

Leíró nyelvtan – adatbázisból

Bódis Zoltán, Kleiber Judit, Szilágyi Éva, Visket Anita

PTE BTK Nyelvtudományi Tanszék
lile@btk.pte.hu

Kulcsszavak: leíró nyelvtan, lexikon, relációs adatbázis, vonzatszótár

Hallgatói kutatócsoportunk egy nyelvészeti adatbázis kifejlesztésére és feltöltésének koordinálására vállalkozott. Az adatbázis, amin dolgozunk, egy nyelvészeti lexikon, amelyben szabad és kötött morféákat tárolunk vonzataikkal és részletesen megadott tulajdonságaikkal együtt.

A lexikonunkat relációs adatbázis formájában építettük fel. Ez a módszer lehetőséget ad arra, hogy a lexikai egységeinket (a morféákat) a mögöttes reprezentáció alkalmazása mellett az unifikációs morfológiában megszokott jegyekkel tároljuk, de dinamikusan bővíthető struktúrában. Továbbá így arra is lehetőségünk nyílt, hogy a morféákhoz rendelt jelentést (intenziót) tetszőleges számú morféma összekapcsolásával definiáljuk. A morféák fonológiai, morfológiai, szintaktikai és szemantikai tulajdonságainak jegystruktúrában való tárolása mellett a morféákon belül vagy a morféák között működő szabályokat is rekordként rögzítjük az adatbázisban, így ez a szabálykészlet is dinamikusan bővíthető. A szabályok definiálásának technikájához az Alberti Gábor által fejlesztett GASG (Generatív / Általánosított Argumentum Struktúra Nyelvtan) adja az elméleti keretet.

Fejlesztőkörnyezetnek a Delphit választottuk, mivel ez lehetőséget nyújt tetszetős felhasználói felület készítésére, és a keresés hatékonyságának növelésére programozhatunk benne gépi kódban is.

Az adatbázis a tanszékünkön folyó számítógépes nyelvészeti alkalmazások fejlesztésének támogatása mellett egyéb célokat is szolgál. Részben oktatási segédanyagként szánjuk ezt a szótárat. Mind a felső-, mind a közoktatásban szükséges a nyelvtani ismeretek átadásának módszertanát megújítani, és ezt kívánjuk támogatni az adatbázisunkra épülő oktatási szoftverekkel. Részben pedig alapját képezi egy későbbi leíró nyelvtan elkészítésének, amelyet nem szabályok halmazának képzelünk, hanem az egyes, egyedileg definiált morféákon vagy szavakon működő jelenségek halmazának. A „nyelvtan” megszokott fogalmához képest a programunkban sokkal kevesebb az általánosítás, és az esetleges általánosítások (külön rekordonként tárolt és meghívható eljárások) által érintett elemek köre pontosan definiált, ahogy ez a számítógépes nyelvészeten általában szokásos. A leíró nyelvtan statisztikai alapon „generálható” a működő szabályainkból: a legnagyobb elemkészleten működő eljárások alkotják a „megnevezendő” szabályokat, a kisebb halmazokon működők pedig a kivételeket.

Az adatbázis feltöltéséhez az anyaggyűjtésünk nem korpusz-, hanem kompetencia-alapú, így felveti a nyelvi regiszter problémáját, ahogy ez a leíró nyelvtanok és a számítógépes alkalmazások esetében törvényszerű.