

## Jelentés-egyértelműsítés – egyértelmű jelentésítés?

Héja Enikő<sup>1</sup>, Kuti Judit<sup>1,2</sup>, Sass Bálint<sup>1</sup>

<sup>1</sup>MTA Nyelvtudományi Intézet, Nyelvtechnológiai Kutatócsoport  
1068, Bp., Benczúr u. 33.

<sup>2</sup>ELTE BTK, Nyelvtudományi Doktori Iskola, Germanisztika Alprogram  
{eheja, kutij, sass.balint}@nytud.hu

**Kivonat:** Az alábbiakban bemutatott esettanulmányunkban azt vizsgáljuk, hogy a magyar nyelvre létező, különböző típusú adatbázisok közül melyek mennyire alkalmas igei Wsd-célokra; emberi annotátorok között milyen mértékű egyértelműsítést lehet elérni. A vizsgált adatbázisok az introspektív, illetve disztribúciós alapon készülő jelentéstárak közötti spektrumot hivatottak képviselni. Az eredmények arra utalnak, hogy a magyarra még nem létezik olyan, jelenlegi formájában késznek tekinthető jelentéstár, amely alapján kapott IAA-érték megfelelő viszonyítási alapot képezhet gépi WSD számára. A jelenlegi adatbázisok további, WSD-orientált fejlesztést igényelnek.

### 1 A feladat

A nyelvtechnológia egyik központi feladata megfelelő jelentés-egyértelműsítő rendszerek kialakítása. A jelentés-egyértelműsítés (a továbbiakban az angol terminus rövidítését használva WSD) számos alkalmazás számára elengedhetetlen; a legfontosabbak ezek közül a gépi fordítás, az információkivonatolás, illetve az információkinyerés. A jelentés-egyértelműsítés feladatát általánosan két alapvető lépésre bontjuk: (1) valamilyen jelentéstár kiválasztása, illetve létrehozása (2) a jelentéstárban szereplő jelentések hozzárendelése a kívánt szóalakokhoz valamilyen algoritmus segítségével. A jelentés-egyértelműsítéssel foglalkozó kutatások általában az utóbbira helyezik a hangsúlyt: azt vizsgálják, hogy a már létező jelentéstárakat milyen algoritmusokkal lehetne a lehető legjobban jelentés-egyértelműsítésre használni (pl. Latent Semantic Analysis (LSA), Hyperspace Analogue to Language (HAL)).<sup>1</sup>

A jelentéstár kiválasztása, minőségének ellenőrzése ehhez képest minimális figyelmet kap. A legkülönbözőbb jelentés-egyértelműsítő részfeladatokhoz (célszó jelentés-egyértelműsítése, automatikus kulcsszókinyerés, szemantikai szerepek címkézése stb.) nagy százalékban a WordNet különböző verzióit használják jelentés-egyértelműsítésre, míg más jelentéstárak használata (pl. FrameNet, VerbNet) háttérbe szorul. (A Senseval versenyeken használt jelentés-egyértelműsített korpuszok több mint fele valamilyen WordNet-típussal lett annotálva). Köztudott ugyan, hogy a felhasznált adatbázisok jellemzően nem kifejezetten a jelentés-egyértelműsítés céljából

<sup>1</sup> Agirre és Edmonds *Word Sense Disambiguation* c. könyvének alcíme nem véletlenül "Algorithms and Applications"

készülnek, a jelentés-egyértelműsítéshez szükséges adatbázisok szerkesztési elveivel és ezek létrehozásának módszereivel jóval kevesebb cikk foglalkozik, mint az algoritmusok tárházával.

A jelentés-egyértelműsítés mint komplex feladat, valamint a jelentések egzakt meghatározása sem tekintendő tehát megoldottnak.<sup>2</sup> Ennek híján viszont jelentésmegkülönböztető adatbázisok előállításánál a fejlesztők sokszor elsősorban saját intuíciójukra vannak utalva.

Ebből fakadóan az enumeratív lexikonok WSD-feladatokra való alkalmassága is megkérdőjelezhető. Véronis *Sense tagging: Does it make sense?* c. cikkében [5] a jelentés-egyértelműsítés viszonylagos megoldatlanságáért az intuíción alapuló jelentéstárak és a szerkesztési mód miatt fellépő inkonzisztenciákat teszi felelőssé. A következetlenségek kikerülésének — szerinte — egyetlen módja, ha a lexikon létrehozása során, a jelentések elkülönítésénél elsősorban megfigyelhető disztribúciós jelenségekre támaszkodunk.

