Introduction to Knowledge Organization Systems

G. Falquet

CUI – Université de Genève

Contents

Knowledge Engineering
and
Knowledge Representation

Knowledge engineering

KE is an engineering discipline that involves integrating knowledge into computer systems in order to solve complex problems normally requiring a high level of human expertise (Feigenbaum & McCorduck, 1983).

Goal: building, maintaining and developing knowledge-based systems

Issues:

- knowledge representation
- knowledge acquisition methodologies
- mathematical logic and reasoning

Data-Information-Knowledge (Tergan & Keller, 1998)

Data: Uninterpreted symbols (no meaning)

Information: Data that has been given meaning through interpretation by way of relational connection and pragmatic context

- provides answers to “who”, “what”, “where”, “why”, “when”
- express facts, opinions, objective information (based on facts), subjective information (cognitive analysis of facts), primary/secondary information
- different format (verbal, print, visual, ...)

Knowledge: cognitively processed information, integrated into a human knowledge structure

Knowledge

information is outside the brain ("knowledge in the world"),

knowledge is inside ("knowledge in the head")

provides answers to "how"

can be

- **explicit:** can be expressed (words, numbers, formulae), shared, transmitted
- **tacit:** highly personal, hard to formalize (insights, intuitions, hunches), based on beliefs, perceptions, emotions, mental models, ...

Aspects of domain knowledge

- **conceptual** knowledge: propositional representation of abstract concepts and their semantic relations
- **episodic** (memory of everyday events)
- **analogical** representations (mental models)
- **procedural** (condition-action pairs)
- **situated** (socio-cultural context)

Knowledge, in general language

from en.wikipedia.org

1. expertise, and skills acquired by a person through experience or education; the theoretical or practical understanding of a subject;
2. what is known in a particular field or in total; facts and information;
3. awareness or familiarity gained by experience of a fact or situation

Knowledge (1. and 3.) is related to a person.

Knowledge, From (Brachman & Levesque, 2004)

"Mary **knows** that Paul is a physicist"

A relation between an agent (knower) and a proposition

A propositional attitude, among others

"Mary **doubts** that Paul is a physicist"

"Mary **hopes** that Paul is a physicist"

"Mary **fears** that Paul is a physicist"

Knowledge, in philosophy (to know something)

In philosophy: knowledge is "justified true belief." (Plato)

S knows that a proposition P is true if, and only if:

1. P is true
2. S believes that P is true,
3. S is justified in believing that P is true

... challenged by Gettier's argument

... repaired by Nozick

... re-challenged ...

Knowledge Representation

The field of study concerned with using formal symbols to represent a set of propositions believed by some putative agent. (Brachman & Levesque, 2004)

Knowledge representation

Defined by its roles (Davis, Shrobe, and Szolovits, 1993)

1. a surrogate
   -- replaces real knowledge
2. a set of "ontological commitments"
   -- a view of the world
3. a fragmentary theory of intelligent reasoning
   -- logical, psychological, biological theory
4. a medium for efficient computation
5. a medium of human expression

Logical Pluralism

Logical Pluralism and Syntactic Heterogeneity (or plurality) of ontology languages is thus clearly an important issue, (Kutz & al., 2010)

Carnap, in 'Die logische Syntax der Sprache'[35, Sect. 17], famously put forward his principle of logical tolerance as follows:

> It is not our business to set up prohibitions, but to arrive at conventions. [. . .] In logic there are no morals. Everyone is at liberty to build up his own logic, i.e. his own language, as he wishes. All that is required of him is that, if he wishes to discuss it, he must state his methods clearly, and give syntactical rules instead of philosophical arguments.

⇒ several languages/logics to "fully" describe some reality

- Modal Logics
- First-order Logic
- Many-sorted First-order Logic.
- Common Logic.
- Relational Schemes.
- Description Logics
- Higher-Order Logics.
- Temporal logic
- Interval temporal logic
- Epistemic logic
- Dynamic logic
- Description logic
- Spatial logic
- Intuitionistic logic

Knowledge organization systems

Schemes for organizing knowledge representation elements

- thesauri
- terminologies
- ontologies
- knowledge graphs
- ...

References

- Feigenbaum, E. A.; McCorduck, P. (1983), The fifth generation (1st ed.), Reading, MA: Addison-Wesley, ISBN 978-0-201-11519-2, OCLC 9324691
- Brachman, R., Levesque, H. (2004) Knowledge Representation and Reasoning. Morgan Kaufmann Series in Artificial Intelligence.
- Davis, R., Shrobe, H., Szolovits, P. (1993). What is a Knowledge Representation? AI Magazine, 14(1):17-33, 1993.
- Kutz, O., Mossakowski, T., & Lücke, D. (2010). Carnap, Goguen, and the Hyperontologies: Logical Pluralism and Heterogeneous Structuring in Ontology Design. Logica Universalis, pp. 255-333. Retrieved from http://www.informatik.uni-bremen.de/~till/papers/Hyperontology.pdf
- Tergan, S.-O., Keller, T. (1998) Knowledge and Information Visualization. LNCS 3426, Springer.

Thesauri and SKOS

Thesaurus

Organized list of controlled and normalized terms that represent the concepts of a knowledge domain. Relations replace definitions

Main use : indexing document corpora with a **controlled vocabulary**

3 types of terms :

- **descriptors:** may be used to index documents
- **non-descriptors:** may not index documents, refer to the correct descriptor to use.
- **auxiliary words:** cannot be use alone, must be coordinated with other descriptors.

Semantic relations

- Broader/Narrower Term (BT/NT)
  - To create a hierarchical structure
  - May represent different conceptual relations
    - generic - specific
    - whole - part
    - class - instance
- Association (RT related term)
  - Any other kind of relation

Terminological relations

- Equivalence (Used For/Use)
- Preferred term (descriptor)
- Synonyms (non-descriptor)

Examples

Agriculture Organization's AGROVOC Thesaurus
https://agrovoc.uniroma2.it/agrovoc/agrovoc/en/

Thésaurus Urbamet
https://2ans-urbamet.developpement-durable.gouv.fr/exl-php/vue/mpd_thesaurus_urbam2_consultation

GEMET
http://www.eionet.europa.eu/gemet .

FORUM on Information Standard Heritage
http://www.heritage-standards.org.uk/fish-vocabularies/

SKOS

- A vocabulary to represent different families of knowledge organization systems, including
  - thesauri,
  - classification schemes,
  - subject heading systems,
  - taxonomies
- These are widely recognized and applied in both modern and traditional information systems. They have much in common.

Elements

- **concept scheme** comprising a set of concepts
- **concepts** can be labeled with any number of lexical (UNICODE) strings
- **notations**, which are lexical codes used to uniquely identify the concept (a bridge to existing systems)
- concepts can be
  - documented with **notes** of various types
  - linked to other SKOS concepts via **semantic relation properties**
  - **mapped** to other SKOS concepts in different concept schemes (hierarchical, associative, close equivalent and exact equivalent)

No formal interpretation

- concepts described through natural language
- concepts may be arranged and organized into various structures
  - hierarchies (most commonly) or association networks.
- structures have no formal semantics
  - cannot be interpreted as either formal axioms or facts about the world.
- structures serve only to provide a convenient and intuitive
  - map of some subject domain,
  - which can then be used as an aid to organizing and finding objects, such as documents, which are relevant to that domain.

