Introduction to WordNet

Vincenzo Pallotta
MEDIA group
DI-LITH EPFL

Lausanne, 19 June 2001

Post-graduate course on
Language and Speech Engineering

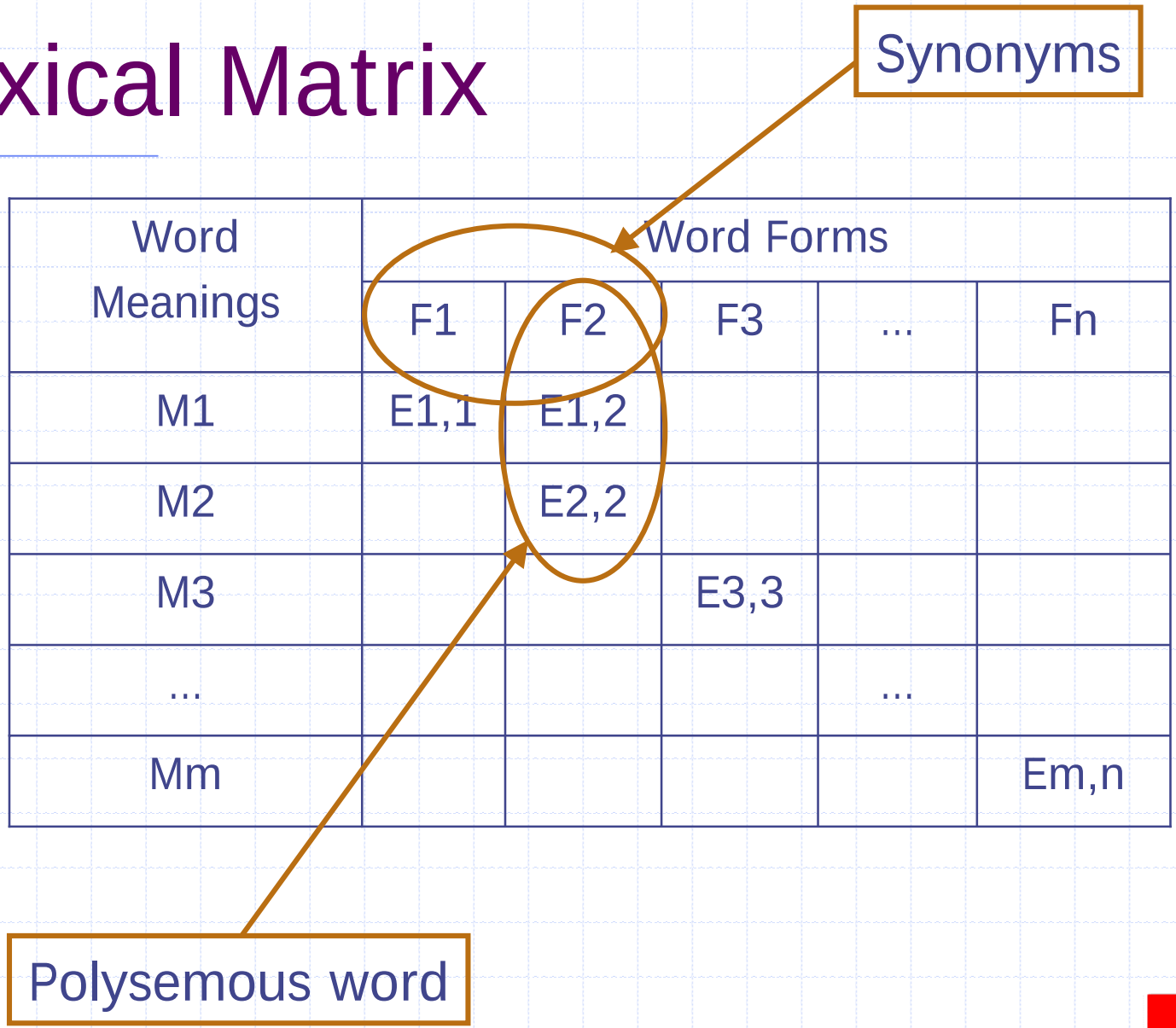
Goals and motivations

- A dictionary based on psycholinguistic principles
  - Synchronic properties of mental lexicon that can be exploited in lexicography.
- A lexical database based on conceptual look-up
  - Organizing concepts in a semantic network.
- Organize lexical information in terms of word meaning, rather than word form
  - Wordnet can also be used as a thesaurus.

Semantic model of WordNet

- A WORD is a conventional association between:
  - A lexicalised concept, and
  - A word form that plays a syntactic role.
  - A practical way of organizing lexicalised concepts that words can express.
- Lexical Matrix: a contingency matrix between:
  - Word forms (the columns)
  - Word meanings (the rows).
  - An entry in the matrix indicates that the form in that column can be used to express (in the appropriate context) the meaning in that row.

Lexical Matrix



Lexical representation theory

- Constructive
  - The representation should contain sufficient information to support an accurate construction of the concept
- Differential
  - Meanings can be represented by any symbol that enable the theorist to distinguish among them

Concepts in Wordnet

- Hypothesis:
  - A *synonym* is often sufficient to identify the concept.
- Differential approach
  - Word meaning can be represented by a list of the word forms that can be used to express it: the **synset**.

