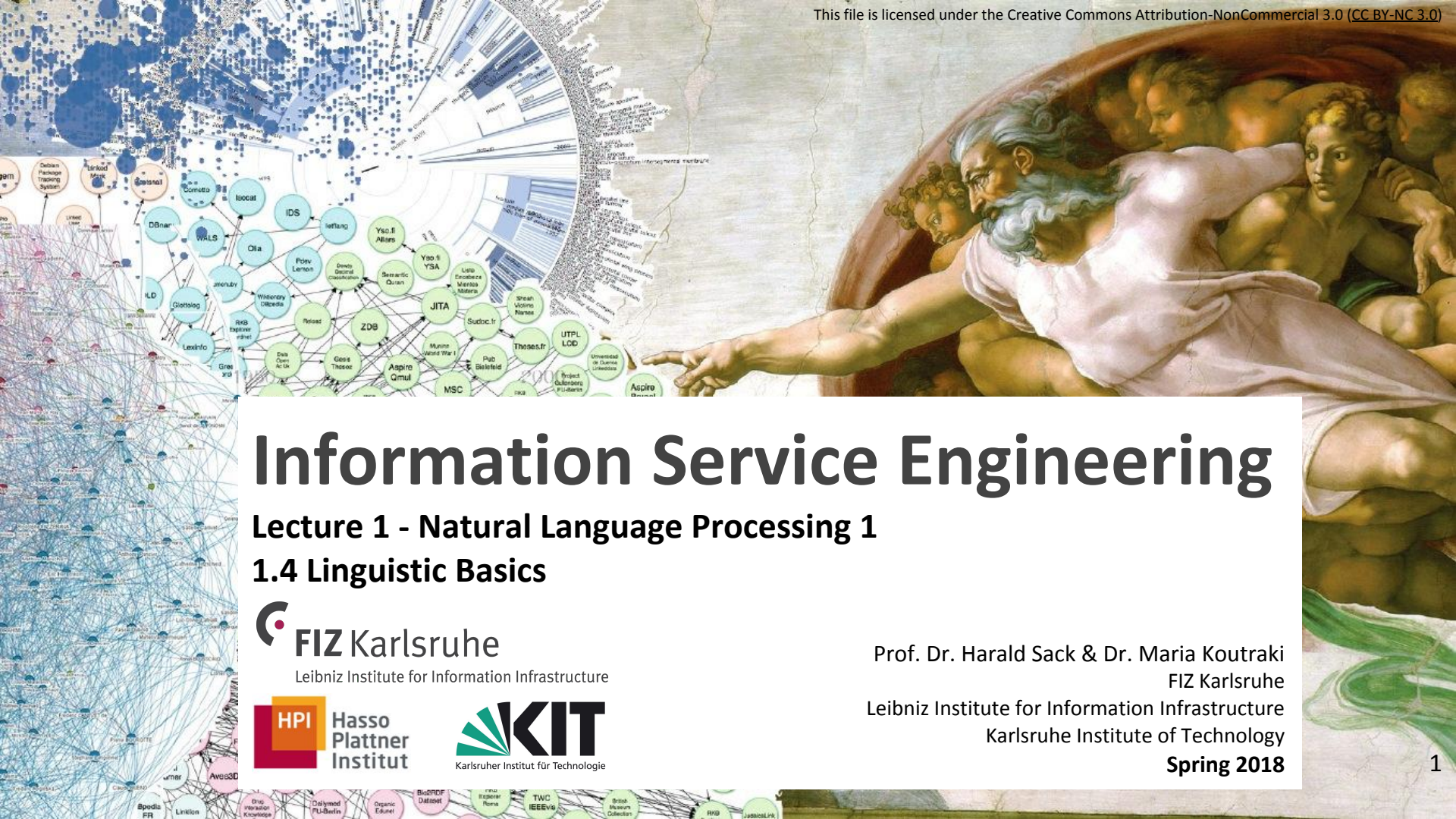




Information Service Engineering

Lecture 1 - Natural Language Processing 1

1.4 Linguistic Basics



FIZ Karlsruhe

Leibniz Institute for Information Infrastructure



Hasso
Plattner
Institut



Karlsruher Institut für Technologie

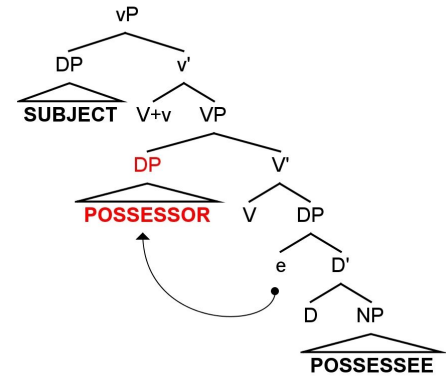
Prof. Dr. Harald Sack & Dr. Maria Koutraki
FIZ Karlsruhe