Example

```
<A> rdf:type skos:Concept ;
skos:prefLabel "love"@en ;
skos:altLabel "adoration"@en ;
skos:broader <B> ;
skos:inScheme <S> .

<B> rdf:type skos:Concept ;
skos:prefLabel "emotion"@en ;
skos:altLabel "feeling"@en ;
skos:topConceptOf <S> .

<S> rdf:type skos:ConceptScheme ;
dct:title "My First Thesaurus" ;
skos:hasTopConcept <B> .
```

Semantic relations

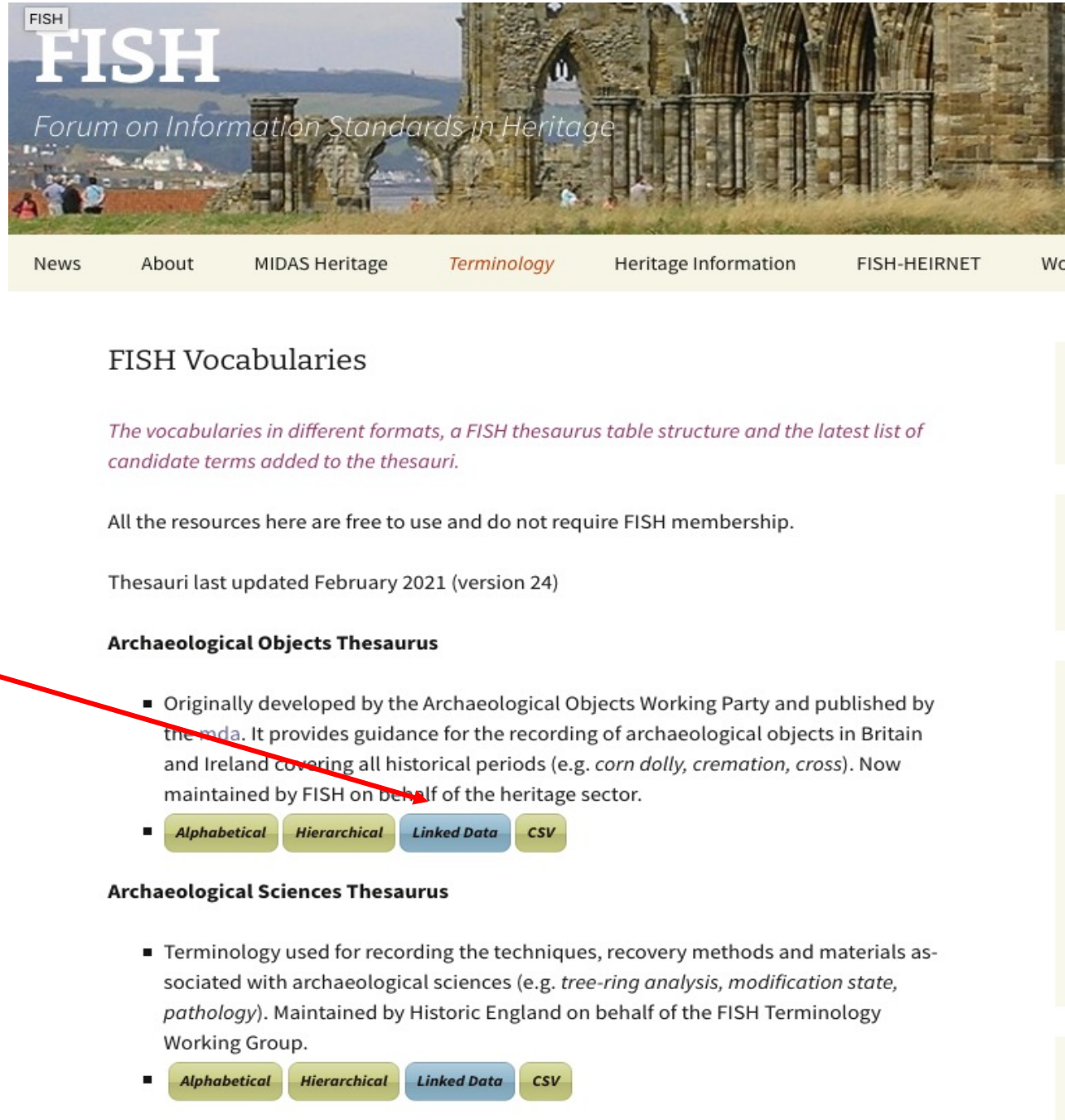
```
skos:semanticRelation
|
+- skos:related
|
+- skos:broaderTransitive
|     |
|     +- skos:broader
|
+- skos:narrowerTransitive
      |
      +- skos:narrower
```

Broader relations

- broader is not transitive but broaderTransitive is
- since broader $\leq$ broaderTansitive
  - if a broader b and b broader c then a broaderTransitive c .
  - but not a broader c
- broader is not functional, poly-hierarchies are allowed
- related is not transitive

Examples

- http://www.heritage-standards.org.uk/fish-vocabularies/
- In SKOS



Schema interconnection (mappings)

Interconnecting concepts that belong to different schemes

closeMatch concepts that are sufficiently similar that they can be used interchangeably in some information retrieval applications

exactMatch high degree of confidence that the concepts can be used interchangeably

broadMatch hierarchical mapping

narrowMatch hierarchical mapping

relatedMatch associative mapping link

The matching relations are subrelations of their internal counterpart.

- broadMatch $\sqsubseteq$ broader $\sqsubseteq$ broaderTransitive
- narrowMatch $\sqsubseteq$ narrower $\sqsubseteq$ narrowerTransitive
- relatedMatch $\sqsubseteq$ related

Entailment

```
<A> skos:broadMatch <B> .
```

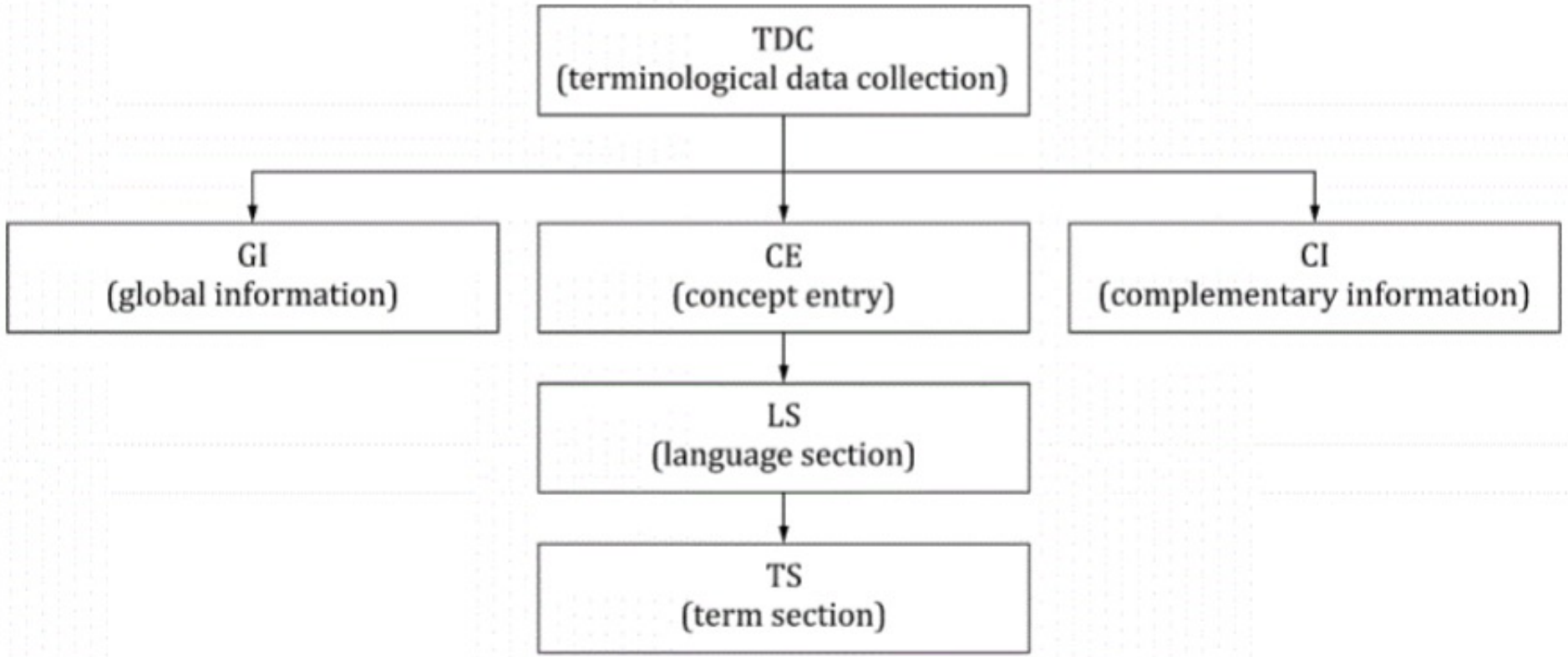
entails

```
<A> skos:mappingRelation <B> .
<A> skos:broader <B> .
<A> skos:broaderTransitive <B> .
<A> skos:semanticRelation <B> .
<A> rdf:type skos:Concept .
<B> rdf:type skos:Concept .
```