The structure of Wordnet

- 95.600 word forms
  - 51.500 simple words
  - 44.100 collocations
- 70.100 word meanings
- Wordnet Relations
  - Lexical relations (between word forms)
    - Synonymy
    - Antonymy
  - Semantic relations (between word meanigns)
    - Hyponymy/Hyperymy
    - Meronymy/Holonymy
    - Entailment

Synonymy

Two words are synonymous if they have the same sense:

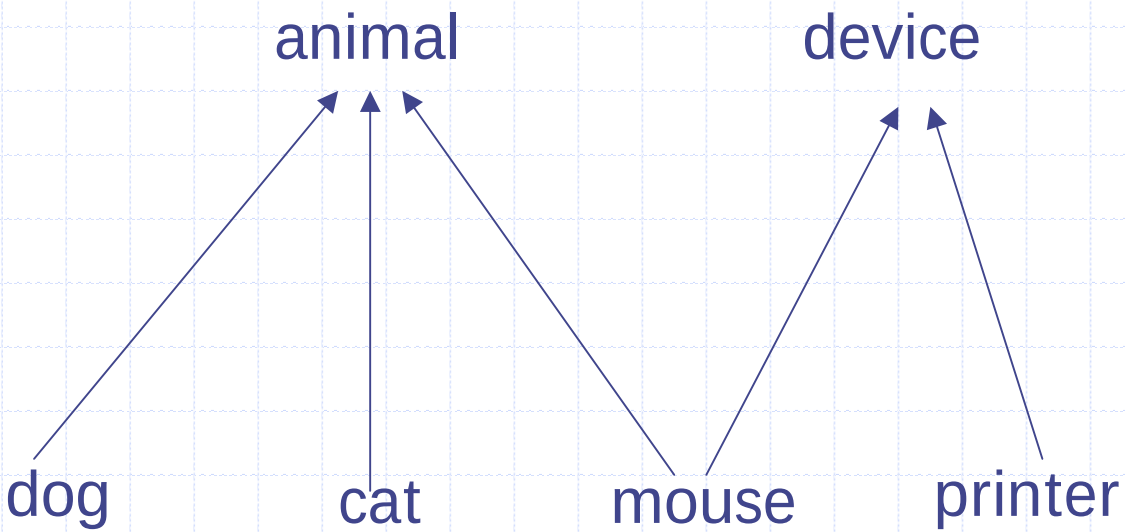
- they have the same values for all their semantic features
- they map to the same concept
- they satisfy the Leibniz's substitution principle
  - If the substitution of one for the other never changes the truth value of a sentence in which the substitution is made

Synset is the set of word forms that share the same sense

- Synsets do not explain what the concepts are, they signify that concepts exists

Hyponymy

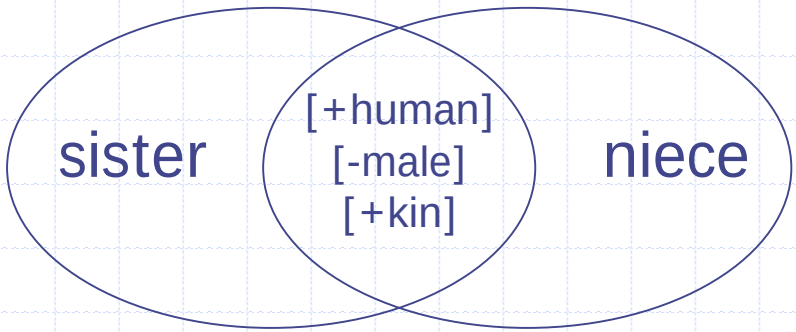
An hyponym is a word whose meaning contains the entire meaning of another, known as the *superordinate*.



Overlap

Two words *overlap* in meaning if they have the same value for some (but not all) of the semantic features.

- Hyponymy is a special case of overlap where all the features of the superordinate is contained by the hyponym.



Meronymy/Holonymy

- A word w1 is a meronym of another word w2 (the holonym) if the relation is-part-of holds betwen the meaning of w1 and w2.
  - Meronymy is transitive and asymmetric
  - A meronym can have many holonyms
  - Meronyms are distinguishing features that hyponyms can inherit.
    - Ex. If “beak” and “wing” are meronyms of “bird”, and if “canary” is a hyponym of “bird”, then (by inheritance), “beak” and “wing” must be meronyms of “canary”.
  - Limited transitivity:
    - Ex. “A house has a door” and “a door has a handle”, then “a house has a handle” (?)

Different type of part-whole relationships

- Component-object (branch/tree)
- Member-collection (tree/forest)
- Portion-mass (slice/cake)
- Stuff-object (aluminium/airplane)
- Feature-activity (paying/shopping)
- Place-area (Lausanne/Vaud)
- Phase-process (addolescence/growing up).

