

# Verso un'ontologia per la rappresentazione della conoscenza linguistica in Senso Comune

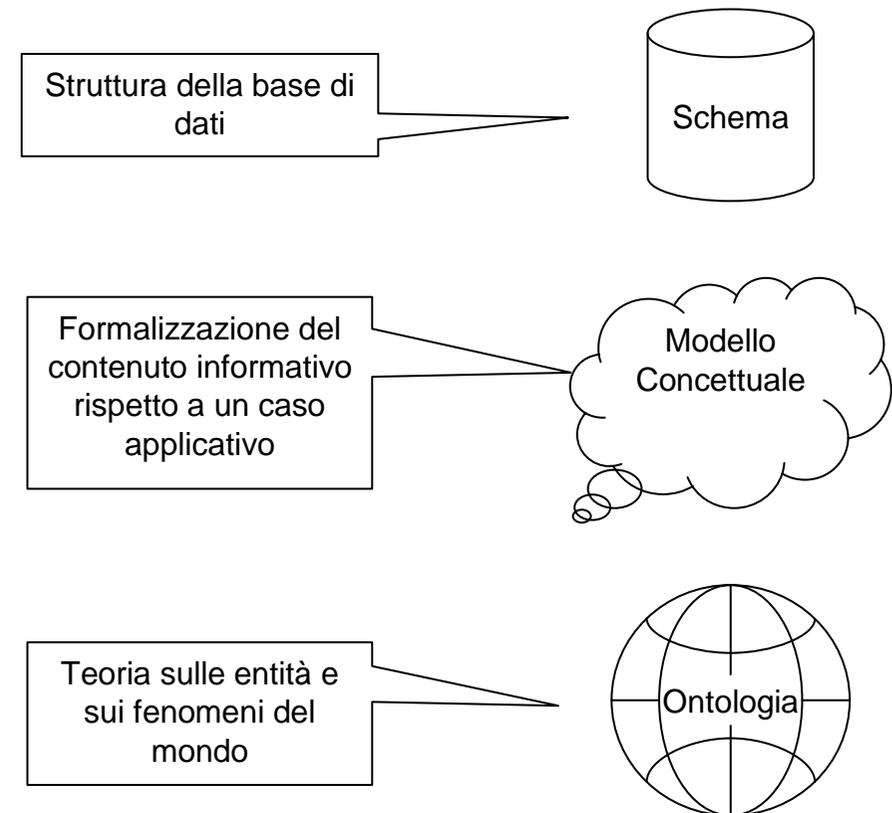
Guido Vetere  
Senso Comune  
IBM Centro Studi Avanzati





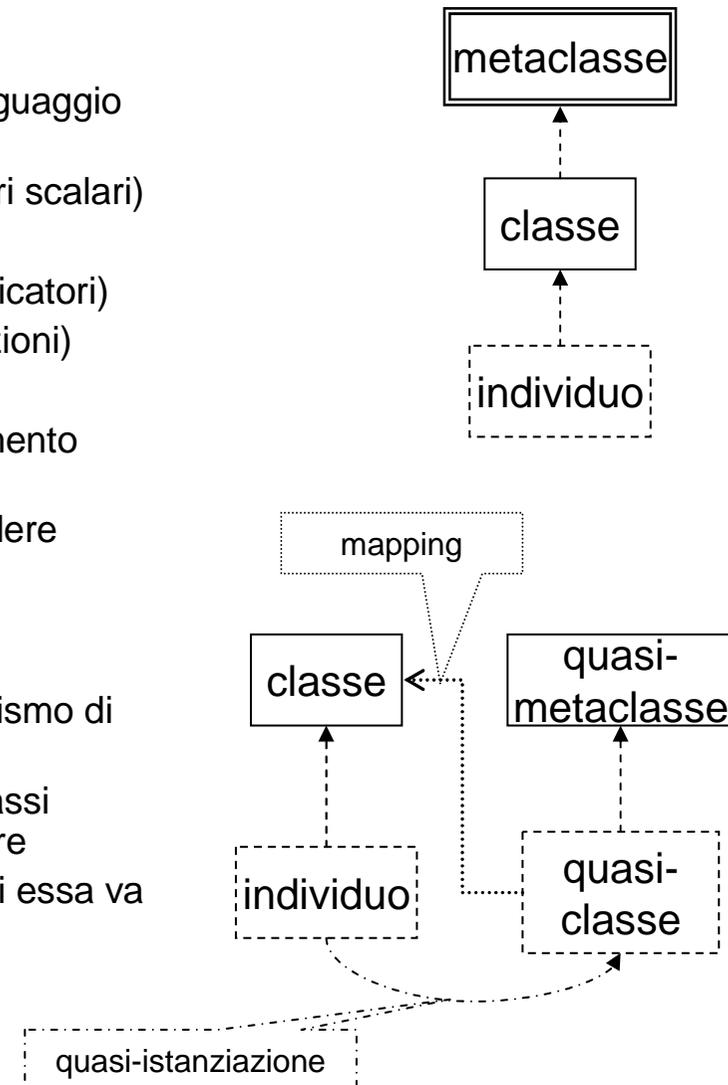
## Rappresentazione della conoscenza linguistica

