



Jelentésmodellezés tudásbázisok segítségével

# Homonímák (azonos alakúság)

- Azonos szóalakokkal, de egymástól független jelentéssel bíró szavak (pl. nyúl, vár)
  - Homonímák adódhatnak morfológiai többértelműségből is (pl. szemét, csódhullám)
- Megfelelő kezelésük számos feladat szempontjából kívánatos
  - Információvisszakeresés (pl. bat care)
  - Gépi fordítás

# Poliszéma (többértelműség)

- Ugyanannak a szóalaknak a jelentésárnyalatai (pl. király)
  - Nekem akkor is Jimmy marad a király.
  - III. András volt az utolsó Árpád-házi király.
  - Ez aztán király volt!

# Poliszéma (többértelműség)

- Ugyanannak a szóalaknak a jelentésárnyalatai (pl. király)
  - Nekem akkor is Jimmy marad a király.
  - III. András volt az utolsó Árpád-házi király.
  - Ez aztán király volt!
- **Metoníma** (névcsere): jelentések közötti kváziekvivalencia
  - Szeretem Arany Jánost. vs. Arany Jánost a Kerepesi temetőben helyezték örök nyugalomra.
  - Az épület-intézmény jelentéspárok tipikusan ilyenek még
    - De szép ez az **iskola**. vs. Az **iskola** igazgatója nem ivott ma még kávé.

# Zeugma teszt

- A **zeugma** kb. tömörítést jelent
  - Egy szó több jelentéssel rendelkezik, ha zeugmákban tud „természetellenesen” is viselkedni
    - Indulás előtt érdemes mély levegőt és kiskabátot **venni**.
    - Öt hete **vezeti** a lejátszási listát és az új autóját.
    - Az órát föl, a cipőmet lehúztam.
- A szavak többsége rendelkezik több jelentésárnyalattal

# Szinonímák

- Erős definíció: két szó akkor szinonimája egymásnak, ha bármely előfordulásukban kicserélhetők egymásra
  - Verda vs. Személygépjármű
    - „Feltartott kézzel szálljon ki a verdájából!”\*
    - „Nem semmi személygépjármű, mikor vetted?”\*
- Lényegében nem is a szavak, hanem inkább a jelentésosztályaik egyezőségéről érdemes beszélni

# Antonímák

- Ellentétes jelentést hordozó jelentéspárok
  - Bizonyos értelemben hasonló jelentéssel bírnak, csak épp a polaritásuk ellentétes (pl. kicsi vs. nagy)
    - Szóvektorok tanulása során is tapasztaljuk mindezt
  - Igék is rendelkezhetnek antonímákkal (pl. jön vs. megy)

# Hiponímák és homonímák

- (A,B) szópár hiponím relációban áll egymással, ha az A szó a B szónak egy specializált alfajtája, pl. (emlős, állat)
  - A hiperníma a hiponím reláció megfordítottja, ilyenkor A szó általánosítója a B szónak, pl. (emlős, gorilla)
- Az-egy (IsA) taxonómiák alapját képezik
- Általában tranzitívak
  - Hiperním(gorilla, emlős) és Hiperním(emlős, állat) → Hiperním(gorilla, állat)



# WordNet

- Nagy szemantikus jelentésrepresentációs háló
- <https://wordnet.princeton.edu/>
- Online lekérdezőfelület:  
<http://wordnetweb.princeton.edu/perl/webwn>
- NLTK-s támogatottság:  
<http://www.nltk.org/howto/wordnet.html>

# A house szó első 3 főnévi szinszetje

- Noun

- S: (n) **house** (a dwelling that serves as living quarters for one or more families) "he has a house on Cape Cod"; "she felt she had to get out of the house"
- S: (n) firm, **house**, business firm (the members of a business organization that owns or operates one or more establishments) "he worked for a brokerage house"
- S: (n) **house** (the members of a religious community living together)

# A house szó igei szinszetjei

- Verb
  - S: (v) house (contain or cover) "This box houses the gears"
  - S: (v) house, put up, domiciliate (provide housing for) "The immigrants were housed in a new development outside the town"

