

Jelentésrepresentáció ontológiában⁸⁷

Szóts Miklós¹

¹ Alkalmazott Logikai Laboratórium,
Hankóczy u. 7, Budapest 1202
szots@all.hu

Az a kutatás, amelyről beszámolunk, nem nyelvészeti ihletből született, hanem az ontológiák tanulmányozásából. Ennek megfelelően a nyelvészeti vonatkozások még kidolgozatlanok, de a „vizió” működését már demo szintű rendszerrel is bemutatjuk (l. Gröbler Tamás benyújtott előadását).

A kutatás célja, hogy egy szöveg jelentését formálisan reprezentáljuk. Ennek több praktikus felhasználási területe lehet, a jelenlegi projekt orvosi szabad szövegeket formális kórlapstruktúrába interpretál, de az igazi célkitűzés szemantikus kereső (text mining eszköz) fejlesztése.

Egyelőre a nyelvnek csak a leíró funkcióját kezeljük, tehát kijelentő mondatokból álló szövegeket elemzünk. Egy szöveg (szövegegység) **jelentése a szöveg által leírt szituáció reprezentálása egy világmodellben**⁸⁸. A módszer a következő modulokból álló tudásbázison alapul:

a világról szóló tudás, azaz a **világmodell**,

a **nyelvi tudás**,

a kettő közti összefüggést reprezentáló **leképzés**.

Mind a világmodell, mind a nyelvi tudást ontológiával reprezentáljuk.

A **világmodell**t reprezentáló ontológia relációkkal bőven felfegyverzett. Célunkhoz mérve⁸⁹ az ontológia szerkezetének legfontosabb elemei az eseményszerűségek, és a tulajdonságok. Az eseményszerűségek jellemzésére a szereprelációkat használjuk – lásd a 2005-ös előadást. A tulajdonságok reprezentálásánál ezek időfüggését, ill. más paramétertől való függést láthatóvá és kezelhetővé tesszük az ontológiában. Különösen jelentős az idő és a hely kezelése.

A **nyelvi tudást** tároló ontológia, – nevezzük ezt lexikonnak, – „fogalmi” az önálló jelentéssel bíró nyelvi egységek, – lehetnek ragok, szavak, vagy több szóból álló frázisok, – nevezzük ezeket lexémáknak. A lexikon szerkezetét egy, a módszerhez jól illő, – feltehetően valamilyen lexikális felépítésű, – nyelvtan adná.

A módszer a világmodell, és nyelvi ontológia közti **leképzés**en alapul. Ennek a leképzésnek kell meghatároznia a nyelvi elemek által **referált** világmodellbeli egyede- ket. Természetes módon a lexémákhoz fogalmakat, vagy fogalmak közti relációkat rendel – ezek előfordulásaira referálnak a lexémák. Ugyanakkor azokhoz a szintakti-

⁸⁷ a kutatás részint az NKFP-2/042/04, részint a GVOP-3.1.1-2004-05-0363/3.0 projekt keretében folyt.

⁸⁸ természetesen világmodell alatt egy jelenségkör egy adott szempontból felépített modelljét értjük.

⁸⁹ és meggyőződésünk szerint általában

kai viszonyokhoz, amelyek lexémákból mondatot fűznek össze, hozzárendel ontológiai relációkat.

Ez utóbbi adja annak lehetőségét, hogy a mondatokból az általuk referált szituáció leírását kapjuk. Legnyilvánvalóbb az igék vonzatkerete. Az ige⁹⁰ eseményszerűsége referál, a vonzataik pedig megfelelnek a szereprelációknak. Mivel a vonzatok jelentése nem univerzális, minden igénél külön a vonzatoknak megfelelő leképzés jelöli ki az ige referátumának megfelelő szereprelációt.

Mivel ez sarkalatos kérdés, hadd hozzunk már itt egy példát: az „ad” ige az ADÁS fogalomra referál, nominatívusz vonzata pedig az ADÁS-ból kiinduló *aktora* relációra. A „kap” ige ugyancsak az ADÁS fogalomra referál, viszont nominatívusz vonzata az ADÁS-ból kiinduló *recipiense* relációra.

A módszer nem kíván teljes lexikont: akkor is képes jelentésrepresentációt építeni, ha a szövegnek nem minden lexémája szerepel a lexikonban, – természetesen ekkor a jelentésrepresentáció is hiányos lesz.

Az elemző algoritmus jelenleg készen kapja a mondat alaktani és szintaktikus elemzését. Az alaktani elemzéssel szemben igényesek vagyunk, de a szintaktikai elemzés hiányosságait a reprezentációt építő algoritmus át tudja hidalni, akár mozaik elemzésből is felépíti a helyes reprezentációt. A fontos az, hogy a szintaktikai elemzés helyes lexémacsoportokat fogjon össze, és jól határozza meg a fejet. Ugyanis a jelenlegi algoritmus a fejből indul ki: az ez által referált fogalomhoz a vonzatoknak megfelelő relációkkal fűzi a fej által uralt lexémák referáltjait. A szintaktikai többletműiséget kezelni tudjuk: a vonzat - szerepreláció megfeleltetés kiszűri a szintaktikusan helyes, de értelmetlen elemzéseket.

A jelenlegi állapotban az alaktani, szintaktikai és szemantikai még el van választva, de olyan továbbfejlesztésről álmodunk, amikor ezek egymásat segítve összedolgoznak.

Annak ellenére, hogy nem nyelvészeti ihletésre indult kutatásunk, nyilvánvalóak nyelvészeti vonatkozásai:

egy DR-hez hasonló struktúrát építünk, – a módszer jellegzetessége az, hogy ezt egy ontológiába építi be. Ezáltal a szemantika önálló életre lép, és kontrollálhatja, vezérelheti az elemzést.

a szereprelációkat a tematikus szerepek ihlették, még ha nem is pontosan azok; módszerünk közel áll a nem transzformációs nyelvtanokhoz, pl. a konstrukciós, vagy a fejvezérelt fázisstruktúra nyelvtanhoz, az előadásban erre részletesebben kitérünk.

⁹⁰ a létige kivétel – azt nem vesszük „igének”.