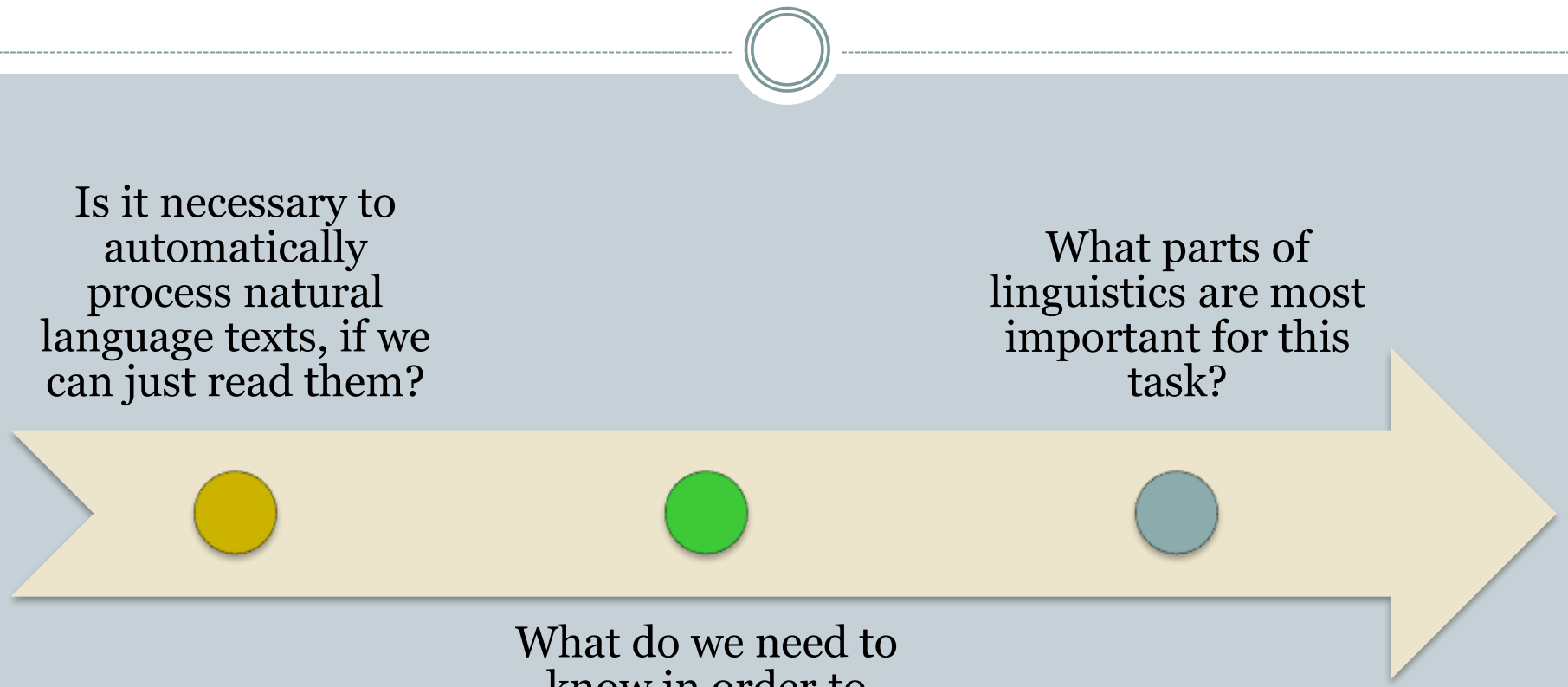
Computational Linguistics

INTRODUCTION
LECTURE 2

Outline

Topical Questions



Computer-Assisted Realia

- create
- proofread
- store
- manage
- search

Alexander Gelbukh, 2014:

The great challenge of the problem of intelligent automatic text processing is to use unrestricted natural language to exchange information with a creature of a totally different nature: the **computer**.