Az intuitív jelentésfogalom problematikus voltát, illetve az enumeratív lexikonok jelentés-egyértelműsítés céljára való alkalmatlanságát Véronis két kísérlettel próbálta meg alátámasztani. Az első kísérletben megmutatta, hogy az annotátorok közötti egyetértés már abban a kérdésben is alacsony (igék esetén 0.37), hogy egy szóalak egy- vagy többjelentésű-e. A második kísérletben 60 szó 3724 előfordulásához kellett hozzárendelni a kísérleti személyeknek a *Petit Larousse* értelmező szótárban felsorolt jelentések közül a kontextusnak megfelelőt. Az annotátorok között ebben a kísérletben is alacsony egyetértés volt (igék esetén 0.41), ami a feladat nehézségét támasztja alá.

## 2 A kísérlet

Esettanulmányunkban a fent említett második kísérletet végeztük el magyarra, kiegészítve azzal, hogy három különböző jelentéstár használatával nyert eredményeket hasonlítottuk össze. Kísérletünkben igei jelentésekkel foglalkoztunk. 15 ige 30 előfordulását annotáltattuk be 5-5 kísérleti személlyel a Magyar Értelmező Kéziszótár (a továbbiakban ÉKSz), a Magyar WordNet [3] (a továbbiakban HuWN) jelentéseivel, illetve az "Igei szerkezetek gyakorisági szótára" [4] (a továbbiakban ISZGYSZ) adatbázis igei szerkezeteivel. A kísérleti személyek az adatbázisokban megadott kategóriacimkék választásán kívül "nincs" és "nem tudom" választ is adhattak.

Utóbbi adattár automatikusan gyűjtött gyakori, különböző specifikusságú ige + főnévi csoport szerkezeteket tartalmaz a vonzatkeretektől a komplex igéken át a szólásokig (ige + esetragok / névutó + leggyakoribb lemmák). Az adatbázist előállító algoritmus szigorúan disztribúciós alapon gyűjt, és a bővítményi szavak eloszlásából képes megállapítani, hogy adott bővítmény kötött vagy szabad. Az adatbázis definíciókat nem tartalmaz, viszont minden szerkezethez ad korpuszból gyűjtött példákat,

2 Agirre és Edmonds így ír a célzott WSD-ről a *Word Sense Disambiguation* c. könyvük bevezetőjében: "... explicit WSD has not yet been convincingly demonstrated to have a significant positive effect on any application."

melyek a szerkezet jelentését hivatottak megvilágítani. A magyar igei WordNet egy, az angol nyelvű Princeton WordNet 2.0 verziójára épülő, de annak struktúrájához nem mereven ragaszkodó lexikális adatbázis, amelynek alapegysége a fogalom / jelentés, nem pedig a tradicionális szótárak alapegysége, a szó / lexéma. A magyar igei WordNet, amellyel hogy a PWN-ben tárolt szemantikai relációkat kódolja mintegy 3000 fogalom között, néhány új reláció bevezetésének segítségével igyekszik lehetővé tenni az igék aspektuális jellemzőinek kódolását is. Fontos továbbá, hogy a magyar igei WordNet készítésekor már vonzatkeretekre vonatkozó, automatikusan kinyert információkat is figyelembe vettünk. Ezért a WordNet-beli jelentésmegkülönböztetések egyrészt nem pusztán introspekción, másrészt nem az angol nyelvű PWN jelentésmegkülönböztetéseire alapulnak.<sup>3</sup> A HuWN ily módon módszertanilag az introspektív jelentésmegkülönböztetéseken alapuló ÉKSz., és a pusztán disztribúciós alapokon nyugvó ISZGYSZ között helyezkedik el.

Esettanulmányunkban azt vizsgáltuk, hogy a magyarban rendelkezésre álló jelentéstárak alapján milyen fokú egyetértést lehet emberi annotátorok között elérni, illetve hogy van-e az annotátorok közötti egyetértésben különbség az adatbázis fajtájától függően. Választ vártunk arra a kérdésre is, hogy milyen tulajdonságokkal kell rendelkeznie egy olyan jelentéstárnak, amelyet *kifejezetten* jelentés-egyértelműsítés céljából készítenek.

