

# Mélyesetek a 41angben

Makrai Márton  
makrai.marton@nytud.mta.hu

MTA {SZTAKI, Nyelvtudományi Intézet} Nyelvtechnológiai Kutatócsoport

MSZNY 2014

# Áttekintés

A 41ang

A mélyesetek

Az egyes mélyesetek

Összehasonlítás

## A 41ang fogalmi szótár

- ▶ a SZTAKI-ban készült, általános szövegértés céljával
- ▶ többnyelvű szótár
  - ▶ 4 nyelv emberi munkával
  - ▶ 50 automatikusan gyűjtve (?)
- ▶ absztrakt tételek
  - ▶ szófajok között nem teszünk különbséget: 'fut' és 'futás' (reprezentációja) ugyanaz
  - ▶ absztrakt fogalmak, monoszémia
- ▶ a fogalmak jelentését leíró definíciók
  - ▶ formális nyelven,
  - ▶ emberi munkával íródtak

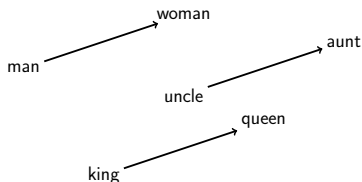
## A 41ang fogalmi szótár

- ▶ a SZTAKI-ban készült, általános szövegértés céljával
- ▶ többnyelvű szótár
  - ▶ 4 nyelv emberi munkával
  - ▶ 50 automatikusan gyűjtve (?)
- ▶ absztrakt tételek
  - ▶ szófajok között nem teszünk különbséget: 'fut' és 'futás' (reprezentációja) ugyanaz
  - ▶ absztrakt fogalmak, monoszémia
- ▶ a fogalmak jelentését leíró definíciók
  - ▶ formális nyelven,
  - ▶ emberi munkával íródtak

	definiáló szókincs	összesen
mélyesete van	238	624
összesen	1103	3478

## Teszt, demó

- ▶ a 41angból vektoros nyelvmodellt készítettünk (*continuous vector space model*)
- ▶ a szójelentés kompozicionalitása, *vector offset method*, (?; ?)



$$\mathbf{v}(\text{king}) + \mathbf{v}(\text{female}) = \mathbf{v}(\text{queen})$$

- ▶ mélyeseteket a pénztáros demóban (?) használtunk először, de még nem a most bemutatott esetkészletet

# Áttekintés

A 41ang

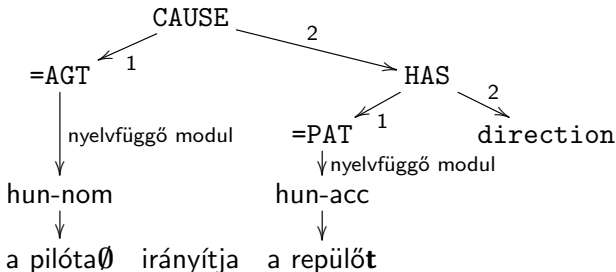
A mélyesetek

Az egyes mélyesetek

Összehasonlítás

# Mélyesetek

- ▶  $\approx$  tematikus szerepek
- ▶ irányít



- ▶ felszíni eset: adpozíció, szórend is
- ▶ általánosítás szemantikai és szintaktikai alapon

## A mélyesetek nem...

- ▶ a mélyeseteknek nem kell minden (rendszeres) szemantikai különbséget tükrözniük, pl. *hide/mislay*, *pour/spill*, ... ágensforma
- ▶ a szokásos szemantikai szerepektől eltérve nem akarjuk különböző igék argumentumait összhangba hozni
  - ▶ János=<sub>AGT</sub> elad egy könyvet=<sub>PAT</sub> Péternek=<sub>DAT</sub>
  - ▶ Péter=<sub>AGT</sub> vesz egy könyvet=<sub>PAT</sub> Jánostól=<sub>FROM</sub>



## Az absztraktság következményei a vonzatkeretre

- ▶ egy (ige)tőnek egy reprezentációja van
- ▶ pl. nincs külön tárgyas és tárgyatlan
- ▶ bizonyos bővítmények szerepelnek a régens reprezentációjában, másokat konstrukciók termelnek be (enyhén konstrukcionista megközelítés)
  - ▶ pl. fest valamit **valakinek**
- ▶ a definícióban szereplő bővítmények sem kötelezők szintaktikailag
- ▶ leginkább a szó jelentésének leírásához szükséges szereplők jelennek meg
- ▶ argumentumok öröklődnek

*bite: cut, INSTRUMENT tooth*

# Áttekintés

A 4lang

A mélyesetek

Az egyes mélyesetek

Összehasonlítás

## Az ágens és a páciens

- ▶ ágens: szándékos okozó,... (=AGT)
- ▶ páciens (=PAT)
  - ▶ az igétől függetlenül keveset tudunk róla mondani
  - ▶ egyargumentumú igéknél is (unakkuzatív hipotézis), *süllyed, fürdik*

