

# Computerlinguistik II / Sprachtechnologie

Vorlesung im SS 2018  
(M-GSW-10)

**Prof. Dr. Udo Hahn**

Lehrstuhl für Computerlinguistik  
Institut für Germanistische Sprachwissenschaft  
Friedrich-Schiller-Universität Jena

<http://www.julielab.de>

# Computational Lexicons, Terminologies & Ontologies

- **Computational Lexicons**
  - Language-specific information (English, Spanish, German, etc.), cover common-sense knowledge
  - Cover, at best, all linguistic description levels for a lexical item but usually don't
  - Undetermined towards formalization, yet electronically available
- **Terminologies**
  - Language-independent (though verbally encoded!), cover domain-specific, expert-level knowledge
  - Cover lexico-semantic information only (semantic relations)
  - Informal, computational issues are (usually) of no concern
- **Ontologies**
  - Language-independent, cover domain-specific, expert-level knowledge
  - cover conceptual information (semantic relations, semantic integrity constraints, rules, etc.)
  - Formal specifications, computational issues are a major concern
  - Formal reasoning: inferences

# Examples: Computational Lexicons, Terminologies & Ontologies

- **Computational Lexicons**
  - WordNet (English) & EuroWordNet
  - GermaNet
  - FrameNet
- **(Biomedical) Terminologies**
  - Unified Medical Language System (UMLS)
  - BioPortal
  - Open Biological Ontologies (OBO)
    - Gene Ontology (GO)
- **Ontologies**
  - Formal reasoning (for text understanding)

# *WordNet*

- **English WordNet (V3.0)**
  - semantic (relation) lexicon of English (general language)
    - no morphology!, no syntax!, no etymology
  - groupings of words into sets of synonyms (**synsets**)
  - English definitions for lexical entries/synsets (**glosses**)
  - defines semantic relations between synsets
  - covers (base forms of) nouns, verbs, adjectives, adverbs
  - Size: more than 155,000 lexical entries

<http://wordnet.princeton.edu/>

# *WordNet*

- EuroWordNet

- Portuguese, Spanish, Spanish Catalan-Basque
- French
- Italian
- German (licence required), Dutch
- Russian, Czech, Hungarian, Slovene, ...
- ...

- Global WordNet

<http://globalwordnet.org>

- Arabic
- Mandarin-Chinese
- Hindi
- ....

# WordNet SynSets and Glosses

- Nouns

S: (n) **jump**, **leap** (**a sudden and decisive increase**) *"a jump in attendance"*

direct hyponym / full hyponym

S: (n) quantum leap, quantum jump (**a sudden large increase or advance**) *"this may not insure success but it will represent a quantum leap from last summer,,*

direct hypernym / inherited hypernym/ sister term derivationally related form

S: (n) **leap**, **jump**, **saltation** (**an abrupt transition**) *"a successful leap from college to the major leagues"*

S: (n) **jump** ((**film**) **an abrupt transition from one scene to another**)

S: (n) **startle**, **jump**, **start** (**a sudden involuntary movement**) *"he awoke with a start"*

S: (n) **jump**, **parachuting** (**descent with a parachute**) *"he had done a lot of parachuting in the army"*

S: (n) **jump**, **jumping** (**the act of jumping; propelling yourself off the ground**) *"he advanced in a series of jumps"; "the jumping was unexpected"*

