

Building and Maintenance of Semantic Networks

Tomáš Čapek

March 21, 2011

Natural Language Processing

Efforts to make computers understand our language

Philosopher's stones:

Machine translation on general domain

HCI based on natural language

Crisis in NLP

Often unclear goals and results (word juggling)

No real major progress after decades of research in:

Syntax, WSD, MT

What we're good at: morphology, KR, corpora, CAT, ...

Main Approaches in NLP

Statistical (Praha)

Rule-based (Brno)

Both use language resources to study languages

Language Resources in NLP

Corpora

Dictionaries

Knowledge Representation schemes

Ontologies, Semantic Networks

Applications of NLP

Dictionary creation (corpora)

Information Retrieval (google)

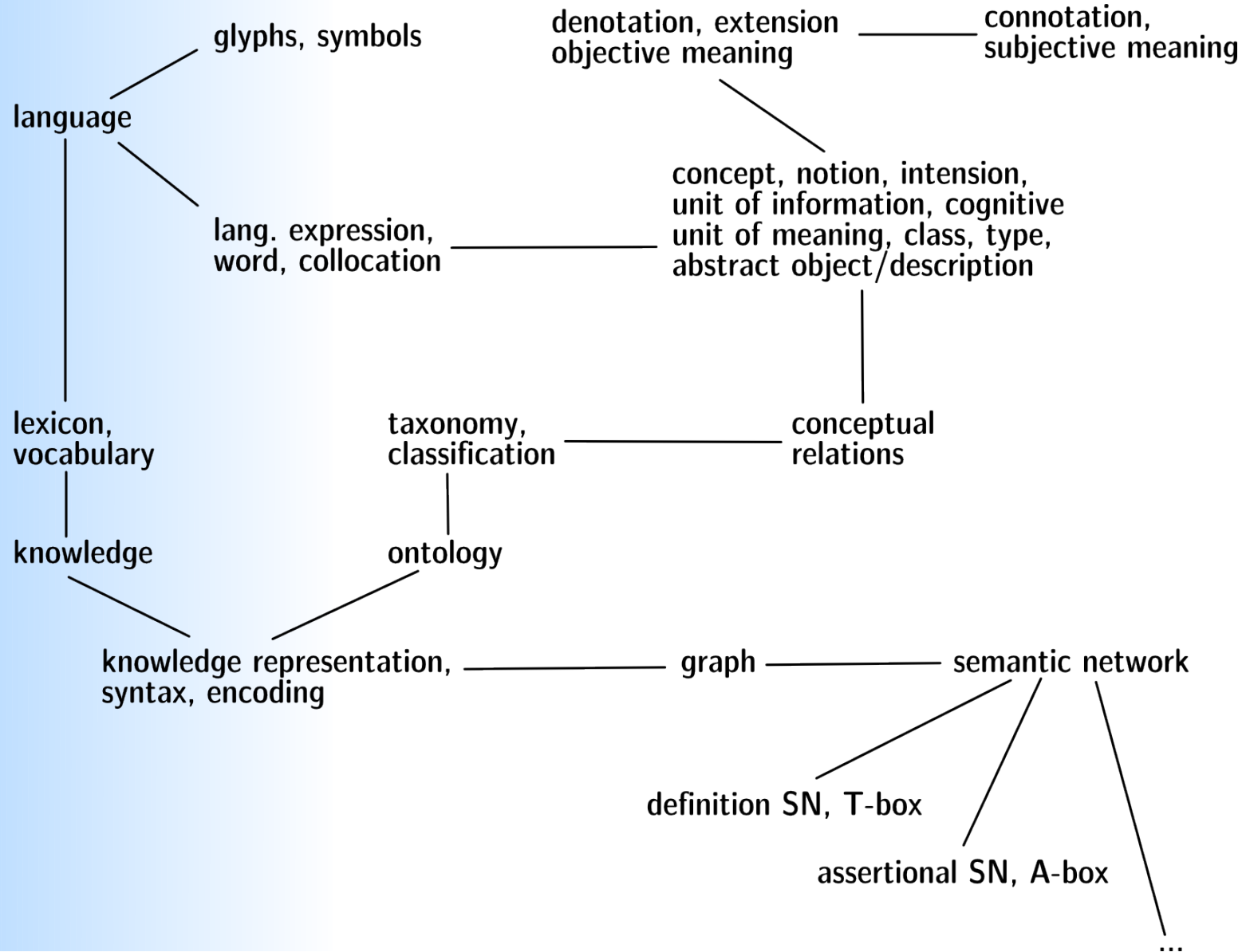
Information Extraction (summarization)

Text Categorization, plagiarism detection

Word Sense Disambiguation

Machine Translation, Dialogue Systems

Semantics in NLP



Role of Annotated Resources

Most valuable for both „rules“ and „statistics“

Most expensive to create by hand

Often „set in stone“ and thus
unreliable and compromising

Issues in Semantic Networks

Style guide (nonexistent, too big, vague)

Balance in the data

Maintenance (manual, random, automated)

Accumulated errors of all kind

Application-unfriendly

→ GIGO

My Proposed Contribution

Application-driven development (merge model)

Clusters of data separate of ontologies

Heuristic tests for automated maintenance

Focused on precision rather than recall

Chain of Succession

morphology → MWE/NE recognition →
(syntax) → knowledge representation (SN) →
semantic annotation → WSD → MT