

## *Sobre codificació semàntica*

Mercè Lorente  
Seminari IULATERM  
12 de juny de 2003

## Guió-Parts

1. Què vol dir, ara per ara, codificar la semàntica en aplicacions informàtiques?
2. El projecte CLIPS (ILC-CNR, Itàlia)

## Codificar el significat

- Què es codifica?
- En quines aplicacions? Estan connectades les aplicacions?
- Amb quina finalitat es pot codificar la semàntica?
- Projectes en marxa
- Hi ha estàndards?
- Models semàntica vs. sintaxi

## Què se sol codificar?

- Lemes de diccionaris computacionals (d'abast general)
- Acepcions de cada lema
- Classes semàntiques
- Relacions semàntiques
- Vinculació de cada acepció a la sintaxi
- Dominis temàtics

## Aplicacions

- En fases de Preprocés, en el PLN
- Diccionaris computacionals per al PLN
- Ontologies conceptuais
- Jerarquies lèxiques
- Bases de dades terminològiques

– Ara per ara, sense connexió en la majoria de casos

## Finalitats

- Disponibilitat de recursos de referència
- Marcatge de corpus
- Extracció d'informació
- Considerat principal: Accés multilingüe a la informació
- Secundaris: MT, CLIR

– Sembla que, si es tracta de RI de documents, hi ha poc interès per costos alts



## Estàndards

- ISLE Computational Lexicon Working Group  
[http://lingue.ilc.cnr.it/EAGLES964isle/ISLE\\_Home\\_Page.htm](http://lingue.ilc.cnr.it/EAGLES964isle/ISLE_Home_Page.htm)
  - Is committed to the consensual definition of a standardized infrastructure to develop multilingual resources for HTL applications, with particular attention to the needs of MT and CLIR (Atkins et al. 2002) →
  - MILE (Multilingual ISLE Lexical Entry)



## Models

- Els tipus per a la codificació semàntica deriven de la sintaxi (Levin 1993)
- La semàntica és autònoma de la sintaxi (FrameNet, Baker et al. 1998)
- La sintaxi i la semàntica són codificades autònomament però hi ha vinculació (Simple, Clips)

## Segona part

*El projecte CLIPS*  
ILC-CNR  
Pisa (Itàlia)

## El projecte CLIPS

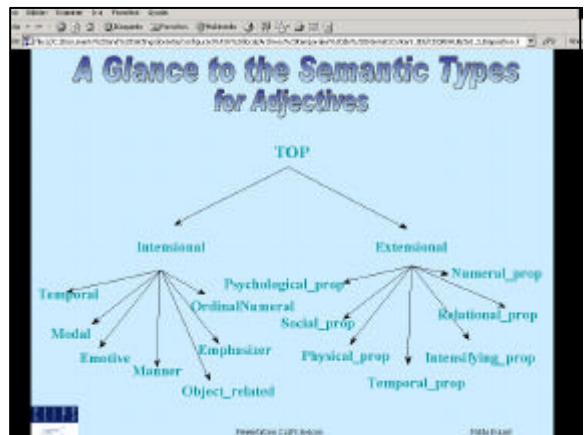
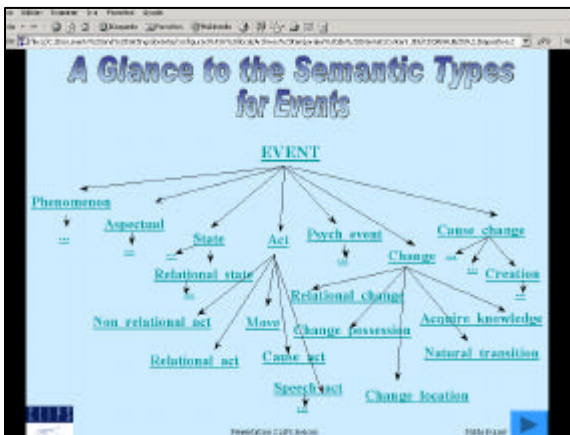
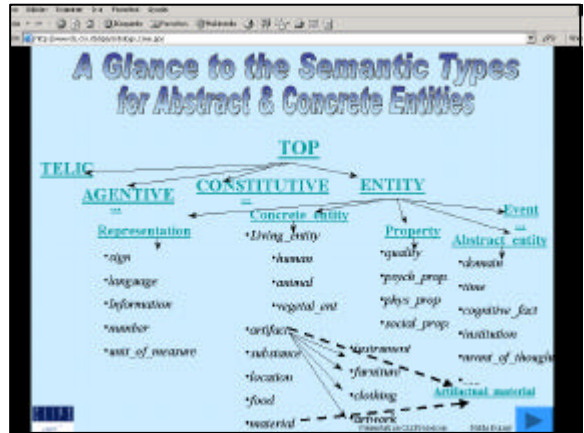
- Istituto di Linguistica Computazionale (CNR, Itàlia)
- Directora: Nicoletta Calzolari
- Directora científica: Nilda Ruimy
- Grup: 2 investigadors, 2 tècnics, 3/6 codif
- Finançament: Estat italià
- Durada: 3 anys (2000-2002) + revisió 2003 + continuació?
- Antecedent: Projectes PAROLE i SIMPLE
- Objectiu: Lèxic computacional de referència de l'italià amb codificació fonològica, morfològica, sintàctica i semàntica (60.000 lemes)