Az annotátorok közötti egyetértést (inter-annotator agreement — IAA) Fleiss-féle multi  $\pi$  érték szerint számoltuk, Artstein és Poesio (2008: 563-564) alapján [1]. A Cohen-féle  $\kappa$ -val szemben e mértéknek előnye, hogy képes elvonatkoztatni az egyes annotátorok esetlegességeitől. A Fleiss-féle multi  $\pi$  az összes annotátor adataiból becsült átlageloszlásból számolja a várható egyetértés mértékét, és azt mutatja meg, hogy a tapasztalt egyetértés hol helyezkedik el a várható egyetértés (0) és a teljes egyetértés (1) által meghatározott skálán. A mérték negatív értéket is felvehet, ha az egyetértés kisebb a véletlenszerűen elvárnál. Minél közelebb van tehát a kapott érték az 1-hez, annál nagyobb a valószínűsége, hogy az annotátorok közti egyetértés nem véletlen. A Fleiss-féle multi  $\pi$  érzéketlen olyan plusz kategóriákra, amelyeket soha egyetlen annotátor sem választott, azaz az eredményben nem jelenik meg, hogy hány kategóriából választhattak eredetileg az annotátorok.

### 3 Kiértékelés

Az alábbi táblázat az egyes adatbázisok szerinti IAA-értékeket mutatja az egyes igékre lebontva. A táblázatban szereplő eredmények számításakor a teljes értékű válasznak tekinthető "nincs" válaszokat önálló értékként kezeltük. Ugyanígy kezeltük a "nem tudom" válaszokat is, melyek előfordulási aránya mindössze 2-6% volt.

---

<sup>3</sup> Az adatbázis építésének főbb módszertani lépéseire ld. [3].

1. táblázat: A Fleiss-féle multi  $\pi$  mérték átlagolt értéke a három adatbázisra vonatkozóan.

|                                  | ÉKSz.        | HuWN  | ISZGYSZ | ÉKSz 2 (fő-jelentések) | választható jelentésszám<br>ÉKSz. / HuWN / ISZGYSZ // ÉKSz2 |
|----------------------------------|--------------|-------|---------|------------------------|---|
| emel                             | <b>0.450</b> | 0.753 | 0.170   | 0.848                  | <b>13 / 10 / 16 // 5</b>                                    |
| feltesz                          | <b>0.493</b> | 0.693 | 0.265   | 0.745                  | <b>14 / 7 / 8 // 7</b>                                      |
| fizet                            | <b>0.157</b> | 0.61  | 0.259   | 0.278                  | <b>12 / 1 / 23 // 5</b>                                     |
| használ                          | <b>0.210</b> | 0.954 | 0.336   | 0.611                  | <b>8 / 2 / 22 //</b>  |
| köt                              | <b>0.449</b> | 0.637 | 0.237   | 0.535                  | <b>29 / 21 / 19 // 12</b>                                   |
| lép                              | <b>0.346</b> | 0.595 | 0.443   | 0.601                  | <b>12 / 11 / 31 // 7</b>                                    |
| megold                           | <b>0.137</b> | 0.197 | 0.255   | 0.449                  | <b>6 / 2 / 12 // 4</b>                                      |
| mutat                            | <b>0.187</b> | 0.153 | 0.284   | 0.365                  | <b>13 / 4 / 27 //5</b>                                      |
| okoz                             | <b>0</b>     | 0.59  | 0.286   | 0                      | <b>2 / 3 / 26 // 2</b>                                      |
| rendelkezik                      | <b>0.195</b> | 0.469 | 0.471   | 0.474                  | <b>6 / 3 / 15 // 4</b>                                      |
| segít                            | <b>0.112</b> | 0.371 | 0.434   | 0.173                  | <b>7 / 4 / 19 // 5</b>                                      |
| szolgál                          | <b>0.279</b> | 0.516 | 0.548   | 0.509                  | <b>15 / 8 / 16 // 7</b>                                     |
| tárgyal                          | <b>0.840</b> | 0.543 | 0.407   | 0.840                  | <b>3 / 2 / 16</b>   |
| választ                          | <b>0.452</b> | 0.935 | 0.444   | 0.713                  | <b>6 / 2 / 24 // 4</b>                                      |
| vállal                           | <b>0.207</b> | 0.311 | 0.275   | 0.623                  | <b>6 / 3 / 26 // 3</b>                                      |
| átlagolt Fleiss-féle multi $\pi$ | 0.300        | 0.483 | 0.340   | 0.517                  |   |