mechanical assistance

intellectual assistance

Intellectual Assistance

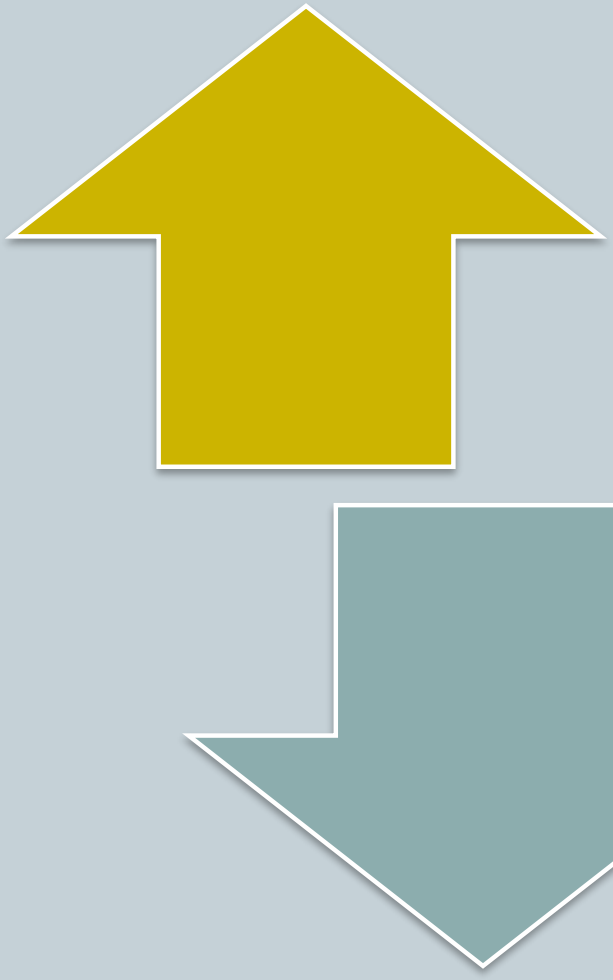
read an unprepared text

test the text for correctness

execute the instructions contained in the text

comprehend the text well enough to produce a reasonable response based on its meaning

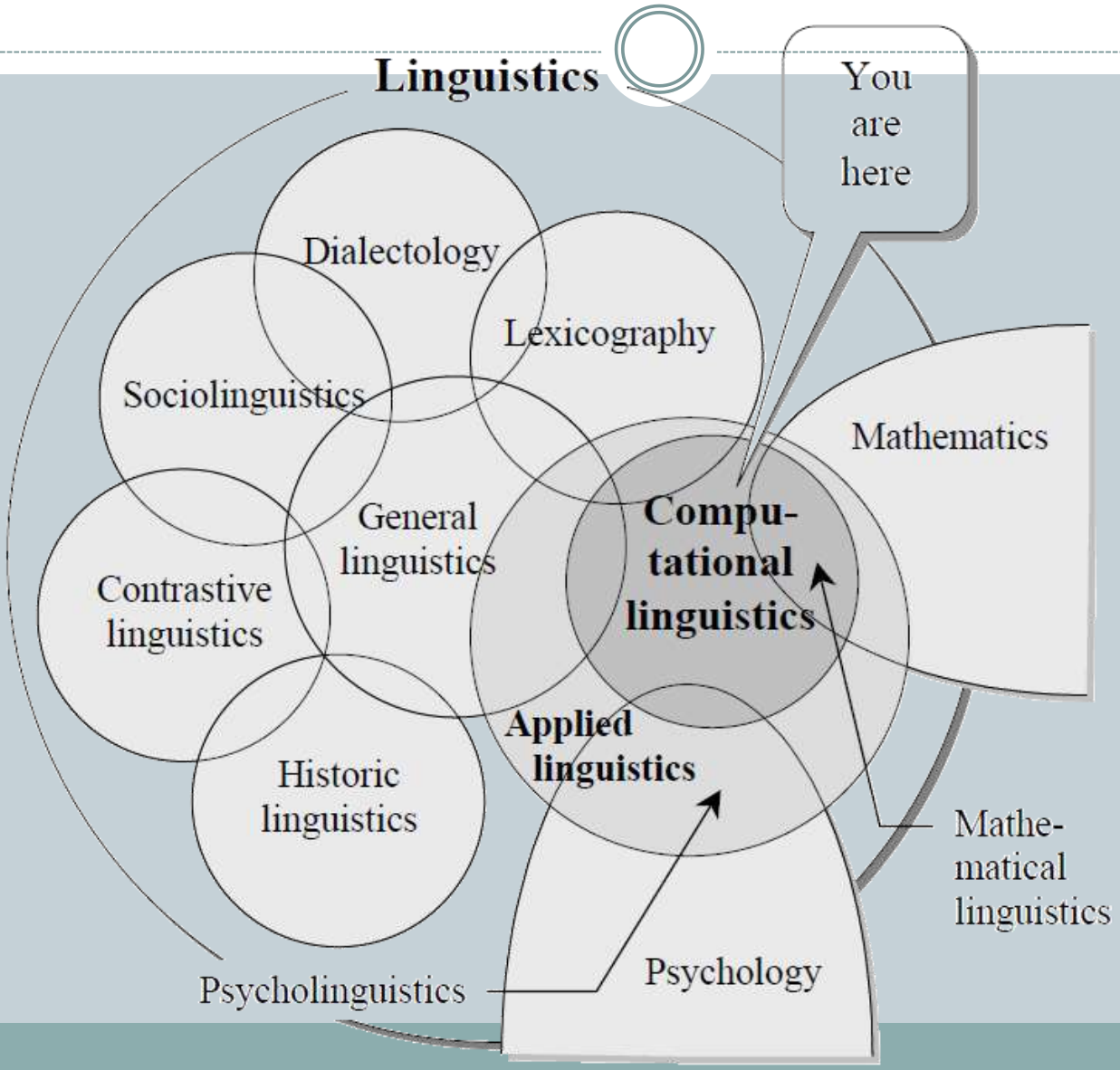
Automatic Text Processing Necessity Circumstance



