

Egy magyar WordNet felé

Miháltz Márton, Dr. Prószéky Gábor
MorphoLogic



Magyar Tudományos Akadémia
Nyelvtudományi Intézet



a
b
c
d
e
f
g
h
i
j
k
l
m

WordNet

- Mentális lexikon modell, Princeton 1985-
- Lexikális szemantikai hálózat
- Szójelentések, szinonimák, **synsetek**, fogalmak

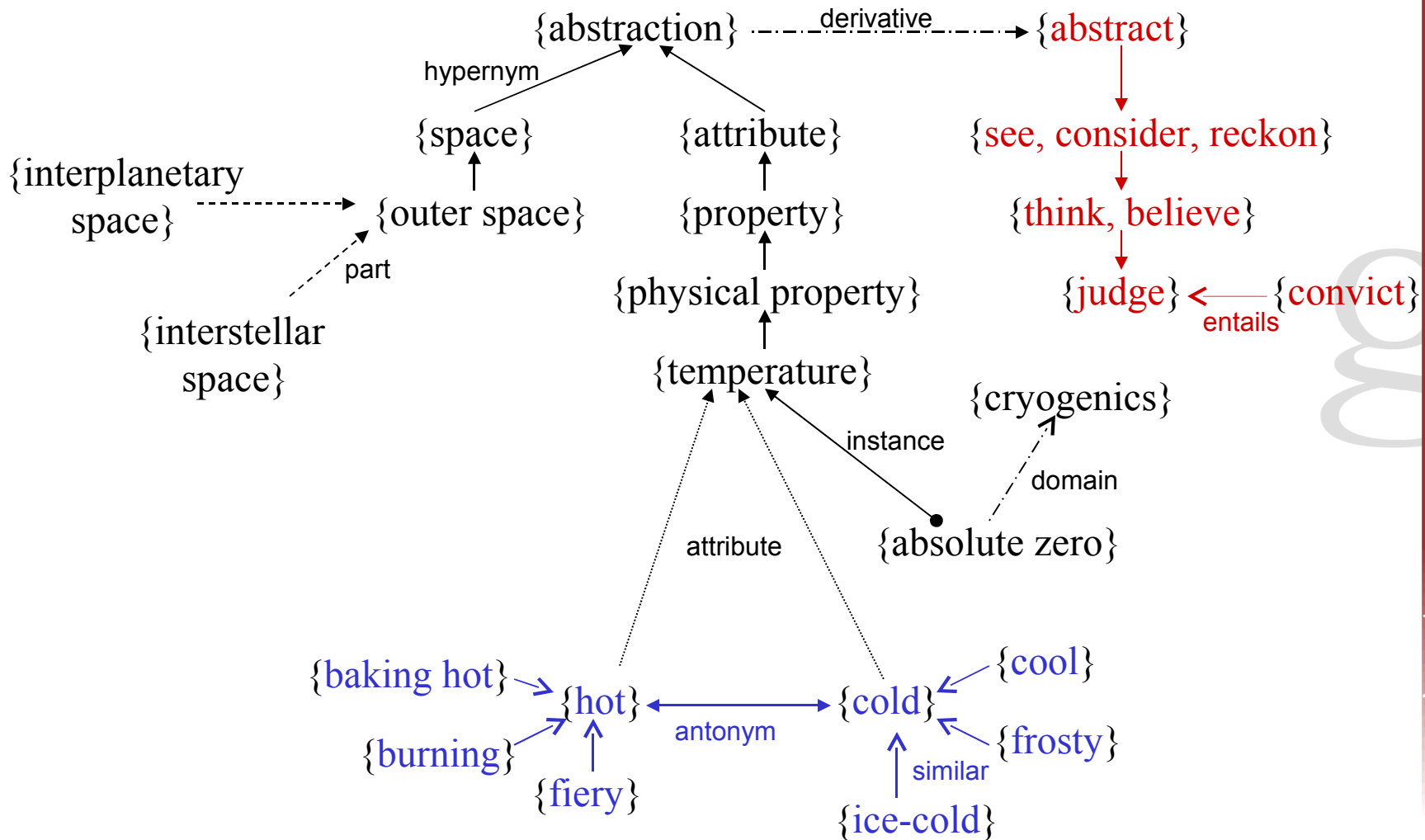
{ ló:1, háziló:1 } vs. { ló:2, huszár:2 }