Leibniz Institute for Information Infrastructure
Karlsruhe Institute of Technology
Spring 2018

Information Service Engineering

Lecture 1: Natural Language Processing

- 1.1 Information and Knowledge
- 1.2 Language and Understanding
- 1.3 NLP Techniques and Challenges
- 1.4 Linguistic Basics**
- 1.5 Morphology
- 1.6 Evaluation, Precision and Recall
- 1.7 Regular Expressions



- is the **scientific study of language and its structure**
- and involves an analysis of
 - language **form** (*phonetics, syntax and grammar*),
 - language **meaning**, and
 - language in **context**

Natural Language Processing

Information Service Engineering - Lecture 1 Natural Language Processing - 1.4 Linguistic Basics

- is a field of computer science, artificial intelligence, and computational linguistics and
- is concerned with the **interactions between computers and human (natural) languages** and, in particular,
- is concerned with **programming computers to fruitfully process large natural language corpora.**
- Specifically the task **to extract meaningful information from natural language input or to produce natural language output.**

- is concerned with the **systematic organization of sounds** in languages, i.e. the abstract, grammatical characterization of systems of sounds (or signs)
- at all levels of language **where sound is considered to be structured for conveying linguistic meaning.**

Phone

- **any distinct speech sound**, regardless of whether the exact sound is critical to the meanings of words

Phoneme

- **smallest (abstract cognitive) sound unit** in a language that is able of conveying a distinct meaning
- **Example:**
 - “s” and “r” in “sing” and “ring”
 - “ss” and “ll” in “kiss” and “kill”

- The study of **internal structures (formation) of words** and **how they can be modified**
- Parsing complex words into their components

⇒ **What is a word?**

Linguistic Basics

Information Service Engineering - Lecture 1 Natural Language Processing - 1.4 Linguistic Basics

A **word** (w_i) is the smallest **independent** unit of language

“Independent”?

- do not depend on other words
- can be separated from other units
- can change position



A **word** (w_i) is the smallest **independent** unit of language

“Independent”?

- do not depend on other words
- can be separated from other units
- can change position

- **Example:**

*The man looked at the **horses***

- **s** is the plural marker, dependent on the noun *horse* to receive meaning
- **Horses** is a word: can occur in other positions or stand on its own

Linguistic Basics

Information Service Engineering - Lecture 1 Natural Language Processing - 1.4 Linguistic Basics

A **vocabulary** consists of a set of **words** (w_i)



Linguistic Basics

Information Service Engineering - Lecture 1 Natural Language Processing - 1.4 Linguistic Basics