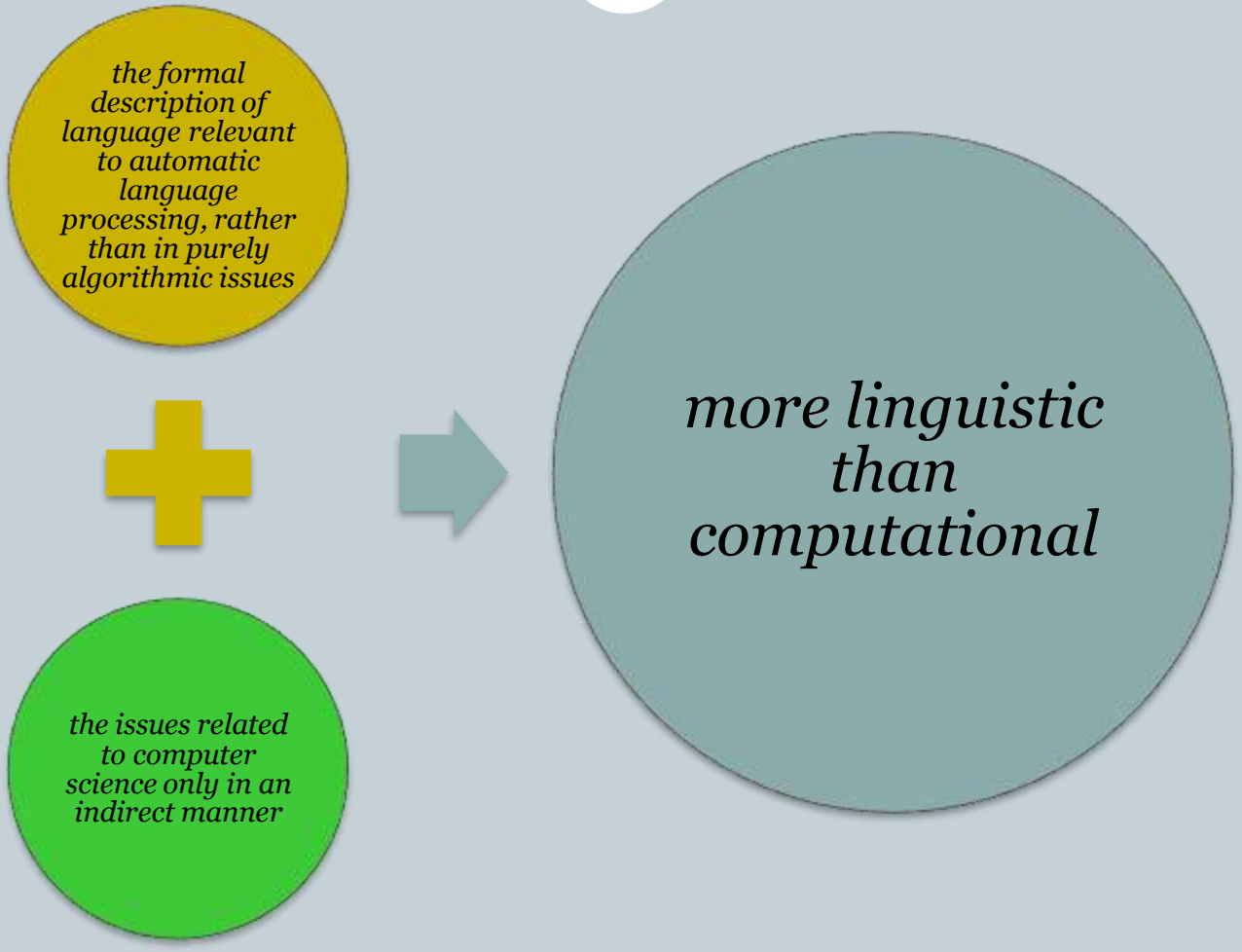
lack of knowledge and education, time and wish to meet the modern standards of document processing

