

# Computation linguistic

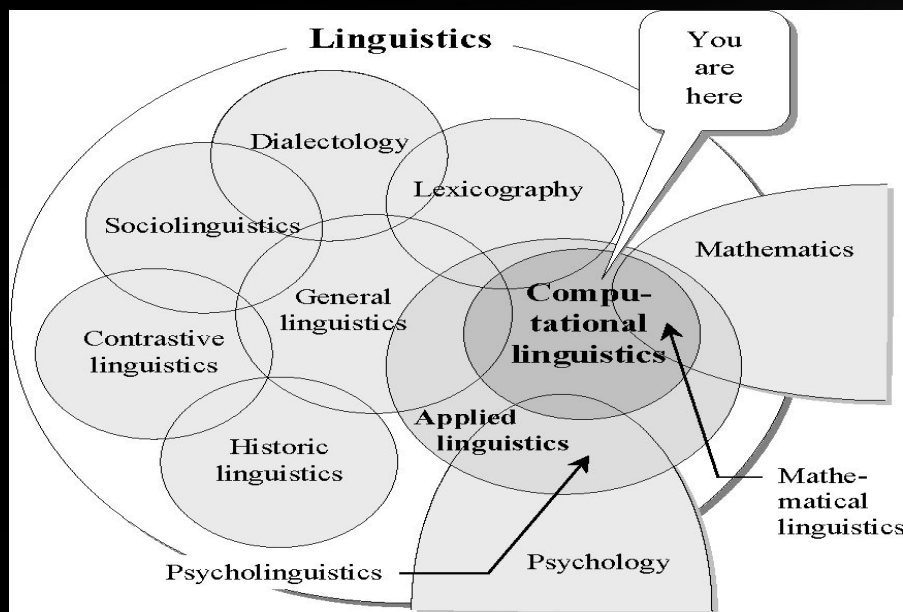
*SI – 4*

*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

Computational linguistics has theoretical and applied components.

Theoretical computational linguistics focuses on issues in theoretical linguistics and cognitive science, and applied computational linguistics focuses on the practical outcome of modeling human language use.

Computational and quantitative methods are also used historically in attempted reconstruction of earlier forms of modern languages and subgrouping modern languages into language families. Earlier methods such as **lexicostatistics** and **glottochronology** have been proven to be premature and inaccurate.

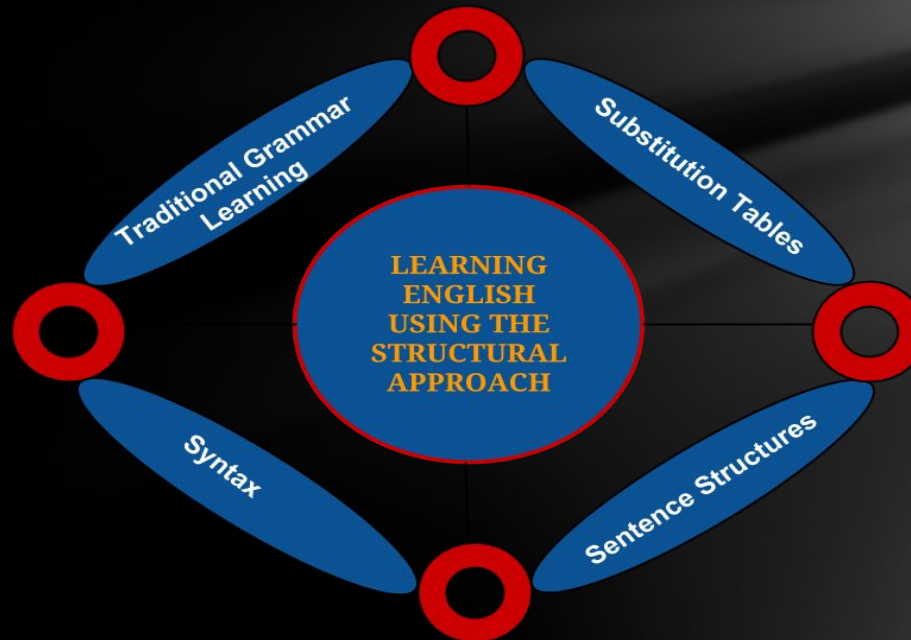
# Developmental approaches

Language is a cognitive skill which develops throughout the life of an individual. This developmental process has been examined using a number of techniques, and a computational approach is one of them. Human language development does provide some constraints which make it harder to apply a computational method to understanding it

Attempts have been made to model the developmental process of language acquisition in children from a computational angle, leading to both statistical grammars and connectionist models.

# Structural approaches

One of the most important pieces of being able to study linguistic structure is the availability of large linguistic corpora, or samples. This grants computational linguists the raw data necessary to run their models and gain a better understanding of the underlying structures present in the vast amount of data which is contained in any single language.



# 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.



# 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

