

I Workshop Per-Fide: Construção, exploração e aplicação de Corpora Paralelos
Universidade do Minho, 16 de Setembro de 2010

Um lexicógrafo e uma ontologista no meio da escada: sobem ou descem?

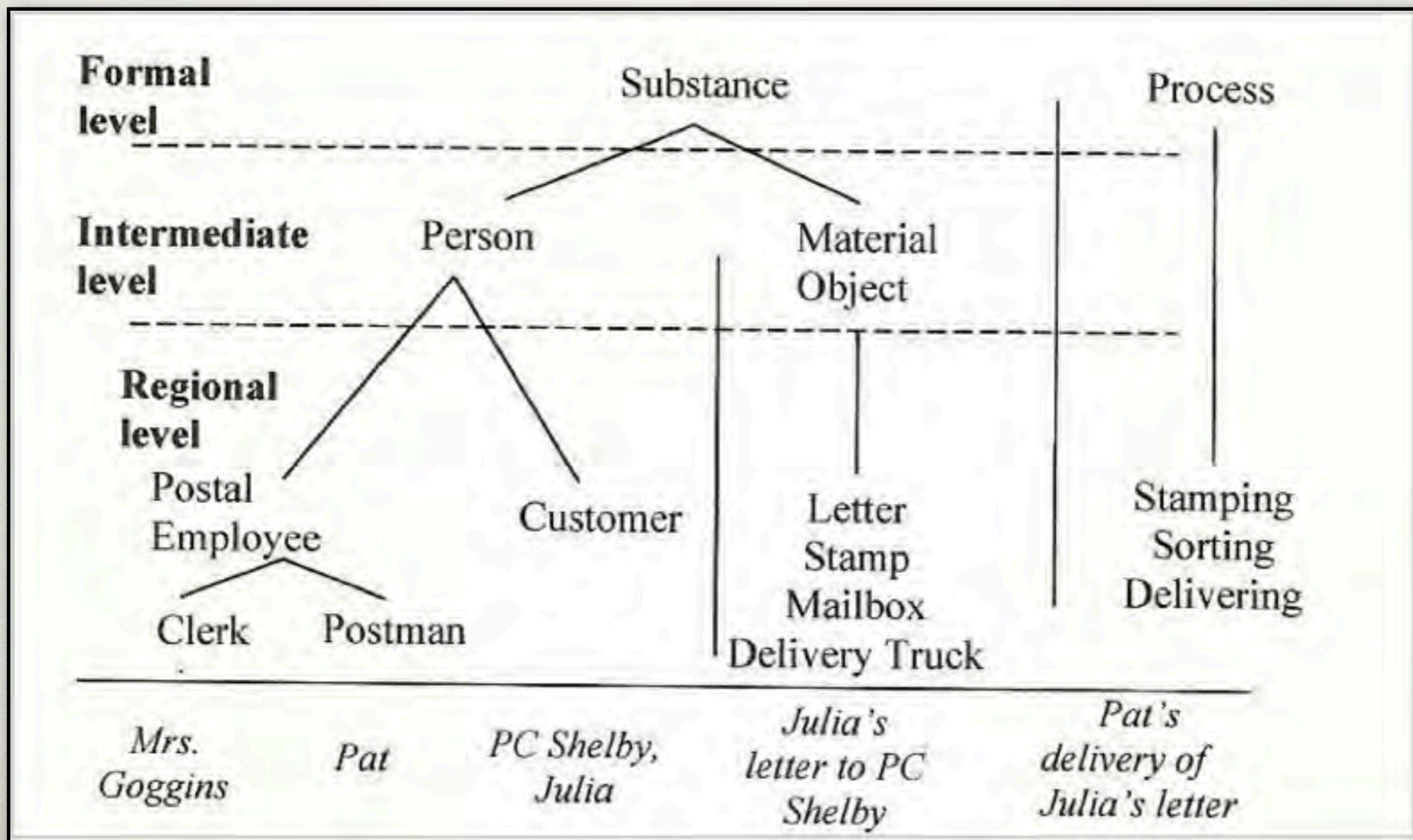
Álvaro Iriarte Sanroman
Patrícia Cunha França