## Codificació CLIPS (1)

- Vincles entre semàntica i sintaxi
  - Cada accepció amb una estructura de constituents determinada
  - Codificació de predicats (diverses accepcions, amb estructures de constituents diferents, poden estar vinculades a un únic predicat
    - Nombre d'arguments
    - Paper semàntic de cada argument
    - Restriccions de selecció lèxica (amb lemes, USem o tipus semàntics)

## Codificació CLIPS (2)

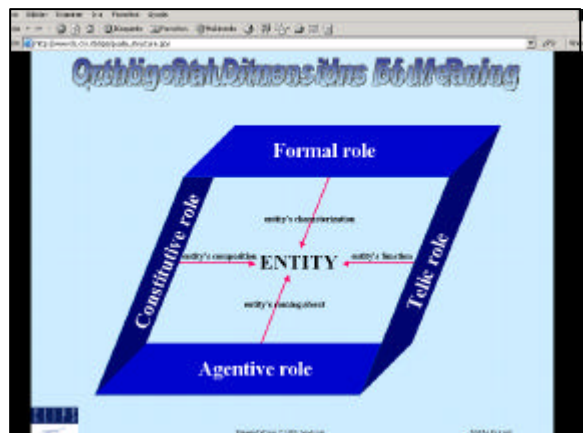
- Semàntica
  - Aceptació = USem
  - Tipus vinculat a una ontologia
  - Classe semàntica vinculada a una segona ontologia (LexiQuest)
  - Domini d'ús
  - Glossa
  - Polisèmia regular (control)
  - Relació amb derivats morfològics
  - Sinònims
  - Qualia (Pustejovsky 1995)