A **text** is composed of a sequence of **words** from a **vocabulary**

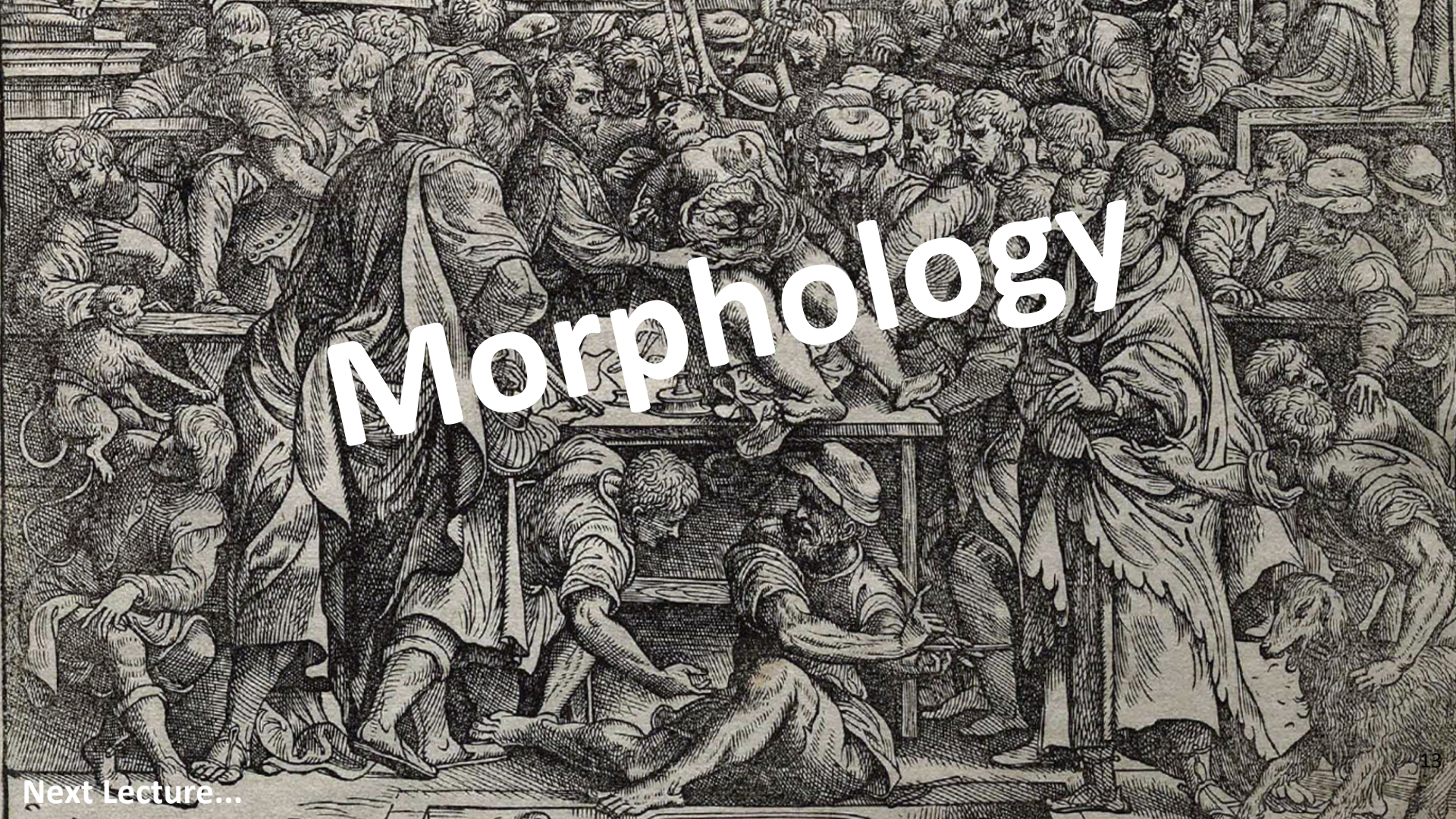
Shall I compare thee to a summer's day?
Thou art more lovely and more temperate:
Rough winds do shake the darling buds of May,
And summer's lease hath all too short a date:
Sometime too hot the eye of heaven shines,
And often is his gold complexion dimm'd,
And every fair from fair sometime declines,
By chance, or nature's changing course untrimm'd:
But thy eternal summer shall not fade,
Nor lose possession of that fair thou ow'st,
Nor shall death brag thou wander'st in his shade,
When in eternal lines to time thou grow'st,
 So long as men can breathe, or eyes can see,
 So long lives this, and this gives life to thee.

Linguistic Basics

Information Service Engineering - Lecture 1 Natural Language Processing - 1.4 Linguistic Basics

A **language** is
constructed of a
set of all possible texts





Morphology

Next Lecture...

Picture References:

- P.8: <https://pixabay.com/p-2041816/>
- P.10: Scrabble game in progress, https://commons.wikimedia.org/wiki/File%3AScrabble_game_in_progress.jpg
- P.12: Carl Spitzweg, Der Bücherwurm (1850), https://commons.wikimedia.org/wiki/File:Carl_Spitzweg_021.jpg
- P.13: Frontispiece of Vesalius' De humani corporis fabrica (1543), https://commons.wikimedia.org/wiki/File:Vesalius_Fabrica_fronticepiece.jpg