- Dal punto di vista tecnico, la conoscenza linguistica di Senso Comune è contenuta in un database relazionale, attraverso un apposito schema logico
- Lo schema è stato ottenuto da un modello concettuale che si è andato perfezionando negli anni
- Tale modello incorpora le comuni nozioni linguistiche e lessicografiche, comprese le articolazioni grammaticali e la struttura semasiologica
- In più, mira a connettere le unità significative del linguaggio con specifiche rappresentazioni del *piano del contenuto* (dette *ontologie*)
- Da questa esigenza di integrazione nasce l'idea di guardare al *piano dell'espressione* e alla sua relazione col *piano del contenuto* in una prospettiva "ontologica"
- Si tratta cioè di guardare ad alcune nozioni della linguistica computazionale (e non) con un chiaro *impegno* verso i fenomeni che dovrebbero modellare
- Il lavoro (non privo di difficoltà) è nelle fasi preliminari



## Il metalinguaggio

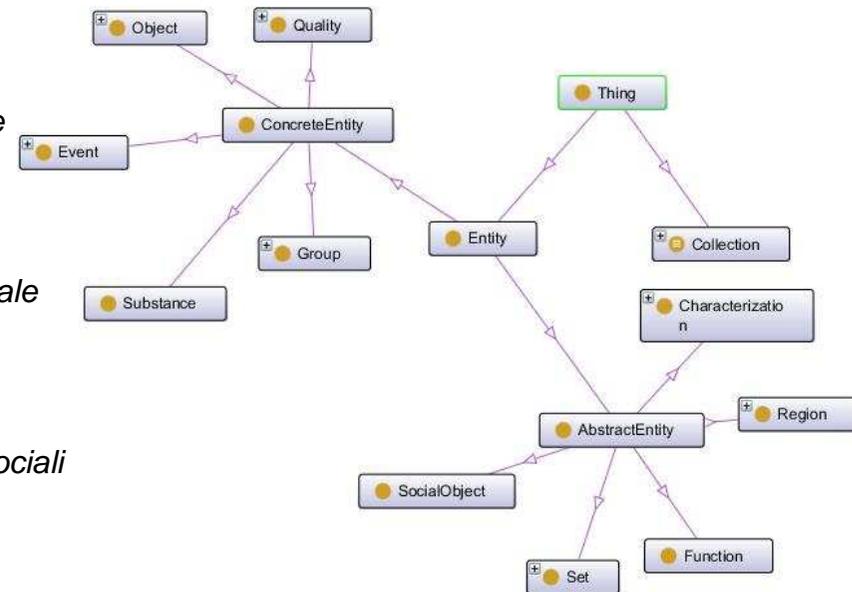
- Per la formalizzazione dell'ontologia si è scelto di usare un linguaggio logico descrittivo di primo ordine (Ontology Web Language)
  - Individui (insieme numerabile di identificatori univoci, valori scalari)
  - Predicati (unari, binari)
  - Operatori (congiunzione, disgiunzione, negazione, quantificatori)
  - Asserzioni terminologiche (equivalenza, inclusione, restrizioni)
  - Asserzioni individuali (classificazioni)
- Questi linguaggi permettono il ricorso a procedure di ragionamento automatico ben conosciute e supportate
- Alcune proprietà e relazioni linguistiche sembrano però richiedere logiche più potenti (secondo ordine)
  - es. *sostantivo(casa)*, *femminile(casa)*, *classe(sostantivo)*, *genere(femminile)*, *attributo(sostantivo,genere)*
  - Questa espressività è approssimata mediante un meccanismo di “reificazione”
    - I predicati possono essere associati ad individui di classi opportune (metaclassi), su cui si può dunque predicare
  - La reificazione è logicamente opaca, il ragionamento su di essa va gestito con procedure ‘ad hoc’
  - Allo studio l’uso del “type punning” (OWL2)





