

Főnevek szemantikai jegyei és kódolásuk a MetaMorpho projektben

Orosz Kata

MorphoLogic Kft., 1126 Budapest Orbánhegyi út 5.
MTA Nyelvtudományi Intézet – Korpusznyelvészeti Osztály
oroszk@morphologic.hu

Kivonat: A cikk a MetaMorpho gépi fordító rendszer magyar-angol változatában a lexikonban található főnevek szemantikai jegyeit, és a jegyek kódolásának folyamatát mutatja be. A cikk első fele a főnévi szemantikai jegyeket osztályozza, valamint tárgyalja, hogy mi a funkciója ezeknek a jegyeknek a MetaMorpho projektben. Ezután a jegyek definiálásával kapcsolatos problémákról, illetve a jegyek egymáshoz való viszonyának meghatározásáról esik szó. A dolgozat a jegykódolási folyamat bemutatásával, a kódolás során felmerülő, részben elméleti, részben gyakorlati problémák illusztrálásával, és az eddig elért eredmények összegzésével zárul.

1 Szemantikai jegyek a MetaMorpho projektben

A MetaMorpho projekt⁶⁴ keretében fejlesztés alatt álló gépi fordító rendszer kétféle szemantikai jegyet használ: igei és főnévi jegyeket. (Az igeik szemantikai jegyeiről ebben a dolgozatban nem esik szó.) A főnévi jegyek is két csoportba oszthatók: az igei vonzatkeret-leírásoknál használt jegyekre, és olyan jegyekre, amelyeket az igei vonzatkeret-leírások nem, csak a vonzatkereteket működtető szintaxis [1] használ.

1.1 A szintaxis által használt szemantikai jegyek

Az igei vonzatkereteket működtető szintaxis két főnévi szemantikai jegyet használ: az **enent** és a **gender** jegyeket. Az **enent** nevű jegy a magyar főnevek angol fordításának megszámlálhatóságáról hordoz információt. Ez a jegy az angol fordítás generálásánál játszik szerepet, amennyiben ez alapján rendeli hozzá a szintaxis az angol főnévhez a megfelelő kvantort (pl. *many* vagy *much*). A jegy kétféle értéket vehet fel: [YES] (megszámlálható) és [NO] (megszámlálhatatlan). A **gender** jegy szintén a generálásnál játszik szerepet; az angol főnevekhez tartozó, az adott szó grammatikai nemének megfelelő személyes névmás kiválasztását teszi lehetővé. Ez a jegy az angol

⁶⁴ A MetaMorpho projektről bővebben ld. Tihanyi László: A MetaMorpho projekt története. *I. MSzNy*, Szeged (2003), A MetaMorpho projekt 2004-ben. *II. MSzNy*, Szeged (2004), A MetaMorpho fordítóprogram projekt 2005-ben. *III. MSzNy*, Szeged (2005).

nyelv grammatikai nemeinek megfelelően háromféle értéket vehet fel: [M] (hímnem), [F] (nőnem), [N] (semlegesnem).

1.2 Az igei vonzatkeret-leírásoknál használt szemantikai jegyek

Az igei vonzatkeretek leírásánál 11 szemantikai jegyet használunk. Ezek (ábécérendben): **abstract**, **animate**, **bodypart**, **dynamic**, **company**, **currency**, **human**, **mass**, **measure**, **time**, **weather**. Ezeknek a jegyeknek az a funkciója⁶⁵, hogy a fordítandó, magyar nyelvű mondatok elemzése során segítsék a megfelelő igei vonzatkeret kiválasztását, és ezzel csökkentsék a nem megfelelő elemzések számát. (Ebből következik, hogy a kétféle – azaz az igei vonzatkeret-leírásoknál használt, illetve a szintaxis által használt – főnévi szemantikai jegycsoport között az az alapvető különbség, hogy az egyik csoportba tartozó jegyek **/encnt**, **gender**/ a generálásban játszanak szerepet, míg a többi 11 jegy az elemzést támogatja.)