Sumário

- Introdução
- Metodologias para elaboração de ontologias modernas
 - Top-down
 - Bottom-up
 - Middle-out
- Ontologias, *corpora* e linguagem natural
 - Ontologias linguísticas e suas limitações
- Propostas para tratamento das estruturas lexicais nas ontologias e outras bases de dados lexicais
- Conclusões

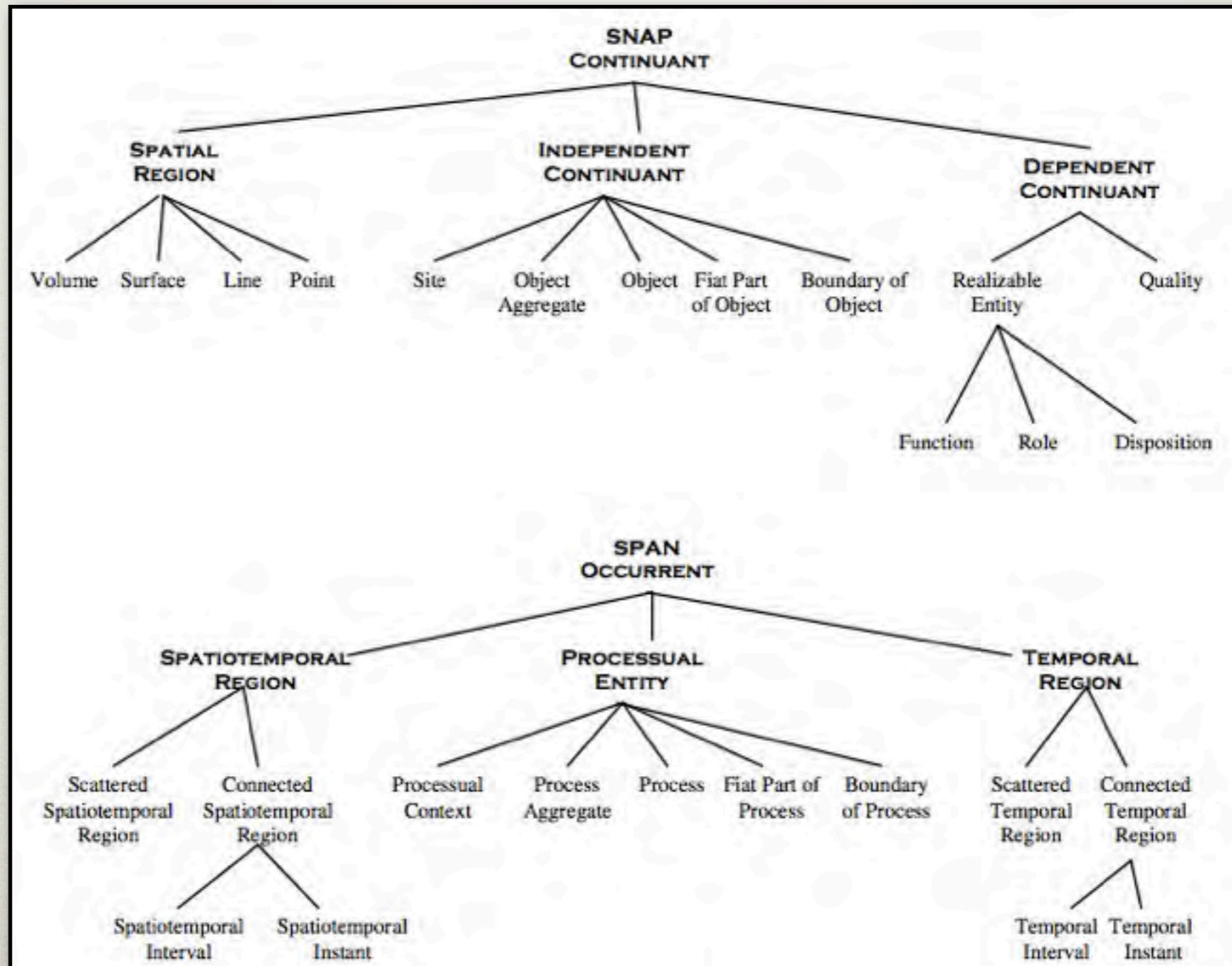
Metodologias para elaboração de ontologias modernas

