Computation linguistic

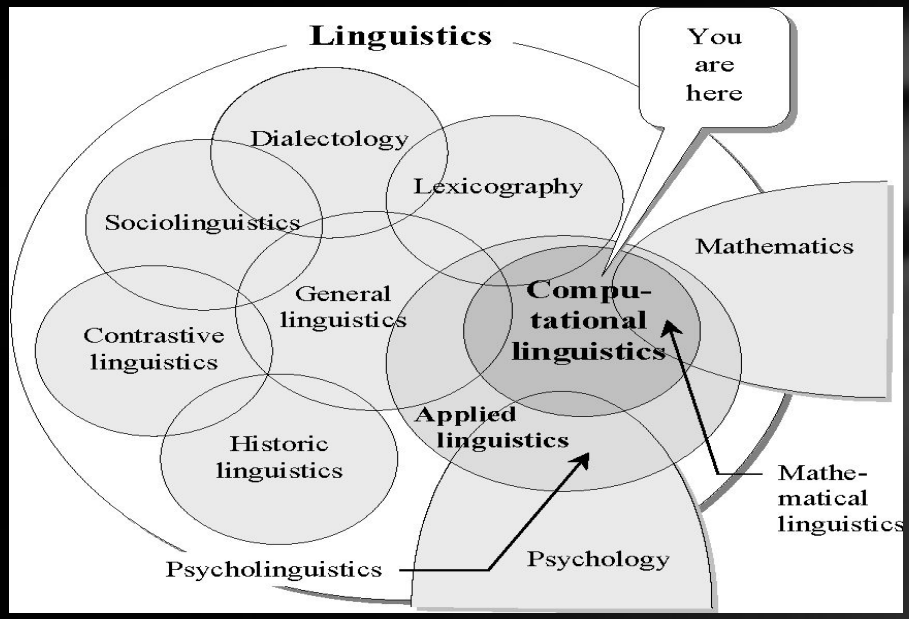
SI – 4

Daria Startseva & Alyona Gordeichuk

What is computational linguistics?

The Association for Computational Linguistics (ACL) describes computational linguistics as the scientific study of language from a computational perspective.

Computational linguistics (CL) combines resources from linguistics and computer science to discover how human language works.



Computational linguists create ***tools*** for important practical tasks such as Machine translation, Natural language interfaces to computer systems, Speech recognition, Text to speech generation, Automatic summarization, E-mail filtering, Intelligent search engines .

CL vs. NLP

Why say "Computational Linguistics (CL)" versus "Natural Language Processing" (NLP)?

Computational Linguistics

- ✔ The science of computers dealing with language
- ✔ Some interest in modeling what people do

Natural Language Processing

- ✔ Developing computer systems for processing and understanding human language text

Why is computation linguistics hard?

Human languages:

- are highly ambiguous at all levels
- are complex , with recursive structures and reference
- subtly exploit context to convey meaning
- are fuzzy and vague
- require reasoning about the world for understanding
- are part of a social system: persuading, insulting, amusing…

Computational linguistic students study subjects such as :

- semantic
- computational semantics
- syntax
- models in cognitive science
- natural language processing systems and applications
- morphology
- linguistic phonetics
- phonology.

Also study: sociolinguistics, psycholinguistics, corpus linguistics, machine learning, applied text analysis, grounded models of meaning, data-intensive computing for text analysis, and information retrieval.