Terminologies and TBX

TermBase eXchange (TBX)

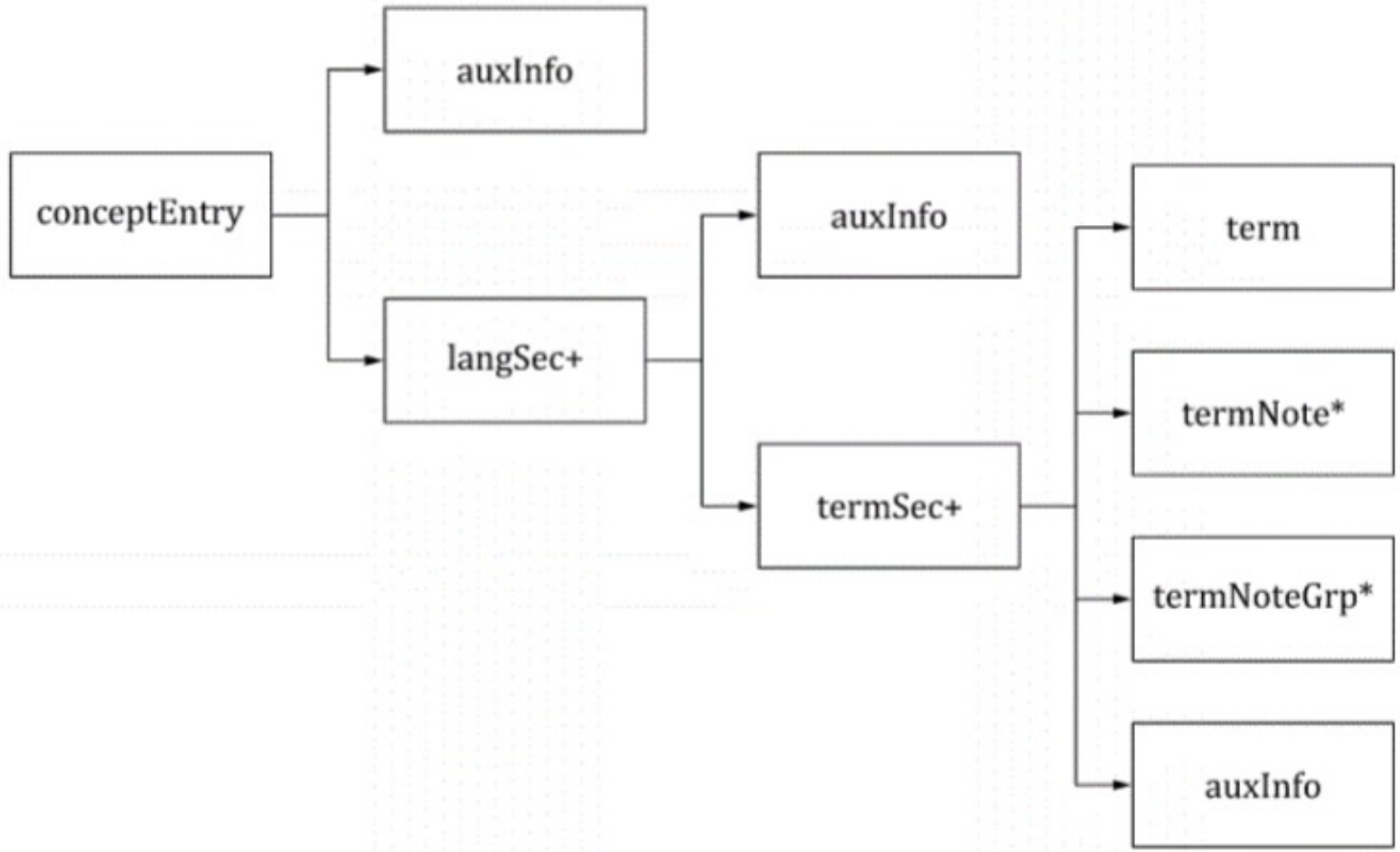
- international standard for representing and exchanging information about terms, words, and other lexical data
- ISO standard 30042
- XML format
- https://www.tbxinfo.net



source: ISO 30042 standard

Concept Entry

- Information about the concept as a whole, as designated by all the terms across all the language sections of the entry, includes:
  - domain to which the concept belongs (required)
  - link to an image illustrating the concept (optional)
  - definitions can be at the concept rather than language level


conceptEntry
auxInfo
langSec+
auxInfo
termSec+
term
termNote*
termNoteGrp*
auxInfo

Language Section

- All the terms in a particular language for a given concept

- information about the language section as a whole:
- language of the terms in this language section (required)
- definition of the concept (optional).
-

Term section

- Information about the term, such as:
- - term type (full form, acronym, abbreviation, etc.)
- - part of speech (noun, verb, etc.)
- - contextual example of the term in a sentence or paragraph
- - customer code (if this term is specific to a particular customer)
- - project code (if this term is specific to a particular project)
- - responsibility (if more than one person works on this termbase)
- - cross-reference to another term (if applicable)
- - usage note (usage notes can, for example, indicate regional variation).

Term Component

- To describe the components (words) of multiword terms
  - e.g. “uninterruptible power supply”

- linguistic information for each word
- part of speech (noun, verb, adjective, adverb, preposition)
- lexical gender
- how words are hyphenated, inflected and pronounced.

```
<conceptEntry id="c1">
    <langSec xml:lang="en">
        <descripGrp>
            <basic:definition>A group of stars formed together in the spiral
                arms of a galaxy.</basic:definition>
            <basic:source>Oxford2007</basic:source>
        </descripGrp>
        <termSec>
            <term>open cluster</term>
            <min:partOfSpeech>noun</min:partOfSpeech>
        </termSec>
        <termSec>
            <term>galactic cluster</term>
            <descripGrp>
                <basic:context>In contrast, the galactic cluster represented by
                    M29 is a grouping of stars—a “knot” of stars in the stellar
                    backdrop.</basic:context>
                <basic:source>http://www.eastbayastro.org/2000/0900/r0900-8.htm</basic:source
            </descripGrp>
            <note>Another name for an open star cluster. They are often termed Galactic
                Clusters because they are found mainly in the plane of our galaxy. If
                you were to view our galaxy from afar, you would find that all the
                open/galactic clusters lie within the spiral arms of the galaxy.</note>
            <note>N-Source:
                http://www.delscope.demon.co.uk/astronomy/glossary.htm#G</note>
        </termSec>
    </langSec>
```

```
    <langSec xml:lang="es">
        <descripGrp>
            <basic:definition>Una acumulación de una docena hasta varios miles
                de jóvenes estrellas en el plano de las galaxias.</basic:definition>
            <basic:source>http://www.espacioprofundo.com.ar/diccionario/Glosario_de_Astronomia/
        </descripGrp>
        <termSec>
            <term>cúmulo abierto</term>
            <min:partOfSpeech>noun</min:partOfSpeech>
            <descripGrp>
                <basic:context>M7 es uno de los cúmulos abiertos estelares más
                    prominentes del cielo. El cúmulo, dominado por brillantes estrellas
                    azules, puede ser visto a ojo desnudo en un cielo oscuro en la cola
                    de la constelación del Escorpión (Scorpius).Contiene unas 100
                    estrellas en total, tiene una edad aproximada de 200 millones de
                    años, mide 25 años luz de ancho, y se encuentra a unos 1000 años luz
                    de distancia.</basic:context>
                <basic:source>http://observatorio.info/2009/11/m7-cumulo-abierto-en-scorpius-2/
            </descripGrp>
        </termSec>
    </langSec>
</conceptEntry>
```

Conceptual Relations

- hierarchical relations
- broader (superordinate),
- narrower (subordinate)
- sibling (coordinate)

- partitive relations
- part of (pupil is-part-of eye)

- associated relations
  (for instance, the relation between “pitcher” and “baseball”).

