

A lexikalista szintaxis rangja(i)

Szilágyi Éva, Kleiber Judit, Alberti Gábor¹

PTE BTK Nyelvtudományi Tanszék;
Pécs 7624 Ifjúság útja 6.
realis@mail.btk.pte.hu

Bevezetés

Az MSZNY konferencián többször szerepelt GeLexi [5] és Lile [6] projektek örökén szerveződött a ReALIS projekt a PTE BTK Nyelvtudományi Tanszékén 2006-ban. Alapvető célunk továbbra is az intelligens nyelvészeti alapokon történő fordítás, totálisan lexikalista keretben [1]. A lexikont egy relációs adatbázis képviseli, amit MS SQL2005-ben valósítunk meg.

Egy relációs adatbázis lényege a „relációk” definíciójában áll, amely a reprezentálható tények leírása mellett, magát az adatbázist alkotja. A leírás az egyedkapcsolat modellen alapul. Minden egyes entitás egy rendezett n -es: elemei valamilyen relációt alkotnak — ez egy rekord. A reláció maga a tábla, ahol minden egyes sor (rekord) egy rendezett n -es, és minden oszlop egy attribútum. A relációk egyedei közötti kapcsolatok állhatnak fel [7]. A számtalan adatbázis-kezelő rendszerből és SQL-megvalósítás közül választásunk az Microsoft SQL Server 2005 rendszerre esett: ezáltal egy komplett relációsadatbázis-kezelő keret rendelkezésünkre áll, és számos olyan szolgáltatással is rendelkezik, amelyet céljaink eléréséhez ki tudunk használni: elsődlegesen a kliens-szerver architektúrát, amelynek köszönhetően az adatbázis egyidejűleg több felhasználó és programmodul által is elérhető, vagyis a felhasználók a rá épített webes felületen keresztül, ingyenesen használhatják a rendszert.

A morfofonológiai komponenst alapvetően a Lile adatbázisából vettük át. Jelenleg a szintaxis „szabályainak” a lexikonba építésén dolgozunk, erről szól a poszterünk is. A fő komponenst az ez után fejlesztendő szemantika jelenti, amit a ReALIS [4] dinamikus szemantikai rendszer implementációjaként képzelünk el.

Predikátumok és argumentumok, régensek és vonzatok

A totális lexikalizmus eszméjével összhangban a fent említett „szabályok” nem szabályok, hanem az egyes lexikai elemekre jellemző tulajdonságok, amelyek végül unifikálód(hat)nak. Egy lexikai egységről tárolnunk kell azt, hogy milyen argumentumokat követel, illetve azokat milyen vonzatként kéri megvalósítani (akár több lehetséges változatban). Ezek mellett fontos tényező még a szórend kialakulása, amely-

¹ Ezúton szeretnénk köszönetet mondani az OTKA-nak (OTKA K60595) a támogatásért.

ről az általunk használt nyelvtan-modellben rangparaméterek adnak számot. A rangparaméterek egy „speciális” megjelenése számot tud adni az olyan esetekről is, amikor a mondatban a semleges szórendet megvariáló (az írott magyar mondatokban másképpen nem észrevehető) fókusz vagy egyéb operátor van jelen.

A lexikai egységek által követelhető (követelt) argumentumstruktúrákat² egy egyednek tekintjük, ennek elemeit egy olyan készlet adja, mely megmondja, milyen típusú argumentumok fordulhatnak elő. Ezek jellemzését egy olyan számparaméter adja, amely az ágens-pacientív skálán helyezi el az argumentumot.³ Az argumentumok így egy argumentumstruktúra, valamint az argumentumtípus kapcsolatából képződnek.

A másik oldalt a hasonlóan felépített vonzatkeretek jelenik. Egy egyednek képeznek a lexikai egységek által követelhető (követelt) „vonzatkeretek”: ezt kétirányúan tároljuk, a régens és a vonzat felől is⁴, elemei pedig az ezek típusait definiáló készletből származnak. Ez a készlet nemcsak esetragokat tartalmaz, hanem az olyan alakzatokat, mint a névutók vagy az infinitívusz, illetve az olyan állandósult kifejezéseként használt szerkezetek, amelyekre egy-egy esetragos alak cserélhető. Például: *Péter elárul pár dolgot Mariról / Marival kapcsolatban*. Az egyes vonzatokat tehát a vonzatkeret és egy vonzattípus kapcsolatai jelentik.

