

# Jelentésegértelműsítés – 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

<sup>2</sup>ELTE BTK, Nyelvtudományi Doktori Iskola, Germanisztika Alprogram

{eheja,kutij,sass.balint}@nytud.hu

## Motiváció

A jelentésegértelműsítés (WSD) két alapvető lépése:

(1) valamilyen jelentéstár létrehozása, illetve kiválasztása;

(2) a jelentéstárban szereplő jelentések hozzárendelése a szóalakokhoz.

Általában (2) – a különféle algoritmusok kutatása – van előtérben. A megoldások sok esetben a WordNet különböző verzióit használják, annak ellenére, hogy a WordNet-et nem kifejezetten WSD céljára fejlesztették [4].

Mi (1)-gyel foglalkozunk.

**Hosszútávú célunk:** milyen tulajdonságokkal bíró jelentéstár alapján képesek a humán annotátorok egyértelműen (nagy egyetértéssel) megadni a jelentéseket, azaz elvégezni a WSD feladatot?

**Humán annotátorok közötti egyetértés (IAA) fontossága:**

– Előfeltevés: adott jelentéstár esetében az automatikus WSD nem lehet jobb, mint az emberi  $\Rightarrow$  magas IAA értékkel bíró jelentéstárat kell használni.  
– Magas IAA érték szükséges egy megbízható tesztkorpusz létrehozásához.

**Nehézség:** a jelentés fogalmára nem létezik olyan széleskörű és operatív módon használható definíció, amely lehetővé tenné az egyéni intuíciónak kiküszöbölését a jelentések megkülönböztetése során  $\Rightarrow$  A (WSD-re használt) jelentéstárak szükségszerűen inkonzisztensek [7].

**Véronis kísérlete:** Jelentéstár: Petit Larousse. 6 kísérleti személynek 60 szó 3724 előfordulásához kellett hozzárendelni a jelentéstárban megfelelő jelentését 1 bekezdés kontextusában.  $\Rightarrow$  Az alacsony IAA (igék esetén 0.41) a feladat nehézségét támasztja alá.

**Megoldás:** A lexikon létrehozása során a jelentések elkülönítésekor az intuíción helyett **megfigyelhető felszíni disztribúciós adatokra** támaszkodni [7].

## Három adatbázis

Három magyar igei adatbázist hasonlítottunk össze az annotátorok közötti egyetértés alapján. A adatbázisok az introspektív illetve disztribúciós alapon készülő jelentéstárak közötti skálán helyezkednek el.

1. Magyar Értelmező Kéziszótár (ÉKSz) [1] – nagyrészt introspekción alapuló lexikográfiai mű, kb. 75.000 címszó magyarázata.
2. Magyar igei WordNet (HuWN) [5] – a Princeton WordNet 2.0-ra épülő lexikális adatbázis, alapegysége a fogalom. Készítésekor automatikusan kinyert vonzatkeret információkat is figyelembe vettünk, így a jelentésmegkülönböztetések nem pusztán introspekción alapulnak. Kb. 3000 jelentés.
3. Igei szerkezetek gyakorisági szótára (ISZGYSZ) [6] – automatikusan gyűjtött gyakori, különböző specifikusságú ige + esetragos/névutós névszói csoport szerkezeteket tartalmaz a vonzatkeretektől a komplex igéken át a szólásokig. Az adatbázist előállító algoritmus szigorúan disztribúciós alapon gyűjt. Kb. 8000 szerkezet.

**Előzetes elvárás:** a Véronis kísérlet alapján:  $IAA_{ÉKSz} \ll IAA_{HuWN} \ll IAA_{ISZGYSZ}$

## A kísérlet

**Kísérleti feladat:** az MNSZ 300 leggyakoribb igéje közül 15 választottunk ki kézzel úgy, hogy mindhárom adatbázisban szerepeljenek. Ezen igék 30-30 előfordulását annotáltattuk 5-5 kísérleti személlyel az ÉKSz és a HuWN jelentéseivel, illetve az ISZGYSZ adatbázis igei szerkezetével. Minden kísérleti személy pontosan egy adatbázis alapján dolgozott.

A kísérleti személyek a jelentések beazonosításakor az alábbi opciók közül választhattak:

– Az adott adatbázis által megkülönböztetett jelentésegységek közül egy.

– Nem tudom eldönteni. („nem tudom”)

– Nem szerepel a megfelelő jelentés az adatbázisban. („nincs”)

Az IAA mértékek azt mutatják meg, hogy a tapasztalt egyetértés hol helyezkedik el a várható egyetértés (0) és a teljes egyetértés (1) által meghatározott skálán. Minél közelebb van 1-hez, annál nagyobb a valószínűsége, hogy az annotátorok közti egyetértés nem véletlen. Az általunk alkalmazott **Fleiss-féle multi  $\pi$**  [3] ...

– a várható egyetértés számításakor elvonatkoztat az egyes annotátorok egyéni torzításaitól ( $\leftrightarrow \kappa$ );

– érzéketlen az olyan kategóriákra, amelyeket soha egyetlen annotátor sem választott, azaz értékét nem befolyásolja, hogy hány kategóriából választhattak eredetileg az annotátorok, csak az, hogy hányat választottak ténylegesen.