Word categories

- Nouns
  - Organised as topical hierarchies with lexical inheritance (hyponymy/hyperymy and meronymy/holonymy).
- Verbs
  - Organised by a variety of entailment relations
- Adjectives
  - Organised on the basisi of bipolar opposition (antonymy relations)
- Adverbs
  - Like adjectives
- Function words
  - Currently omitted, stored separately as part of syntactic component of language

Nouns in Wordnet

- 57.000 nouns and 48.800 synsets.
- Many are compounds
- No proper nouns
- Definition given
  - By its hypernym
    - Ex. “Tree is_a Plant”
  - By its distinguishing features
    - Attributes (modification) – from the gloss,
    - Parts (meronymy) – currently implemented,
    - Functions (predication) – from the gloss
    - Ex. “A tree is a large woody, perennial plant with a distinct trunk”.

Hierarchical semantic organization

- Generated by the Hyponym/Hypernym relation.
  - A systematic effort has been made to connect hyponyms with their hypernyms (and vice versa).
- Implemented by use of labeled pointers between synsets.
  - Ex. An entry for “tree” contains a pointer ‘,@’ to an entry for “plant” and pointers ‘,’ to entries such as “conifer”,”alder”,…
    - Synset for “tree” = {tree, plant,@ conifer, alder, …}
- Each word inherits the distinguishing features of all its hypernyms.

Building the noun hiearchy

- Hyponymy:
  - Transitive
  - Asymmetric
  - Generates a hierarchy (there is normally a single hypernym).
- Semantic primes
  - Select a (relatively small) number of generic concepts and treat each one as the unique beginner of a separate hierarchy.

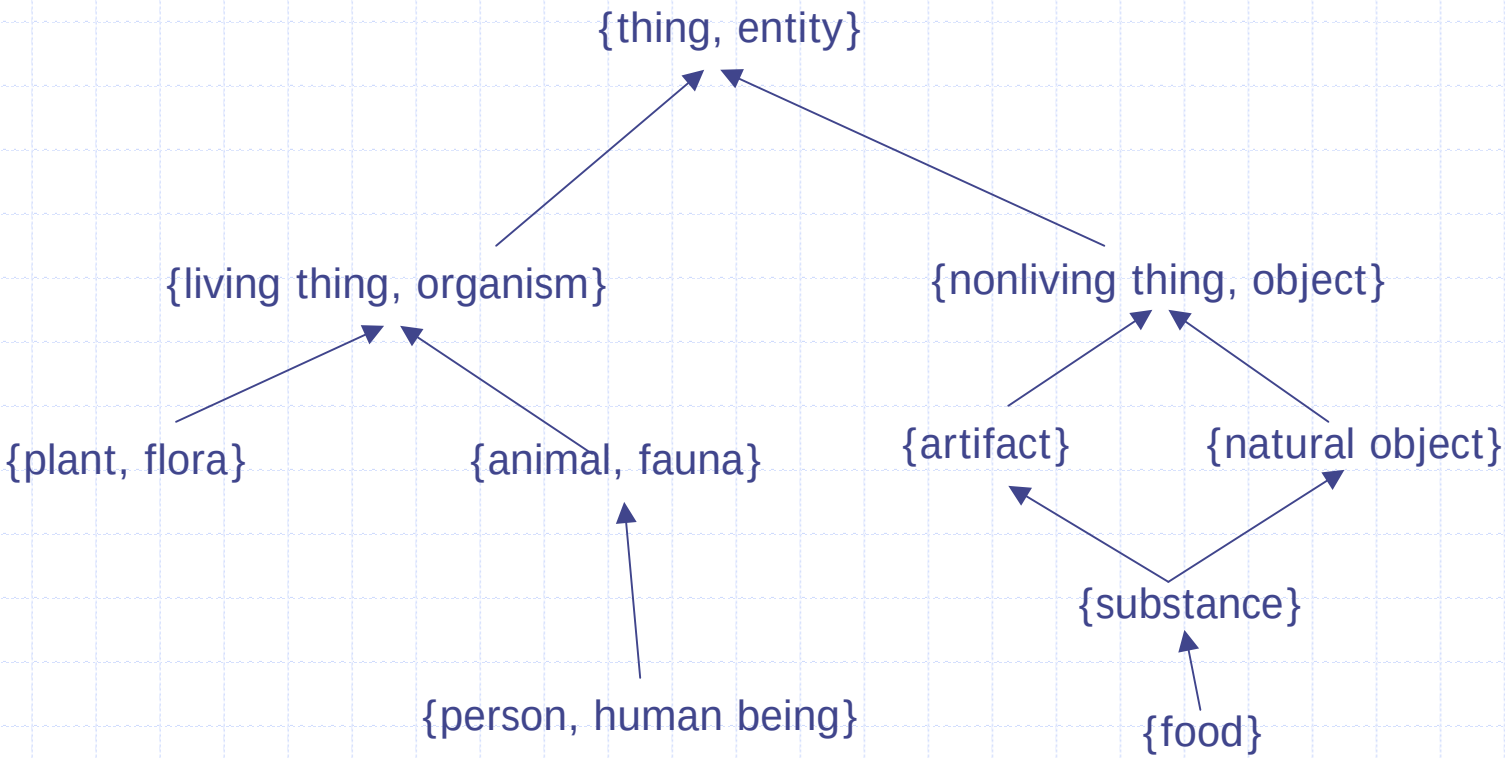
Unique beginners

- 25 unique beginners hierarchies
- Not mutually exclusive
  - Cross-referencing is required
- They cover distinct conceptual and lexical domains