Az igei vonzatkeret-leírásoknál használt szemantikai jegyek háromféle értéket vehetnek fel: [YES], [NO] és [NIL]. A főnevek szemantikai jegyének értéke határozza meg, hogy az adott főnév milyen vonzatkereteket képes kiválasztani. Az adott jegyre nézve [YES] vagy [NO] értékkel rendelkező főnevek képesek olyan igei vonzatkeretek kiválasztására, amelyek az adott jegyre nézve a főnévével megegyező értékű szemantikai megkötést tartalmaznak. (A kiválasztás természetesen csak abban az esetben történik meg, ha az igei vonzatkeret-leírás által megkövetelt jegyre nézve [YES] vagy [NO] értékkel rendelkező főnév a fordítandó mondatban a potenciálisan választható igei vonzatkeret meghatározott vonzataként szerepel.) Például:

(1)

***VP=fel|ad**

HU.VP = SUBJ + TV(:lex="fel|ad") + **OBJ**(pos=N, case=ACC, **abstract=NO**, human=NO)

EN.VP = SUBJ + TV[lex="post"] + OBJ

Vmi felad valamit. (János feldta a levelet.)

(2)

***VP=fel|ad**

HU.VP = SUBJ + TV(:lex="fel|ad") + **OBJ**(pos=N, case=ACC, **abstract=YES**)

EN.VP = SUBJ + TV[lex="give"] + PART[lex="up"] + OBJ

Vmi felad abstractvmit. (János feladta a hobbiját.)

⁶⁵ A MetaMorpho projektben használt szemantikai jegyek funkciójának definíciója, és az igei vonzatkeretek kiválasztásában játszott szerepük elméleti és technikai kidolgozása Merényi Csaba érdeme.

A fenti példákban a HU.VP-vel kezdődő sorok az adott igei vonzatkeret-leíráshoz tartozó, mmd-formalizmussal [1] kódolt elemzősorok; az EN.VP-vel kezdődő sorok a generálósorok; a dőlt betűvel szedett sorok a vonzatkeret-leírások értelmezését segítő példamondatok (egy általános és egy konkrét példamondat). A főnévi lexikonban a *csomag* [NO], a *remény* [YES] értékkel rendelkezik az **abstract** jegyre nézve. Tegyük fel, hogy a gépi fordítónak az alábbi mondatok fordításához kell kiválasztania a megfelelő igei vonzatkereteket:

- (4) Tibi feladta a csomagot.
 (5) Tibi feladta a reményt.

A *csomag* és a *remény* is tárgyi vonzata (OBJ) a *felad* igének, tehát elméletileg mindkét főnév választhatná mindkét igei vonzatkeretet, így a következő fordítások születnének:

- (4/1) Tibi posted the package.
 (4/2) *Tibi gave up the package.
 (5/1) *Tibi posted hope.
 (5/2) Tibi gave up hope.

Azonban az (1) és a (2) példában a tárgyi vonzat szerepét betöltő főnévre nézve a leírások tartalmaznak egy szemantikai megkötést: az (1) példában csak [abstract=NO] jeggyel rendelkező főnév, a (2) példában csak [abstract=YES] jeggyel rendelkező főnév választhatja ki magának az adott vonzatkeretet. Tehát amikor a gépi fordítónak a *Tibi feladta a csomagot.* mondathoz kell kiválasztania a megfelelő vonzatkeretet, az elemzés során észleli, hogy a tárgyi vonzat szerepét betöltő főnév az [abstract=NO] jeggyel rendelkezik, és ezért az (1) példában szereplő vonzatkeretet választja ki; így csak a helyes fordítást generálja (*Tibi posted the package.*)

Egy főnév akkor vesz fel [NIL] értéket, ha adott szóalaknak poliszémia vagy homonímia következtében több jelentése (és esetleg, de nem szükségszerűen több fordítása) lehetséges. Például:

- (6) Az anyák kiborultak.

