8. september 2022

# IN1140 Gruppe 3

# Denne økta

Korpora

Lingvistiske nivåer

Morfologi
- Oppsummering
- Tidligere eksamensoppgave

Gjennomgang av oblig 1a

Python-repetisjon



# Tekstkorpora

Betydning innenfor språkvitenskap og IT:
"(innsamlet) materiale til (språk)vitenskapelig undersøkelse" [1]

Samling av dokumenter

Elektronisk lagret

Brukes som:

- Empirisk data til lingvistikkstudier
- Treningsdata til språkteknologiske modeller

# Tekstkorpora

Automatisk eller manuelt annotert

## Ordklasser (parts-of-speech tagging)

```
[("disse", "at"), ("unike", "jj"), ("områdene", "nns"), ("kan", "vb"),
 ("bli", "vb"), ("nye", "jj"), ("nasjonalparker", "nns")]
```

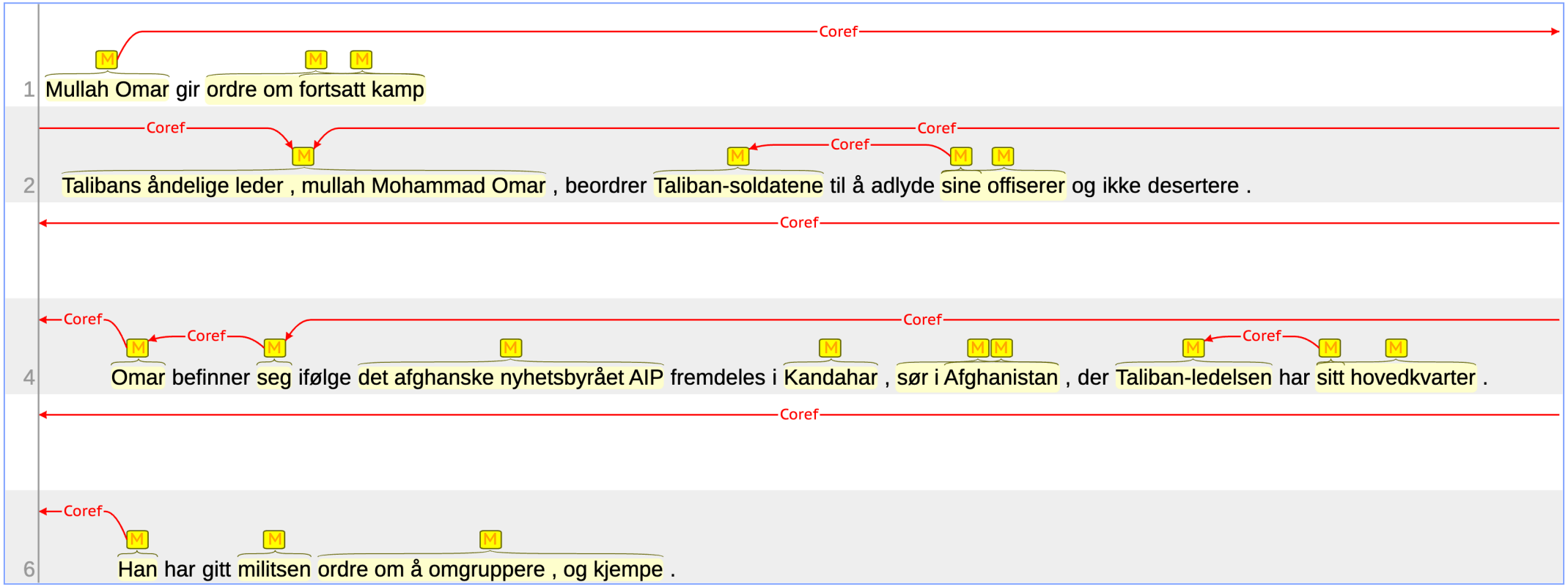
## Sentiment (positive og negative følelser)

```
[("Denne bilen er klønete utformet.", "NEG"), ("Kaken er frisk og god!", "POS"),
 ("Ungene er glade!", "POS"), ("Oliven smaker vondt.", "NEG")]
```