Data-category name	Data type	Target
antonymConcept	basicText	entry
associatedConcept	basicText	entry
broaderConceptGeneric	basicText	entry
broaderConceptPartitive	basicText	entry
conceptPosition	plainText	conceptSysDescrip
coordinateConceptGeneric	basicText	entry
coordinateConceptPartitive	basicText	entry
relatedConcept	basicText	entry
relatedConceptBroader	basicText	entry
relatedConceptNarrower	basicText	entry
sequentiallyRelatedConcept	basicText	entry
spatiallyRelatedConcept	basicText	entry
subordinateConceptGeneric	basicText	entry
subordinateConceptPartitive	basicText	entry
superordinateConceptGeneric	basicText	entry
superordinateConceptPartitive	basicText	entry
temporallyRelatedConcept	basicText	entry

A Lexical Ontology: WordNet

http://wordnetweb.princeton.edu

http://wordnetweb.princeton.edu/perl/webwn

WordNet (overview)

- a lexical ontology
- conceptual level + linguistic level
- limited set of relations
- concept = "synset" (set of synonyms)

- Contains
- all the English words
- all the meanings of these words

Definitions

- vocabulary = set of words
- word = (f, s)
  - f : lexical form = character string
  - s : sense
- syntactic categories : names, adjectives, verbs, adverbs

Relations

- Morphological
  - between forms
  - inflection, derivation, composition

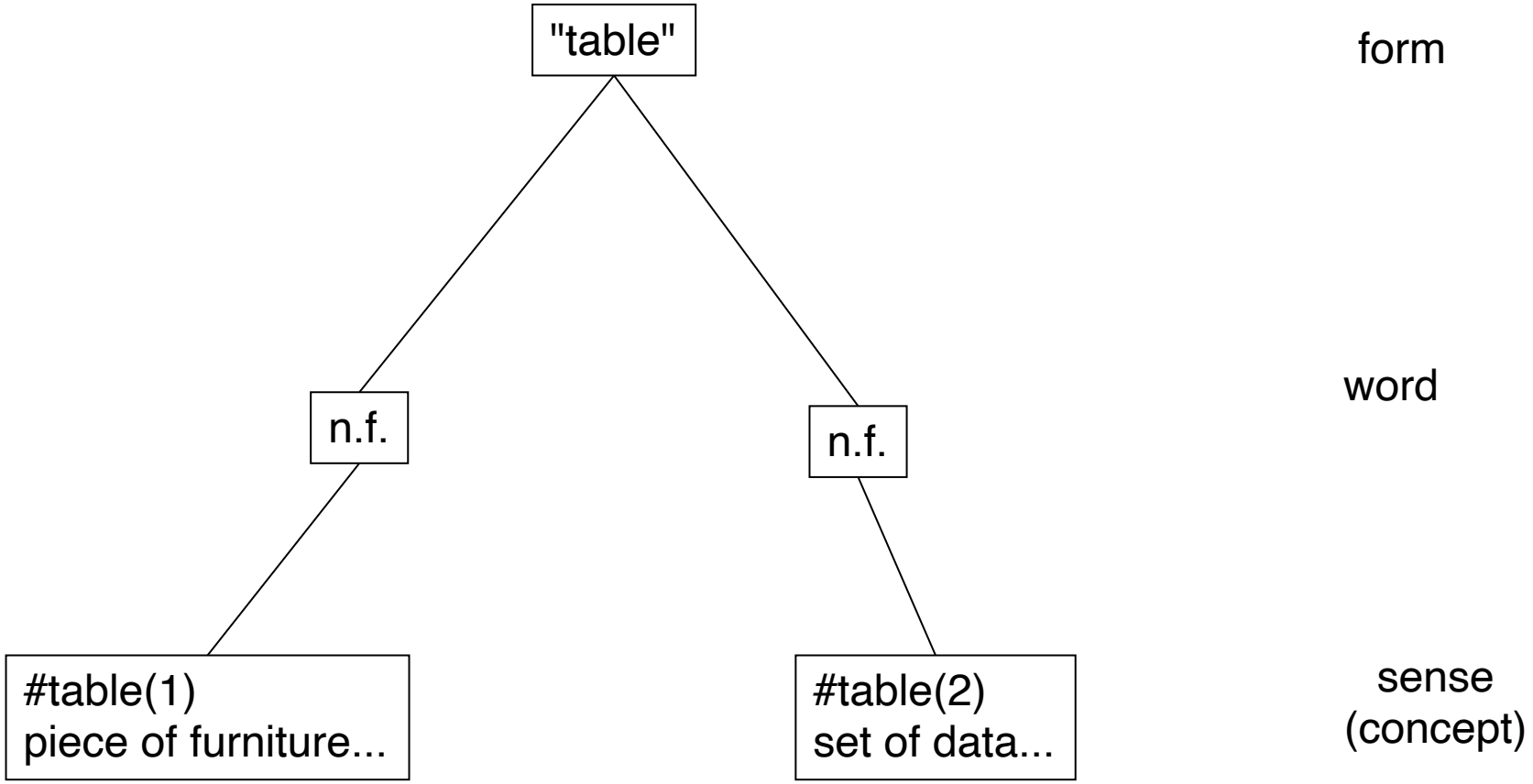
- Lexical semantics
  - relations between the word senses
  - semi-formalized definitions

Base WordNet (English)