Az *anya* szónak a (6) mondatban két jelentése is lehet: édesanya (**anya1**, angolul *mother*), és anyacsavar (**anya2**, angolul *screw nut*). Attól függően, hogy melyik értelemben szerepel az *anya* szó, a *kiborul* igének is két különböző jelentése – és ebben az esetben két különböző fordítása – lehetséges: leesik, szétszóródik (**kiborul1**, angolul *spill*) és ideges lesz (**kiborul2**, angolul *become upset*). A **kiborul1** igei vonzatkeretben az alanyi vonzaton az [animate=NO], a **kiborul2**-ben az [animate=YES, human=YES] szemantikai megkötés szerepel. Az *anya* szó a lexikonban (erős egyszerűsítéssel) a következőképpen fog szerepelni:

- (7)

***NX=anya**

HU.NX (animate=NIL) = **N**(lex="anya")

EN.NX (animate=YES, human=YES) = **N[lex="mother"]**

EN.NX (animate=NO) = **N#1[lex="screw"] + N[lex="nut"]**

Amikor a gépi fordító elemzi a (6) mondatot, látja, hogy az alany szerepét betöltő főnév [NIL]-es értékű, ezért mindkét (**kiborul1** és **kiborul2**) vonzatkeretet kiválasztja. Azonban nem mind a négy lehetséges fordítást adja meg (köztük két helytelen fordítással, nevezetesen: **The mothers spilled.* és **The screw nuts became very upset.*), hanem a főnevek megfelelő fordítását a több lehetséges angol generációsor közül (**EN.NX**) már a vonzatkeret-leírások megkötéseinek megfelelően választja ki, és ezeknek megfelelően csak két fordítást generál: *The mothers became upset.* és *The screw nuts spilled.*

A főnevek szemantikai jegyei akkor tudják igazán növelni az elemzés hatékonyságát (t.i. akkor tudják kiválasztani a megfelelő igei vonzatkeretet), ha a vonzatkeretek leírásánál használt szemantikai megkötések, és a lexikon főnevein kódolt szemantikai jegyek összhangban vannak egymással. Az összehangoláshoz pedig szükség van a jegyek minél pontosabb definiálására.

1.3 A szemantikai jegyek definiálásának problémája

A szemantikai jegyek definiálásánál több probléma is felmerült. Az egyik ilyen dilemma az, hogy a gépi fordító hatékonysága szempontjából nagyságrendileg hány jegyet célszerű bevezetni. A másik, hogy pontosan mik legyenek azok a jegyek, amikre az igei vonzatkeretek hivatkoznak.

A második probléma megoldása statisztikai alapon történt. Az igei vonzatkeretek megírásával foglalkozó munkatársak⁶⁶ lettek felkérve, hogy még a pontos jegydefiniciók megalkotása előtt írjanak össze olyan, intuitív (ontológiai) definíciókon alapuló jegyeket, amiket alkalmasnak találnak azonos az azonos alakú, de eltérő jelentésű és/vagy eltérő vonzatkerettel rendelkező igék szemantikai alapú elkülönítésére. A gépi fordító angol-magyar változatában még csak két, intuitív definíción alapuló jegy játszott kitüntetett szerepet: az **animtype** (élő) és **humantype** (akaratlagos cselekvésre képes) jegyek. A magyar-angol változat készítésekor ez a két jegy **animate** és **human** néven tovább élt, azonban további jegyek bevezetésére mutatkozott igény. A vonzatkeret-leírások írói által javasolt jegyek közül végül 9 – **abstract** (elvont fogalom), **bodypart** (testrész), **dynamic** (valamilyen „működésre” képes gép), **company** (intézmény), **currency** (pénznem), **mass** (anyagnév), **measure** (mértékegység), **time** (idő-mértékegység), **weather** (időjárással kapcsolatos jelenség) – került azon az alapon elfogadásra, hogy az igék feldolgozásának akkori fázisában hány igei vonzatkeret-leírásnál érezték szükségesnek az adott jegyre való hivatkozást. Az **1. táblázat** azt mutatja, hogy 2006 júliusában az addig megírt közel 20000 igei vonzatkeret közül hány hivatkozott az adott jegyre.

⁶⁶ Az igei vonzatkeretek megírásával a Nyelvtudományi Intézet munkatársai foglalkoznak. (A MetaMorpho projekt a MorphoLogic Kft., a Nyelvtudományi Intézet és a Szegedi Egyetem alkotta konzorcium keretei között folyik.)

