkezet része

Az algoritmus minden esetben csak az ablakot, tehát az éppen elemzett szót követő két tokent használja fel a döntéshez. Az elemzés ezen része tehát lokálisan, az első fázisban történik, az állítmányi szerep azonosítása azonban nem oldható meg így. Az ablak által nyújtott információ alapján csak nagyon ritkán lehet megállapítani, hogy az adott névszó predikatív-e, ehhez nagyobb egységekre (az NP-kre) és – a magyarra jellemző alany-állítmány szórend miatt – általában inkább a kérdéses szót megelőző mondatrészre van szükség.

A névszók predikatív tulajdonságának vizsgálata tehát a második fázisban, a bal kontextus, azaz az AnaGrammában tározónak nevezett adathalmaz alapján történik. A tározóban a megelőző kontextus elemei találhatóak, elemzett vagy részlegesen elemzett formában. Ebből tehát minden információ rendelkezésre áll, ami az addig elhangzott szavakból kiderülhetett; tudjuk, mely mondatrészek vannak már meg, és mi az, amit még keresünk. Ha a mondat több tagmondatból áll, ezek elhatárolása is menet közben történik, tehát a kérdéses névszóhoz érkeve már azt várjuk, hogy ismert a névszó tagmondatának eleje. Fontos megjegyzés még, hogy ekkor már csak azokról a névszókról kell dönteni, amelyek státuszát a nom-or-what algoritmus függőben hagyta.

Összefoglalva tehát a cél egy olyan szabályalapú algoritmus létrehozása, amely képes egy névszóról a bal kontextusa (és esetleg az öt közvetlenül követő szavak) alapján megállapítani, hogy állítmányi szerepben van-e.

2 Elméleti háttér

Az algoritmus részletes ismertetése előtt két elméleti kérdésre szeretnék kitérni. Egyrészt tisztázni kell, pontosan melyek azok a névszók, amelyeknek az esetleges predikatív „esetét” egyértelműsíteni akarjuk, másrészt az is némi magyarázatra szorul, hogy mit kell érteni az „eset-egyértelműsítés” alatt.

2.1 Predikatív névszók

A kutatás fókuszában a nominális mondatok állnak, azaz azok, amelyekben a predikátum szerepét nem egy (lexikális) ige, hanem egy névszó tölti be. A predikatív névszók lehetnek főnevek, melléknevek, ritkábban számnevek (l. [1]) és az ezeket helyettesítő névmások.

- [1] a. *Ez nekem sok.*
b. *A tegnapi eredmény 2:0 volt.*

Fontos még tisztázni, hogy a kopulás mondatok közül is csak a predikatívokkal (a tipológiát l. Higgins 1973) foglalkozunk, az azonosító mondatokkal (pl. *Én vagyok az apád.*) egyelőre nem, ezekben ugyanis nem mindig egyértelmű, hogy melyik névszót vagy névmást kell predikatívnek tekinteni, és melyiket alanynak, vagy hogy egyáltalán feltételezzünk-e alany-állítmányi viszonyt a két DP között. Ennek a mondattípusnak a számítógépes kezelése további kutatásokat igényel.

Jelen tanulmány nem foglalkozik a másodlagos predikátumokkal (pl. *Tibi okosnak látszik.*) sem, mivel ezek datívuszban állnak, a kutatás pedig csak az esetrag nélküli névszókat célozza.

Az algoritmusnak tehát akkor kell egy névszót predikatívnek ítélni, ha az az alany valamilyen attribútumát jelöli. Ha a névszói állítmány frázis, akkor a program a fejnek ítéli meg a predikatív jegyet, amit [2]-ben a vastag betű jelöl. Ha a további elemzés során az NP-összecsomagolás megtörténik, akkor a fej predikatív jegye természetesen az egész frázisé lesz, hasonlóan az esetragokhoz.

- [2] a. *Judit **ügyvéd.***
b. *Judit megfontolt, kezdő, ugyanakkor nagyon dörzsölt **ügyvéd.***

2.2 Milyen eset a névszói állítmány?

Az elérhető korpuszok általában minden esetrag nélküli névszót nominatívusznak jelölnék, a számítógépes elemzót viszont az segítené, ha csak az alany lenne NOM-ként címkézve, a többi funkció más jelölést kapna. A nom-or-what fentebb ismertetett kategóriái közül a (ii)–(iv) valóban kevésbé tekinthető nominatívusznak. A predikatív névszók esetében azonban nem ilyen egyértelmű a helyzet.