- ≥ 166 000 words (f, s)
- 118 000 forms
- 90 000 senses

- 17% of polysemous words
- 40% of the words have synonyms

Exemple : polysemy



http://wordnet.princeton.edu/perl/webwn

Enter a word to search for: table Search WordNet

KEY: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

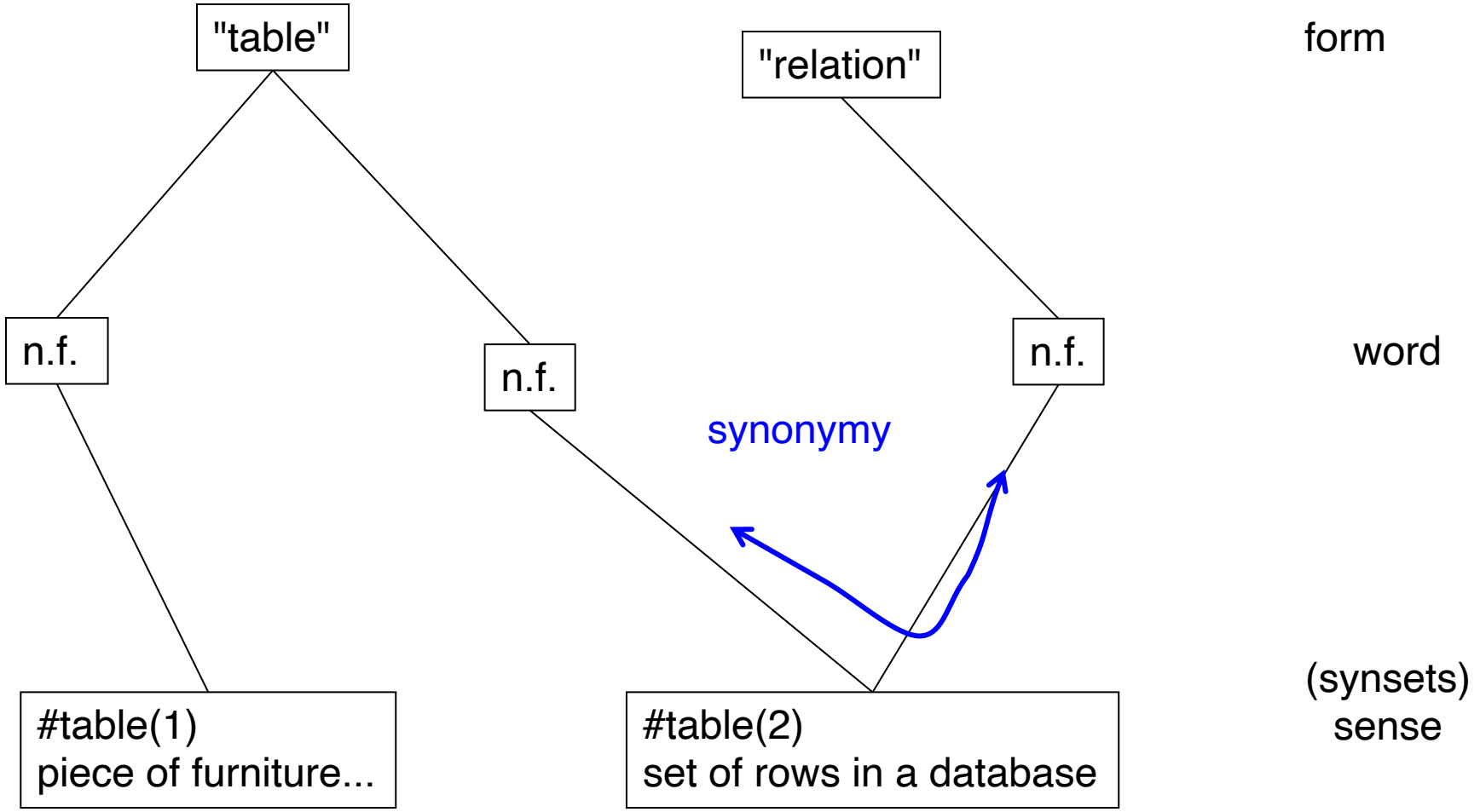
Noun

- S: (n) **table**, tabular array (a set of data arranged in rows and columns) *"see table 1"*
- S: (n) **table** (a piece of furniture having a smooth flat top that is usually supported by one or more vertical legs) *"it was a sturdy table"*
- S: (n) **table** (a piece of furniture with tableware for a meal laid out on it) *"I reserved a table at my favorite restaurant"*
- S: (n) mesa, **table** (flat tableland with steep edges) *"the tribe was relatively safe on the mesa but they had to descend into the valley for water"*
- S: (n) **table** (a company of people assembled at a table for a meal or game) *"he entertained the whole table with his witty remarks"*
- S: (n) board, **table** (food or meals in general) *"she sets a fine table"; "room and board"*

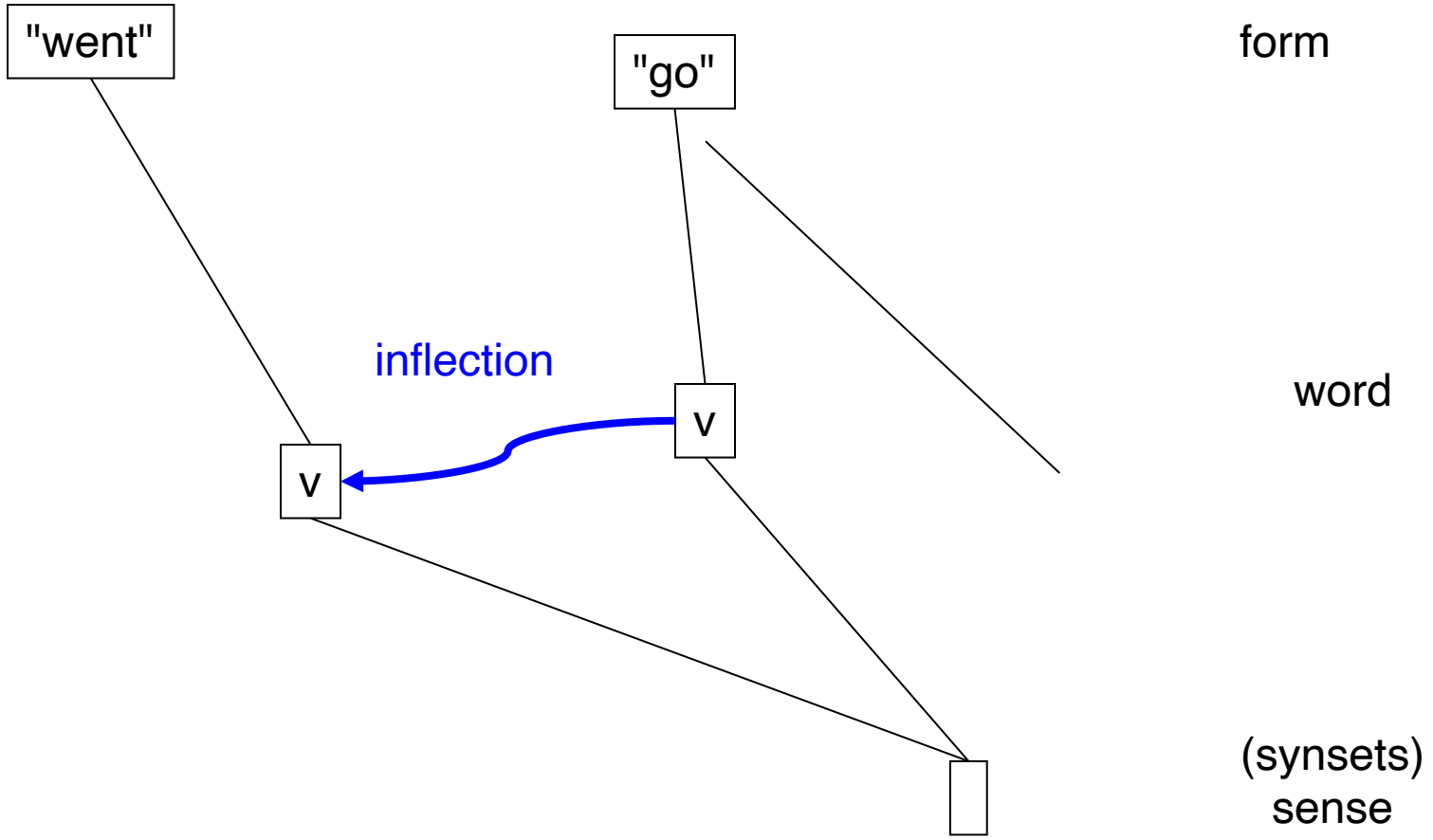
Verb

- S: (v) postpone, prorogue, hold over, put over, **table**, shelve, set back, defer, remit, put off (hold back to a later time) *"let's postpone the exam"*