1. táblázat:

Szemantikai jegyekre való hivatkozás gyakorisága az igei vonzatkeret-leírásokban

A jegy neve	Hivatkozások száma
HUMAN	8682
ABSTRACT	3469
ANIMATE	2836
BODYPART	303
MEASURE	150
TIME	99
DYNAMIC	70
MASS	61
COMPANY	9
CURRENCY	6
WEATHER	1

A hivatkozások gyakorisága mellett fontos szempont volt a hivatkozást tartalmazó igei vonzatkeretek korpuszbeli gyakorisága is – így kerülhetett be a jegyek közé a **company**, a **currency** és a **weather**. Ezek ugyan kevés hivatkozást kaptak, ám az a néhány vonzatkeret, ami mégis hivatkozott rájuk, olyan gyakran használatos a magyar nyelvben, hogy megtartásuk mégis célszerűnek tűnt.

A jegyek listájának megalkotása többé-kevésbé az első kérdésre is megadta a választ; nevezetesen, hogy hány főnévi szemantikai jegyet érdemes bevezetni a gépi fordító hatékonyságának szempontjából. A MetaMorpho projektben érvényesített, az 1.2 pontban vázolt jegyfelfogás nem teszi célszerűvé a jelenleginél nagyságrendileg több főnévi szemantikai jegy bevezetését, hiszen az igei vonzatkeret-leírások szemantikai alapú elkülönítés már a fent említett 11 jeggyel is kivitelezhető. Ezért még ha a projekt folyamán a jegyek listája részben változik is, a jegyek számának radikális növekedése kevésbé valószínű.

A jegyek listájának ilyen módon történő összeállítása egy gyakorlati hátránnyal járt: a jegyek „definíciói” egy kisszámú embercsoport nyelvi intuícióján alapultak, ezért a lexikon főnévi elemeinek kódolásához is csak intuitív támpontot adtak. Ez azonban már átvezet a jegykódolás során felmerült problémákhoz.

2. A jegyek kódolása a magyar-angol lexikonban

A magyar-angol irányú fordítóprogram lexikonába két lépésben kerültek fel a jegyek. Első lépésben a lexikon kb. 72.500 főnévének körülbelül 25%-a lett részben (azaz csak néhány, és nem az összes jegyre) kódolva. Ez nagyrészt az angol-magyar lexikonban **animtype** és **humantype** jeggyel rendelkező szavak magyar megfelelőinek az **animate** illetve a **human** jeggyel való automatikus – számítógépes módszerekkel történő – felruházását jelentette. Szintén az első fázisban, manuális módszerekkel, és lényegében az intuitív jegydefiníciók alapján további 5000 főnév is részleges kódolásra került – ezúttal már a 9 új jegyet is felhasználva.

A második fázisban szükségesnek mutatkozott, hogy a (lehetőségekhez mérten) pontosan definiáljuk, milyen típusú főnevek tartozzanak az adott jegyekhez. A jegyek definiálása végül a WordNet ontológia segítségével, az intuitív jegydefiníciók figyelembevételével történt.

2.1 A jegyek WordNet-alapú definiálása

A WordNet ontológiának mind az angol (Princeton WordNet), mind a jelenleg is fejlesztés alatt álló magyar verzióját [2] felhasználtuk a jegyek definícióinak megalkotásához. A WordNetben olyan csomópontokat kerestünk, melyek az adott jegyhez intuitíve tartozó szavak átfogó kategóriáiként szolgálhattak. Egy jegy jellemzően több, egymással alá-fölérendeltségi viszonyban nem álló csomópontból állt össze. (Néhány jegy WordNet-alapú definícióját a **2. táblázat** tartalmazza.)

A WordNet-alapú jegydefiníálás két szempontból is hasznosnak bizonyult. Egyrészt a WordNet-csomópontokban megadott definíciók alapján egy intuitíve nehezen kódolható szóról a WordNet ontológia segítségével gyorsan és egyszerűen eldönthető, hogy melyik jegy(ek)hez tartozik. A gyakorlatban pl. a VisDic program [2] segítségével ellenőrizhető, hogy adott szó mely WordNet-es csomópontokon „halad át” – ha olyan csomópont alá tartozik, ami egy, a MetaMorpho projektben használt szemantikai jegy definícióját adja, akkor a szó az adott jegyre nézve [YES] értékkel kódolandó.