# A house#N#1 hipernímái

- S: (n) dwelling, home, domicile, abode, habitation, dwelling house (housing that someone is living in) "he built a modest dwelling near the pond"; "they raise money to provide homes for the homeless"
  - S: (n) housing, lodging, living accommodations (structures collectively in which people are housed)
    - S: (n) structure, construction (a thing constructed; a complex entity constructed of many parts) "the structure consisted of a series of arches"; "she wore her hair in an amazing construction of whirls and ribbons"
      - S: (n) artifact, artefact (a man-made object taken as a whole)
        - S: (n) whole, unit (an assemblage of parts that is regarded as a single entity) "how big is that part compared to the whole?"; "the team is a unit"
          - S: (n) object, physical object (a tangible and visible entity; an entity that can cast a shadow) "it was full of rackets, balls and other objects"
            - S: (n) physical entity (an entity that has physical existence)
              - S: (n) entity (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

# A house#N#1 további hipernímái

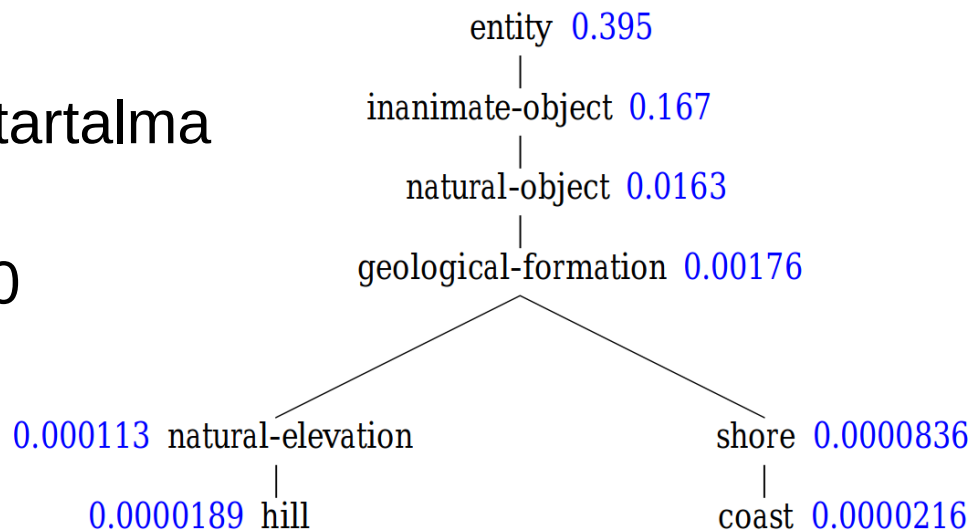
- S: (n) building, edifice (a structure that has a roof and walls and stands more or less permanently in one place) "there was a three-story building on the corner"; "it was an imposing edifice"
  - S: (n) structure, construction (a thing constructed; a complex entity constructed of many parts) "the structure consisted of a series of arches"; "she wore her hair in an amazing construction of whirls and ribbons"
  - S: (n) artifact, artefact (a man-made object taken as a whole)
    - S: (n) whole, unit (an assemblage of parts that is regarded as a single entity) "how big is that part compared to the whole?"; "the team is a unit"
      - S: (n) object, physical object (a tangible and visible entity; an entity that can cast a shadow) "it was full of rackets, balls and other objects"
        - S: (n) physical entity (an entity that has physical existence)
          - S: (n) entity (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

# Szemantikus hasonlóság tezaurusz alapú számítása

- Első ötlet: két szinset közötti távolság a szemantikus gráfban
  - A távolság lehetne a gráfban található legrövidebb út hossza
    - Több szempontból se jó ötlet
  - Vizsgáljuk a két szinset legspeciálisabb (hozzájuk legközelebbi) közös ősüket a gráfban (lowest common subsumer=LCS)
    - Még mindig nem jó ötlet, mert nem minden linknek ugyanakkora a súlya, ugyanakkor az egyes linkek mentén jelentkező „távolság” nem (feltétlen) képezi a tudásbázisok részét

