

Anaforafeloldás menet közben – névmások egy pszicholingvisztikailag motivált elemzőben

No Author Given

No Institute Given

Előadásomban a koreferenciafeloldás számítógépes nyelvészeti kezelésével foglalkozom, közelebbről a személyes-, visszaható- és kölcsönös névmásokkal egy pszicholingvisztikailag motivált elemzőrendszer keretében. A koreferenciaviszonyok további fajtái – mint az ismétlés, tulajdonnév-variánsok, szinonímia, hiponímia stb. – további kutatásokat igényelnek.

Az ANAGRAMMA elemző (Prószéky és Indig, 2015; Prószéky et al., 2016) az emberi szövegfeldolgozás mintájára szigorúan balról jobbra és szavanként elemzi a szöveget, ezért a koreferenciafeloldásnak is "menet közben" kell megtörténnie az elemzés folyamatában. Az emberi szövegfeldolgozás – és ezen belül a koreferenciaviszonyok feloldásának – modellálásakor alaposan meg kell figyelni a koreferenciaviszonyokat kialakító elemek viselkedését.

A ragozásból bizonyos argumentumok egyes jegyei kiszámíthatóak. Az ige vonzatkeretének előhívása után¹ a vonzatkeresők számban és személyben egyező alanyt és tárgyat keresnek. A vonzatigény felmerülésekor egy default zérus csomópont jön létre a megfelelő számmal, személlyel és esetraggal, amelynek helyét az aktuálisan a mondatban szereplő elem veheti át – ha van ilyen. Ennek megfelelően ebben a keretben a zérónévmások kezelése is megoldott.

A pronominalizáció mögött húzódó szabályrendszer (Pléh és Radics, 1976) feltárása hozzájárul ahhoz, hogy a balról jobbra és szavanként történő mondat-elemzés során "menet közben" létrejöhessenek az anaforikus viszonyokat és a koreferenciát jelző függőségi élek. A mögöttes szabályrendszer beépítésével az elemző képes arra, hogy az emberhez hasonlóan kezelje a pronominalizációt és a koreferenciaviszonyokat.

A kölcsönös és visszaható névmások a saját esetragukkal a többi argumentumhoz hasonlóan készen állnak az ige vonzatigényeinek kielégítésére. Az elemzés során az egyik legégetőbb probléma a többértelműségek kezelése. A magyarban a *maga* névmás két jelentéssel bír: (1) egyes szám harmadik személyű visszaható névmás nominatívuszban, és (2) az egyes szám második személyű személyes névmás magázó alakja nominatívuszban².

Előadásomban egy megoldást kínálok a személyes, visszaható és kölcsönös névmások kezelésére az ANAGRAMMA keretében a (Pléh és Radics, 1976) algoritmus felhasználásával. A jelenségekkel kapcsolatos megfigyeléseimet korpuszmérésekkel támasztom alá, amelyeket a Pázmány Korpuszon (Endrédi, 2016) végeztem.

¹ Ezen lépés részleteiért lásd Indig és Vadász (2016)

² Emellett a *maga* névmásnak van egy harmadik használata is a pl. a *maga az ördög* szerkezetben.

Irodalomjegyzék

- Endrédi István. *Nyelvtechnológiai algoritmusok korpuszok automatikus építéséhez és pontosabb feldolgozásukhoz*. Doktori disszertáció, Pázmány Péter Katolikus Egyetem Információs Technológiai és Bionikai Kar, Budapest, 2016.
- Indig Balázs és Vadász Noémi. Windows in human parsing – How far can a preverb go? In: Marko Tadić és Bekavac Božo szerk. *Proceedings of the Tenth International Conference on Natural Language Processing (HrTAL2016)*, Dubrovnik, Croatia. sept, 2016. (Elfogadva, nyomtatás alatt).
- Pléh Csaba és Radics Katalin. "Hiányos mondat", pronominalizáció és a szöveg. *Általános nyelvészeti tanulmányok*, 1976, 11(1):261–277.
- Prószéky Gábor és Indig Balázs. Magyar szövegek pszicholingvisztikai indíttatású elemzése számítógéppel. *Alkalmazott nyelvtudomány*, 2015, 15(1-2):29–44.
- Prószéky Gábor, Indig Balázs és Vadász Noémi. Performanciaalapú elemző magyar szövegek számítógépes megértéséhez. In: Bence Kas szerk. *"Szavad ne feledd!"*: Tanulmányok Bánréti Zoltán tiszteletére. MTA Nyelvtudományi Intézet, Budapest, 2016, 223–232.