2. táblázat:

Példa a MetaMorpho projekt szemantikai jegyeinek WordNet-alapú definíciójára

A jegy neve	WordNet Synonym	WordNet ID
ABSTRACT	psychological feature:1; abstraction:6	ENG20-00020333-n; ENG20-00020486-n
ANIMATE	living thing:1, animate thing:1; cell:2	ENG20-00003009-n; ENG20-00004824-n
BODYPART	body part:1	ENG20-04919813-n
COMPANY	organization:1, organisation:3	ENG20-07523126-n
CURRENCY	medium of exchange:1, monetary system:1; monetary unit:1	ENG20-12615184-n; ENG20-12627781-n

A WordNet-alapú jegydefiníciók másik haszna, hogy segítségükkel minden olyan főnévről, ami a WordNet ontológiában és a MetaMorpho projekt magyar-angol főnévi lexikonában *is* szerepel, automatikusan – gépi heurisztikák alkalmazásával – megállapítható, hogy milyen jegyekre nézve kell [YES]-re kódolni; azaz listákat lehet készíteni azokról a szavakról, amik egy adott jegyre nézve [YES] értékkel rendelkeznek. A szavak listázásnál a számítógépes ontológiákban – így a WordNet-ben is – alkalmazott hipernímia-hiponímia relációt [2], [3] használtuk ki.

A jegykódolásnak ebben a fázisában a magyar WordNet főnévi állománya még viszonylag kicsi volt; ezért döntöttünk úgy, hogy a lényegesen nagyobb szóállomány-

nyal rendelkező angol WordNetből is megpróbálunk szemantikai információt kinyerni, és az angol szólisták automatikus fordításával növeljük az adott jegyre nézve [YES] értékre kódolandó magyar szavak listáját. Ez természetesen problémát okozott a poliszém és a homonim szavaknál, ezért az így nyert listákat minden esetben manuálisan kellett korrigálni.

Az angol WordNet alapján készült listák esetében a poliszémia és a homonímia vezetett nem megfelelő szavak listázásához. Például az angol *bay* szó öblöt és pejlovat is jelent. Utóbbi jelentése az angol WordNet-ben az **animate** definíciója alá tartozik, a másik nem. Amikor viszont a WordNet-ből kinyert angol szavak listája gépi módszerekkel össze lett vetve az angol szavak magyar fordításával, nem csak a *pej ló* összetett főnév került az [animate=YES] értékkel bíró szavak listájára, hanem az *öböl* szó is (hiszen a gépi illesztés csak a szóalak azonosságát vette figyelembe).

Szintén manuálisan kellett eltávolítani a listákról azokat a szavakat, amik a WordNet-definíciók és az intuitív definíciók közötti apró eltérések következtében kerültek fel tévesen a listákra. Például a kitalált személyek a WordNet-ontológiákban az **abstract** jegy alá tartoznak, az igei vonzatkeretek szempontjából viszont Pán Péter épp úgy viselkedik, mint Kovács Béla, hiszen mindketten akaratlagos cselekvésre képes lények, azaz inkább a **human** jegyhez tartoznak.

A poliszémiából, homonímiából, és a definíciós különbségekből adódó hibaarány még így is 15% alatt maradt, ezért a manuális korrekció szükségessége ellenére is elmondható, hogy ez az eljárás a kódolási folyamatot nagy mértékben felgyorsította. A **3. táblázat** azt mutatja, hogy hány, a MetaMorpho projekt főnévi lexikonában szereplő szóról tudtuk megállapítani a WordNet ontológiák segítségével, hogy az adott jegyre nézve [YES] értékkel rendelkezik.