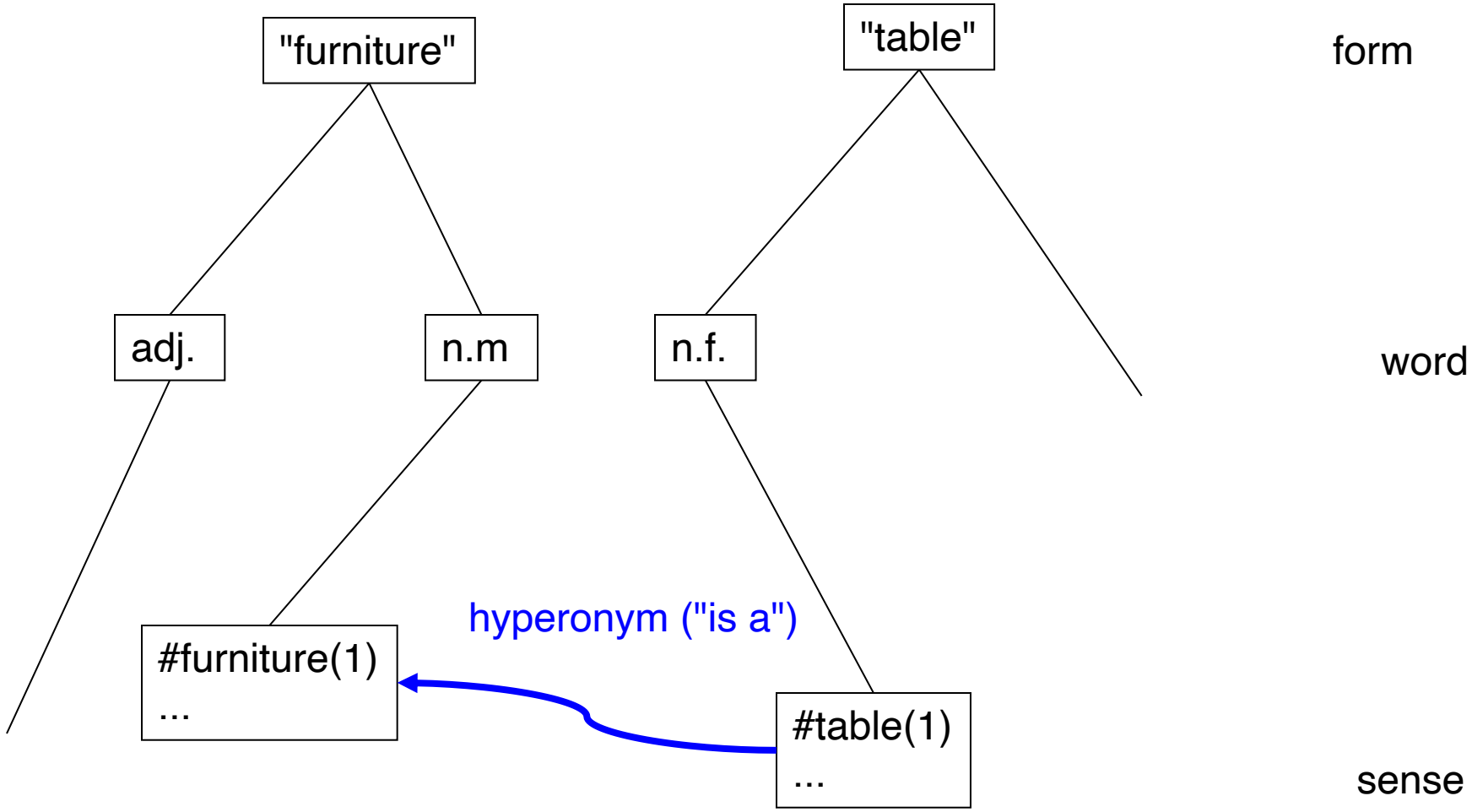
Synsets


"table"
"relation"
form
n.f.
n.f.
n.f.
word
synonymy
#table(1)
piece of furniture...
#table(2)
set of rows in a database
(synsets)
sense

morphological relations



Semantic relations : hyperonymy



- S: (n) administration, governance, governing body, **establishment**, brass, organization, organisation (the persons (or committees or departments etc.) who make up a body for the purpose of administering something) *"he claims that the present administration is corrupt"; "the governance of an association is responsible to its members"; "he quickly became recognized as a member of the establishment"*
  - *direct hyponym* / ***full hyponym***
    - S: (n) Curia ((Roman Catholic Church) the central administration governing the Roman Catholic Church)
    - S: (n) top brass (the most important persons in a governing body)
    - S: (n) executive (persons who administer the law)
      - S: (n) Bush administration (the executive under President George W. Bush)
      - S: (n) Clinton administration (the executive under President Clinton)
      - S: (n) Bush administration (the executive under President George H. W. Bush)
      - S: (n) Reagan administration (the executive under President Reagan)
      - S: (n) Carter administration (the executive under President Carter)
    - S: (n) judiciary, bench (persons who administer justice)
    - S: (n) county council (the elected governing body of a county)
    - S: (n) government officials, officialdom (people elected or appointed to administer a government)
      - S: (n) civil service (government workers; usually hired on the basis of competitive examinations)
        - S: (n) Whitehall (the British civil service)
      - S: (n) bureaucracy, bureaucratism (nonelective government officials)
        - S: (n) Pentagon (the United States military establishment)
    - S: (n) management (those in charge of running a business)
      - S: (n) house (the management of a gambling house or casino) *"the house gets a percentage of every bet"*

Hyponymy

A is a hyponym of B if the sentence

"A is a B"

is true

Encompasses notions such as superclass and instance.