- Szemantikai **relációk**

{ ló:1, háziló:1 } ← { póniló:1, póni:1 }

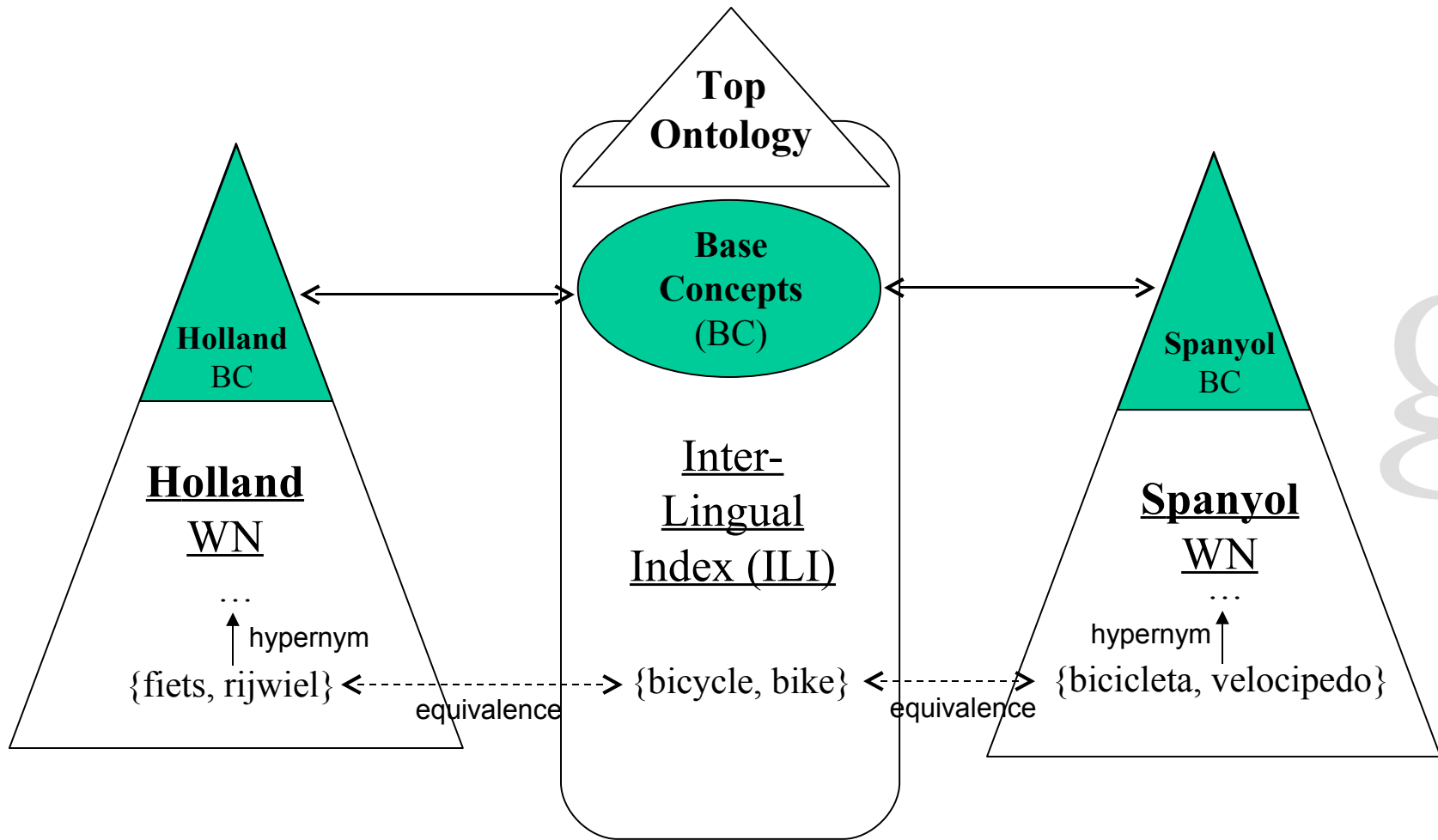
{ ló:2, huszár:2 } → { sakkfigura }

Princeton WordNet



a
b
c
d
e
f
g
h
i
j
k
l
m

EuroWordNet: többnyelvű WN



Magyar WN ontológia (HuWN)

