

SZEMANTIKA, 4. ELŐADÁS (3. ELMARADT SPORTNAP MIATT)

Kornai András

BME 2020/9/30

EMLÉKEZTETŐ

- Szótári információ 12-15 bit/szó
- Mondattani < 2 bit/szó
- *A híd leszakadt a teherautó súlya alatt*
The bridge collapsed under the weight of the truck
- Mi van abban a $2 \cdot 7$ bitben ami a nyelvtani szerkezetre jut?
- ((A híd) leszakadt) ((a teherautó súlya) alatt)
- (I saw)₁ (the man)₂ (on the hill)₃ (with the telescope)₄
- Láttam az embert a hegyen ahol a teleszkóp van (4 3-at módosítja)
- Láttam a teleszkópos embert a hegyen (4 2-t módosítja)
- A teleszkóppal láttam az embert a hegyen (4 1-et módosítja)
- Zárójelezéssel ezt nehéz mert a fák nem mindig síkbarajzolhatóak

SZÓTÁRI INFORMÁCIÓ: MIT, MENNYIT?



- Cél a szótár minimalizálása. Ezt két alapelv segíti:
- 1. Ha kiszámítható, akkor nem írjuk be
- Pl. a szólakokat csak akkor írjuk be ha rendhagyóak: a *remarked* nem szerepel a szótárban, mert a *remark*-ból szabályosan képezhető, de a *went* igen mert a *go*-ból nem kitalálható
- 2. Ha nem nyelvi információ, akkor nem írjuk be
- Pl. a Hold tömege $7.34767309 \cdot 10^{22}$ kg, de nincs az a nyelvtani ismeret amiből ez kikövetkeztethető, és nincs olyan része a nyelvi viselkedés leírásának ahol erre az ismeretre szükség lenne
- Az ilyenek helye nem a *lexikon* hanem az *enciklopédia*

MI AZ AMI BIZTOS (NEM) KELL?

- Maga a szótő kiszámíthatatlan
- Kivéve a hangtörvények alapján pl. szláv *malina, palica* → magyar *málna, pálca* (Horger-törvény)
- Manapság ezt nem vesszük figyelembe, mert a nyelvészajátítónak nem áll rendelkezésére a bemeneti infó
- A szófaj a legtöbbször kiszámítható (a szavak 90+% -a főnév, elég a többit beírni). Ezt a számítógépesek “alapértelmezés”, a nyelvészek “redundanciaszabály” néven emlegetik (default)
- Kiejtés: izgi, de mi nem fogunk foglalkozni vele

arduus

From Longman Dictionary of Contemporary English

ar·du·ous /'ɑ:dʒuəs \$ 'ɑ:rdʒuəs/ adjective  

involving a lot of strength and effort

arduus task/work

 the arduus task of loading all the boxes into the van

arduus journey/voyage

 an arduus journey through the mountains

- Definíció! —arduusly **adverb**

DEFINÍCIÓK

- Ez a szemantika szive: hogy kell kinézzenek azok a definíciók amik megragadják a szójelentést
- A szótár a szavakat más szavakkal definiálja
- Két probléma: érthetlenség és cirkularitás
- hastate 'triangular with sharp basal lobes spreading away from the base of the petiole'
- gentile 'heathen' és heathen 'gentile'
- HF4.1 Írj szócikket az alábbi szavakhoz: bakkecske, hadd, hogy, tud

A DEFINÍCIÓK EGYÜTTES STRUKTÚRÁJA

- Legyenek a V szótárgráf pontjai a szavak, és i -ből j -be fusson él, ha j szerepel i definíciójában. Lesznek (irányított) körök
- Legyen egy részhalmaz W *öndefináló* ha nincs $W \rightarrow V \setminus W$ él
- $D(W)$ legyen a W által definiált szavak halmaza (tehát ha $d \in D(W) \Rightarrow dV \subset W$)
- Iteratívén képezhető $W, D(W), D(D(W)), \dots$
- Ha egy S szótár szavai minden elérhetők W -ből, $\cup_i D^i(W) = S$, akkor azt mondjuk, hogy W *defináló* S -re nézve
- Mi ezt tipikusan nem W -ből kiindulva 'kifelé' keressük, hanem S -ből kiindulva 'befeelé'.

GYAKORLATI MUNKA

- Tipikus szótár: Collins COBUILD 169,612 definíció 65,891 szót használ
- Modern szótár: LDOCE 82,128 szóhoz 2,200 szót használ!
- Mert vigyáztak arra, hogy minimalizálják W -t, ez a híres Longman Defining Vocabulary (LDV)
- De még nekik se sikerült az összes redundanciát kivenni
- A 41ang W -je ennek harmada, 718 elem
- Ezért nagyon meg kell küzdeni

A DEFINÍCIÓK BELSŐ SZERKEZETE

- Olvasmány: *The algebra of lexical semantics*
- Ez egyfajta logika: van egy formális nyelv, ehhez tartozik egy interpretáció (model theory) és egy bizonyításelmélet (proof theory). Ami nincs, az igazságelmélet
- Miért nincs? Mert “minden igaz”
- Vagy legalábbis majdnem minden. Megpróbálunk értelmet tulajdonítani mindennek amit hallunk/olvasunk ‘charitable interpretation’ és az nagyon ritka, hogy ez nem sikerül
- Milyen a háromszögletű kör?

ILYEN