A fenti adatok alapján a következő iránymutató következtetések vonhatók le: annotátorok közti egyetértés nagyságrendje összevethető Véronis kísérletének eredményeivel, minden adatbázis esetében. A szokásos (0.7-0.8) küszöbérték alapján egyik adatbázis szerinti IAA-érték sem lett olyan magas, amely alapján gépi WSD számára megbízható referenciakorpusz készíthető volna. A jelentéstárként használt adatbázis típusa nagyban befolyásolja, hogy milyen IAA-mértéket kapunk – jelenlegi állapotában az igei WordNet alapján kaptuk a legjobb értékeket, azt az ÉKSz-szerinti értékelést kivéve, amelyekben csak az ÉKSz fő jelentéscsoportjait vettük figyelembe, az ezek alatt meghatározott aljelentéseket (az adatbázis legfinomabb jelentésmegkülönböztetéseit) nem (l. a fenti táblázat jobbszélső oszlopát). A jelentések megkülönböztetésének finomsága, úgy tűnik, befolyásolja az IAA-értéket (ld a két ÉKSz-en alapuló IAA érték összehasonlítását (0.300 vs. 0.517)), ám pusztán az ige poliszémiájának mértéke nem tűnik relevánsnak az IAA-mérték szempontjából.

Véronis hipotézise a jelenlegi magyar nyelvre elérhető jelentéstárak alapján nem igazolható: a tisztán disztribúciós alapon készült ISZGYSZ jelenlegi formájában még nem válthatja fel a (legalább részben) introspektív alapon készült jelentéstárakat. Ennek okát az első kvalitatív elemzések alapján abban látjuk, hogy egyes annotátorok tisztán formai jegyek alapján rendeltek szóelőfordulásokhoz szerkezeteket, mások pedig az esetleges lemmákat, vagy akár az esetragokat / névutókat is szemantikailag

reprezentatív tartalommal töltötték meg. Az alábbi két tesztmondatban előforduló *emel* szó különböző annotációi jól illusztrálják ezt:

(1) Ezek az eredmények pedig az érdekképviseltek presztízsét emelik.

(2) A kipattanó labdát Makaay négy méterről a teljesen üres kapu fölé emelte.

Az (1) mondat esetében mind az öt annotátor különböző választ adott: hárman szemantikai értelmezés után az *emel magas-rA -t*, az *emel magas-bA -t*, *emel ár-A-t* kereteket választották, egy annotátor az *emel -t* keret mellett döntött, egy pedig "nincs" választ adott. A (2) mondat esetében egy annotátor az *emel -ba -t* szerkezetet választotta, hárman az *emel fölé -t* keretet, egy pedig "nincs" választ adott.

Esettanulmányunk alapján összegzésként elmondhatjuk, hogy a jelenleg rendelkezésünkre álló igei adatbázisok Wsd-céljára való alkalmazása további, WSD-orientált fejlesztést igényel. A továbbiakban az esettanulmány eredményeként kapott adatok további – kvalitatív és kvantitatív – elemzésével kívánjuk meghatározni, hogy milyen kritériumoknak kell egy WSD-célokra tervezett adatbázisnak megfelelnie.

## Hivatkozások

1. Artstein, R., Poesio, M.: Inter-coder agreement for computational linguistics. *Computational Linguistics* 34(4) (2008) 555-596
2. Agirre, E., Edmonds, Ph.: *Word sense disambiguation. Algorithms and Applications.* (Text, Speech and Language Technology), Springer-Verlag New York, Inc., Secaucus, NJ, (2007)
3. Kuti J., Varasdi K., Gyarmati Á., Vajda P.: Hungarian WordNet and representation of verbal event structure. In: *Acta Cybernetica*, 18(2), (2007) 315-328
4. Sass B., Pajzs J.: FDVC - Creating a Corpus-driven Frequency Dictionary of Verb Phrase Constructions for Hungarian. In: *Abstracts of the eLexicography in the 21st century Conference*, Louvain-la-Neuve, Belgium, (2009) 183-186
5. Véronis, J.: Sense tagging: does it make sense? In Wilson, A., Rayson, P. és McEnery, T. (Ed.) *Corpus Linguistics by the Lune: a festschrift for Geoffrey Leech*. Frankfurt: Peter Lang (2003)