

Kötőszók korpuszalapú vizsgálata

Gábor Kata¹, Héja Enikő¹, Mészáros Ágnes¹

¹ MTA Nyelvtudományi Intézet, Korpusznyelvészeti Osztály
1068 Budapest, Benczúr u. 33
{gkata, eheja, magnes}@corpus.nytud.hu

Kulcsszavak: kötőszó, tagmondathatár, szintaktikai elemzés, MNSZ

Munkánk távlati célja egy magyar szintaktikai elemző létrehozása. Az elemző első lépésben a kötött szórendű frázisokat látja el címkézett zárójelekkel, majd az összetevők közötti viszonyok feltérképezéséhez szükséges szintaktikai és szemantikai jegyekkel. A folyamat második fázisa az összetevők mondattani szerepének felismerése. Kiindulópontunk az igei állítmány vonzatkerete, melyet egy lexikai adatbázis alapján azonosítunk. Nyelvünk sajátosságai miatt az argumentum-szerepű frázisok kiválasztásánál nem hivatkozunk a szórendre: az igei szubkategorizációnak megfelelő frázisokat akkor akarjuk vonzatként megjelölni, ha ugyanabban a *tagmondataban* vannak, mint a lehetséges régenjük. Ezért a vonzatkeret illesztését meg kell, hogy előzze egy lépés, mely kijelöli a *tagmondathatárokat*. Hogy mit értünk tagmondat alatt, azt meghatározza az elemzés választott kerete. Mivel az igei argumentumok azonosítására törekszünk, és a tagmondathatár jelöli ki azt a területet, amin belül az argumentumokat keressük, célszerű elfogadnunk a hipotézist, hogy *egy mondatban két finit ige között mindig van tagmondathatár*. Kérdés, hogyan használhatók a kötőszók a tagmondathatár pontos helyének azonosítására. Mivel nincs minden tagmondathatáron kötőszó, azt vizsgáljuk meg, hogy ahol (két finit ige között) kötőszót találunk, ott mindig van-e határ. Ha a kötőszót olyan szófajként definiáljuk, mely két azonos típusú összetevőt köt össze, akkor azokat az eseteket kell kiszűrniük, amikor a kötőszó mindkét oldalán tagmondat áll. Ha a tagmondatok elhatárolására akkor kerül sor, amikor már felépítettük a kötött szórendű frázisokat, melyek tartalmazhatnak kötőszókat, akkor csak azokat a kötőszókat kell vizsgálnunk, melyek a megjelölt frázisokon kívül esnek.

Megfigyeltük, hogy a kötőszók a tagmondathatárhoz képest eltérő pozíciókat foglalhatnak el. Egyes kötőszók mindig első helyen állnak a tagmondatban, mások az első összetevőt követik, néhányuk pedig szinte bárhol megjelenhet. A kötőszók korpuszalapú vizsgálata azonban azt mutatta, hogy az adatbázisban szereplő 71 kötőszó legtöbb előfordulásában még a lehetséges pozíciójukra vonatkozó információ birtokában sem elégséges a tagmondathatár kijelöléséhez. Ennek okai a kötőszók eltérő disztribúciója és a homonímia. Előadásunkban bemutatjuk a kötőszók disztribúciójának a Magyar Nemzeti Szövegtár adatain alapuló vizsgálatát, és a tagmondathatár felismeréséhez jól hasznosítható csoportosításukra teszünk javaslatot.