As metodologias top-down, bottom-up e middle-out:



Metodologias para elaboração de ontologias modernas

BFO (Basic Formal Ontology)



Ontologias, corpora e linguagem natural

- **Ontologias e bases de dados lexicais**

“An ontology [...] is a set of categories and objects or ideas in the world, along with certain relationships among them; it is not a linguistic object.” (Hirst, 2004: 8)

“All ontologies in information science contain terms. (...) the experts in the various specialized domains of knowledge generally look through the terms. However, an ontology such as WordNet presents a special case, for (if it is to be called an ontology at all) it is an ontology of terms and meaning; it is like a dictionary, not like a taxonomical textbook (...). It is clear that the term ‘cat’ is mentioned and not used in WordNet. Both the scare quotes around the term ‘cat’ and the fact that it is preceded by the term ‘noun’ makes it clear that WordNet contains no talk of real cats.” (Johansson, 2008: 303)

Ontologias, corpora e linguagem natural

Noun

- **S: (n) cat, true cat** (feline mammal usually having thick soft fur and no ability to roar: domestic cats; wildcats)
 - [direct hyponym](#) / [full hyponym](#)
 - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
- **S: (n) guy, cat, hombre, bozo** (an informal term for a youth or man) "a nice guy"; "the guy's only doing it for some doll"
- **S: (n) cat** (a spiteful woman gossip) "what a cat she is!"
- **S: (n) kat, khat, qar, quat, cat, Arabian tea, African tea** (the leaves of the shrub *Catha edulis* which are chewed like tobacco or used to make tea; has the effect of a euphoric stimulant) "in Yemen kat is used daily by 85% of adults"
- **S: (n) cat-o'-nine-tails, cat** (a whip with nine knotted cords) "British sailors feared the cat"
- **S: (n) Caterpillar, cat** (a large tracked vehicle that is propelled by two endless metal belts; frequently used for moving earth in construction and farm work)
- **S: (n) big cat, cat** (any of several large cats typically able to roar and living in the wild)
- **S: (n) computerized tomography, computed tomography, CT, computerized axial tomography, computed axial tomography, CAT** (a method of examining body organs by scanning them with X rays and using a computer to construct a series of cross-sectional scans along a single axis)

Noun

- **S: (n) bird** (warm-blooded egg-laying vertebrates characterized by feathers and forelimbs modified as wings)
 - [direct hyponym](#) / [full hyponym](#)
 - [part meronym](#)
 - **S: (n) beak, bill, neb, nib, pecker** (horny projecting mouth of a bird)
 - **S: (n) furcula** (a forked bone formed by the fusion of the clavicles of most birds)
 - **S: (n) feather, plume, plumage** (the light horny waterproof structure forming the external covering of birds)
 - **S: (n) wing** (a movable organ for flying (one of a pair))
 - **S: (n) pennon, pinion** (wing of a bird)
 - **S: (n) bird's foot** (the foot of a bird)
 - **S: (n) uropygium** (posterior part of a bird's body from which the tail feathers grow)
 - **S: (n) hindquarters, croup, croupe, rump** (the part of an animal that corresponds to the human buttocks)
 - **S: (n) air sac** (any of the membranous air-filled extensions of the lungs of birds)
 - **S: (n) uropygial gland, preen gland** (oil-secreting gland situated at the base of the tail in most birds)
 - **S: (n) syrinx** (the vocal organ of a bird)
 - **S: (n) bird, fowl** (the flesh of a bird or fowl (wild or domestic) used as food)

◦ [member holonym](#)

Ontologias, corpora e linguagem natural

- **Ontologias linguísticas (definição)**

“ [are] large lexical resources that cover most words of a language, while at the same time also providing an ontological structure where the main emphasis is on the relations between concepts; linguistic ontologies can therefore be seen both as a particular kind of ontology.” (Magnini & Speranza, 2002:43)