## Eredmények

Az egyes adatbázisok szerinti IAA-értékek az egyes igékre lebontva:

	Fleiss-féle multi $\pi$ értékek				választható jelentésszám			
	ÉKSz	HuWN	ISZGYSZ	ÉKSz/fő	ÉKSz	HuWN	ISZGYSZ	ÉKSz/fő
emel	0.450	0.753	0.170	0.848	13	10	16	5
feltesz	0.493	0.693	0.265	0.745	14	7	8	7
fizet	0.157	0.610	0.259	0.278	12	1	23	5
használ	0.210	0.954	0.336	0.611	8	2	22	4
köt	0.449	0.637	0.237	0.535	29	21	19	12
lép	0.346	0.595	0.443	0.601	12	11	31	7
megold	0.137	0.197	0.255	0.449	6	2	12	4
mutat	0.187	0.153	0.284	0.365	13	4	27	5
okoz	0.000	0.590	0.286	0.000	2	3	26	2
rendelkezik	0.195	0.469	0.471	0.474	6	3	15	4
segít	0.112	0.371	0.434	0.173	7	4	19	5
szolgál	0.279	0.516	0.548	0.509	15	8	16	7
tárgyal	0.840	0.543	0.407	0.840	3	2	16	3
választ	0.452	0.935	0.444	0.713	6	2	24	4
vállal	0.207	0.311	0.275	0.623	6	3	26	3
átlag	0.300	0.483	0.340	0.517				

A „nincs” és „nem tudom” (utóbbi előfordulási aránya 2-6%) válaszokat önálló értéként kezeltük. 1. oszlop: az összes ÉKSz jelentést tekintetbe véve; 4. oszlop: az aljelentéseket utólagosan összevonva.

## Kiértékelés

**Következtetések:**

– Az IAA nagyságrendje minden adatbázis esetében összevethető Véronis kísérletének eredményeivel.

– **A szokásos IAA küszöbértéket figyelembe véve (0.7-0.8) egyik adatbázis sem alkalmas arra, hogy gépi WSD számára megbízható tesztkorpusz készítésére szolgáljon.**

– A jelentéstár típusa nagyban befolyásolja az IAA-értéket, a sorrend:

**1.** HuWN **2.** ISZGYSZ **3.** ÉKSz

– Az ige polyszemiájának mértéke nem tűnik relevánsnak: a választható jelentésszám nincs összefüggésben a  $\pi$  értékekkel.

– A jelentések megkülönböztetésének finomsága relevánsnak tűnik: az ÉKSz-ben csak a fő jelentéscsoportokat figyelembe véve (4. oszlop) jóval magasabb  $\pi$  értékeket kapunk, mint az összes jelentés esetén (1. oszlop).

**Előzetes elvárásunk nem igazolódott:** A tisztán disztribúciós ISZGYSZ jelen állapotában nem válthatja fel a (legalább részben) introspektív alapon készült jelentéstárakat.

**Lehetséges magyarázat:** 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. Példák:

„A kipattanó labdát Makaay négy méterről a teljesen üres kapu fölé emelte.”

Válaszok:

emel fölé -t (3 db),  
emel -ba -t (1 db),  
„nincs” (1 db)

„Ezek az eredmények pedig az érdeklődő presztízst emelik.”

Válaszok:

emel magas-rA -t (1 db),  
emel magas-bA -t (1 db),  
emel ár-A-t (1 db),  
emel -t (1 db),  
„nincs” (1 db)

**Konklúzió és további feladatok:**

- A létező igei adatbázisok WSD céljára való alkalmazása WSD-orientált fejlesztést (pl. szemantikai információk disztribúciós megragadása) igényel.
- Az esettanulmány eredményeként kapott adatok további elemzése szükséges annak meghatározásához, hogy milyen kritériumoknak kell egy WSD-célokra tervezett adatbázisnak megfelelnie.

## Hivatkozások

[1] Magyar Értelmező Kéziszótár. Budapest: Akadémiai Kiadó, 2006.

[2] Eneko Agirre and Philip Edmonds. *Word sense disambiguation. Algorithms and Applications. (Text, Speech and Language Technology)*. Springer-Verlag New York, 2007.

[3] Ron Artstein and Massimo Poesio. Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4):555–596, 2008.

[4] Christiane Fellbaum. *WordNet An Electronic Lexical Database*. MIT Press, 1998.

[5] Judit Kuti, Károly Varasdi, Ágnes Gyarmati, and Péter Vajda. Hungarian WordNet and representation of verbal event structure. *Acta Cybernetica*, 18(2):315–328, 2007.

[6] Bálint Sass and Júlia Pajzs. FDVC – creating a corpus-driven frequency dictionary of verb phrase constructions. In *Proceedings of eLexicography in the 21st Century Conference*, page (accepted), Louvain-la-Neuve, Belgium, 2009.

[7] Jean Véronis. Sense tagging: does it make sense? In A. Wilson, P. Rayson, and T. McEnery, editors, *Corpus Linguistics by the Lune: a festschrift for Geoffrey Leech*. Frankfurt: Peter Lang, 2003.