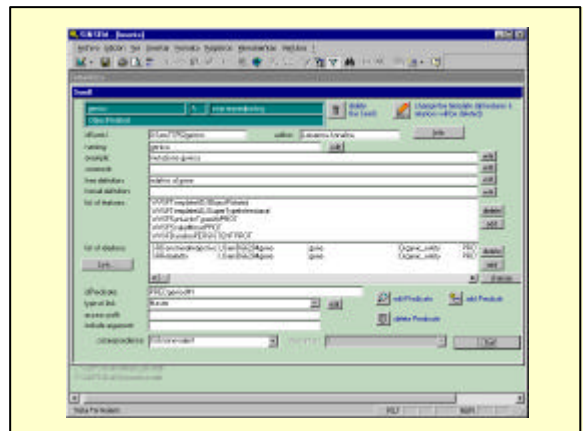
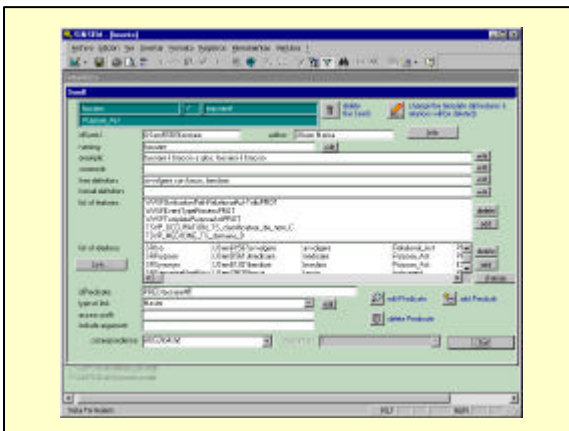
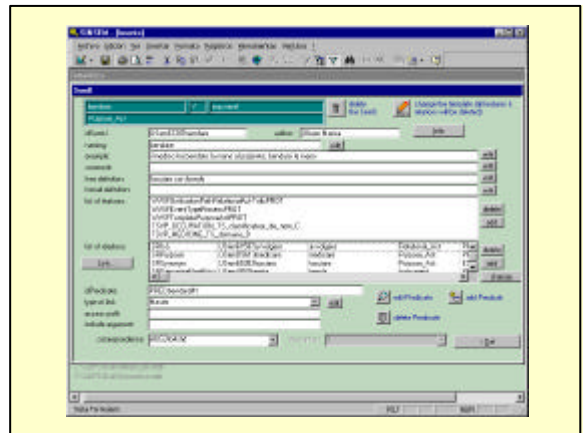
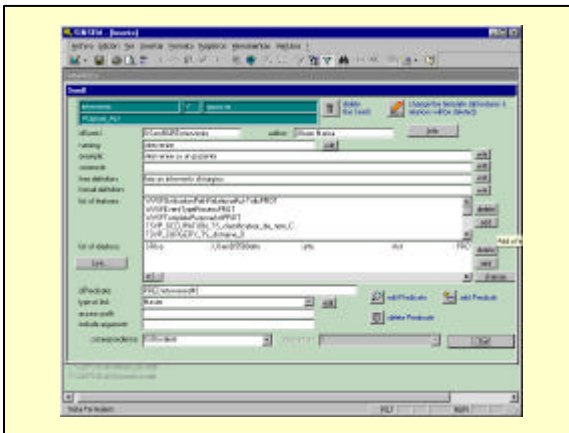
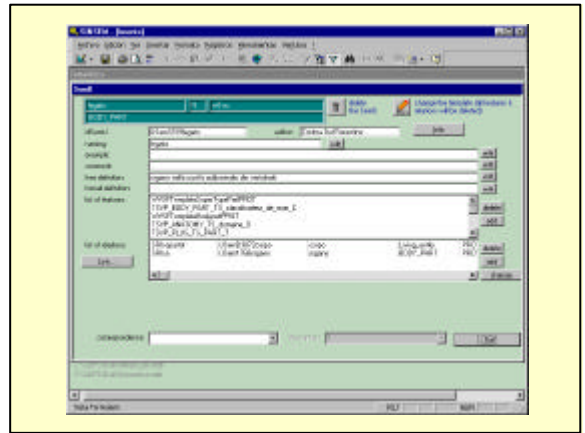
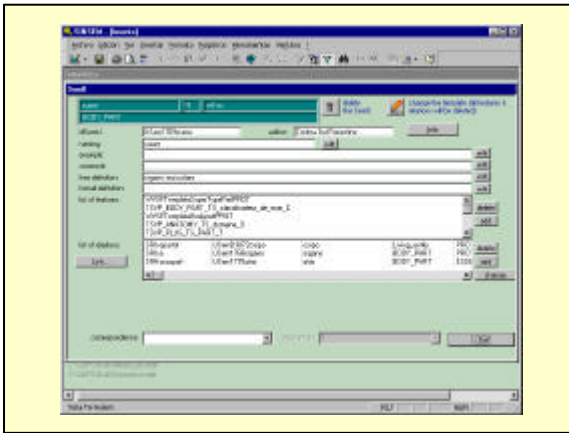
## Estructura de Qualia

Font: Pustejovsky 1995

1. Qualia Formal (V,N,A): Hiperònim (mínim) *is\_a*
2. Qualia Constitutiva (N,A): Relacions ontològiques i meronímiques. De caràcter bidireccional: Ex. *is a part of; has as part*
3. Qualia Agentiva (N,V): Origen. Ex. *created by; result of*
4. Qualia Tèlica (N,V): Funció. Ex. *used as; is the activity of*







## Noció de Template

- “Il Template è uno ‘schema’, una struttura che vincola ciascun tipo semantico ad un corredo strutturato di informazioni considerato cruciale per la definizione di quel tipo (...) Il Template funge da interfaccia tra l’ontologia ed il lessico.” (CLIPS Specifiche, Livello semantico, p. 10)
- Funcions de control d’errors i de generalització (codificació completa prèvia)

## Per continuar ...

- Documentació del projecte CLIPS
- Base de dades de CLIPS
- Publicacions penjades a Internet dels diversos projectes esmentats
- Revisió de casos concrets: *La codificació de verbs sota CLIPS. L’anàlisi dels verbs del genoma.*