# Annotering

Det å markere relevant informasjon i data

Visuelt eksempel med koreferanse:



# De lingvistiske nivåene

## Fonetikk

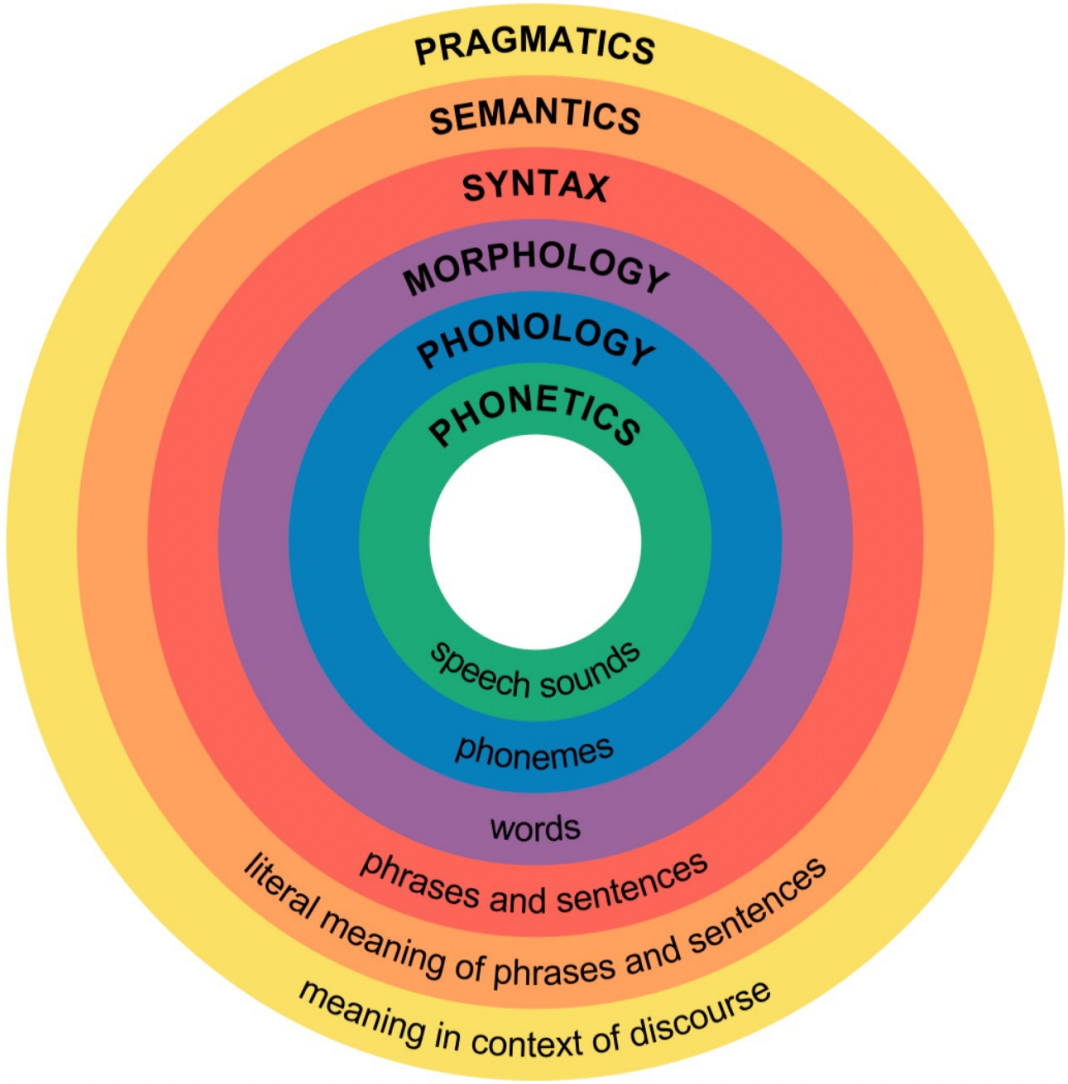
"Læren om språklydenes fysiske og fysiologiske egenskaper." [2]

