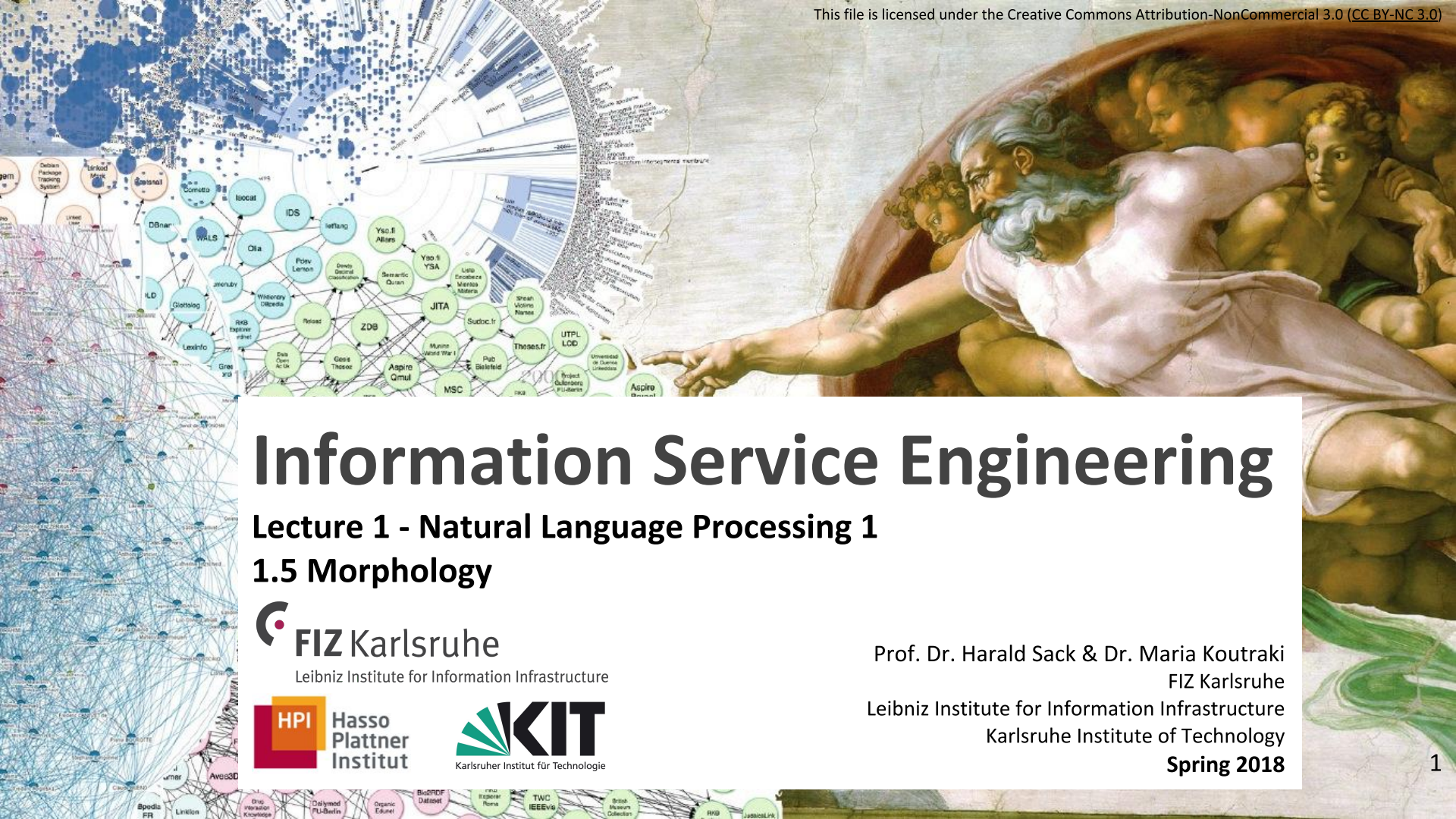




Information Service Engineering

Lecture 1 - Natural Language Processing 1

1.5 Morphology



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

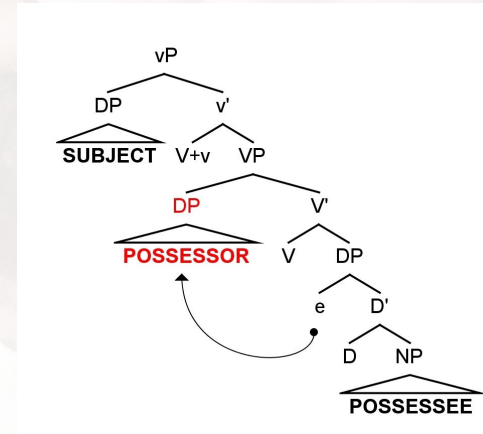
Karlsruhe Institute of Technology

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



Morphology

Information Service Engineering - Lecture 1 Natural Language Processing - 1.5 Morphology

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

Morphemes

Information Service Engineering - Lecture 1 Natural Language Processing - 1.5 Morphology