{act, activity}	{food}	{possesion}
{animal, fauna}	{group, grouping}	{process}
{artifact}	{location}	{quantity, amount}
{attribute}	{motivation, motive}	{relation}
{body}	{natural, object}	{shape}
{cognition,knowledge}	{natural phenomenon}	{state}
{communication}	{person, human being}	{substance}
{event, happening}	{plant, flora}	{time}
{feeling,emotion}		

Natural groupings

- Small `Tops'



Adjectives in Wordnet

- 19.500 adjective words
- 10.000 word meanings (sysnsets)
- Types:
  - Descriptive adjectives
    - Clusters based on antonymy
    - Used to ascribe attribute values to a noun
      - “X is Adj” presuppose there is an attribute A s.t. A(x) = Adj.
  - Relational adjectives
    - Similar to nouns used as modifiers
  - Reference modifying adjectives
    - Ex. Former, alleged,...

Adjective Hierarchy

- Noun modifications:
  - Present and past participles of verbs
    - Ex. “The leaning tower”
  - Prepositional phrases
    - Ex. “The man with the telescope”
  - Noun phrases
    - Ex. “My grandfather’s chair”
  - Entire clauses
    - Ex. “The chair that you bought at the auction”.
- Descriptive adjectives
  - Pointers between adjectives and the noun synset by which that attribute is lexicalised.

Antonymy

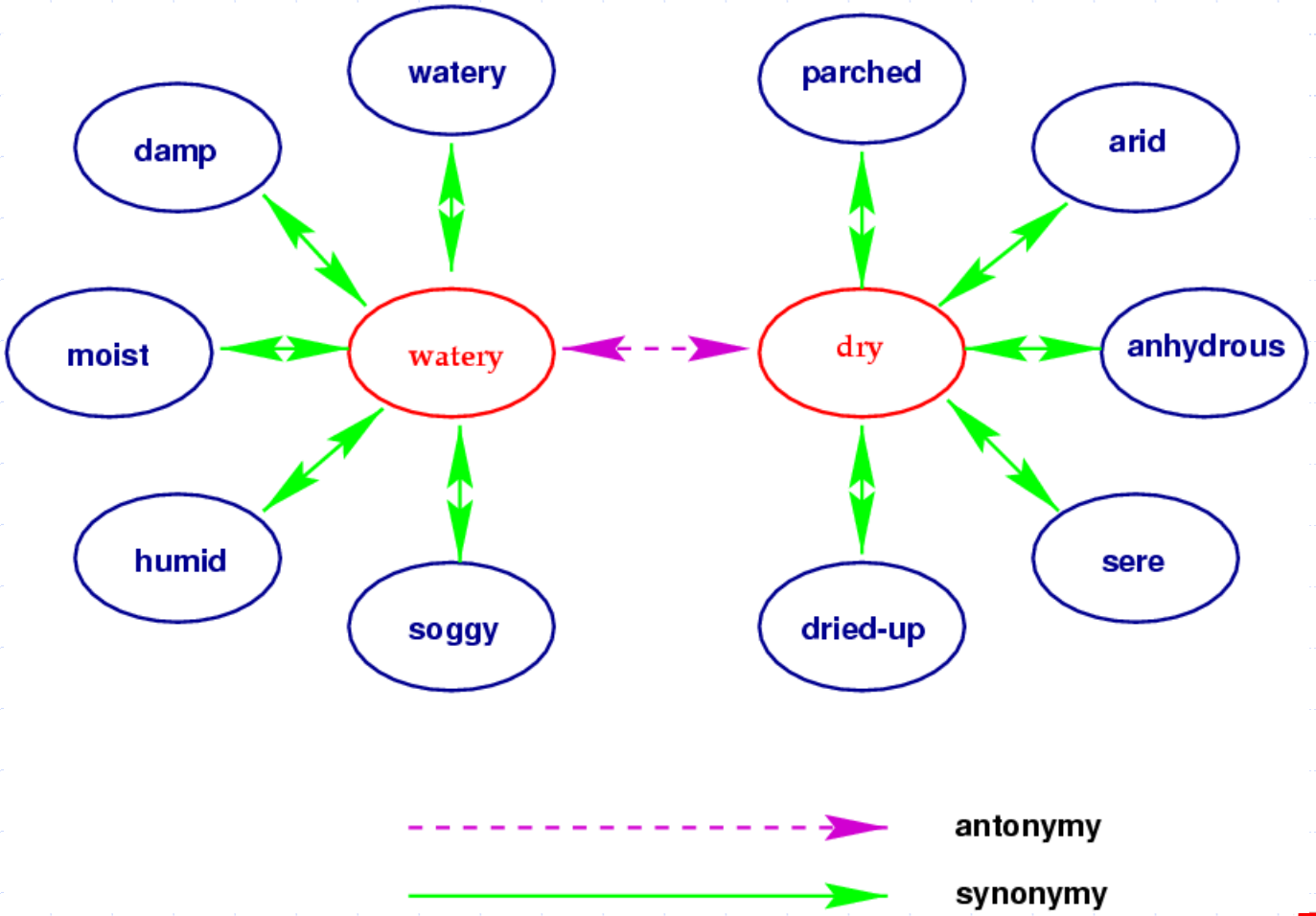
Two words are antonyms if their meanings differ only in the value for a single semantic feature.