## Fonologi

"Læren om språklydenes funksjoner, om hvordan de lydene som menneskene kan frembringe gjennom taleorganene, utnyttes i det enkelte språk." [3]

## Morfologi

"Den grenen av språkvitenskapen der man analyserer ordene i mindre deler som har sin egen betydning eller sin egen funksjon." [4]



# De lingvistiske nivåene

## Syntaks

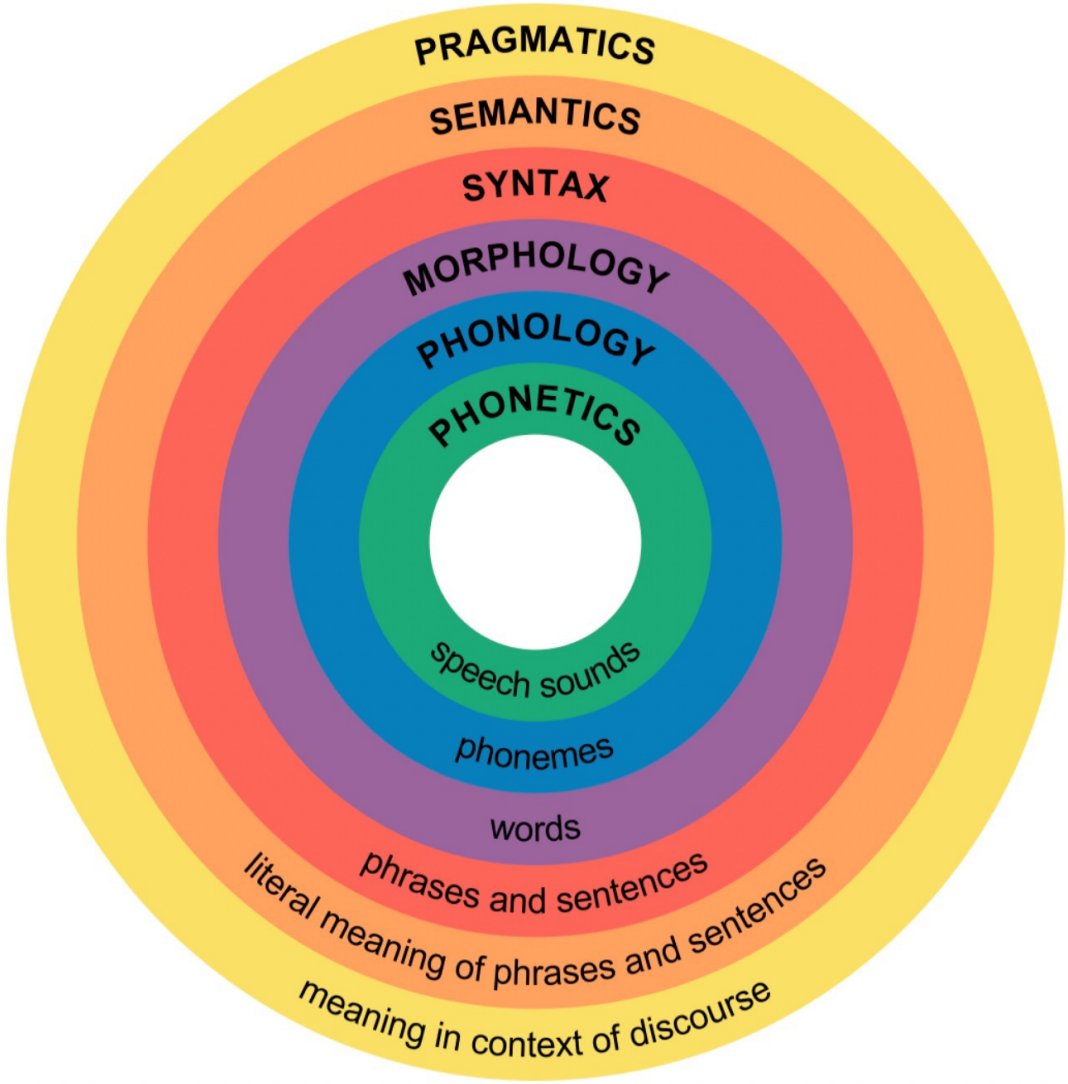
"Reglene som styrer hvordan setninger og fraser kan bygges opp i et språk" [5]