3. táblázat:

A WordNet alapján készült listák szavainak száma, jegyekre lebontva

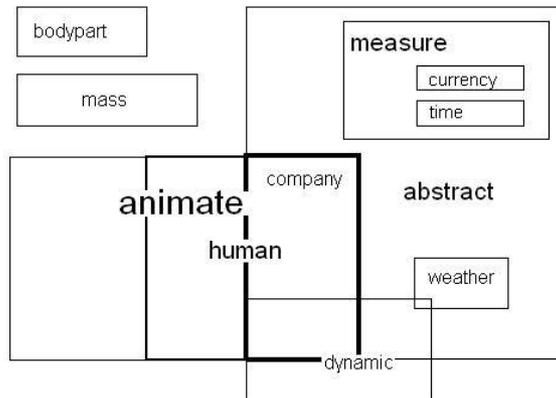
A jegy neve	Angol WordNet-ből kinyert szavak száma	Magyar WordNet-ből kinyert szavak száma
ABSTRACT	6688	372
ANIMATE	3067	121
BODYPART	663	67
COMPANY	514	42
CURRENCY	90	7
DYNAMIC	4	168
HUMAN	1823	36
MASS	3322	223
MEASURE	748	63
TIME	407	35
WEATHER	88	6
	ÖSSZESEN:	18554 szó

Természetesen ez a szám (18554) nem azt jelenti, hogy ennyi szó teljes kódolása megoldódott volna. A szólisták csupán annyit árulnak el, hogy a rajtuk szereplő szavak biztosan [YES] értékűek az adott jegyre nézve; a többi jegy tekintetében viszont

nem hordoznak információt. Pontosabban nem mindegyikük, hiszen ösztönösen is érezhető, hogy a jegyek között átfedés van. A jegyek egymáshoz való viszonyának megállapítása azért vált szükségessé, hogy leírhatóak legyenek az automatikus jegyöröklődések, és ezáltal a jegykódolás folyamata még gyorsabbá és hatékonyabbá váljon.

2.2 A jegyek egymáshoz való viszonyának megállapítása

A jegyek közötti kapcsolatokat a Cruse-féle halmazelméleti szemantikai modellt alkalmazva, az *átfedés*, *inklúzió*, *különbözőség* kategóriái [3] mentén fogalmaztuk meg. A jegyek közti kapcsolat megállapításánál figyelembe vettük az igei vonzatke-
reték leírásánál már használt, intuitív jegydefiníciókat is. A **1. ábrán** látható a főne-
vek szemantikai jegyeinek egymáshoz való viszonya.



1. ábra: A MetaMorpho projektben használt szemantikai jegyek egymáshoz való viszonya.

A jegyek inklúziós kapcsolatából az alábbi automatikus jegyöröklések következ-
nek:

- | | | |
|---------------------|------------------|------------------|
| (8) [human=YES] | → [animate=YES] | |
| (9) [company=YES] | → [human=YES] | → [animate=YES] |
| (10) [company=YES] | → [abstract=YES] | |
| (11) [weather=YES] | → [abstract=YES] | |
| (12) [measure=YES] | → [abstract=YES] | |
| (13) [currency=YES] | → [measure=YES] | → [abstract=YES] |
| (14) [time=YES] | → [measure=YES] | → [abstract=YES] |

2.3 A jegyek kódolása: eredmények, további feladatok

A MetaMorpho projekt magyar-angol gépi fordító rendszerének fejlesztése során a szemantikai jegyek kódolása jelenleg harmadik fázisánál tart. Az első fázisban az angol-magyar változatban használt két jegyet alapul véve a magyar-angol változat főnévi lexikonában közel 18000 szó lett kódolva az **animate** és **human** jegyekre; valamint manuális módszerekkel további 5000 szó lett kódolva az időközben bevezetett 9 jegy valamelyikére. A második fázisban a WordNet ontológiák segítségével további 18500 jegyérték került azonosításra; ez az automatikus jegyörökléseket figyelembe véve mintegy 14700 szót érint. Mivel a WordNet alapján készített listákra már csak olyan szavak kerültek fel, melyek semmilyen jegyre nézve nem tartalmazta korábbi információt a MetaMorpho főnévi lexikonában, némi kerekítéssel azt mondhatjuk, hogy a főnevek több, mint 50%-áról (37700 szó) van már valamilyen (kódolt) szemantikai információnk.