Előfordulnak olyan esetek is, amikor a lexikai egység nem ad számot az argumentum esetéről. Például a *lakik* ige esetén két argumentum szerepel: egy *lakó* és egy *hely*, ahol ő lakik. A vonzatok között szintén kettőt találunk egy NOM esetet, amit a lakó számára tartogat a lexikai egység (amit a vonzat és az argumentum összekötése ad meg; „bármire illeszhető” típusokkal is dolgozva), és egy másik vonzatot, amelynek az esete nem specifikált – ennek ellenére hiánya agrammatikus mondatot eredményez. Erre az esetre az argumentumtípusok és a vonzattípusok is összeköthetőek: a példában az argumentum HELY típusához így szelektálhatók az olyan vonzattípusok, mint a -bAn (*egy szép házban*), az -On (*Szentesen*) vagy a mellett (*az iskola mellett*).

A szintaxisnak számot kell még adnia a szabad határozókról. Szabad határozónak azt tekintjük, ahol a rag kompozicionálisan szerepel, csak úgy mint [8], ám vonzatnak tekintjük az olyan kompozicionális elemeket, amelyeknek a jelenlétét valamelyik másik elem előírja. Ekkor a rag (vagy a lexikai egység, pl. *ott*) árulja el magáról, hogy ő az általa keresett főnévvel együtt szabad határozóként jelenhet meg. A szemantikai struktúra fog a lexikai egységeknél olyan tulajdonságokat megadni, ami alapján például a különböző típusú időhatározók közül a grammatikusakat ki lehet választani. Modellünkben ezek az elemek a mondat eventuális argumentumát keresik.

² Több argumentumkeretnek tekintjük azt a névszókra jellemző jelenséget, hogy lehetnek ők maguk vonzatok, illetve lehetnek névszói állítmányként predikátumok is. Ezek mellé kell még számolnunk a rövid birtokos formát is, ami keres egy birtokos személyjelet. (Pl. *Péter Budapesten lakik. Annak a fiúnak a neve Péter. Péter kalapja*)

³ Az ilyen szempontból semleges argumentumok semleges számparamétert kapnak, mivel a centrális keretben – amely az alanyok és tárgyak viszonyainak alakulását írja le – nincsenek benne.

⁴ A szabad bővítmények csak egyirányúak

Rangparaméterek

A lexikai egységek a követelményeik mellett azt is megmondják, mennyire szükséges azt kielégíteni, és a szomszédosságra vonatkozó követelmény esetén azt is, hogy az azt kielégítő elemet milyen irányban, illetve hogy a szón belül vagy egy másik szóban keressük. Ennek okát az adja, hogy ami szemantikai kapcsolatban áll a régenssel, az annak szomszédja szeretne lenni. Mivel azonban a mondatok lineárisak, a régensnek mindössze két szomszédja lehet: ám sokkal inkább csak egy, hiszen a régensek a nyelvre jellemzően egy irányból veszik fel a vonzataikat⁵. A rangparaméterek a nyelv leírásából derülnek ki, tapasztalat útján; tárolásuk az egyes vonzatoknál történik. Minden nyelv régensei ugyanazt a követelményt támasztják (ezért is egy egyed – egy tábla), ám különféle erősségű rangokkal (amelyek így a nyelvek közötti különbségekről is számot adnak, lásd pl. *scrambling* jelenségek).

A rangparaméterek kétféle kielégíthetőségi követelménnyel léphetnek fel. A *domináns* rangparaméterek esetében elegendő, ha egy ellentmondó, erősebb paraméter ki van elégítve. A szomszédossági viszonyokat megadó *recesszív* paraméterek esetén a követelményt mindenképpen ki kell elégíteni – akár részlegesen is [3].

Az igekötők esetében például kétféle rangparaméter is él, mindkettő recesszív: Az egyik esetben az *el* igekötő az *indul* elé kell, hogy kerüljön, ezt egy erős (-2) rangparaméter mondja meg. Fonológiai szempontból a hangsúly is az igekötőre esik, míg az igéről lekerül, így egy fonológiai szót alkotnak; ez az aspektualizáló argumentum⁶ [2]. A második esetben az igekötő kerülhet az ige mögé is egy gyengébb (+3) rangparaméter szerint, ekkor mindkét elem külön fonológiai szóként realizálódik.

Semleges mondatban az erősebb (-2) rang elégül ki, míg egy fókusz megjelenése esetén, ami erős domináns típusú rangparamétert jelent⁷, az igekötő igét megelőző helyére vonatkozó követelményt nem lehet, és nem is kell kielégíteni. A gyengébb (+3) rangparaméter követelménye viszont továbbra is jelen van, így azt ki lehet és kell elégíteni.

Nemcsak az igekötő lehet aspektualizáló: hasonló a helyzet a már említett *lakik* igénél: a helyre vonatkozó vonzatát – pl. *Budapesten* – az erős (-2) ranggal szeretné maga előtt tudni, vagy egy gyengébb (+3) ranggal maga mögött. Ez speciális eset, ahol az aspektualizáló argumentum a lokatívusz. Bizonyos esetekben igekötős ige esetén is ez az aspektualizáló vonzat: a *megszáll* igével semleges esetben a *Péter Budapesten szállt meg* a grammatikus mondat, míg semmiképpen nem semleges a *Péter megszállt Budapesten*. Az aspektualizáló argumentumot mindig két ranggal vesszük fel.