## Datív, =DAT

- ▶ ditranzitívok recipiense, címzett (hallgató, kapó)
- ▶ *mond valamit* **valakinek**
  - ▶ enged, parancsol, megmagyaráz, megtilt, hálás, bemutat, mutat
- ▶ *ad valamit* **valakinek**
  - ▶ kölcsönad, átenged, bérbe ad, áldoz (istennek), ajánl, tartozik, fizet, elad, adományoz, ad, segít

## Lokatív mélyesetek

- ▶ absztrakció: ha egy szereplő sok nyelvben olyan felszíni esetet kap, mint a mozgás célja (a magyarban főleg határozóragok, az angolban pedig prepozíciók), akkor cél
- ▶ cél (=TO)
  - ▶ *hozzászoktat, összead, vkihez tartozik, belevesz, fülel, emlékeztet*
- ▶ forrás (=FROM)
  - ▶ *elfogad, kölcsönöz, vesz, levág, ered, eltávolít, bérel, kivon, elvesz*
- ▶ statikus hely (=AT)
  - ▶ *valahol helyezkedik el, marad*

## Relációs főnevek

- ▶ birtokos, =POSS
  - ▶ hiány, tanács, légkör, nagynéni, háttér, alap, vminek a feneke
- ▶ cél, =T0
  - ▶ jog, alkalom, készség, tendencia (valamire, for something)
- ▶ az igéből képzett relációs főneveknek nincs saját tételük  
(*valaminek a keresése*)

## Funkciómorfémák vonzatai és ferde esetek: =REL

- ▶ funkciómorfémák
  - ▶ tartalmaz ragok
  - ▶ produktív képzők és adpozíciók
  - ▶ a három szófajt egyformán kezeljük
- ▶ ferde eset
  - ▶ *prefer to something, jobban szeret valaminél*

## Funkciómorfémák vonzatai és ferde esetek: =REL

- ▶ funkciómorfémák
  - ▶ tartalmaz ragok
  - ▶ produktív képzők és adpozíciók
  - ▶ a három szófajt egyformán kezeljük

for -ért pro dla: AT/2744 exchange, =REL[price]
- ▶ ferde eset
  - ▶ *prefer to something, jobban szeret valaminél*





---

		db
ágens	=AGT	383
páciens (igefüggő)	=PAT	311
egyetlen vagy ferde	=REL	81
birtokos	=POSS	52
datív	=DAT	30
cél	=TO	17
forrás	=FROM	11
statikus hely	=AT	2

---

# Áttekintés

A 41ang

A mélyesetek

Az egyes mélyesetek

Összehasonlítás

## Töbnyelvű lexikai erőforrások

- ▶ WordNet
  - ▶ *synsetek*
  - ▶ glossza (szöveges definíció)
  - ▶ példamondat
  - ▶ szemantikai viszonyok, pl. is-a, instance-of, part-of + gloss-related
  - ▶ Miháltz Márton, Sass Bálint: *A Magyar WordNet automatikus kiterjesztése szelekciós preferenciák[...]*ka
- ▶ BabelNet
  - ▶ 50-nyelvű szótár, szemantikai relációkkal
  - ▶ többek közt a WordNetből és a Wikipédiából gyűjtötték automatikusan
- ▶ 41ang: formális, „gold” definíciók

## Tematikus szerepek: VerbNet

- ▶ csak angol
- ▶ a Levin-féle igeosztályok kiterjesztése
- ▶ tematikus szerepek
  - ▶ 23, míg a 41angben 8 absztraktabb
  - ▶ adott ige által kiosztott tematikus szerepek konstrukciónként különböznek, a 41angben azonosak
- ▶ argumentumok szelekciós megszorításai
- ▶ a jelentést leíró predikátumok idői függvényekkel  
hit:

*cause(Agent, E)*  
*manner(during(E), directedmotion, Agent)*  
*!contact(during(E), Agent, Patient)*  
*manner(end(E), forceful, Agent)*  
*contact(end(E), Agent, Patient)*

# Hivatkozások I

<http://hlt.sztaki.hu/resources/4lang/>

## Az ágens és a páciens határa

- ▶ ergatív nyelv  $\approx$  a tárgyatlan igék argumentuma olyan esetet kap, mint a tárgyasok páciense (nem-ágense)

## Az ágens és a páciens határa

- ▶ ergatív nyelv  $\approx$  a tárgyatlan igék argumentuma olyan esetet kap, mint a tárgyasok páciense (nem-ágense)

(?)

	tárgyjelölő	ergatív 1.	ergatív 2.	aktív	lexikalizált aktív	alany- jelölő
Péter írja a levelet.	nom	ag	ag	ag	ag	ag
Péter ír.	nom	nom	ag	ag	ag	ag
Péter sétál.	nom	nom	nom	ag	ag/nom	ag
Péter beteg.	nom	nom	nom	nom	ag/nom	ag

tárgyjelölő	angol (eng), magyar (hun)
ergatív 1.	kabardi (kbd), avar (ava), adige (ady)
ergatív 2.	agul (agx), udi (udi)
aktív	bacbi (bbl)
lexikalizált aktív	grúz (kat), dakota (dak)
alanyjelölő	megrel (xmf), maidu (nmu)