Introduction

Morphology is the study of the way words are built from smaller units: morphemes **un-believe-able-ly**

Two broad classes of morphemes: **stems** (main, meaning) and **affixes** (additional).

Affixes
Prefixes: precide the stem: **un-certain, un-chain**
Sufixes: eat-s

Introduction

Affixes
Prefixes: precide the stem: **un-certain, un-chain**

Sufixes: eat-s

Circumfixes: prefixes and sufixes: **sagen – ge-sag-t**

Infixes: Inserted in the middle of the word: tagalog language, not in formal English (but in dialects: bl**dy,f**king, **abso-bl**dy-lutely**).

Introduction

Agglutinative languages tend to string affixes together
- Turkish, ten or more affixes
- English no more than five

Different ways to combine morphemes:
Inflection: stem + grammatical morpheme
syntactic function: plural and gender in nouns
tense on verbs
Derivation: stem + grammatical morpheme
different class, different meaning
Computerize-computerization

Introduction

Different ways to combine morphemes:

Inflection: stem + grammatical morpheme (syntactic function: plural, gender, tense)

Derivation: stem + grammatical morpheme (different class, different meaning).

Computerize-computerization

Compounding.Combination of multiple stems: d**oghouse**

Cliticization: stem+ clitic (reduced in form): **I've**

Inflection in English is simple (-s,-ed,-ing)

Derivation is more complex (suffixes –ation,-ness,-able, prefixes co-,re-)

Introduction

Morphological parsing is the process of finding the constituent morphemes in a word

cat +N+ pl for cats

To build a morphological parser we need

A lexicon: the list of stems and affixed and basic information about them.

Morphotactics is the model of morpheme ordering that explains the allowable morpheme sequences.

Orthografics rules: spelling rules to model the changes when combining morphemes: city- cities

Introduction

Many constraints on morphotactics can be represented by finite automata

Finite state transducers are an extension of finite-state automata that can generate output symbols.

Finite state transducers are used for: morphology representation, parsing, spelling error detection:

Lexicon and spelling rules can be represented by composing and intersecting transducers