- Dead/alive, above/below, hot/cold, fat/skinny,...
- Binary antonyms (dead/alive: [+/- living])
- Gradable antonyms
  - Hot,...,warm,...,cool,...,cold

Semantic organization for descriptive adjectives

- Natural organization in terms of
  - binary opposition (antonymy)
  - Similarity of meaning (synonymy)
- Clusters of adjectives:
  - Associated by semantic similarity to a focal adjective
    - relates the cluster to a contrasting cluster at the opposite pole of the attribute.

Bipolar clusters of Adjectives



Relational adjectives

- Stylistic variants of modifying nouns
  - Cross-referenced to the noun files.
    - They mean something like:
      - “of X”
      - “relating/pertaining to X”
      - “associated with X”
      - Ex. “dental” relates to “tooth”.
    - They do not relates to an attribute, nor to a property of their head nouns.
    - The adjective and the related noun refer to the same concept, but they differ morphologically.
  - Some nouns give rise to two homonymous adjectives:
    - One relational and the other descriptive
    - Ex. “musical instrument” vs “musical child”.
  - More nouns for the same relational adjective:
    - Ex. “chemical” as in “chemical fertilizer” and “chemical engineer”
  - Prefixed adjectives: point to the unprefixed adjectives
    - Ex. Interstellar, extramural, premedical, etc...

Organization of relational adjectives

- Do not have direct antonyms.
  - Can often be combined with “non”
  - In general do not express the opposite value of an attribute, but “Everything else”.
- Wordnet maintains a separate file with pointers to the corresponding nouns
- Each synset consist of one or more relational adjective, followed by a pointer to the appropriate noun
  - Ex. {stellar, astral, sidereal, noun.object:star}

Caracterization of relational adjectives

- Can occur only in attributive position.
  - Not always.
- Do not combine well with descriptive adjectives in modifying the same head.
  - Ex. “nervous and life-threatening disease”
- Absence of corresponding nominalization
  - “nervous person” -> “the person’s nervousness”
  - “nervous disorder” -> *“disorder’s nervousness”
- Relational adjectives are not gradable (like noun modifiers)
    - *”the extremely atomic bomb”
    - *”the very baseball game”

Verbs in Wordnet

- 21.000 verbs word forms
  - 13.000 are unique strings
  - 8.400 word meanings (synsets)
  - Includes phrasal verbs
- Divided into semantic domains
  - Bodily care and functions, change, cognition, communication, competition, consumption, contact, creation, emotion, motion, perception, possession, social interaction, and weather verbs.
    - Events or actions
- States
  - Does not constitute a semantic domain, do not share semantic properties (other than they refer to states).
  - Ex. Suffice, belong, resemble, …

Semantic relations for verbs

- Synonymy
  - Apparently synonymous verbs exhibit subtle meaning differences:
    - Different selectional restrictions
    - Verb synsets often contains *periphrastic expressions*, rather than lexicalised synonyms.
    - Ex. {hammer, (hit with a hammer)}
      - Gloss breaks down a synonymous verb into an entire VP that indicates the *basic action*, the *role of the noun* (material or instrument) with which the action is performed.
    - {whiten, (turn white)}: changes expresed as “become + adjective”.
    - {swimm, (travel through water)}: manner elaborations of a more basic verb.

Verb taxonomies (1)

- Verbs cannot easily arranged into the kind of tree structure onto which nouns are mapped
  - Using semantic relations like:
    - Entailment
    - Temporal inclusion
    - Causation
- Within a single semantic field not all verbs can be grouped under a single unique beginner

Verb Taxonomies (2)

- Some semantic fields must be represented by several independent trees.
  - Motion verbs have two tops nodes: {move, (make a movement)} and {move, travel}.
  - Possession verbs can be traced up to the verbs:
    - {give, transfer}, {take, receive} and {have, hold}.
  - Verbs of bodily care and functions consist of a number of independent hierarchies that form a coherent semantic field.
    - Most of verbs (wash, comb, shampoo, make-up) select for the same kinds of noun argument (body parts).
  - Communication verbs are headed by the verb communicate but immediately divide into verbs of verbal and nonverbal communication
    - Verbal communication divides into actions denoting communication of:
      - Spoken language vs Written language

Lexical entailment

- A verb V1 logically entails a verb V2 when the sentence « Someone V1 » (logically) entails the sentence « Someone V2 ».
  - Ex. “snore” lexically entails “sleep”.
  - The first sentence “presuppose” the second.
- Negation reverses the direction of entailment:
  - Ex. Not sleeping entails not snoring.
- Lexical entailment is a non-symmetric relation:
  - Only synonymous verbs can be mutually entailing
    - Ex. “A defeated B” and “A beat B”.