Szécsényi (2000) amellet érvel, hogy az állítmányi szerepű névszó az alannal való egyeztetés útján maga is nominatívuszi esetet kap. Ezzel szemben Kádár (2011) szerint egyáltalán nincs is értelme a névszói predikátumok esetéről beszélni, hiszen ezek nem argumentumok, így nincs szükségük absztrakt esetre. Az, hogy alakilag a nominatívuszi névszókkal egyeznek meg, azért van, mert a magyarban morfológiai szempontból ez az alapértelmezett (Schütze 2001).

Jelen tanulmány ebben a kérdésben nem kíván állást foglalni. A gépi elemzés szempontjából a predikatív tulajdonság mint egyfajta funkciójegy jelzi, hogy az adott elem egy alanyt és esetleg egy kopulát vár maga mellé. Az absztrakt eseteknek tehát itt valójában nincs jelentősége, a névszói állítmányt így egyszerűen esetrag nélküli, predikatív jeggyel bíró névszónak tekintjük.

3 A névszók eset egyértelműsítése

Bár a nom-or-what eset-egyértelműsítésről beszél (az általa egyébként részletesen nem tárgyalt predikatív névszók esetében is), valójában az „eset” megnevezés, amint láttuk, nem feltétlenül követi az elméleti nyelvészet esetfogalmait. Inkább a morfoszintaktikai címkében jelölendő tulajdonságokról van szó, amelyek szükségesek ahhoz, hogy az elemző egyrészt meg tudja határozni az NP-ket (nom-or-what), másrészt felismerje a nem igei állítmányt (predikatív névszók azonosítása).

A kutatás céljával tűzött predikatív névszókat azonosító algoritmus az „is-pred” nevet viseli, és a kétfázisú elemzési folyamat második fázisában valósul meg. Az elsőben, amint az 1.2 alfejezetben említettem, a nominatívusz, a genitívusz, az NP-módosító és a névutó előtti névszó azonosítása történik lokálisan, az ablak alapján. A második fázisban az is-prednek már csak azokról a névszókról kell döntenie, amelyeket a nom-or-what nem tudott egyértelműen azonosítani. Az első fázis jelenléte rendkívül fontos, enélkül kevés esély lenne arra, hogy a predikatív tulajdonságot a bal kontextus alapján fel tudjuk ismerni, amint azt [3] példái mutatják.

- [3] a. *Zsófi orvos.* – predikatív
 b. *Zsófi orvos barátja...* – NP-módosító
 c. *Zsófi orvos nélkül is kigyógyult az influenzából.* – névutós szerkezet része

Ez lényegében azt jelenti, hogy bár az is-pred algoritmus a pszicholingvisztikai indíttatás szerint alapvetően a már elhangzott elemeket veszi figyelembe, azért van valamennyi „rálátása” a kérdéses névszót követő szavakra is.

Az algoritmus bemenete tehát az eldöntendő szó, az azt megelőző mondatrész (tározó) és az azt közvetlenül követő két token (ablak). A kimenet pedig nagyjából a háromértékű logika értékeinek felel meg: *Pred*, ha a névszó egyértelműen predikatív, *Nonpred*, ha egyértelműen nem az, és *Undefined*, ha a meglévő információkból nem lehet biztos döntést hozni. Az algoritmus magas pontosságra törekszik, és kevesebb hangsúlyt fektet a fedésre: amennyiben a teljes mondat ismerete nélkül nem egyértelmű a predikativitás kérdése, inkább *Undefined* választ ad, minthogy hibázzon, hiszen később, a további szavak beolvasásával a bizonytalan ítélet pontosítható lesz.

A döntést segítő szabályok elsősorban a szófajcímkékre alapoznak, a lexikális információkat kevésbé veszik figyelembe. Az algoritmus tehát alapvetően a mondat szerkezet alapján azonosítja a predikatív névszókat. A szabályok kialakítása nyelvészeti megfontolások és intuíciók alapján történt, melyeket minden esetben korpuszmérésekkel ellenőriztem.