## Senso Comune

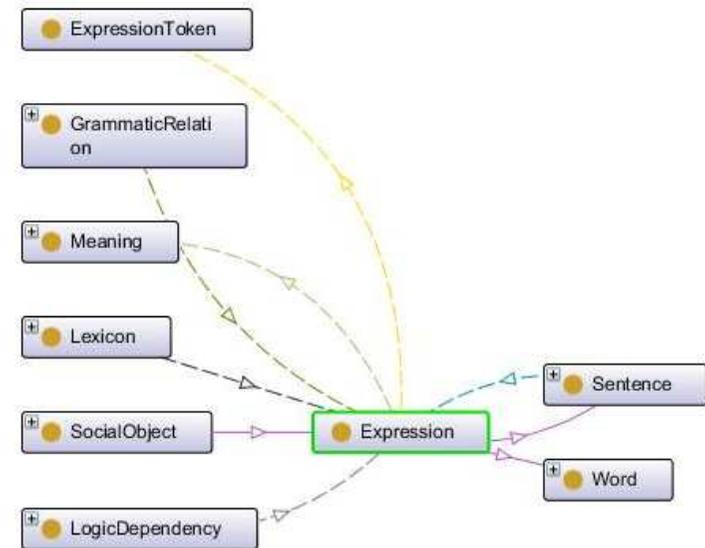
- Il Top Level di Senso Comune si può considerare una versione nominalistica (senza universali) di DOLCE
  - **Entity**
    - *Qualsiasi cosa identificabile*
    - DOLCE: particular
  - **Collection**
    - *Somma arbitraria di entità*
    - DOLCE: arbitrary-sum + set
  - Entity > **ConcreteEntity**
    - *Entità con una definita collocazione spazio-temporale*
    - DOLCE: spatio-temporal particular
    - es. questo tavolo, questa riunione
  - Entity > **AbstractEntity**
    - *Entità senza una definita collocazione spazio-temporale*
    - DOLCE: abstract + non-physical endurant
    - es. il contenuto di questa presentazione
  - ...> AbstractEntity > **SocialObject**
    - *Entità astratte la cui identità è stabilita nei processi sociali*
    - DOLCE: non-physical-object, DnS: social-object
    - es. un segno linguistico
  - ...> AbstractEntity > **Characterization** (quasi-metaclassa)
    - *Funzione di corrispondenza (mapping) tra n-uple di entità e valori di verità*
    - hasInstance some (Entity or Collection)
    - classMapping some URI (mappatura con nomi di classi)





# Significanti

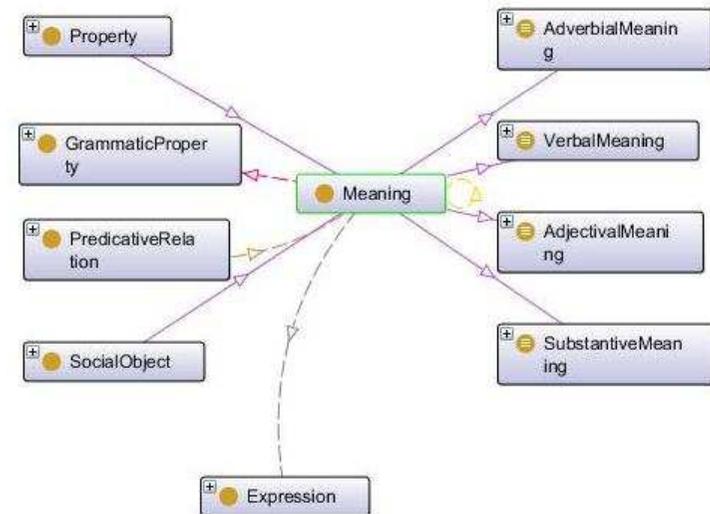
- Il *piano dell'espressione* (parole, frasi, proprietà e relazioni grammaticali) è collocato nelle astrazioni (type)
  - .. > SocialObject > **Expression**
    - *Astrazione dei segmenti verbali o scritti combinabili, componibili e associabili a un contenuto di esperienza (significato)*
    - hasMeaning some Meaning
    - hasInstance some (Graph or Utterance)
  - .. > Expression > **Word**
    - *Espressione autonoma minima, che può essere analizzata solo in unità dipendenti (morfemi)*
    - hasAttribute some WordAttribute
    - hasClass max 1 WordClass
  - .. > Expression > **Sentence**
    - *Espressione autonoma che può essere analizzata nei termini di altre espressioni*
    - hasMember min 1 Expression
  - ..> Characterization > Property > **GrammaticProperty**
    - *Proprietà che caratterizzano le espressioni*
    - hasInstance only Expression
  - ..> GrammaticProperty > **WordClass**
    - *Caratterizzazione del ruolo che le parole svolgono nella frase (parti del discorso)*
  - ..> GrammaticProperty > **WordAttribute**
    - *Caratterizzazione delle proprietà grammaticali relative alla classe della parola (genere, numero, ecc)*





