



**4lang**

# A fogalmi szótár

- feladatai:
  - az alapszókinccs meghatározása
  - a definíciós tevékenység támogatása
  - szövegmegértési feladatok támogatása
- a konceptuális szótár lemmái többnyelvűek (az eredeti változatban angol, magyar, latin, lengyel – ezért 4lang)
- fogalom: idea, az emberek fejében megjelenő konceptum; a 4lang gyökerei Platón idea- elméletétől a skolasztikus fogalomfelfogáson át Locke-ig (1689) és Fregéig (1879) vezethetők vissza nyelvfilozófiai értelemben
- az ideák a formális modellek tárgyai

# A szókincs

- gerincét a Longman Defining Vocabulary adja (2000 szó; ezt kiegészítették még Diederich (1939) 300 leggyakoribb latin szóra vonatkozó listájával, Whitney (1845) szanszkrit gyöklistájával, ill. több magyar és angol gyakorisági vizsgálat leggyakoribb szavaival), kb. 2700 szót tartalmaz
- ez önmagában nem alkalmas fogalmi szótárnak, mert ehhez meg kell oldani, hogy a szavak ugyanabban az értelemben forduljanak elő a definiendum és a definiens oldalán, ill. a szavak minden előforduló kombinációját is definiálni kellett minden olyan esetben, ahol definiensben is előfordul
- a körmentességet a a definíciók formális nyelvi eszközökkel való megragadása és gépi elemző építése tette lehetővé

- körmentesség alatt az értendő, hogy definiendán kívüli elem ne legyen egyetlen definiensben sem
- ez nem zárja ki, hogy legyenek primitíveknek tekintett elemek, ill. olyan párok (vagy n-esek), amelyek kikerülhetetlenek egymás definíciójában
- a konceptuális szótárnak nem kell eldöntenie, hogy a tyúk vagy a tojás volt-e előbb :)
- a definíciós módszerek garantálják, hogy ne legyen szükség a 4lang körén kívül eső elemre
- nemcsak szavakat, hanem kötött morféimákat is tartalmaz (ezt már a Longman Defining Vocabulary is tartalmazza), továbbá olyan szabályokat, amelyek a morfológiai összetétel során kialakuló jelentést is származtatni tudják az összetevők jelentéséből

# A fogalmak és a szavak viszonya

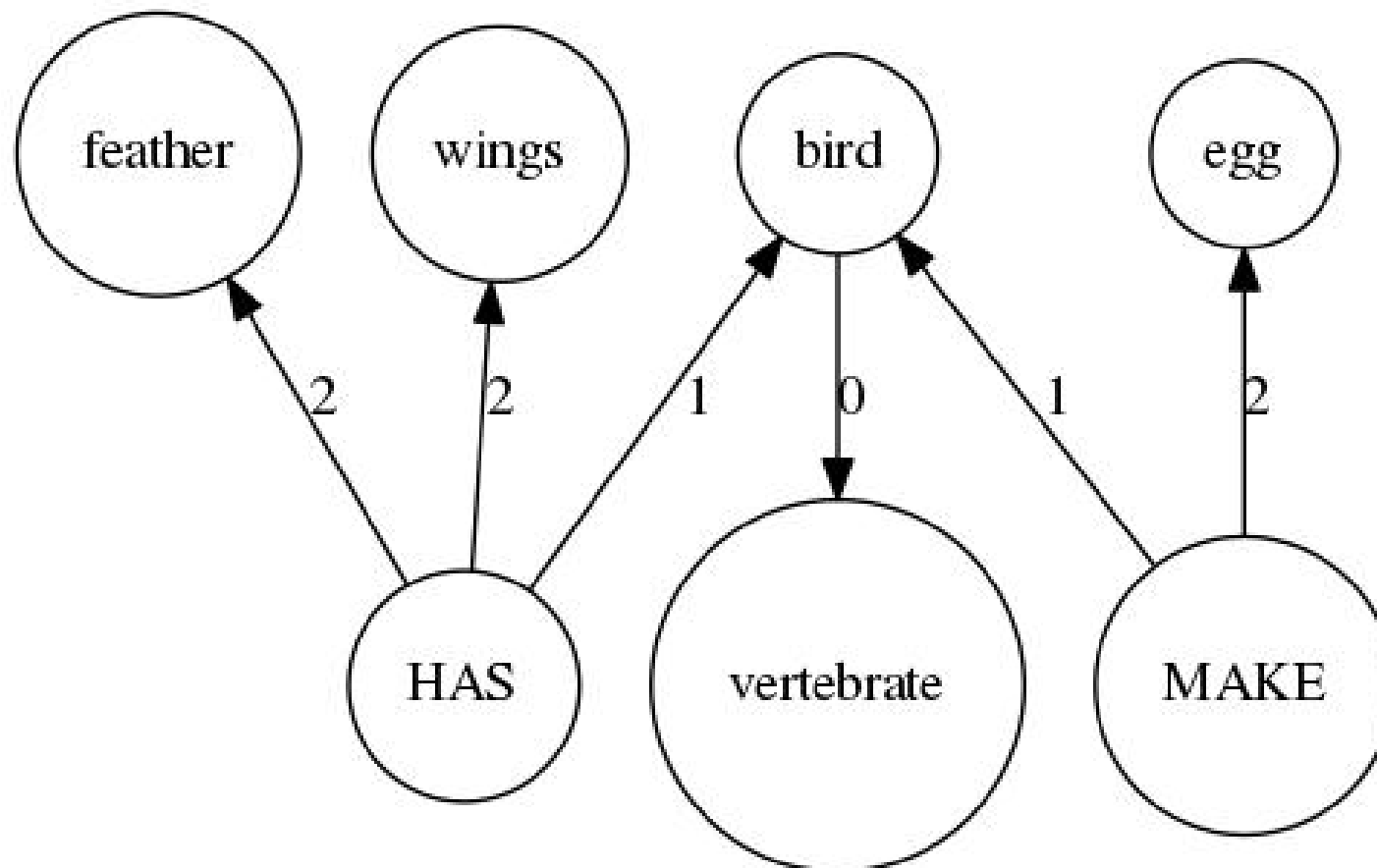
- a fogalmi szótár központi eleme a definíció (ezek emberi munkával készültek)
- példa (itt lehet megtalálni a definíciókat: <https://github.com/kornai/4lang/blob/master/4lang>):  
*cat macska felis kot 1575 u N mammal, <AT/2744 house>, CATCH/828 vertebrate, <pet/966>*
- a definíciót támogatja a grammatikai információ, ez egyrészt a szófajból, másrészt a definícióban előforduló mélyesetekből (argumentumstruktúrára vonatkozó információ) olvasható ki
- nem foglalkozik a fonológiai információval, stiláris értékkel, etimológiával stb.
- a definíció a fogalmak közti belső kapcsolatokat rögzíti