A felhasznált szabályok sorrendben a következők:

1. Ha van a tározóban vagy az ablakban ragozott ige, amely nem (potenciális) kopula → *Nonpred*
2. Ha a szó címkéje *ROMAN* (római szám) → *Nonpred*

3. Ha a következő szó címkéje EKSZ (-e kérdőszó) → Pred
4. Ha az ablakban van 1–2. számú létige → Pred
5. Ha a szó DP (része) → Nonpred
6. Ha a tározóban van egyértelműsített nominatívusz, és a szó magában áll vagy frázis vége → Pred
7. Ha a szó melléknév vagy számnév, és magában áll vagy frázis vége → Pred
8. Ha a szó többes számú melléknév → Pred
9. Egyébként → Undefined

A szabályok sorrendje fontos, a 7. és 8. például csak akkor működik, ha az 5. már kizárta azt a lehetőséget, hogy az adott melléknév alany legyen. Az 5. szabály egyébként DP-nek tekinti a határozott névelős főnévi csoportokat, a tulajdonneveket, a személyes és mutató főnévi névmásokat és bizonyos birtokos szerkezeteket. Ez utóbbiak közül az számít DP-nek, amelyek alany vagy azonosító, és nem tulajdonságot jelöl (l. [4]). A DP-k tehát (a birtokosokon kívül) a tározó információi alapján elméletileg egyértelműen azonosíthatók. Mivel a jelen algoritmusnak nem célja az NP-sítés, a tesztkorpuszban ezek kézzel vannak jelölve.

- [4] a. *Ádám tagja a szövetségnek.* – predikatív
 b. *Ádám a szövetség legifjabb tagja.* – azonosító

Azt, hogy a szó magában áll vagy frázis vége-e, az ablak alapján próbálja eldönteni az algoritmus. Ehhez egy reguláris kifejezéseket illesztő függvényt használ, amely olyan POS-tag-mintákat keres, amelyek egyértelműen jelzik a frázis végét: például hogyha a névszó után névelő vagy ragozott ige következik. Ennek kidolgozásában megint csak a nagy pontosság volt a cél: a függvény csak akkor jelez frázisvéget, ha teljes biztonsággal megállapítható, hogy a szót tartalmazó frázis már nem fog folytatódni.

[5]-ben a 6. és 7. szabályokra adok egy-egy példát. (A többi szabály egyértelmű.) Vastag betű jelöli a kérdéses névszót, az utána következő mondatrész ebben az elemzési fázisban nem látható.

- [5] a. *ez/DET_NM.NOM a/DET szombati/MN.0 buszjárat/FN.NOM **vice**/FN.pred* → predikatív a 6. szabály alapján
 b. *Eredeti/MN.0 formájában/FN.PSe3.INE nem/HA **azonos**/MN.pred a/a/DET kereszténység/FN.GEN istenfogalmával/FN.PSe3.INS ./SPUNCT* → predikatív a 7. szabály alapján

Az algoritmust, egyelőre teszt céllal, Python3-ban implementáltam.

4 Az algoritmus tesztelése

4.1 Tesztadat, módszer

Az is-pred algoritmus teszteléséhez az MNSz.²-ből (Oravecz–Váradi–Sass 2014) készítettem tesztkorpuszt. Ez 300 tagmondatot tartalmaz, melyek közül 200-ban van predikatív névszó. A tesztmondatok között nem szerepel azonosító mondat, hiszen mint 2.1-ben említettem, ezek elemzése nem feltétlenül követi a predikatív mondato-

két. Egyéb tekintetben a korpusz random összetételű. Az MNSz.² morfológiai annotációjához képest annyi módosítást tettem, hogy a tulajdonnevek saját címkét kaptak. Ezen kívül speciális jelzéseket adtam a DP-knek (hogy az esetlegesen hibás NP-chunkolás ne zavarja az algoritmus értékelését), illetve szükség esetén kézzel javítottam a POS-tagek hibáit.