## Significati

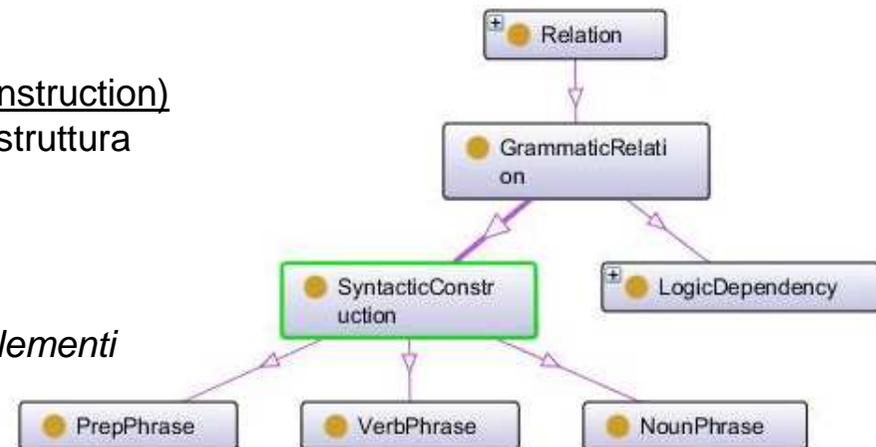
- L'articolazione del *piano del contenuto* si basa sulla reificazione del rapporto tra espressioni e entità (oggetti dell'esperienza)
  - ..> SocialObject > **Meaning**
    - *Caratterizzazione socialmente costruita mediante espressioni linguistiche*
    - meaningOf some Word
    - hasDefinition some Sentence
    - hasConstraint some GrammaticProperty
    - hasMeaningRelation some Meaning
      - ..> synonym, hyponym, ..., extension