Phonetics and phonology

Phonetics studies the sounds of a language

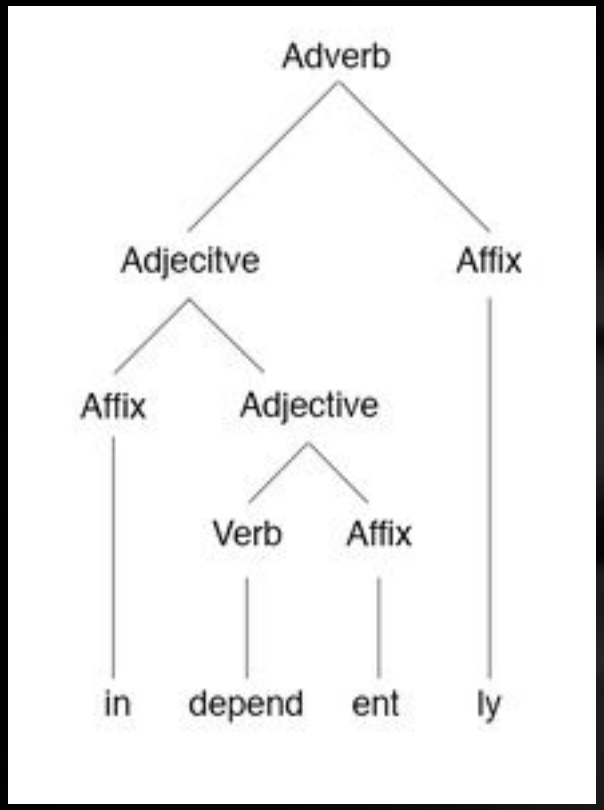
- [t] and [d] differ in voice onset time
- English aspirates stop consonants in certain positions (e.g., [t hop] vs. [stop])

Phonology studies the distributional properties of these sounds

- the English noun plural is [s] following unvoiced segments and [z] following voiced segments
- English speakers pronounce /t/ differently (e.g., in water)

Morphology

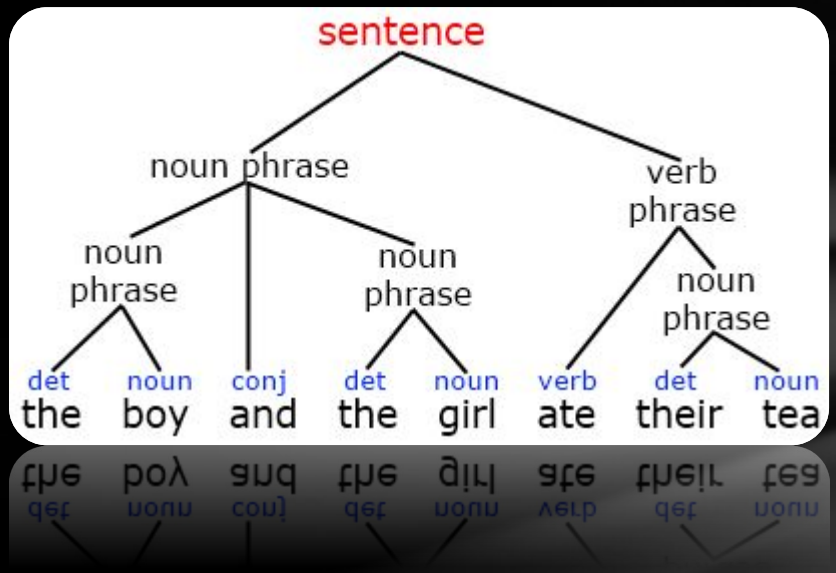
- Morphology studies the structure of words



The suffix usually determines the syntactic category of the derived word

Syntax

Syntax studies the ways words combine to form phrases and sentences



Syntactic parsing helps identify who did what to whom, a key step in understanding a sentence

Semantics and pragmatics

Semantics studies the meaning of words, phrases and sentences

E.g., I ate the oysters in/for an hour.

Pragmatics studies how we use language to do things in the world

E.g., Can you pass the salt?

.

Machine translation

Input: a sentence (usually text) f in the source language

Output: a sentence e in the target language

Challenges for Machine Translation:

- the best translation of a word or phrase depends on the context
- the order of words and phrases varies from language to language
- there's often no single "correct translation"

Why are the results so poor?

- ✔ Language understanding is complicated
- ✔ The necessary knowledge is enormous
- ✔ Most stages of the process involve ambiguity
- ✔ Many of the algorithms are computationally intractable

Companies

- Alelo
- Apple
- Expert System
- Facebook
- Google
- Intel
- Lingsoft
- Lionbridge
- Microsoft
- North Side
- Nuance
- Oracle
- SDL