## Semantikk

"Læren om språkets innhold, sammenhengen mellom ord, fraser og setninger og deres betydning eller mening." [6]

## Pragmatikk

"Læren om hvordan kontekst bidrar til meningsinnholdet i en ytring." [7]



# Hva er et ord?

Når vi skal telle antall ord …

**Enkeltord (token)**
Tell alle ord.

**Ordtyper**
Tell alle unike ord.

**Leksem**
Tell alle unike leksikalske oppslag.

Illustrasjon fra IN4080 - Natural Language Processing

| Substantiv | | | |
|---|---|---|---|
| Entall | | Flertall | |
| Ubestemt | Bestemt | Ubestemt | Bestemt |
| gutt | gutten | gutter | guttene |
| jente | jenta | jenter | jentene |
| barn | barnet | barn | barna |

Hver linje er et leksem

Lemma *(grunnformen i ordboka)*

# Hva er et ord?

## Innholdsord

Objekter, handlinger, egenskaper, ideer o.l. med semantisk innhold
Åpen klasse (får stadig nye ord, f.eks. "elsparkesykkel")

## Funksjonsord

Konjunksjoner, preposisjoner, artikler, pronomen o.l. uten semantisk innhold
Lukket klasse (får sjeldent nye ord)

## Usikker på en ordklassen til et ord?

Det norske akademis ordbok (bokmål)
Ordbøkene (bokmål & nynorsk)