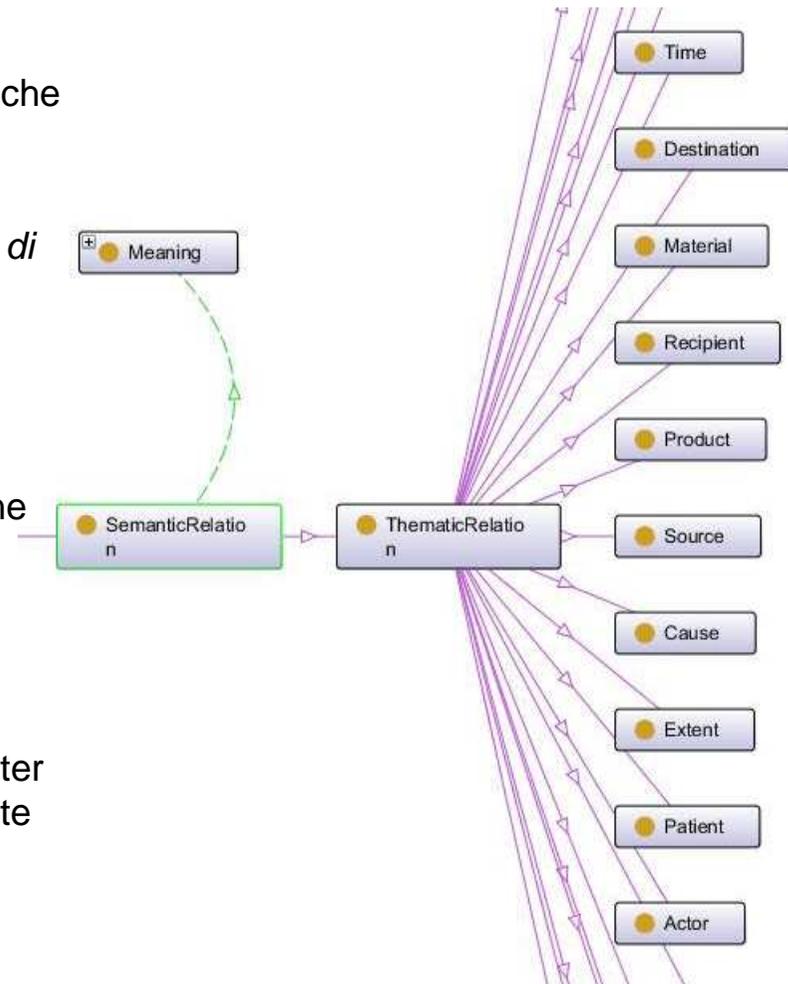
## Relazioni grammaticali

- Sul *piano dell'espressione* si individuano costruzioni e dipendenze
  - ..> GrammaticRelation > **SyntacticConstruction**
    - *Caratterizzazione della frase basata sulla realizzazione di regole sintattiche*
    - head max 1 WordClass
    - constituent some (WordClass or SyntacticConstruction) (nb: non può pienamente rappresentare una struttura sintagmatica)
    - NP, VP, NP, etc
  - ..> GrammaticRelation > **SyntacticDependency**
    - *Caratterizzazione di dipendenze binarie tra elementi (anche complessi) di una frase*
    - head exactly 1 (WordClass or Word or SyntacticConstruction or Sentence)
    - dependant some (WordClass or Word or SyntacticConstruction or Sentence)
    - Subject, Specification, Complement, Adposition, etc



## Relazioni semantiche

- Sul *piano del contenuto*, sono individuate le relazioni tematiche tra i sensi evocati nella frase
    - ..> Relation > **SemanticRelation**
      - *Relazioni tra sensi nella loro occorrenza all'interno di frasi*
      - predicate exactly 1 Meaning
      - argument some Meaning
      - grammaticalFacet some GrammaticalRelation
        - questa proprietà consente (per ora) l'associazione sintattico-semantic
    - ..> SemanticRelation > **ThematicRelation**
      - *Relazioni tra sensi verbali e sostantivi*
      - predicate only VerbalMeaning
      - argument only SubstantiveMeaning (qui si deve poter far riferimento a tutti i sensi mappati con determinate classi ontologiche, es. Person, Artifact, ecc)
- ..> Actor, Patient, Cause, ecc.





## Discussione

- L'ontologia di Senso Comune è un tentativo di caratterizzare il fenomeno linguistico come rapporto articolato tra espressioni linguistiche e loro contenuto
- L'approccio logico-descrittivo garantisce formalità, decidibilità e computabilità (al prezzo di un'espressività limitata)
- L'approccio ontologico rende chiaro il rapporto tra la rappresentazione e il suo denotato
- Il raggiungimento dell'obiettivo rappresenta un avanzamento rispetto alla rappresentazione di dati linguistici (es. PAROLE+SIMPLE, LMF) e a risorse come WordNet, Framenet, VerbNet
  - Semantica formale
  - Biplanarità (= capacità di rappresentare adeguatamente sia il piano dell'espressione sia quello del contenuto)
  - Separazione tra ontologia e semantica, e loro raccordo
  - Separazione tra relazioni sintattiche e argomentali (es. tematiche), e loro raccordo



## Problemi aperti (alcuni)

- Dipendenze tra argomenti (filler) dei ruoli tematici
  - es. Mangiare
    - Agent:Person, Object:Food, Instrument:Artifact
    - Agent:Animal, Object:Food
  - Rilassamento della ‘soundness’?
    - Agent:Animal, Object:Food, Instrument:Artifact
  - Introduzione della nozione di ‘struttura argomentale’?
- Ereditarietà delle strutture argomentali (es. attraverso i troponimi)
- Caratterizzazione dei ruoli per ‘centralità’, ‘preferenza’, ecc.
- Relazione tra strutture sintattiche e semantiche
  - Problema della co-referenza



## Conclusione

- L'adeguatezza del modello di Senso Comune come teoria di quello che c'è nel linguaggio si potrà valutare solo rispetto ad una vasta sperimentazione
- Il lavoro è all'inizio ed è presto per parlare di risultati
- Un primo passo sarà quello di rappresentare il contenuto attuale della risorsa mediante il modello (tecnicamente significa: serializzare il dizionario Senso Comune in OWL)
- Inoltre, si verificherà il modello delle relazioni tematiche su un campione sufficientemente vasto di casi concreti
- Team: Isabella Chiari, Carlo Ferrarini, Aldo Gangemi, Elisabetta Jezek, Alessandro Oltramari, Guido Vetere, Fabio M. Zanzotto
- Qualsiasi apporto è benvenuto