- the **sm.**
i.e. the

A bound morpheme that is part of a complex word but does not belong to any lexical category (i.e., is not a verb, a noun, an adjective)

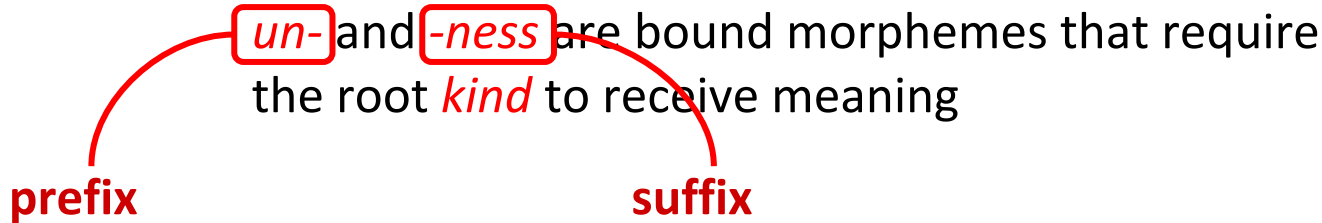
- We distinguish:
 - **Simple words:** consist of a single morpheme
E.g. *work, build, run*, etc.
 - **Complex words:** have internal structure, i.e. exist of 2 or more morphemes
E.g. *worker*, **affix** *-er* added to **root** *work*

core part of a complex word, the part that carries the major component of its meaning

Free vs. Bound Morphemes

Information Service Engineering - Lecture 1 Natural Language Processing - 1.5 Morphology

- **Free morpheme:** a simple word, consisting of one morpheme,
 - e.g. *house, work, high, chair, wrap.*
- **Bound morpheme:** morphemes that must be attached to another morpheme to receive meaning,
 - e.g. *unkindness*

The diagram illustrates the morphological structure of the word 'unkindness'. The root word 'kind' is written in red. A red curved line connects the prefix 'un-' (enclosed in a red box) to the root, with the label 'prefix' in red below it. Another red curved line connects the suffix '-ness' (enclosed in a red box) to the root, with the label 'suffix' in red below it. The text 'are bound morphemes that require the root *kind* to receive meaning' is positioned between the two boxes.

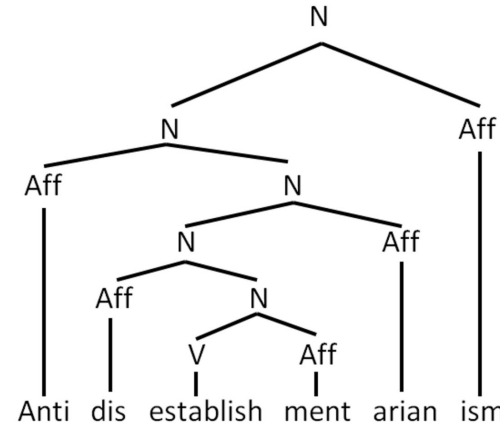
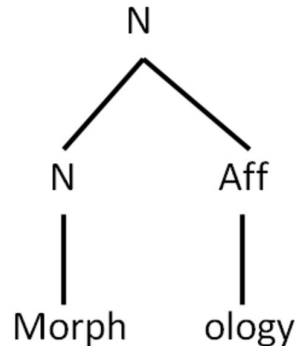
un- and **-ness** are bound morphemes that require the root *kind* to receive meaning

prefix **suffix**

Morphological Parsing

Information Service Engineering - Lecture 1 Natural Language Processing - 1.5 Morphology

- the process of determining the morphemes (and their purpose) from which a given word is constructed
- can be visualized in a tree diagram (**morphology tree**)



Morphological Rules

Information Service Engineering - Lecture 1 Natural Language Processing - 1.5 Morphology

- Languages build more complex words out of morphemes via
 - Derivation
 - Compound
 - Inflection

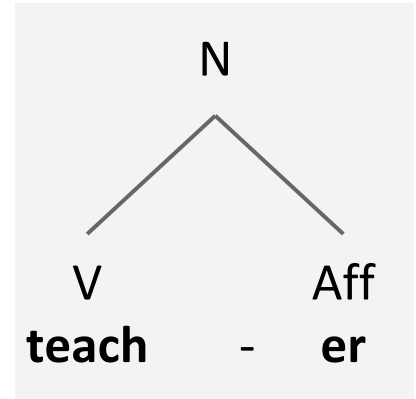
Derivation

Information Service Engineering - Lecture 1 Natural Language Processing - 1.5 Morphology

- the process of forming a new word from an existing word by **adding affixes**
- the meaning of the resulting word is different from that of its base
- very often there is a change in **word category** involved

- **Example**

○	<i>teach</i>	<i>-er</i>	<i>teacher</i>
	root	affix	resulting word
	(verb)		(noun)



Compound

Information Service Engineering - Lecture 1 Natural Language Processing - 1.5 Morphology

- Combination of already existing words into a new ones
- There is no affixation but each of the parts can be assigned to a certain word category

- Examples:

- N + N → N : *lawn mower*
- P + N → N : *up shot*
- N + V → V : *blow dry*
- P + Adj → Adj : *over grown*

Head

Inflection

Information Service Engineering - Lecture 1 Natural Language Processing - 1.5 Morphology

- modification of a word to **express different grammatical categories** such as **tense, case, voice, aspect, person, number, gender, and mood**.
- In English inflection is predominantly expressed by affixation

English has only eight inflectional affixes:

- noun plural {-s} *He has three dessert**s**.*
- noun possessive {-s} *This is Betty'**s** dessert.*
- verb present tense {-s} *Bill usually eat**s** dessert.*
- verb past tense {-ed} *He bak**ed** the dessert yesterday.*
- verb past participle {-en} *He has always eat**en** dessert.*
- verb present participle {-ing} *He is eat**ing** the dessert now.*
- adjective comparative {-er} *His dessert is larg**er** than mine.*
- adjective superlative {-est} *Her dessert is the larg**est**.*

Inflection vs. Derivation

Information Service Engineering - Lecture 1 Natural Language Processing - 1.5 Morphology

- **Derivation** often changes the category of the base, **inflection** never does that.
- **Derivation** changes the meaning of the base, **inflection** does not
- **Derivation** applies before **inflection**

Inflection vs. Derivation

Information Service Engineering - Lecture 1 Natural Language Processing - 1.5 Morphology

1. The farmer's cows escaped.
2. It was raining.
3. Those socks are inexpensive
4. Jim needs the newer copy.
5. The strongest rower continued.
6. The pitbull has bitten the cyclist.
7. She quickly closed the book.
8. The alphabetization went well.

Inflection vs. Derivation

Information Service Engineering - Lecture 1 Natural Language Processing - 1.5 Morphology

1. The farmer's cows escaped.
2. It was raining.
3. Those socks are inexpensive
4. Jim needs the newer copy.
5. The strongest rower continued.
6. The pitbull has bitten the cyclist.
7. She quickly closed the book.
8. The alphabetization went well.

Inflection vs. Derivation

Information Service Engineering - Lecture 1 Natural Language Processing - 1.5 Morphology

1. The farmer's cows escaped.
2. It was raining.
3. Those socks are inexpensive
4. Jim needs the newer copy.
5. The strongest rower continued.
6. The pitbull has bitten the cyclist.
7. She quickly closed the book.
8. The alphabetization went well.

Inflection vs. Derivation

Information Service Engineering - Lecture 1 Natural Language Processing - 1.5 Morphology

1. The farmer's cows escaped.
2. It was raining.
3. Those socks are inexpensive
4. Jim needs the newer copy.
5. The strongest rower continued.
6. The pitbull has bitten the cyclist.
7. She quickly closed the book.
8. The alphabetization went well.

Inflection vs. Derivation

Information Service Engineering - Lecture 1 Natural Language Processing - 1.5 Morphology

1. The farmer's cows escaped.
2. It was raining.
3. Those socks are inexpensive
4. Jim needs the newer copy.
5. The strongest rower continued.
6. The pitbull has bitten the cyclist.
7. She quickly closed the book.
8. The alphabetization went well.

Inflection vs. Derivation

Information Service Engineering - Lecture 1 Natural Language Processing - 1.5 Morphology

1. The farmer's cows escaped.
2. It was raining.
3. Those socks are inexpensive
4. Jim needs the newer copy.
5. The strongest rower continued.
6. The pitbull has bitten the cyclist.
7. She quickly closed the book.
8. The alphabetization went well.

Inflection vs. Derivation

Information Service Engineering - Lecture 1 Natural Language Processing - 1.5 Morphology

1. The farmer's cows escaped.
2. It was raining.
3. Those socks are inexpensive
4. Jim needs the newer copy.
5. The strongest rower continued.
6. The pitbull has bitten the cyclist.
7. She quickly closed the book.
8. The alphabetization went well.

Inflection vs. Derivation

Information Service Engineering - Lecture 1 Natural Language Processing - 1.5 Morphology

1. The farmer's cows escaped.
2. It was raining.
3. Those socks are inexpensive
4. Jim needs the newer copy.
5. The strongest rower continued.
6. The pitbull has bitten the cyclist.
7. She quickly closed the book.
8. The alphabet-ization went well.

Stemming vs. Lemmatization

Information Service Engineering - Lecture 1 Natural Language Processing - 1.5 Morphology

- **Stemming**

the process of reducing inflected or sometimes derived words to their **word stem**

- *Example:* **cats** → **cat**

morphological parse of cats: **cat** + **N** + **PL**

- **Lemmatization**

the process of grouping together the inflected forms of a word so they can be analysed as a single item, identified by the word's **lemma**, or **dictionary form**

- *Example:* **better** → **good**



Evaluation, Precision, and Recall

Next Lecture...

Picture References:

- P.23: Rembrandt, Belshazzar's Feast (1635-38), <https://en.wikipedia.org/wiki/File:Rembrandt-Belsazar.jpg>