

Glossary(I)

- **Morphology** is the study of the way words are built from smaller units or **morphemes**.
- **Stems** morphemes containing main meaning
- **Affixes** morphemes containing additional meaning
- **Prefixes affixes** preceding the stem: ***un-certain***
- **Suffixes affixes** following the stem: ***eat-s***
- **Circumfixes**: prefixes and suffixes: ***sagen – ge-sag-t***
- **Infixes**: Inserted in the middle of the word

Glossary(II)

- **Inflection**: stem + morpheme. Same class.
- **Derivation**: stem + morpheme. Different class:
Computerize-computerization
- **Compounding**. Combination of stems: *doghouse*
- **Cliticization**: stem+ clitic: I've

Glossary(III)

Morphological parsing: finding the constituent morphemes

- **Morphotactics** is the model of morpheme ordering
- **Orthographic rules** model the changes when combining morphemes
- **Finite state transducers** extension of finite-state automata that can generate output symbols
 - Lexicon and spelling rules can be represented by composing and intersecting transducers

Glossary(IV)

Superficial level: preparing the text for further processing

- **Lexical level:** processing the words
- **Lexicon:** dictionary
- **Tokenization:** segmentation
 - Word tokenization. Separating a sentence in words
 - Sentence tokenization. Separating a text in sentences.