



Szemantika: modalitás, kompozicionalitás



Bevezetés

- **Szemantika: jelentéssel foglalkozó nyelvészeti részterület**
- **Mi a jelentés?**
 - Térben és időben változhat (*leó, király*)
 - Egyéni különbségek (*fogas*)
- **Szavak jelentése: lexikális szemantika**
- **Mondatjelentés: predikátumlogika**
- **Szövegek jelentése: kijelentéslogika**

Formális szemantika a nyelvészetben

- **Kijelentéslogika**

$p \vee q$

$\neg p$

- **Szavak jelentése**

Kutya 'kutya'

Ugat 'ugat'

- **(Elsőrendű) predikátumlogika**

$\neg \exists x(\text{kutya}(x) \ \& \ \text{ugat}(x) \ \& \ \text{harap}(x))$

Mennyire adtuk meg a tényleges jelentést?





Számítógépes szemantika

- Tartalmi elemzés automatikusan?
- Szöveg jelentésének megragadása távoli cél...
- Lexikális szemantika (jövő hét):
 - Jelentés-egyértelműsítés
 - Szójelentés megadása

Modalitás

Kompozicionalitás

Modalitás

- **Az események lehetnek**
 - igazak
 - hamisak (tagadás)
 - bizonytalanok
- **Nagyon fontos őket elkülöníteni és sajátos kezelésben részesíteni (alkalmazások: IE, IR, MT, OM...)**
- **Fókuszban: bizonytalanság azonosítása**



A feladat

- Bizonytalanságot jelző nyelvi elemek (kulcsszavak) azonosítása a szövegben és egyértelműsítése

*In **May**, there **may** be heavy storms.*

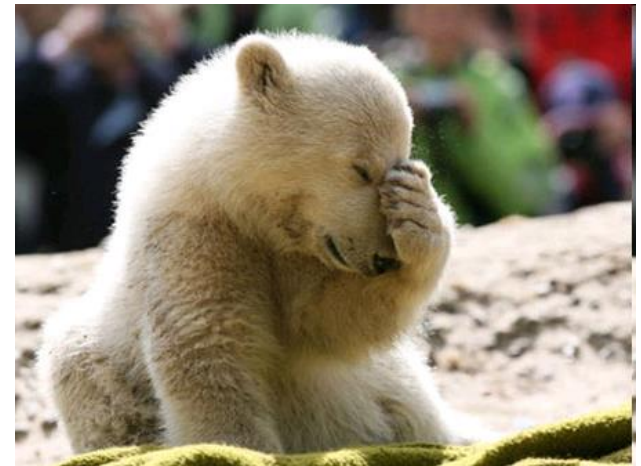
- Hatókör azonosítása / a kulcsszavak és események összerendelése
- Alkalmazásnak megfelelően az adott szövegrészek kiszűrése, törlése, kigyűjtése stb.





Mi a bizonytalanság?

- Uncertainty
- Hedge
- Speculation
- Factuality
- Polarity
- Weasel
- Uncertain
- Speculative
- Probable
- Possible
- Doubtful





Korpuszok

- BioScope (Vincze et al. 2008)
- FactBank (Saurí and Pustejovsky 2009)
- WikiWeasel (Farkas et al. 2010)
- MPQA (Wiebe, Wilson, and Cardie 2005)
- Biológiai cikkek (Medlock & Briscoe 2007)
- PubMed-absztraktok (Settles, Craven, and Friedland 2008)
- Genia Event (Kim, Ohta, and Tsujii 2008)
- 10K biológiai mondat (Shatkay et al. 2008)
- E. Coli (Thompson et al. 2008).
- Genia Pathway (Nawaz, Thompson, and Ananiadou 2010)
- 112 újságcikk (Rubin, Liddy, and Kando 2005; Rubin 2010)
- Orvosi zárójelentések (Uzuner, Zhang, and Sibanda 2009)





Egységesítés

- **Egységes definíció:**
- **Bizonytalanság = információ hiánya: a befogadó nem lehet biztos valamely információban**
- **vs. tény = a befogadó biztos afelől, hogy az információ helytálló (és a világban is helytálló)**
- **vs. tagadás = a befogadó biztos afelől, hogy az információ téves (és a világban is téves)**



A bizonytalanság fajtái

- **Szemantikai bizonytalanság: nem rendelhető igazságérték a propozícióhoz**
Ha huzamos ideig nem találtak élelmet, az egész csoport éhen halhatott.
- **Diskurzusszintű bizonytalanság: a propozíciónak van igazságértéke, a bizonytalanság a forrás/viszonyítási pont hiányából fakad**
Számos filozófus szerint az emberi lét értelmét vesztené a halál ismerete nélkül.



Szemantikai bizonytalanság

- **EPISZTEMIKUS:** **Lehet**, hogy esik.
- **HIPOTETIKUS:**
 - **DINAMIKUS:** Mennem **kell**.
 - **DOXASZTIKUS:** Azt **hiszi**, hogy a Föld lapos.
 - **VIZSGÁLAT:** A felvétel manipuláltságáról **vizsgálatot folytattak**.
 - **FELTÉTELES:** **Ha** esik, itthon maradunk.