A fókusz, ami ugyan (a magyarban) hangalakot nem ölt, lexikai egységként jelenik

⁵ Az irányultságot a szám 0-hoz viszonyított iránya fejezi ki, vagyis negatív számok jelentik a megelőző elemként való keresést, és pozitív számok a következő elemre vonatkozó keresést. Az erősséget a paramétert jelentő szám abszolút értéke adja.

⁶ Mindig van egy olyan vonzat, amit aspektualizálásra használ az ige (Rendszerint az igekötőt, alkalmanként saját magát, pl. *Péter csalódik Mariban*).

⁷ Amit először ki kell elégíteni, hogy ez után kiirtódhassanak a dominánssal ellentétes követelmények; és ez után következnek a recesszívek.

meg az adatbázisban. A mondatban két elemet keres: a fókuszált elemet és az igét. Az igének olyan rangparamétert oszt, hogy annak mögötte kell lennie, a legerősebb (+1) domináns rangban. Hasonlóan kezelhető a telikus szituációkra vonatkozó progresszív alak működése is.

Megfigyelhetünk olyan elemeket is, amelyek jelenlétében mindig található fókusz is, ilyen például a *csak*, illetve a mondatrésztágadást jelző *nem*. Ezen elemek erős domináns ranggal vonzzák a fókusz maguk mögé, ami a szórendi változásról gondoskodik.

Összegzés

Rendszerünk előnye és újszerűsége abban rejlik, hogy a lexikonban ugyanazon kerekék között kezelhető a szintaxis körébe (is) tartozó számos tényező (predikátum-argumentum, illetve régens-vonzat viszonyok, szabad határozók, szórend).

A rangparaméterek elegánsan számot adnak az egy nyelven belüli szórendi variációkról (a szón belül a morfémák sorrendjéről), valamint a nyelvek közötti különbségekről is. A domináns rangparaméterek használatával a szórendet megvariálók, sokszor láthatatlan elemek (fókusz, progresszív) is kezelhetőek.

Projektünk jelenleg a fent vázolt rendszer megvalósításán dolgozik. A következő lépés a fő komponensként elképzelt szemantika implementálása lesz, mert hisszük, hogy intelligens alkalmazások csak valós nyelvészeti alapokon készíthetők, ahhoz pedig alaposan kidolgozott szemantikai rendszerre van szükség.

Hivatkozások

- [1] Alberti G.: GASG: The Grammar of Total Lexicalism; IN: Working Papers in the Theory of Grammar 6/1. Theoretical Linguistics Programme, Budapest University and Research Institute for Linguistics, Hungarian Academy of Sciences, 1999
- [2] Alberti G.: Az aspektus szintaxisa a magyarban; IN: Újabb tanulmányok a strukturális magyar nyelvtan és a nyelvtörténet köréből, Osiris, Budapest, 2001; pp.145-164.
- [3] Alberti G., Balogh K., Kleiber J., Viszket A.: A totális lexikalizmus elve és a GASG nyelvtan-modell; IN: Maleczki Márta (szerk.): A mai magyar nyelv leírásának újabb módszerei V. SZTE, Szeged, 2002; pp. 193-218
- [4] Alberti G., Dóla M., Farkas J., Kántor Gy., Kleiber J., Ohnmacht M. (2007): ReALIS: a "reális" interpretációs rendszer; IN: Alberti G. és Fóris Á. (szerk.) A mai magyar nyelvtudomány formális műhelyei, Nemzeti Tankönyvkiadó, Budapest, pp. 139-156.
- [5] Alberti G., Kleiber J., Viszket A.: GeLexi projekt: Gépi fordítás totálisan lexikalista alapon; : Dr. Alexin Z., Csendes D. (szerk.): MSZNY2004 - II. Magyar Számítógépes Nyelvészeti Konferencia; Informatikai Tanszékcsoport, Szeged, 2004; pp.73-80
- [6] Bódis Z., Kleiber J., Szilágyi É., Viszket A.: Lile projekt: Adatbázis mint "dinamikus korpusz"; IN: Dr. Alexin Z., Csendes D. (szerk.): MSZNY2004 - II. Magyar Számítógépes Nyelvészeti Konferencia; SZTE Informatikai Tanszékcsoport, Szeged; pp.11-18
- [7] Halassy B.: Az adatbázis-tervezés alapjai és titkai. Avagy az út az adattól az adatbázison át az információig; IDG, Budapest 1994
- [8] Gábor K. - Héja E.: Predikátumok és szabad határozók, IN: Kálmán László (szerk.) KB 120 A titkos kötet, MTA Nyelvtudományi Intézet Tinta Könyvkiadó, Bp., 2006; pp.134-152.