A tesztkorpuszon először lefuttattam a nom-or-what implementációját, majd az eredményt ismét kézzel javítottam, ahol szükséges volt. Ez után 896 egyértelműsítendő névszó maradt, amelyből 208 volt predikatív. Ezeket kézzel annotáltam a kérdéses névszók valódi funkciója szerint, a teljes mondatot figyelembe véve (Pred vagy Nonpred), így állt elő a kiértékeléshez szükséges gold standard. A kiértékelési módszer sajátossága, hogy az Undefined választ semmilyen esetben nem tekinti jónak, akkor sem, ha egyébként az adott elemzési ponton valóban nem lehet még dönteni a névszó predikativitásáról, azaz az algoritmus tulajdonképpen jól működött, de az eredeti kérdésünkre nem adott választ. A teszt tehát nem az *elvárt*, hanem a *hasznos* működést értékeli.

4.2 Eredmények és értékelés

Az is-pred script gold standardhoz viszonyított eredményei az 1. táblázatban láthatók. A sorok jelzik az algoritmus válaszait, az oszlopok a várt (kézi annotáció szerinti) értékeket. A tévesztési mátrix mellett a pontosság és fedés értékek is szerepelnek.

	Pred	Nonpred	Pontosság (%)
Pred (db)	124	2	98,4
Nonpred (db)	0	546	100
Undefined (db)	84	139	-
Fedés (%)	59,6	79,5	

1. táblázat. Az is-pred algoritmus eredményei

Mint látható, az algoritmus magas pontossággal dolgozik, az Undefined válaszokat nem számítva csak kétszer tévedett. Ezek a hibák [6]-ban láthatók.

- [6] a. *Ebből egy ötödik, a férfi pólósoké rendkívüli családás.*
 b. *A három fiú közül az egyik nagydarab volt, mint egy súlyemelő.*

A [6a]-ban az algoritmus az *ötödik* számnevet predikatívnak ítélte a 7. szabály alapján, az ablakban található determináns ugyanis a frázis végét jelezte a program számára. Vagyis ezt a hibát nem az algoritmus szabályainak alkalmatlansága, hanem a frázis végét jelző függvény pontatlansága okozta, ez ugyanis nem számolt az ehhez hasonló magyarázó szerkezetekkel. A [6b]-ben pedig a 6. szabály alapján ítélte a program a *súlyemelő* szót predikatívnak, az ilyen jellegű összehasonlításokra ugyanis nem volt felkészítve.

Az 1. táblázatból az is látszik, hogy a predikatív névszókat gyakrabban specifikálja alul a program, mint a nem predikatívokat, azaz előbbieket esetében alacsonyabb a fedés. Úgy tűnik tehát, hogy az algoritmus a „nem” döntést könnyebben meg tudja hozni, mint az „igen”-t. Azt, hogy a névszó biztosan nem predikatív, közel 80% biztonsággal meg tudta mondani a program, míg az egyértelműen predikatív válasz esetén

ez a szám nem érte el a 60%-ot. Ennek az lehet oka, hogy a Nonpred választ adó szabályok feltételeinek igazsága könnyebben megállapítható a rendelkezésre álló információkból. Mindemellett a viszonylag alacsony fedés a teljes elemzés szempontjából nem olyan nagy probléma, a későbbiekben, a további szavak beolvasásával könnyen korrigálható.

A fedésértékek egyébként pszicholingvisztikai szempontból is érdekesekek, hiszen azt mutatják, hogy (legalább) ekkora arányban „valós időben”, azaz csak a szó elhangzásáig rendelkezésre álló információból eldönthető, hogy a névszó predikatív-e. De emberi értékeléssel ennél valószínűleg még jóval nagyobb számokat is kaphatnánk, hiszen az ember lexikális információkra is támaszkodik, amire a program egyáltalán nem.

A 2. táblázat azt mutatja, hogy melyik szabályt hányszor alkalmazta a program.

1. ha van ige → Nonpred	97
2. ha ROMAN → Nonpred	3
3. ha EKSZ következik → Pred	1
4. ha 1–2. személyű létige van az ablakban → Pred	2
5. ha DP (része) → Nonpred	446
6. ha a tározóban van nominatívusz → Pred	65
7. ha melléknév vagy számnév → Pred	44
8. ha többes számú melléknév → Pred	14
9. egyébként → Undefined	224

2. táblázat. A szabályok alkalmazásának gyakorisága

Úgy tűnik, a leggyakoribb nom-or-what által azonosítatlan névszó tulajdonnév vagy DP része (5. szabály). Ezek az elemzés második fázisában, balról jobbra elemezve a determinánsok segítségével már könnyebben egyértelműsíthetők. A predikativitás felismerésében legtöbbször a nominatívusz jelenléte segített (6. szabály), de a melléknévek és számnevek esetében sokszor enélkül is megállapítható az állítmányi szerep (7. és 8. szabály). Amint az előzőekből is kiderült, az algoritmus viszonylag nagy arányban, az esetek negyedében nem tud egyértelmű döntést hozni.