- a tiszta homonímiákat nehéz elkülöníteni a poliszémiától (megoldás: rugalmasabb szófajkezelés, ill. ahol lehet, általános, absztrakt jelentést feltételez)
- fontos elkülöníteni a lexikai és az enciklopédikus információt – az enciklopédikus tudás rögzítésével szükségtelen poliszémia állna elő
- a köznapi tudás reprezentációjára törekszik: akkor is tudták az emberek, hogy mi az a víz, amíg nem ismerték a kémiai összetételét
- ahol mégsem lehet kikerülni a tudományos definíciót, ott a wikipédiára mutató kereszthivatkozásokat használ
- a definíciók szintaxisa arra a feltevésre épít, hogy a primitívek listájára nem szükséges kettőnél több argumentumú elemet felvenni, mert ezeket mindig lehet egyszerűbben definiálni

- intranzitív elemek: a rájuk analitikusan jellemző predikátumok konjunkciója, pl. *happy boldog fortunatus szcze1s1liwy 326 A emotion, good*
- a primitív tranzitív elemek legnagyobb része grammatikai jellegű, pl. a főnevek birtoklásánál kikerülhetetlen a HAS: *box doboz capsa pudel1ko 478 u N container, HAS lid, cube*
- a tranzitívnál bonyolultabb argumentumstruktúrát egymásba ágyazott tranzitív relációkkal kezeli: *disappoint kia1bra1ndi1t fallo rozczarowac1 1324 V =AGT CAUSE[=PAT[sad]], before(=PAT EXPECT other), expect[fail] event struct*
- többszófajú elemek problémája, pl. 'fagy': az igénél és a főnévnél ugyanarról a fogalomról van szó, arról a folyamatról, amiben a víz szilárd lesz, vagy ennek okáról: *freeze fagy gelu mro1z 715 N cold CAUSE, before(liquid), solid[after, <ice>]*

# Gráfok

- online demo: <http://4lang.hlt.bme.hu/>
- példa: *bird* *mada1r* *avis* *ptak* 1576 u N *vertebrate*, *MAKE* *egg*, *HAS* *feather*, *HAS* *two(wing)*





- csomópontok: a fogalmak; jellemzőjük, hogy nincs semmilyen grammatikai tulajdonságuk (szófaj, eset, szám stb - nem tesz különbséget pl. a freeze(N) és a freeze(V) között)
- 0-ás él: tulajdonságokat, az *IS\_A* relációt és unáris predikációt reprezentál; így pl. a water freezes és a frozen water egyformán jelenik meg: water → freeze.
- 1-es és 2-es élek: bináris predikátumokat kapcsolnak az argumentumaikhoz (a fogalmi szótárban a binárisokat nagybetűvel jelölik): az 1-es él az alany, a 2-es él a tárgy

# Felhasznált irodalom

- Kornai András, Makrai Márton: A 4lang fogalmi szótár  
In: Tanács Attila, Vincze Veronika (szerk.)  
IX. Magyar Számítógépes Nyelvészeti Konferencia:  
MSZNY 2013. 376 p.
- Recski Gábor disszertációja: Computational methods  
in semantics