need to take into consideration a quantity of texts thousands times larger than one person is physically able to read in a lifetime

Computational Linguistics and Linguistics Models



Course Objectives



Natural Language

- *Ferdinand de Saussure* considers natural language as a structure of mutually linked elements, similar or opposed to each other.
- *Leonard Bloomfield* claims for a fully "objective" description of natural languages, with special attention to superficially observable facts.
- Thus, sentences are split into the so-called *immediate constituents*, or *phrases*, which are in their turn split into *subconstituents*, etc., down to single words.
- Such a method of syntactic structuring was called the ***phrase structure***, or ***constituency approach***.

Noam Chomsky Study

The generative grammars produce strings of symbols, and sets of these strings are called *formal languages* (texts)

The phrase structures were formalized as *context-free grammars* (CFG) and became the basic tool for description of natural languages

Context-Free Grammar for Generating a Simple English Sentences

- initial symbol ***S*** of a sentence to be generated and several other ***non-terminal*** symbols:
  - the noun phrase symbol ***NP***
  - verb phrase symbol ***VP***
  - noun symbol ***N***
  - verb symbol ***V***
  - determinant symbol ***D***

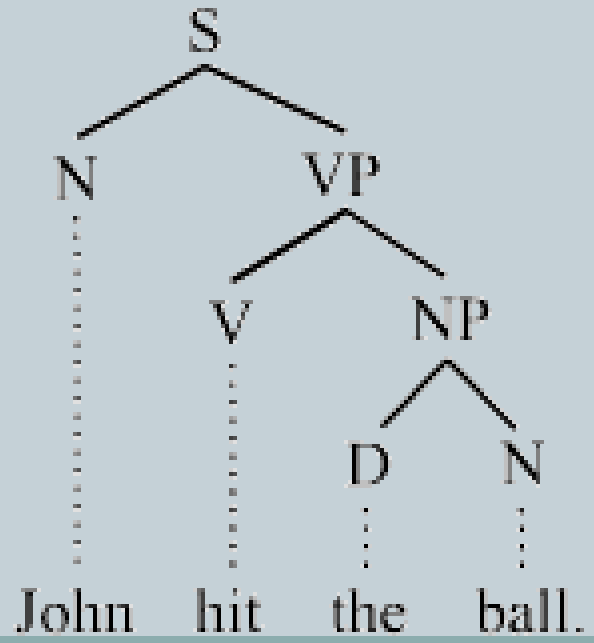
All non-terminal symbols are interpreted as ***grammatical categories.***

Production Rules

- $S \rightarrow NP\ VP$
- $VP \rightarrow V\ NP$
- $NP \rightarrow D\ N$
- $NP \rightarrow N$

Constituency Tree

- Syntactic structure of a sentence was identified with the so-called ***constituency tree*** *(*a nested structure subdividing the sentence into parts, then these parts into smaller parts, and so on*)*.



Constituency-based parse tree

Transformational Grammar

- A transformational grammar is a set of rules for such insertions, permutations, movements, and corresponding grammatical changes. Such a set of transformational rules functions like a program. It takes as its input a string constructed according to some context-free grammar and produces a transformed string [Gelkbuch, 2014].

Transformational Grammar