Temporal inclusion

- A verb V1 will be said to temporally include a verb V2 if there is some stretch of time during which the activities denoted by the two verbs co-occur, but no time during which V2 occurs and V1 does not
  - Ex. “snore” entails “sleep” and is properly included by it.
- If V1 entails V2 and if a temporal inclusion relation holds between V1 and V2, then people will accept a *part-whole* statement relating V2 and V1.

Troponymy

- The troponymy relation between two verbs V1 and V2 can be expressed by the formula:
  - To V1 is to V2 in some particular “manner”.
  - Ex. Troponyms of communication:
    - Encode the speaker’s intention like in
      - Examine, confess, preach, ...
    - Encode the medium of communication like in
      - Fax, email, phone, telex, …
  - Troponymy is a particular kind of entailment:
    - Every troponym V1 of a (more general) verb V2 also entails V2.
    - The activity referred by a troponym and its more general hypernym are always temporally coextensive.
    - Obs. “snore” is not a troponym of “sleep” (because of proper temporal inclusion).

Backward presupposition

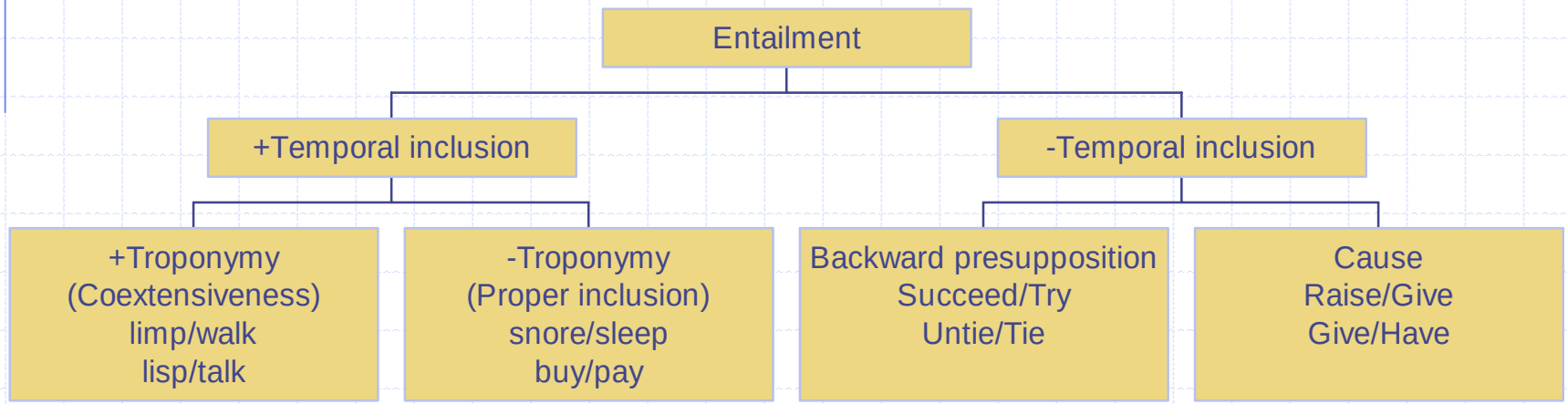
- The activity denoted by the entailed verb always preceeds in time the activity denoted by the entailing verb.
  - Succeed-try
  - Untie-tie

Causation relation

- The causation relation relates two verbs concepts:
  - One causative (like give)
  - One resultative (like have).
- Constraints:
  - The subject of the causative verb usually has a referent that is distinct from the subject of the resultative verb.
  - The subject of the resultative verb must be an object of the causative verb (which is therefore necessarily transitive).
  - Causation is anti-symmetric:
    - For someone to have something does not entail that he was given it.
- Causation is a specific case of entailment:
  - If V1 necessarily causes V2, then V1 also entails V2.
  - Causal entailment lacks temporal inclusion.

Entailment Relations for Verbs

Entailment relations among verbs



Sentence frames for verbs

- Wordnet includes for each verb synset one or several sentence frames
  - Which specify subcategorization features of the verbs
  - Indicating in the synset a list of verb frames illustrating the types of simple sentences in which the verbs in the synset can be used.
- Why?
  - To cover the most important syntactic aspects of verbs
  - Distinctive syntactic behaviour of verbs arises from their semantic components
  - Importance of correlations between verb's semantic makeup and its syntax.