# WordNet Synsets and Glosses

- Verb

- S: (v) **jump**, **leap**, **bound**, **spring** (move forward by leaps and bounds) *"The horse bounded across the meadow"; "The child leapt across the puddle"; "Can you jump over the fence?"*
- S: (v) **startle**, **jump**, **start** (move or jump suddenly, as if in surprise or alarm) *"She startled when I walked into the room"*
- S: (v) **jump** (make a sudden physical attack on) *"The muggers jumped the woman in the fur coat"*
- S: (v) **jump** (increase suddenly and significantly) *"Prices jumped overnight"*
- S: (v) **leap out**, **jump out**, **jump**, **stand out**, **stick out** (be highly noticeable)
- S: (v) **jump** (enter eagerly into) *"He jumped into the game"*
- S: (v) **rise**, **jump**, **climb up** (rise in rank or status) *"Her new novel jumped high on the bestseller list"*
- S: (v) **jump**, **leap**, **jump off** (jump down from an elevated point) *"the parachutist didn't want to jump"; "every year, hundreds of people jump off the Golden Gate bridge"; "the widow leapt into the funeral pyre"*
- S: (v) **derail**, **jump** (run off or leave the rails) *"the train derailed because a cow was standing on the tracks"*
- S: (v) **chute**, **parachute**, **jump** (jump from an airplane and descend with a parachute)
- S: (v) **jump**, **leap** (cause to jump or leap) *"the trainer jumped the tiger through the hoop"*
- S: (v) **jumpstart**, **jump-start**, **jump** (start (a car engine whose battery is dead) by connecting it to another car's battery)
- S: (v) **jump**, **pass over**, **skip**, **skip over** (bypass) *"He skipped a row in the text and so the sentence was incomprehensible"*
- S: (v) **leap**, **jump** (pass abruptly from one state or topic to another) *"leap into fame"; "jump to a conclusion"; "jump from one thing to another"*
- S: (v) **alternate**, **jump** (go back and forth; swing back and forth between two states or conditions)

# *WordNet Relations*

- Nouns

- Hypernyms

- „Y is a hypernym (more general term) of X, if every X is a (kind of) Y“

- Hyponyms

- „Y is a hyponym (more specific term) of X, if every Y is a (kind of) X“
    - Y is hyponym of X  $\leftrightarrow$  X is a hypernym of Y

- Coordinate terms

- „Y is a coordinate term of X, if X and Y share a hypernym“

- Holonyms

- „Y is a holonym (whole) of X, if (every/some?) X is a part of Y“

- Meronyms

- „Y is a meronym (part) of X, if (every/some?) Y is a part of X“
    - Y is a meronym of X  $\leftrightarrow$  X is a holonym of Y



# *WordNet Relations*

- Verbs
  - Hypernyms
    - „the verb Y is a hypernym (more general term) of the verb X, if the activity X is a (kind of) Y“
      - e.g., *travel* to *movement*
  - Troponyms
    - „the verb Y is a troponym of the verb X, if the activity Y is doing X in some manner“
      - e.g., *lisp* to *talk*
  - Entailment
    - „the verb Y is entailed by the verb X, if by doing X you must be doing Y“
      - e.g., *snoring* by *sleeping*
  - Coordinate terms
    - „Y is a coordinate verb of X, if X and Y share a hypernym“

# *WordNet Relations*

- **Adjectives**
  - Related nouns
  - Participle of verb
- **Adverbs**
  - Root adjectives

# *WordNet V3.0 Statistics*

POS	Unique Strings	SynSets	Word-Sense Pairs (word - #synset pairs)
Noun	117,100	82,100	146,300
Verb	11,500	13,800	25,000
Adj	21,500	18,200	30,000
Adv	4,500	3,600	5,600
$\Sigma$	155,300	117,700	206,900

# *WordNet V3.0 Statistics*

POS	Monosemous Words / senses	Polysemous Words	Polysemous senses
Noun	101,900	16,000	44,400
Verb	6,300	5,300	18,800
Adj	16,500	5,000	14,400
Adv	3,700	700	1,800
$\Sigma$	128,400	27,900	79,500

# *WordNet V3.0 Statistics*

POS	Average polysemy* (incl. monosemous words)	Average polysemy* (excl. monosemous words)
Noun	1.24	2.79
Verb	2.17	3.57
Adj	1.40	2.71
Adv	1.25	2.50

\* Number of synsets that contain the word

<http://wordnet.princeton.edu/wordnet/man/wnstats.7WN.html>

# GermaNet

- Lexikalische Einheiten sind nach Lesarten gruppiert:

- “Bank” has 2 readings
  - Reading 1: [*Bank*, {*Sitzbank*}] (*bench*)
  - Reading 2: [*Bank*, {*Geldinstitut*}] (*financial institution*)
- “Leiter” has 3 readings
  - Reading 1: [*Leiter*, {*Steiggerät*}] (*ladder*)
  - Reading 2: [*Leiter*, {*Verantwortlicher*, *Anführer*}] (*leader*)
  - Reading 3: [*Leiter*, {*stromleitender Stoff*}] (*electric conductor*)

# GermaNet: SynSets

## Synsets in GermaNet

- Semantic concepts are represented by *synsets*
- A synset is a set of (near-)synonymous words