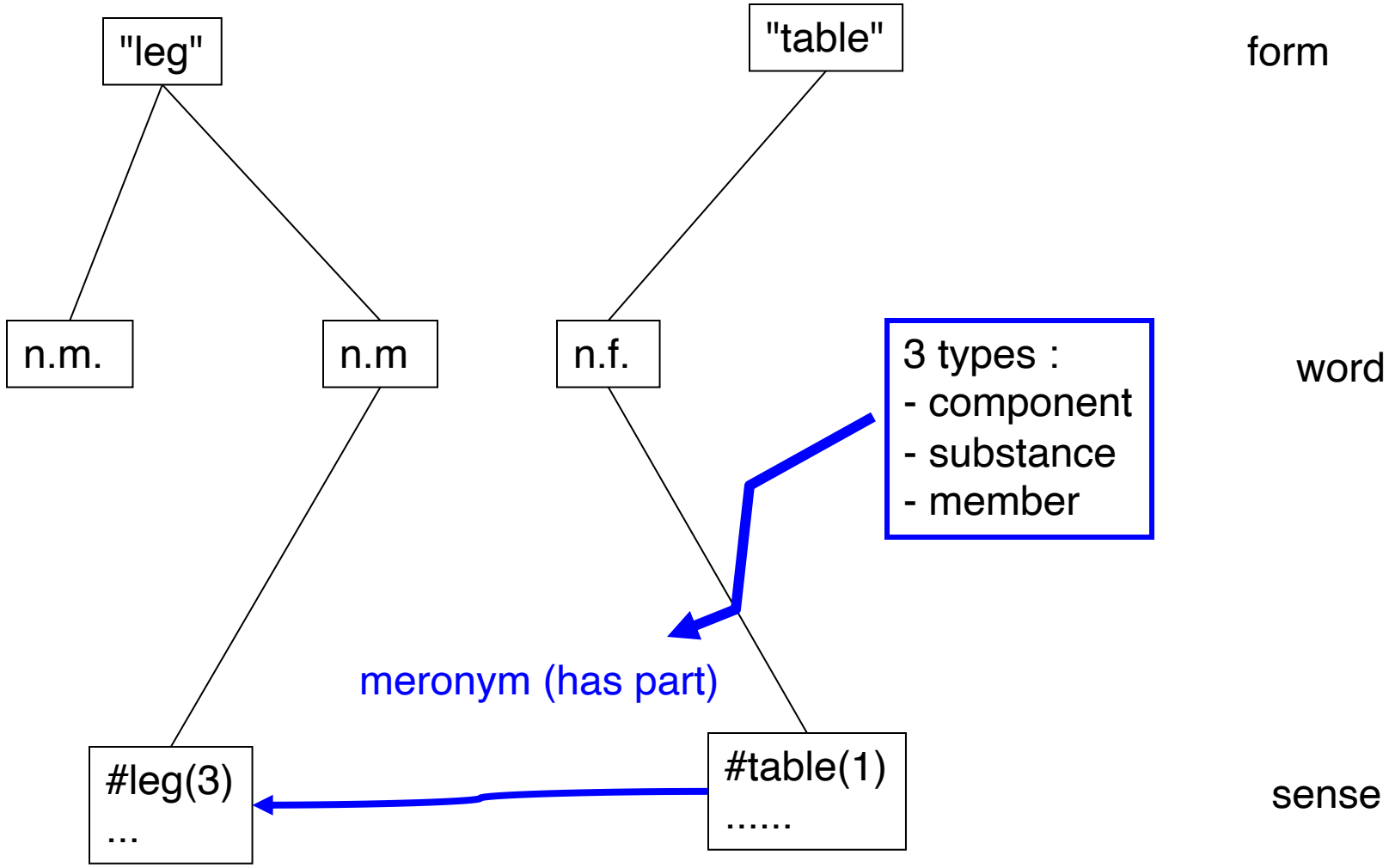
a lion is a feline

Barcelona is a city

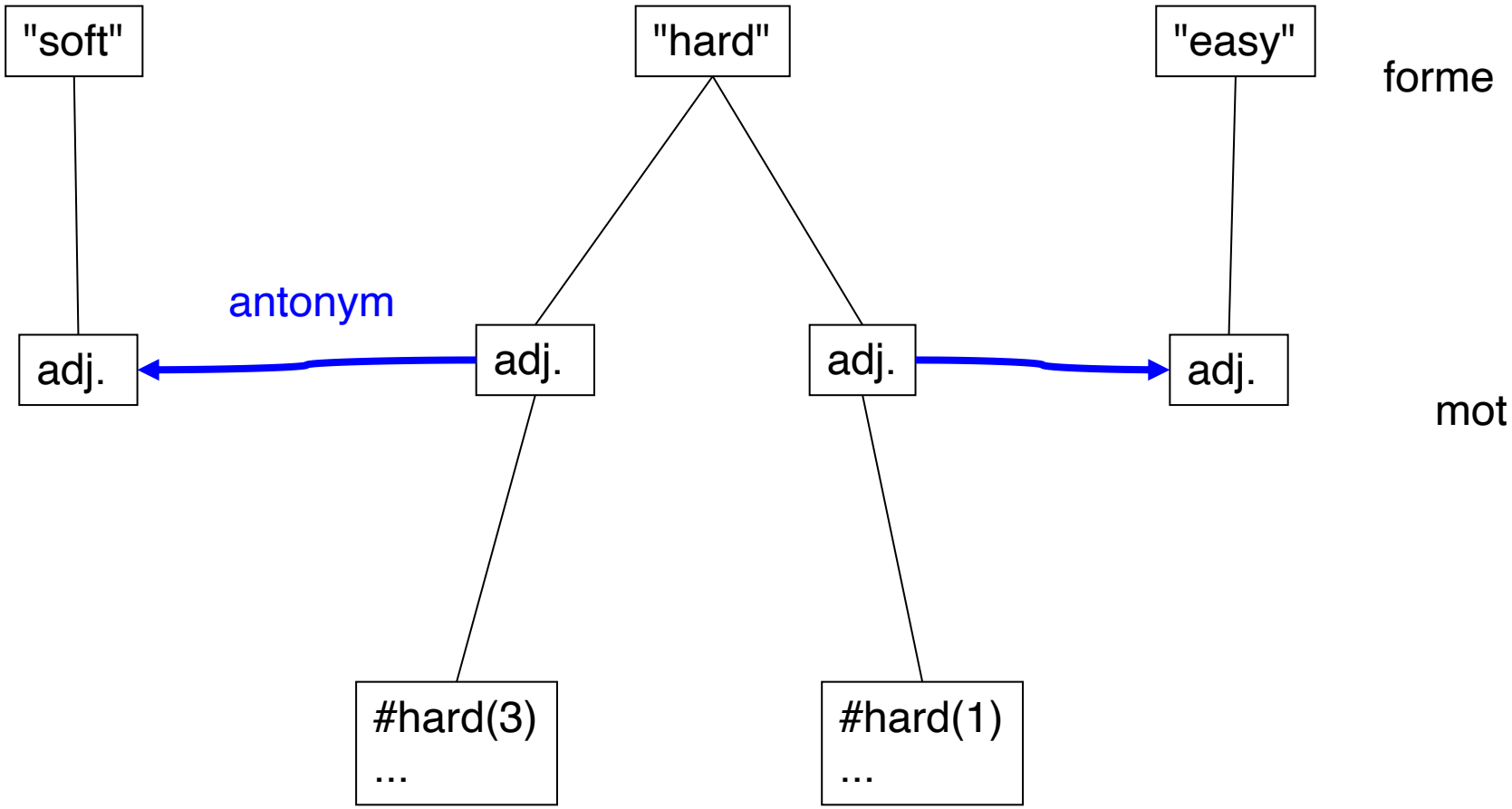
Wordnet also has an instance relation.

- S: (n) **Lobito** (a seaport on the Atlantic coast of Angola)
  - *part holonym*
  - ***instance***
    - S: (n) city, metropolis, urban center (a large and densely populated urban area; may include several independent administrative districts) *"Ancient Troy was a great city"*
    - S: (n) port (a place (seaport or airport) where people and merchandise can enter or leave a country)

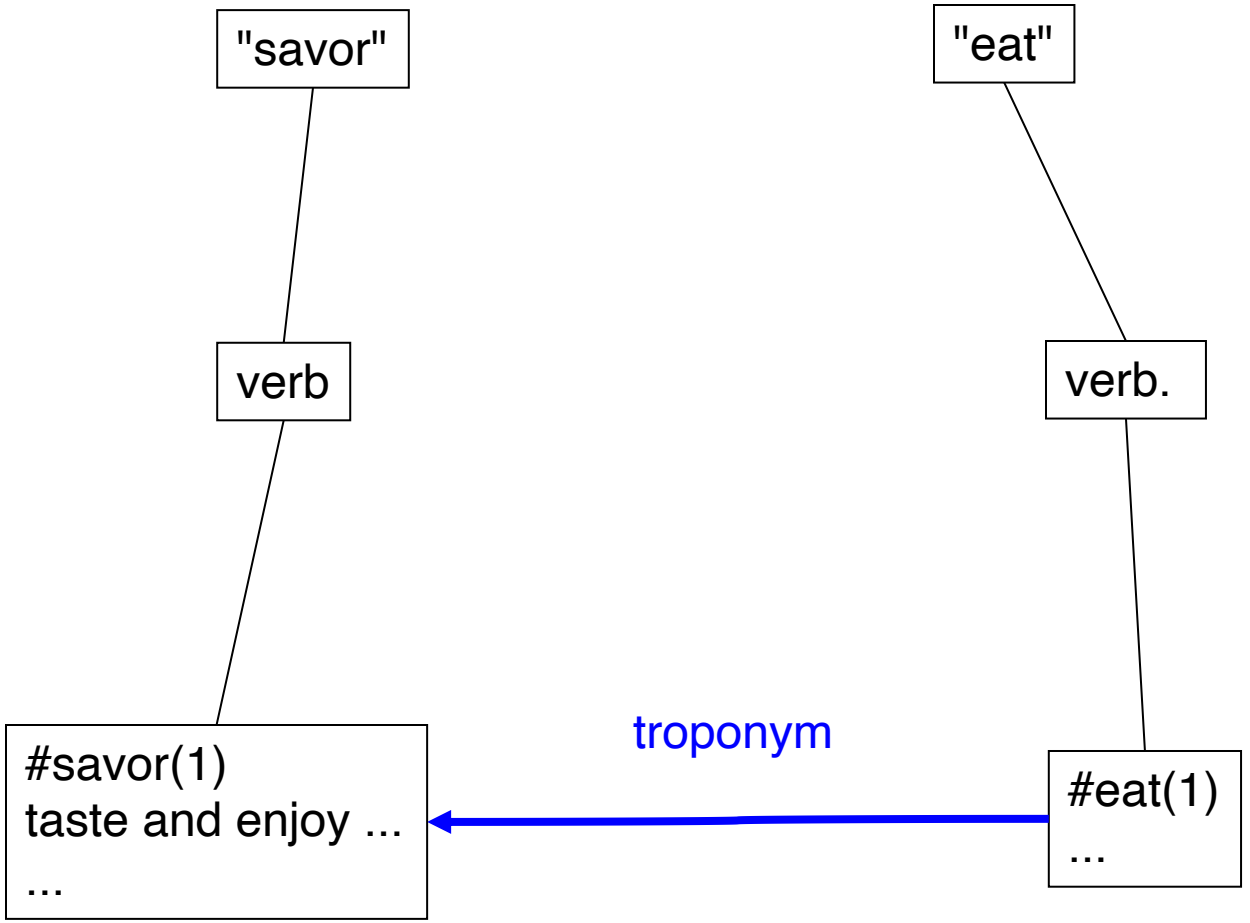
R. S. : meronymy (parthood)



