

Computational Linguistics (INF2820 — Morphology)

{ eat, eats, eating, ate, eaten }

Wilhelm Stephan Oepen & Jan Tore Lønning

Universitetet i Oslo

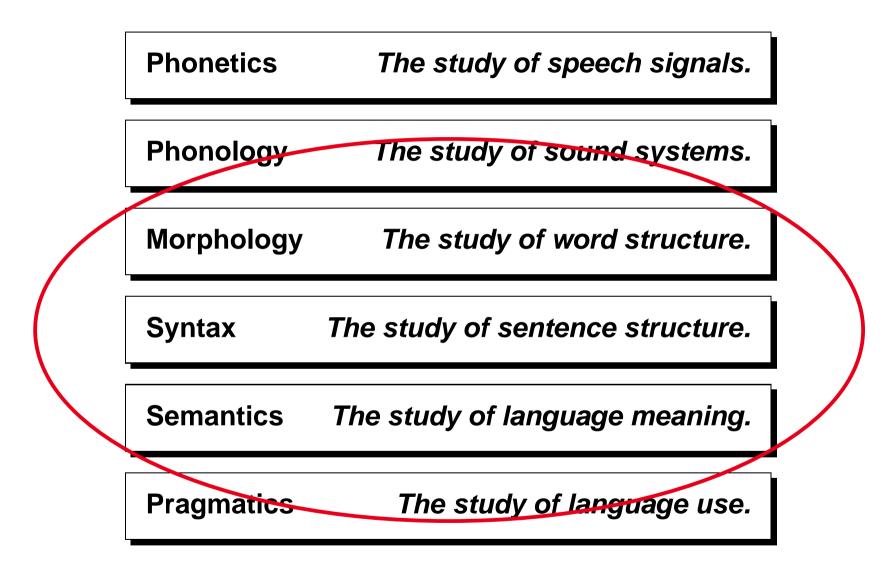
oe@ifi.uio.no

Some Areas of Descriptive Grammar

| Phonetics | The study of speech signals. |
|------------|----------------------------------|
| Phonology | The study of sound systems. |
| Morphology | The study of word structure. |
| Syntax | The study of sentence structure. |
| Semantics | The study of language meaning. |
| Pragmatics | The study of language use. |



Some Areas of Descriptive Grammar





- INF2820 - 5-APR-11 (oe@ifi.uio.no)

Tokenization: Finding the Basic Building Blocks

Mr. Browne, who's prime minister, arrived.



- INF2820 - 5-APR-11 (oe@ifi.uio.no) -

Finite-State Machines — Morphology (3)

Tokenization: Finding the Basic Building Blocks

Mr. Browne, who's prime minister, arrived.

He eats chocolate, candy (i.e. sugar), etc.



— INF2820 — <u>5</u>-APR-11 (oe@ifi.uio.no) -

Finite-State Machines — Morphology (3)

Analyzing Complex Morphology: Esperanto

- (1) La alta knabo malsaniĝis.'The tall boy fell ill.'
- (2) Ĉu li grandigis la grandecon de la dormejo?'Did he increase the size of the dormitory?'
- (3) Ankaŭ malaltaj knabinoj povas esti belaj.'Short girls, too, can be beautiful.'
- (4) Mia patro estas sana ĉar li ne trinkas vinon.'My father is healthy because he doesnt drink wine.'
- (5) La bonaj monaĥinoj volis preĝi en la preĝejo.'The good nuns wanted to pray in church.'

| mal | san | iĝ | is | = became not healthy |
|-----|---------|--------|------|----------------------|
| not | healthy | become | past | |



INF2820 - 5-APR-11 (oe@ifi.uio.no) -

Morphological Categories (1 of 3)

Parts of Speech (PoS)

| noun (N) | cat, dog, neighbours, |
|-----------------|------------------------------|
| verb (V) | barks, chased, was, |
| adjective (Adj) | fierce, angry, black, young, |
| adverb (Adv) | quickly, probably, not, |
| determiner (D) | a, the, my, that, |
| preposition (P) | of, by, on, at, under, |
| pronoun (Pro) | she, mine, those, what, |
| conjunction (C) | and, neither nor, because, |
| | |

How to discover the inventory of categories?



- INF2820 — 5-APR-11 (oe@ifi.uio.no)

A Quick Tour of English Morphology



- INF2820 - 5-APR-11 (oe@ifi.uio.no) -

Finite-State Machines — Morphology (6)

Morphological Categories (2 of 3)

Parts of Speech (PoS)

| noun (N) | cat, dog, neighbours, |
|-----------------|------------------------------|
| verb (V) | barks, chased, was, |
| adjective (Adj) | fierce, angry, black, young, |
| adverb (Adv) | quickly, probably, not, |
| determiner (D) | a, the, my, that, |
| preposition (P) | of, by, on, at, under, |
| pronoun (Pro) | she, mine, those, what, |
| conjunction (C) | and, neither nor, because, |

- **Paradigm** set of word forms, e.g. {*bark*, *barks*, *barking*, *barked* };
- **Unit Categories** dimensions structuring a paradigm *internally*;
- Paradigm Categories properties *common* to all paradigm units.



 ${\tt INF2820}$ — 5-APR-11 (oe@ifi.uio.no) –

Morphological Categories (3 of 3)

Number — Person — Case — Gender

That dog barks. — Those dogs bark. I bark. — You bark. — They bark. — Sam shaved himself. We bark. — You bark. — Those dogs bark. I saw her. — She saw me. — My dog barked.

...

How many distinct verb forms according to number and person?

Tense — Aspect — Mood

The dog barks. — The dog barked — The dog will bark. The dog has barked. — The dog is barking. If I were a carpenter, ...



- INF2820 — 5-APR-11 (oe@ifi.uio.no) -

Finite-State Machines — Morphology (8)