Ontologias, corpora e linguagem natural

Algumas limitações apontadas à WordNet:

- ausência de informação semântica ao nível textual ou do discurso (“*tennis problem*”) (Fellbaum, 1999: 10);
- incapacidade de ligar o significado (“*meaning*”) ao uso (Hanks, 2003);
- exemplos fabricados, pouco naturais (Hanks, 2003);
- ausência de indicação de frequência relativa ou de importância dos diferentes sentidos (Hanks, 2003);
- fraseologia inconsistente e confusa (ex.: os *phrasal verbs*: ‘*deal with*’ não é a mesma coisa que ‘*deal*’) (Hanks, 2003);
- inclusão de sinónimos que não são sinónimos (‘*idea*’ = ‘*content*’) (Hanks, 2003);
- excesso de hierarquia (Hanks, 2003);
- por vezes os exemplos não se aplicam a todos os membros dos *synsets* a que estão ligados (Hanks, 2003).
- ausência de opções de preferências sintagmáticas (ex.: [[Person]] abandon [[Process]]: *abandon a project*; [[Person 1]] abandon [[Person 2 | Animate (=Pet)]]: *abandon a child | dog*. (Hanks, 2003).

Propostas para tratamento das estruturas lexicais nas ontologias e outras bases de dados lexicais

do conceito ao termo?



... ou do termo ao conceito?

HAIMAN, J. (1980): «Dictionaries and Encyclopedias»,
FRAWLEY, W. (1981): «In Defense of the Dictionary: A
Response to Haiman»,

«**Horse**: an animal called 'horse'. ...
Elephant: an animal of the species 'elephant'...»
(Haiman, 1980: 330).

Semiótica (Morris)	Linguística
sintaxe	fonética fonologia morfologia sintaxe semântica
semântica	semântica
pragmática	pragmática

Linguística “estrutural”

fonética
fonologia
morfologia
sintaxe

semântica intensional

~~semântica extensional~~

~~pragmática~~

Linguística “estrutural”	dicionário
fonética fonologia morfologia sintaxe semântica intensional	
	<input checked="" type="checkbox"/>
	<input checked="" type="checkbox"/>
	<input checked="" type="checkbox"/>
	<input checked="" type="checkbox"/>
semântica extensional	<input checked="" type="checkbox"/>
pragmática	<input checked="" type="checkbox"/>

informação recolhida num dicionário

informação gramatical

informação semântica (“estrutura interna do significado”)

informação referencial

combinatória lexical

informação sobre o uso

informação recolhida num dicionário

informação gramatical

informação semântica (“estrutura interna do significado”)

informação referencial

combinatória lexical

informação sobre o uso

O conceito **colocação** é utilizado em linguística com dois sentidos diferentes:

a) «las combinaciones específicas, relevantes, cotidianas que regularmente son empleadas para enunciar los acontecimientos típicos alrededor del vocábulo» (Irsula, 1992);

ou

b) combinação aparentemente livre de palavras, gerada a partir das regras da língua, mas onde actua qualquer tipo de restrição lexical determinada pela norma.

O meu gato de estimação

Mi gato de compañía

My pet cat

estimação = estimación = estimation ?

estimação = companhia = pet?

alinhamento à palavra (NATools)?:

□	+	■	/	◇	+	○
↓		↓		↓		↓
□'	+	■'	/	◇'	+	○'

ter mau há lito	tener mal aliento
tomar fôlego	tomar aliento
palavras de alento	palabras de aliento

to have a miscarriage	ter um aborto
to have a abortion	fazer um aborto

ser	o	braço direito	de
-----	---	---------------	----

ser	o	melhor	amigo	de
-----	---	--------	-------	----

*reactor nuclear
cópia de segurança
dar um salto
dar uma volta
ter bom ouvido,*

ser

o

braço direito

de

jogar

com

o

braço

direito

animal de estimação

pet

Casar o homem. Vxorem ducere. **casala mulher.**s.darlhe marido. Elo loco, as, donuptum
casar a mulher. Nubo, is, nupsi. **casalo homem.**s.darlhe mulher. Vxorendo.