# Morfemer

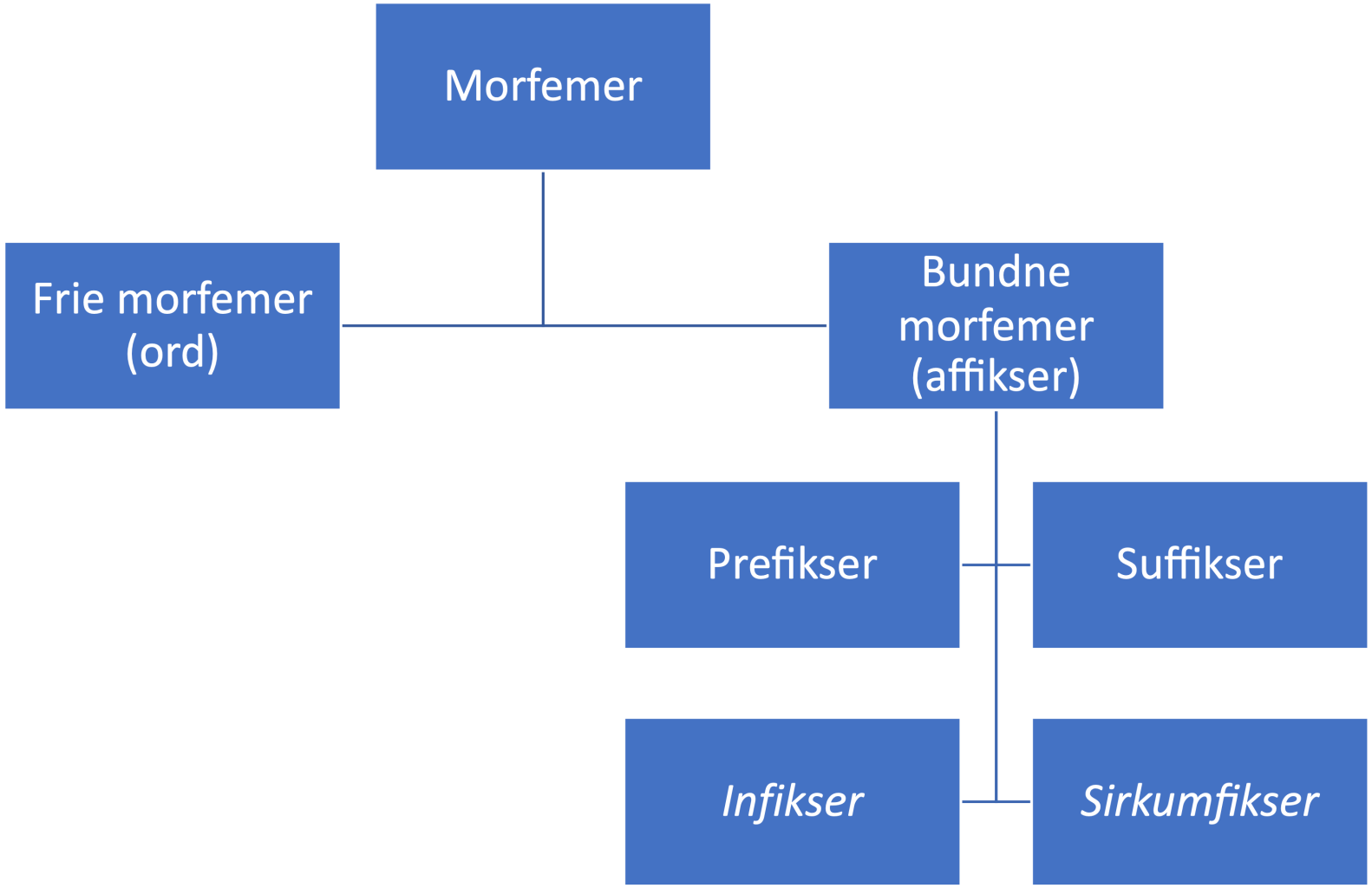
De minste meningsbærende delene av ord

Ett morfem: gutt, bil, hus

To morfemer: bil-er, gå-gate, u-venn

Tre morfemer: an-streng-else, u-venn-er

# Oversikt over ulike morfemtyper



# Eksempler for affikser

På norsk har vi prefikser og suffikser. Disse kalles også *for-* og *etterstavelser*.

### Prefiks
**u**-venn, **opp**-tre

### Suffiks
bil-**er**, be-tyde-**lig**

### Sirkumfiks (tysk)
**ge**-lieb-**t**, **ge**-schrieb-**en**

### Infiks (nicaraguansk spansk)
Vict**ít**or, Osqu**ít**ar

# Orddannelse



# Orddannelse

## Bøying

Bevarer ordklassen

Bevarer stammen

gå → går

grønn → grønnere

bil → biler

## Avleding

Kan endre ordklassen

Endrer stammen

barn → barnslig

oppnå → oppnåelig

mørk → mørkne

## Sammensetting

løkke + skrift → løkkeskrift

sparke + sykkel → sparkesykkel

under + vurdere → undervurdere

# Ordklasser (eksamen 2017)

Her skal vi jobbe med følgende setning:

`Jeg ser mannen med kikkerten.`

Gitt ordklassene i Tabell 1 under, tildel ordklasser til alle ordene i setningen. Du må velge ett alternativ for hvert tilfelle.

| | |
|---|---|
| CC | konjunksjon |
| DET | determinativ |
| JJ | adjektiv |
| NN | substantiv |
| PR | preposisjon |
| PO | pronomen |
| RB | adverb |
| SB | subjunksjon |
| VB | verb |

Table: 1

=>

`Jeg/PO ser/VB mannen/NN med/PR kikkerten/NN`

# Referanser

[1] https://naob.no/ordbok/korpus (7. sep. 2022)
[2] https://snl.no/fonetikk (8. sep. 2021)
[3] https://snl.no/fonologi (8. sep. 2021)
[4] https://snl.no/morfologi_-_grammatikk (8. sep. 2021)
[5] https://snl.no/syntaks (8. sep. 2021)
[6] https://snl.no/semantikk (8. sep. 2021)
[7] https://snl.no/pragmatikk_-_spr%C3%A5kvitenskap (8. sep. 2021)