- **BalkaNet projekt erőforrások használata**
 - „Mag” rész: BN Concept Set (8 516 synset 13 nyelv alapján)
 - BN Interlingual Index (PWN 2.0 + SUMO hierarchia)
 - VisDic editor
- **Kiterjesztéses modell (+ más)**
 - Angol synsetek fordítása, relációk átvétele
 - Alapos kézi ellenőrzés és javítás
- **Fél-automatikus módszerek**
 - Korábban kifejlesztett fordító heurisztikák
 - 70% körüli pontosság (főnevek)
- **Meglévő erőforrások integrációja**
 - Magyar Értelmező Kéziszótár meghatározásai
 - NYTI igei vonzatkeret-adatbázis

English WordNet 2.0.c

[n] dog:1, domestic dog:1, Car

View Tree RevTree BCS1,2 E

POS: n ID: ENG20-02001223-n
 BCS: 3
 Synonyms: dog:1, domestic dog:1, Canis familiaris:1
 Definition: a member of the genus Canis (probably descended from the common wolf) that has been domesticated by man since prehistoric times; occurs in many breeds
 Usage: the dog barked all night
 Domain: zoology
 SUMO/MILO: + Canine
 -->> [hypernym] *[n] canine:2, canid:1
 -->> [holo_member] +[n] Canis:1, genus Canis:1
 -->> [holo_member] +[n] pack:6
 <<-- [hyponym] [n] pooch:1, doggie:1,

Hungarian BCS

kutya [n] Canis familiaris:1, házikutya:1, kutya:, eb: [n] korcs kutya:1, keverék kutya: [n] őrző-védő kutya:

View Tree RevTree Edit Nouns Verbs Verl

Definiton
 A ragadozók közé tartozó, ház- és nyájé
 Non Lexicalized
 Usage
 A szomszéd kutyaja egész éjsz
 Note
 kutya_1_1
 Part of Speech
 n
 Synonyms: Literal, Sense, LNote
 Canis far 1 + -
 házikutya 1 + -
 kutya 1 + -

lagyar Ertelmezo Keziszotar (MorphoLogic Tezaurusz (r)

kutya [fn] kutya_1_1 [fn] kutya_1_2 [fn] kutya_1_3 [a] borzasztó, éktelen, c [n] kutya, eb, kutyus, ku [n] senki, senki fia, senki

View XML

POS: fn ID: kutya_1_1
 Headword(s): kutya
 Definition: A ragadozók kö
 tartozó, ház- és nyájörzés
 vadászatra stb. haszn.
 vagy kedvtelésből tartott
 háziállat.
 -->> [hypernym] háziállat
 POS: n ID: TEZ-06661-n
 Synonyms:
 kutya
 eb
 kutyus
 kutyuli

HuWN: igék

- Problémák

- Homályos jelentésbeli megkülönböztetések
- Inkonzisztens angol WN

Thematikus szerepek, metaforikus jelentések, szelekciós megkötések stb.

- Megoldás

- „Vegyes” metodológia:

BCS fordítás + MNSZ vonzatkeret-gyakoriság alapján kiválasztott igék, saját rendezés

- Specifikus magyar relációk

Igekötők, -képzők kezelése stb.

Eddigi eredmények

- BCS lefordítása
 - Gépi fordítás (fedés: 50%)
 - Synsetek és relációk kézi ellenőrzése, kiegészítése
- Számok

8,600	Synset
14,700	Szó
21,100	Szójelentés
25,200	Reláció



További munka

- Mag rész kiterjesztése
 - MNSZ és ÉKSz korpuszgyakoriságok alapján
- Ontológia további bővítése
 - Főnevek, melléknevek:
 - Iteratív koncentrikus bővítés PWN alapján
 - ÉKSz-ben feltárt szemantikai relációk alapján
 - Igék:
 - MNSZ vonzatkeret-gyakoriság alapján
 - PWN alapján
 - **2007: kb. 40K synset**

Köszönjük a figyelmet!

a
b
c
d
f
h
i
k
l
m