(et passim).

Jerónimo Cardoso (ca. 1562) *Dictionarium ex lusitanico latinum sermonem*

Cantar el tordo o zorzal. trutilo.as.avi.

Cantar el estornino. pisito.as.avi.

Cantar la perdiz. cacabo.as.avi.

Cantar la tórtola o paloma. grino.is.

Cantar la paloma torcaz. plausito.as.

(et passim)

Cabestrar bestia. capistro.as.

Caçar aves. aucupor.aris.

Caçar fieras o montar. venor.aris

Capitanear gente. duco exercitum

Antonio de Nebrija (ca. 1494) *Vocabulario de romance en latin*

Conclusões

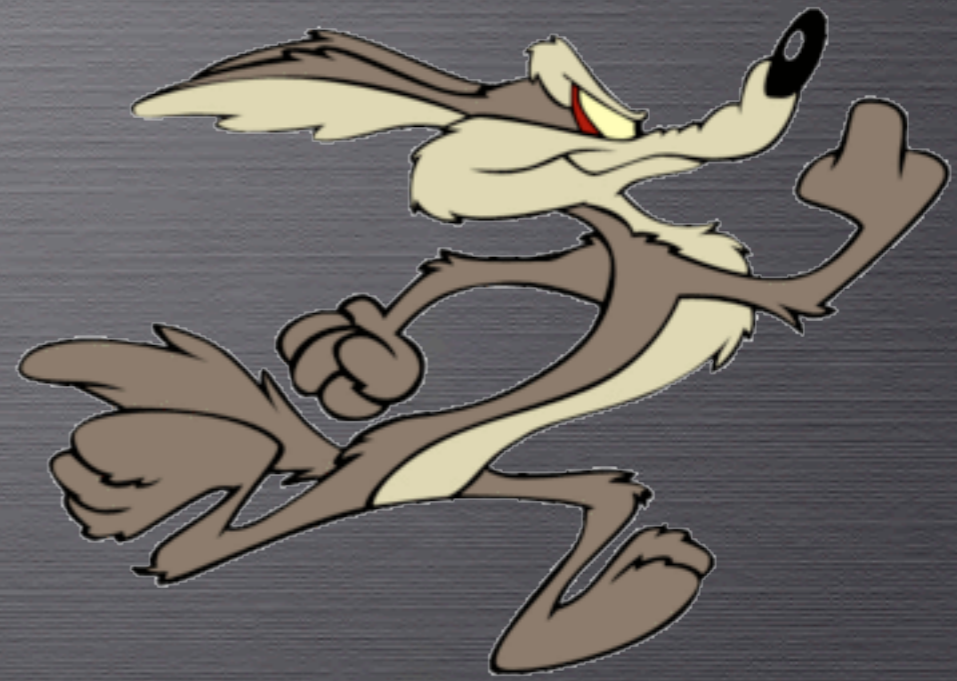
- As bases de dados lexicais deverá basear-se num sistema referencial que seja capaz de relacionar estruturas lexicais com estruturas de conhecimento

Conclusões

- As bases de dados lexicais deverá basear-se num sistema referencial que seja capaz de relacionar estruturas lexicais com estruturas de conhecimento
- Dicionários onomasiológicos (orientados ao conceito) \neq dicionários combinatórios, mas...

Conclusões

- As bases de dados lexicais deverão basear-se num sistema referencial que seja capaz de relacionar estruturas lexicais com estruturas de conhecimento
- Dicionários onomasiológicos (orientados ao conceito) \neq dicionários combinatórios, mas...
- Os dicionários, as bases de dados lexicais, os *corpora* são repositórios de algumas “amostras” de significados, isto é, dos *sentidos* que as palavras adquirem em determinados contextos. São, portanto, dicionários de *parole*, não de *langue*



alvaro@ilch.uminho.pt

patrticiacunhafranca@gmail.com