Diskurzusszintű bizonytalanság

- **WEASEL:** homályos, félrevezető kifejezések, forrás nélküli mondatok (kik?)
Egyesek szerint inkább megszállást kellene mondani.
- **HEDGE:** mennyiséget/minőséget módosítják (elmossák a határokat) (pontosan mennyi?)
A belga lakosság kb. 10%-a él Brüsszelben.
- **PEACOCK:** alá nem támasztott vélemények, túlzások, szubjektív elemek
Apafi négy évet keserves tatár fogságban töltött.
- **Az információ megbízhatósága kérdéses**





Bizonytalanság automatikus azonosítása

- Angolra jól működő gépi tanuló rendszerek a világban (Szegeden is) – szemantikai bizonytalanság
- Diskurzusszintű bizonytalanság: vannak próbálkozások (Szegeden is)
- Magyarra folyamatban mindez...

Kompozicionalitás

- Egy összetett kifejezés jelentése meghatározható tagjainak jelentéséből és azok kapcsolódási módjából

piros autó

kenyeret eszik

- De nem minden nyelvi kifejezés kompozicionális...





Többszavas kifejezések (MWE)

- **Több szövegszóból álló lexikai egységek**
- **Szintaktikai, szemantikai, pragmatikai vagy statisztikai szempontból sajátos viselkedést mutatnak**
- **Kompozicionalitás (teljes) hiánya jellemzi őket**

MWE típusai

- **Összetett főnevek:** *középiskola, farkasétvágy*
- **Összetett melléknevek:** *római katolikus, nagyotmondó*
- **Igekötős igék:** *berúg*
- **Közmondások, szólások:** *alulról szagolja az ibolyát*
- **Félig kompozicionális főnév + ige szerkezetek (FX):** *előadást tart, csődbe megy*
- **Egyéb:** *status quo, ilyen az élet*



MWE-k a számítógépes nyelvészetben

- Sajátos bánásmód alkalmazások szintjén: *racing car* – *versenyautó* (MT), *előadást tart* - **tartás* – *előadás* (IE)
- Azonosítani kell őket szövegekörnyezetben (*give a ring, tevékenységet folytat*)
- Annotált korpuszok (Wiki50, magyar FX-korpuszok...)



On August 20, 2005, in a private ceremony, Thompson's ashes were fired from a cannon atop a tower of his own design (in the shape of a double-thumbed fist clutching a peyote button) to the tune of Norman Greenbaum's "Spirit in the Sky" and Bob Dylan's Mr. Tambourine Man. Red, white, blue, and green fireworks were launched along with his ashes. As the city of Aspen would not allow the cannon to remain for more than a month, the cannon has been dismantled and put into storage until a suitable permanent location can be found.

According to his widow Anita, Thompson's funeral was financed by actor Johnny Depp, a close friend of Thompson. Depp told the Associated Press, "All I'm doing is trying to make sure his last wish comes true. I just want to send my pal out the way he wants to go out."

Other famous attendees at the funeral included U.S. Senator John Kerry and former U.S. Senator George McGovern; 60 Minutes correspondent Ed Bradley and Charlie Rose; actors Jack Nicholson, Bill Murray, Benicio del Toro, Sean Penn, and Josh Hartnett; singers Lyle Lovett, John Oates and numerous other friends. An estimated 280 people attended the funeral.

The plans for this monument were initially drawn by Thompson and Ralph Steadman and were shown as part of an Omnibus program on the BBC entitled Fear and Loathing in GonzoVision (1978). It is included as a special feature on the second disc of the 2003 Criterion Collection DVD release of Fear and Loathing in Las Vegas (labeled on the DVD as "Fear and Loathing on the Road to Hollywood"). The video footage of Steadman and Thompson drawing the plans and outdoor footage showing where he wanted the cannon constructed were planned prior to the unveiling of his cannon at the funeral.



Angol MWE-k gyakorisága

Category	Occurrence	Unique	Avg. frequency
Noun Comp.	2929	2405	$0.0263 \pm 2.1E-4$
Adj. Comp.	78	60	$0.0008 \pm 1.1E-6$
VPC	446	342	$0.0038 \pm 8.9E-6$
LVC	368	338	$0.0030 \pm 4.9E-6$
Idiom	19	18	$0.0002 \pm 1.2E-7$
Other	21	17	$0.0002 \pm 8.1E-8$
MWE sum	3861	3180	$0.0342 \pm 2.0E-4$
PER	4093	1533	$0.0352 \pm 5.8E-4$
ORG	1498	893	$0.0133 \pm 2.0E-4$
LOC	1558	705	$0.0150 \pm 2.5E-4$
MISC	1827	952	$0.0166 \pm 2.3E-4$
NE sum	8976	4083	$0.0801 \pm 7.5E-4$



MWE-k kezelése

- **Automatikus azonosításukra szabályalapú és gépi tanulási módszerek angol és magyar nyelvre**
- **Annotált korpuszok több doménre és nyelvre**



Alkalmazások szintje

- Egy MWE egy egységként tárolódik (pl. fordítási egység)
- Speciális szabályok:
 - IE: FX-eknél a főnév jelentése számít
 - Nem kompozicionálisoknál nem használható egyik alkotóelem sem - lexikon
- Gépi fordításban nagyon nagy jelentősége van az MWE-knek (*fűbe harap – to kick the bucket*)