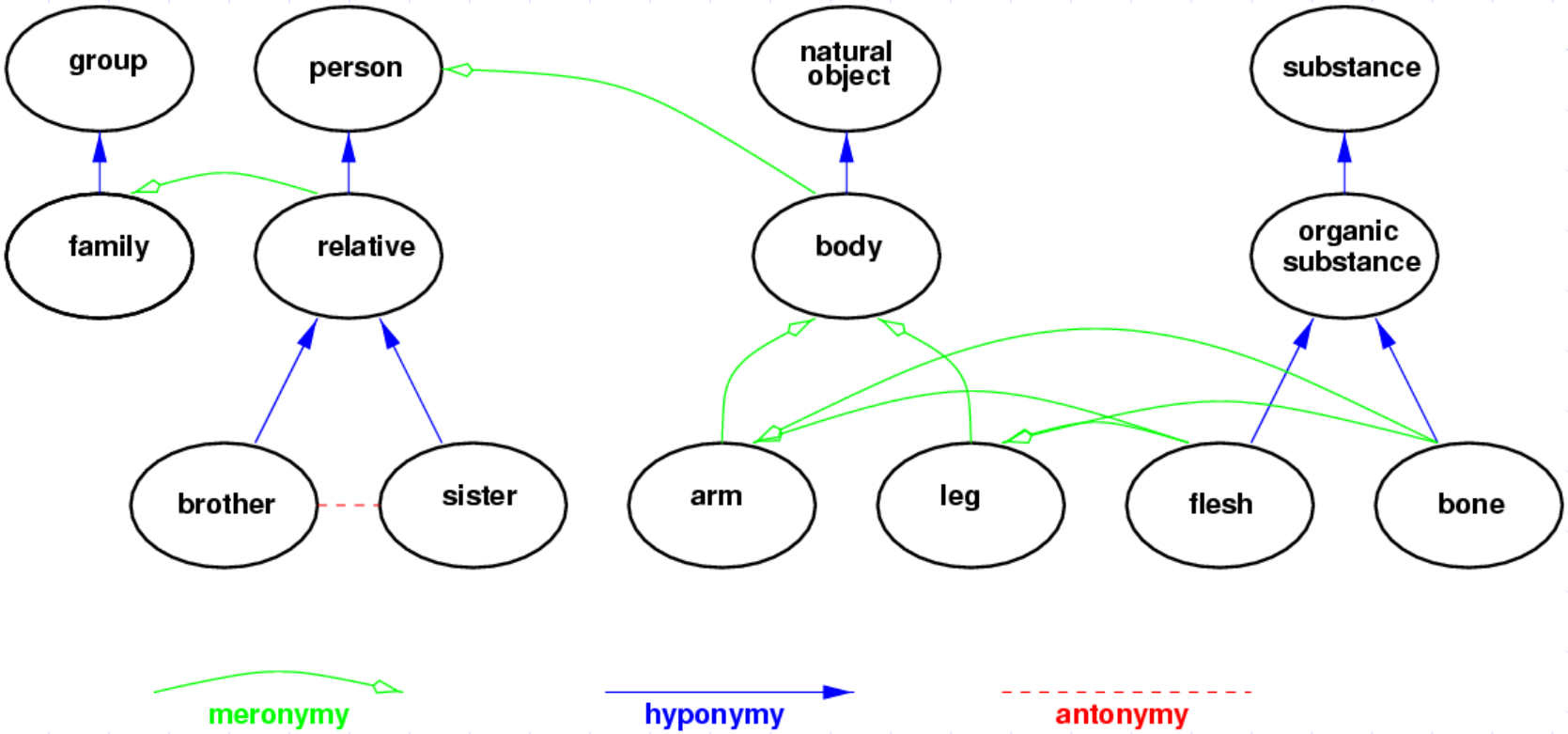
Verbs sentence frames

1 Something s
2 Somebody s
3 It is s-ing
4 Something is s-ing PP
5 Something s something Adjective/Noun
6 Something s Adjective/Noun
7 Somebody s Adjective
8 Somebody s something
9 Somebody s somebody
10 Something s somebody
11 Something s something
12 Something s to somebody
13 Somebody s on something
14 Somebody s somebody something
15 Somebody s something to somebody
16 Somebody s something from somebody
17 Somebody s somebody with something
18 Somebody s somebody of something
19 Somebody s something on somebody
20 Somebody s somebody PP
21 Somebody s something PP
22 Somebody s PP\\
23 Somebody's (body part) s
24 Somebody s somebody to INFINITIVE
25 Somebody s somebody INFINITIVE
26 Somebody s that CLAUSE
27 Somebody s to somebody
28 Somebody s to INFINITIVE
29 Somebody s whether INFINITIVE
30 Somebody s somebody into Ving something
31 Somebody s something with something
32 Somebody s INFINITIVE
33 Somebody s VERBing
34 It s that CLAUSE
35 Something s INFINITIVE