Sajnos ez nem azt jelenti, hogy ezzel a kb. 37700 szóval már ne kellene foglalkozni, hiszen a jegyöröklődések csak egy irányban működnek; azaz azoknál a szavaknál, amelyekről csak az **animate** vagy **abstract** jegyek vonatkozásában vannak információink, manuálisan kell egyértelműsíteni, hogy milyen értékkel rendelkeznek a **human**, **company**, **measure**, **time** és **currency** jegyek tekintetében. Ilyen szóból pedig közel 33500 van. Ami azt jelenti, hogy jelenleg a teljes főnévi állományból csak 4200 szónak rendelkezünk a teljes szemantikai leírásával (vagyis adott szóról mind a 11 jegy esetében tudjuk, hogy [YES] vagy [NO] értéket vesz-e fel) – ez a teljes szótári állománynak csupán 6%-a.

A kódolási folyamat második szakaszának fontos eredményeként könyvelhető el ugyanakkor, hogy a szemantikai jegyek WordNet-alapú definiálásával, illetve a jegyek egymáshoz való viszonyának feltérképezésével létrehoztunk egy olyan elméleti keretet, ami a gyakorlatban is megkönnyíti a kódolási folyamatot. A WordNet-ontológiákhoz készített gépi heurisztikák pedig a későbbiekben is felhasználhatóak; tehát a WordNet ontológiák állományának bővülésével újabb és újabb szemantikai információk nyerhetők ki, amik felhasználhatók a jegykódolási folyamatban.

A jegykódolás jelenleg is zajló, harmadik szakaszában a manuális jegykódolás jut főszerephez; először azokat a szavakat ellenőrizzük, amikről az első két kódolási fázis eredményeképpen már rendelkezünk valamilyen szemantikai információval⁶⁷; továbbá zajlik azoknak a szavaknak a (szintén manuális) kódolása, amelyek még teljesen „érintetlenek” ebből a szempontból.

A [NIL] értékű szavak feldolgozása is megkezdődött. A poliszém illetve homonim főnevekről a projekt egy korábbi szakaszában már készültek listák. Az elsődleges feladat ezeknek a listáknak a kibővítése; majd az egy szóalakhoz tartozó összes lehetséges jelentés, és a különféle jelentéseknek megfelelő fordítás megtalálása. Ezután a szójelentéseket egyesével kell kódolni a szemantikai jegyek szempontjából; ezek a szójelentés-sorok végül gépi módszerekkel lesznek „összeadva”, hogy az egyes szójelentések, fordításaik és szemantikai jegyeik a (7) példához hasonló lexikon-elemekké álljanak össze. Pl. az **ülés1** (*seat*) [abstract=NO] és az **ülés2** (*session*) [abstract=YES] sorok egyetlen lexikon-elemmé, az *ülés* [abstract=NIL]-lé fognak összevonódni.

⁶⁷ A jegykódolásnak ezt a részét a Szegedi Tudományegyetem Mesterséges Intelligencia Tanszék Kutatócsoportjának munkatársai végzik.

A jegykódolás utolsó szakaszában az addig külön listákon tárolt szemantikai információkat gépi módszerekkel, az mmd-formalizmusnak megfelelően kell kódolni a főnévi lexikonban.

Köszönetnyilvánítás

Szeretnék köszönetet mondani Kiss Gabriellának, Merényi Csabának, Miháltz Mártonnak és Tihanyi Lászlónak, hogy a projektben végzett munkámhoz segítséget nyújtottak, és jelen cikk megírásakor is hasznos tanácsokkal láttak el.

Bibliográfia

1. Merényi, Cs.: A MetaMorpho magyar-angol gépi fordító rendszer igei vonzatkereteit működtető nyelvtan. In: III. Magyar Számítógépes Nyelvészeti Konferencia, Szeged (2005) 108–115
2. Miháltz, M.: Magyar EuroWordNet projekt: bemutatás és helyzetjelentés. In: III. Magyar Számítógépes Nyelvészeti Konferencia, Szeged (2005) 68–78
3. Szakadát, I.: Réteges struktúrák, alaprelációk. In: III. Magyar Számítógépes Nyelvészeti Konferencia, Szeged (2005) 43–55