5 Konklúzió

Az is-pred algoritmus elérte a kitűzött célját: nagy pontossággal képes a pszicholingvisztikai indíttatású elemző elveit követve megállapítani, hogy egy névszó predikatív szerepet tölt-e be. A fedés viszonylag alacsonyabb mértéke részben a korlátozott információ (nem áll rendelkezésre a teljes mondat), részben az algoritmus „óvatossága” miatt van, úgy lett ugyanis felépítve, hogy csak biztos választ adjon, a bizonytalan eseteket hagyja függőben.

Az eredmények alapján elmondható, hogy a nagy pontosságra törekvés mellett is elég nagy számban lehet a névszók állítmányi szerepéről dönteni egy egyszerű, kevés szabályt tartalmazó algoritmussal. Ehhez viszont szükség van arra, hogy a bal kontextus ragtalan névszónak többi funkciója (különösen a nominatívusz) jól legyen azonosítva, illetve jól működjön az elemzőben az NP-sítés. Vagyis a predikatív névszók azonosítása akkor lehet sikeres, ha a kisebb, lokális elemzési feladatok jól vannak megoldva. Ez a már említett kétfázisú elemzési modell logikus következménye.

További kutatási feladat a pontosság-fedés arány optimalizálása esetleges további szabályokkal vagy a meglévők finomításával. Ezenkívül komoly elméleti megfontolásokat is igénylő kérdés az itt nem tárgyalt azonosító mondatok kezelése.

Köszönetnyilvánítás

Jelen kutatás az FK~125217 és a PD~125216 számú projekt keretében az FK17 és a PD17 pályázati program finanszírozásában a Nemzeti Kutatási Fejlesztési és Innovációs Alap által biztosított támogatással valósult meg.

Irodalom

- Frazier, L., Fodor, J. D. 1978. The sausage machine: A new two-stage parsing model. *Cognition* 6/4. 291 – 325.
- Higgins, R. 1973. *The Pseudo-Cleft Construction in English*. New York: Garland Press.
- Kádár, E. 2011. *A kopula és a nominális mondatok a magyarban*. Budapest: Akadémiai Kiadó.
- Ligeti-Nagy, N., Vadász, N., Dömötör, A., Indig, B. 2018. Nulla vagy semmi? Esetgyértelműsítés az ablakban. In: Vincze, V. (szerk.): *XIV. Magyar Számítógépes Nyelvészeti Konferencia* (Szeged, 2018. január 18–19.) Szeged: Szegedi Tudományegyetem. 25–37.
- Nivre, J. 2014. *Nonverbal Predication and Copulas in UD v2*. Elérhető: <http://universaldependencies.org/v2/copula.html> Letöltés dátuma: 2018. febr. 8.
- Oravecz, Cs., Váradi, T., Sass, B. 2014. The Hungarian Gigaword Corpus. In: Calzolari, N. et al. (szerk.): *Proceedings of the 9th International Conference on Language Resources and Evaluation* (Reykjavik, Izland, 2014. május 26–31.) European Language Resources Association. Reykjavik, Izland. 1719-1723.
- Prószéky, G., Indig, B. 2015. Magyar szövegek pszicholingvisztikai indíttatású elemzése számítógéppel. *Alkalmazott Nyelvtudomány*, 15(1-2): 29–44.
- Prószéky, G., Indig, B., Vadász, N. 2016. Performanciaalapú elemző magyar szövegek számítógépes megértéséhez. In: Kas, B. (szerk.): „Szavad ne feledd!”: *Tanulmányok Bánréti Zoltán tiszteletére*. Budapest: MTA-NYTI. 223–232.
- Schütze, C. 2001. On the Nature of Default Syntax. *Syntax*, 4: 205–238.
- Szécsényi T. 2000. Esetgyeztetés a predikatív főnévi csoportban. In: Büky, M., Maleczki, M. (szerk.): *A mai magyar nyelv leírásának újabb módszerei IV*. Szeged: SZTE. 189–202.