Semantic Relations in Wordnet

Relation	Connects	Number	Examples	Comments
HYPERNYM	n_synset-n_synset	61,123	{oak} → {tree}	IS-A, asymmetric,
	v_synset-v_synset	10817	{hit} → {propel, impel}	transitive
HYPONYM	n_synset-n_synset	61,123	{tree} → {oak}	reverse IS-A
	v_synset-v_synset	10817	{propel, impel} → {hit}	
HAS-MEMBER	n_synset-n_synset	11,472	{family, family unit} → {child, kid}	asymmetric, transitive
IS-MEMBER-OF	n_synset-n_synset	11,472	{child, kid} → {family, family unit}	reverse HAS-MEMBER
HAS-STUFF	n_synset-n_synset	366	{tank, army tank} → {steel}	asymmetric, transitive
IS-STUFF-OF	n_synset-n_synset	366	{steel} → {tank, army tank}	reverse HAS-STUFF
HAS-PART	n_synset-n_synset	5,695	{torso, body} → {shoulder}	asymmetric, transitive
IS-PART-OF	n_synset-n_synset	5,695	{shoulder} → {torso, body}	reverse HAS-PART
ENTAIL	v_synset-v_synset	435	{snore, saw wood} → {sleep, slumber}	asymmetric, transitive
CAUSE-TO	v_synset-v_synset	204	{develop} → {grow, become larger}	asymmetric, transitive
PAST-PARTICIPLE	adj-verb	89	developed → develop	asymmetric
ATTRIBUTE	adj_synset-n_synset	636	{hypocritical} → {insincerity}	asymmetric
SYNSET	n_synset-n	107,484	{place, property} → place	from synsets
	v_synset-v	25,768	{travel, journey} → travel	to words
	adj_synset-adj	28,762	{glad, happy} → happy	
	adv_synset-adv	6,203	{well, much} → well	
PERTAINYM	adj-n	3,458	academic → academia	asymmetric
	adj-adj	3539	universalistic → universal	
	adv-adj	2894	reasonably → reasonable	
ANTONYM	n-n	1713	presence → absence	symmetric
	v-v	1025	rise → fall	
	adj-adj	3748	active → passive	
	adv-adv	704	always → never	
SIMILAR-TO	adj-adj	20050	abridge → shorten	symmetric
SEE-ALSO	v-v	840	touch → touch down	symmetric
	adj-adj	2686	inadequate → unsatisfactory	

A Wordnet Semantic network



Related Projects

- WordNet has inspired the construction of semantic networks in other languages:
  - The newly formed Global WordNet Association is a free, public and non-commercial organization that provides a platform for discussing, sharing and connecting wordnets for all languages in the world.
  - The EuroWordNet project, begun in 1996 under the direction of the University of Amsterdam, is building a multilingual lexical database modelled on WordNet. It currently supports Dutch, Italian, and Spanish.

Extensions to Wordnet

- Paul Buitelaar, of DFKI-Language Technology in Germany, has developed **CoreLex**, an ontology and semantic database of 126 underspecified semantic types, covering around 40,000 nouns. CoreLex defines a large number of systematic polysemous classes, derived by a careful, semi-automatic analysis of sense distributions in WordNet.
- Tool for **Knowledge Extension of WordNet** with Prof. Dan Moldovan. It disambiguates all WordNet glosses and transforms them in logical formulae.

Wordnet interfaces

- A **"one-touch" interface** to WordNet 1.6 developed by Greg Peterson at Notre Dame Women's College in Kyoto, Japan.
- A web-based **SQL interface** to WordNet 1.6 developed by Mark Julien at Oxford English Online Limited.
- An **SQL-based interface** developed by Chris Greaves. This interface, containing about 25,000 Chinese entries, allows you to search WordNet alphabetically, and for co-occurrences.
- A **visual navigation tree interface** for WordNet 1.6 has been developed by Jorge J. Gomez Sanz, under the direction of Manuel de Buenaga.
- A **WWW based Python inteface** to WordNet developed by Francios Yvon at ENST.
- A **"one-touch" interface** to WordNet 1.5 allows you to select multiple searches at one time. This was developed by Andrew Daviel at Vancouver Webpages.
- An **interactive CGI interface** from which you can select many searches at once was developed at E.N.S.T in France by Francois Yvon, Didier Verna and several undergraduate students. This interface is in French.
- Noah, a cool PalmPilot reader for WordNet 1.6 was developed by Krzysztof Kowalczyk.
- WordNet TreeWalk, a W95/98/NT GUI interface to WordNet 1.6 based on tree-views developed by Bernard Bou, Lycée Champollion, France.

Programming language interfaces

- A really cool Java/WAP interface that allows you to navigate WordNet from your mobile phone was developed by Joris Van den Bogaert.
- Ken Bowen has developed a direct interface from Prolog to the WordNet database that avoids the problems of loading the fact-based Wordnet Prolog database into memory.
- A new version of the Java library, written by John Didion, has been released.
- A Perl extension module for accessing and manipulating WordNet has been developed by Dan Brian.
- Another Perl interface has been developed by Jason Rennie at the MIT AI Lab.
- A Mac OS X Server and Mac OS X front end developed by Erik Doernenburg. This application is a Mac-like front end for the WordNet database, and requires a separate download of a local copy of the database files.
- A Python based interface to WordNet 1.6 developed by Oliver Steele. This allows the user to type expressions such as hyponyms(N['dog'][0]), closure(ADJ['red'], SYNONYM), and meet(N['dog'][0], N['cat'][0]) to compute lexical relations over the database. He has also developed JWordNet, a Java standalone object-oriented interface.
- A Common Lisp interface to WordNet has been developed by Mark Nahabedian at the MIT AI Lab..
- Another Lisp interface developed by Wheeler Ruml is also available for download. It has been tested in Allegro Common Lisp for HP-UX.