# Resnik tezaurusz-alapú hasonlósága

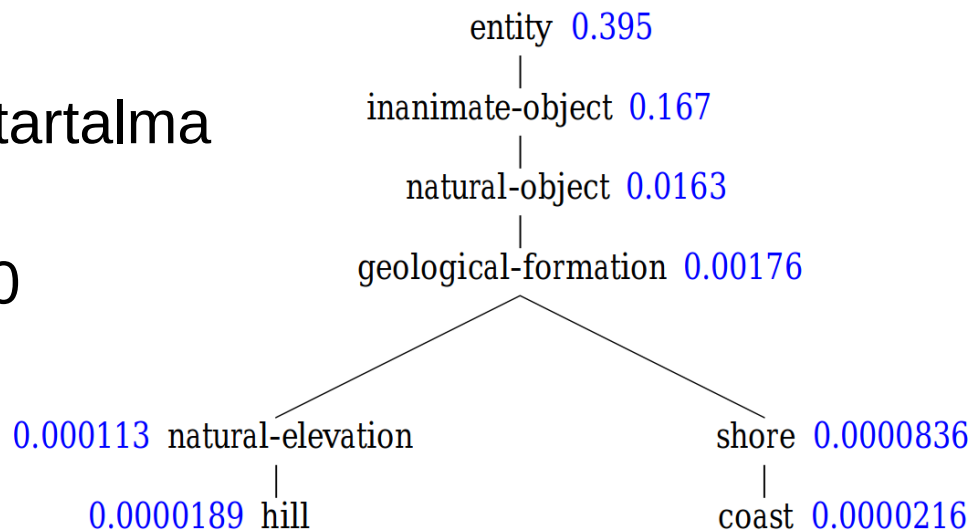
- A tezaurusz mellett egy háttérkorpuszból számított valószínűségi információra is támaszkodunk
  - A szemantikus gráf minden egyes  $c$  csúcsához társítsunk egy  $P(c)$  előfordulási valószínűséget
    - $P(\text{root})=1.0$
  - Legyen egy  $c$  csúcs információtartalma  $IC(c) = -\log P(c) = \log 1/P(c)$ 
    - Ahogy  $P(c) \rightarrow 1$ , úgy  $IC(c) \rightarrow 0$



# Resnik tezaurusz-alapú hasonlósága

- A tezaurusz mellett egy háttérkorpuszból számított valószínűségi információra is támaszkodunk
  - A szemantikus gráf minden egyes  $c$  csúcsához társítsunk egy  $P(c)$  előfordulási valószínűséget
    - $P(\text{root})=1.0$
  - Legyen egy  $c$  csúcs információtartalma  $IC(c) = -\log P(c) = \log 1/P(c)$ 
    - Ahogy  $P(c) \rightarrow 1$ , úgy  $IC(c) \rightarrow 0$

$$\text{Resnik}(c_1, c_2) = -\log P(c_1, c_2)$$





# Lin-féle tezaurusz-alapú hasonlóság

- Kontrasztba állítja a két csúcs megfigyelési valószínűségét és az LCS-ükét
  - $Lin(c_1, c_2) = \frac{2 * \log(P(LCS(c_1, c_2)))}{\log(P(c_1)) \log(P(c_2))}$
  - Amennyiben  $P(c_1) = P(c_2) = P(LCS(c_1, c_2))$  fönnáll, úgy a hasonlóság mértéke a lehető legmagasabb fokú
- Hasonló dolgot számszerűsít a Jiang-Conrath **távolság** is
  - $JC(c_1, c_2) = 2 * \log(P(LCS(c_1, c_2))) - (\log(P(c_1)) + \log(P(c_2)))$

# Resnik hasonlóság

- Ha két szinset jelentésében komoly átfedés mutatkozik, akkor ez definíciójuk szókincsében is visszaköszön
  - A több szóból egymás utáni szóból álló átfedést jobban jutalmazzuk
    - ***drawing paper***: paper that is specialy prepared for use in drafting
    - ***decal***:the art of transferring designs from specialy prepared paper to a wood or glass or metal surface
  - Kiterjeszthető a szinsetek közvetlen definíciójának vizsgálata helyett a velük relációban álló (pl. közvetlen hipernímák) definíciójának figyelembe vételével

# További erőforrások

- ConceptNet
- BabelNet
- (Open)Cyc
- FrameNet
- VerbNet
- ...