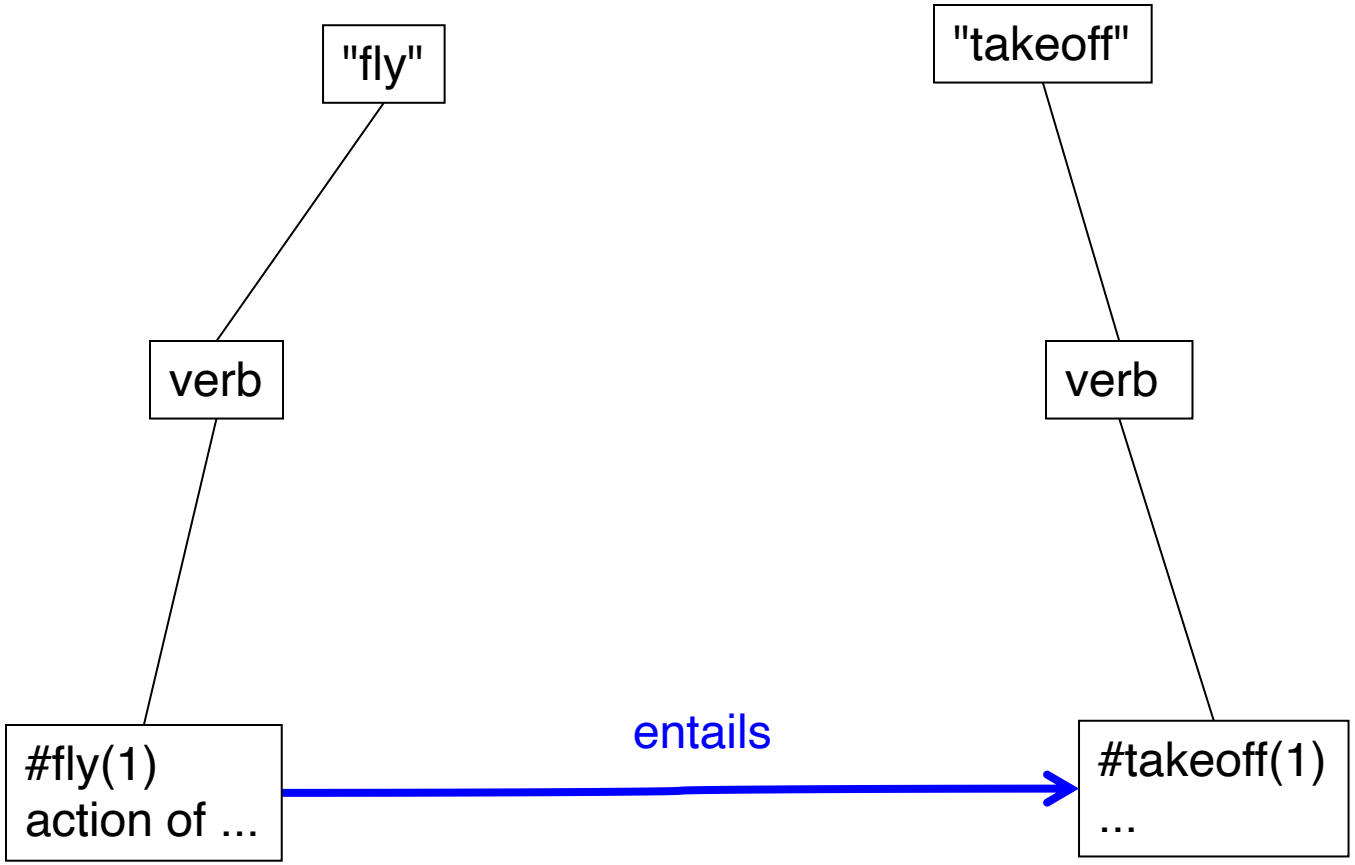
Antonym


"soft"
"hard"
"easy"
antonym
adj.
adj.
adj.
adj.
#hard(3)
...
#hard(1)
...
forme
mot
sens

R. S. Troponymie



Entailment



Applications

Natural language processing

Examples

- normalization (replace synonyms by a standard form)
- disambiguation: make explicit the meaning of each word in a text
  - several techniques (generally based on the word's context)
  - still a research topic

Application in information retrieval

Query Expansion

1. With synonyms

Q : automobile

exp(Q) : automobile OR car OR motorcar

Increases recall

Polysemy problem

How to expand 'car' ?

car, auto, automobile, machine, motorcar (a motor vehicle with four wheels; usually propelled by an internal combustion engine) "he needs a car to get to work"

car, railcar, railway car, railroad car (a wheeled vehicle adapted to the rails of railroad) "three cars had jumped the rails"

cable car, car (a conveyance for passengers or freight on a cable railway) "they took a cable car to the top of the mountain"

car, gondola (the compartment that is suspended from an airship and that carries personnel and the cargo and the power plant)

car, elevator car (where passengers ride up and down) "the car was on the top floor"

==> interactive interfaces

2. With hyponyms

a) increase recall

Q $\rightarrow$ Q OR $hypo_1$ OR $hypo_2$ OR ...

car $\rightarrow$ car OR ambulance OR beach wagon OR compact OR convertible OR coupe OR cruiser ...

b) specification (if too many answers)

Q $\rightarrow$ one select hyponym of Q

car $\rightarrow$ electric car

[requires user interaction]

Critique of WordNet by D. Lenat (CyC)

- Too few semantic relations
  thousands in CyC, 6 dans WN
  inCyC : no. rel type $\sim=$ no. concepts/10

- A richer knowledge base is required for true contextual disambiguation
  "The police arrested the demonstrators because they feared violence"
  "The police arrested the demonstrators because they advocated violence"

  Who is "they" ?
  A general world knowledge is necessary

Other Wordnets

34 Open Wordnets Merged

Wordnet	Lang	Synsets	Words	Senses	Core	Licence	Data	Citation
Albanet	als	4,675	5,988	9,599	31%	CC BY 3.0	als.zip (+xml)	cite:als; (.bib)
Arabic WordNet (AWN v2)	arb	9,916	17,785	37,335	47%	CC BY SA 3.0	arb.zip (+xml)	cite:arb; (.bib)
BulTreeBank Wordnet (BTB-WN)	bul	4,959	6,720	8,936	99%	CC BY 3.0	bul.zip (+xml)	cite:bul; (.bib)
Chinese Open Wordnet	cmn	42,312	61,533	79,809	100%	wordnet	cmn.zip (+xml)	cite:cmn; (.bib)
Chinese Wordnet (Taiwan)	qcn	4,913	3,206	8,069	28%	wordnet	qcn.zip (+xml)	cite:qcn; (.bib)
DanNet	dan	4,476	4,468	5,859	81%	wordnet	dan.zip (+xml)	cite:dan; (.bib)
Greek Wordnet	ell	18,049	18,227	24,106	57%	Apache 2.0	ell.zip (+xml)	cite:ell; (.bib)
Princeton WordNet	eng	117,659	148,730	206,978	100%	wordnet	eng.zip (+xml)	cite:eng; (.bib)
Persian Wordnet	fas	17,759	17,560	30,461	41%	Free to use	fas.zip (+xml)	cite:fas; (.bib)
FinnWordNet	fin	116,763	129,839	189,227	100%	CC BY 3.0	fin.zip (+xml)	cite:fin; (.bib)
WOLF (Wordnet Libre du Français)	fra	59,091	55,373	102,671	92%	CeCILL-C	fra.zip (+xml)	cite:fra; (.bib)
Hebrew Wordnet	heb	5,448	5,325	6,872	27%	wordnet	heb.zip (+xml)	cite:heb; (.bib)
Croatian Wordnet	hrv	23,120	29,008	47,900	100%	CC BY 3.0	hrv.zip (+xml)	cite:hrv; (.bib)
IceWordNet	isl	4,951	11,504	16,004	99%	CC BY 3.0	isl.zip (+xml)	
MultiWordNet	ita	35,001	41,855	63,133	83%	CC BY 3.0	ita.zip (+xml)	cite:ita; (.bib)
ItalWordnet	ita	15,563	19,221	24,135	48%	ODC-BY 1.0	ita.zip (+xml)	cite:iwn (.bib)

Derived project