

# Gruppetime #2

IN1140, gruppe 1

# Tekstkorpora

- Strukturert samling tekster
- 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 [1]

Sentiment

"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 rådata
  - Et enkelt illustrert eksempel med koreferanse

Den siktede sa til sitt forsvar at det ikke var hans skyld .

The diagram illustrates coreference in the sentence "Den siktede sa til sitt forsvar at det ikke var hans skyld .". The words "Den siktede", "sitt forsvar", and "hans skyld" are highlighted in yellow. Blue brackets are placed above "sitt forsvar" and "hans skyld". A red arrow originates from the bracket above "hans skyld" and points to the bracket above "sitt forsvar". Another red arrow originates from the bracket above "Den siktede" and points to the bracket above "sitt forsvar". This indicates that "Den siktede" and "hans skyld" refer to the same entity, and "sitt forsvar" refers to that same entity.

# De lingvistiske nivåene

## Fonetikk

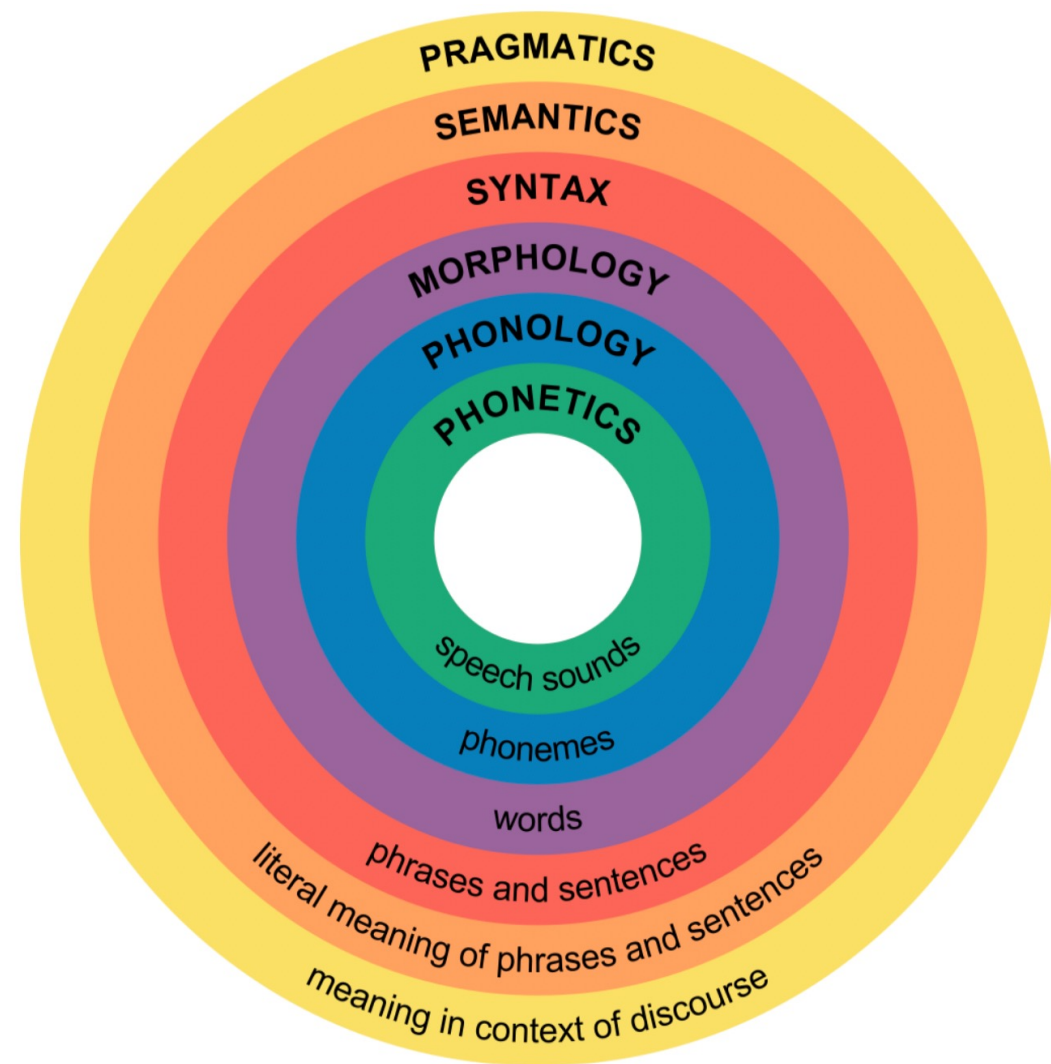
"Læren om språklydenes fysiske og fysiologiske egenskaper." [2]

## 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

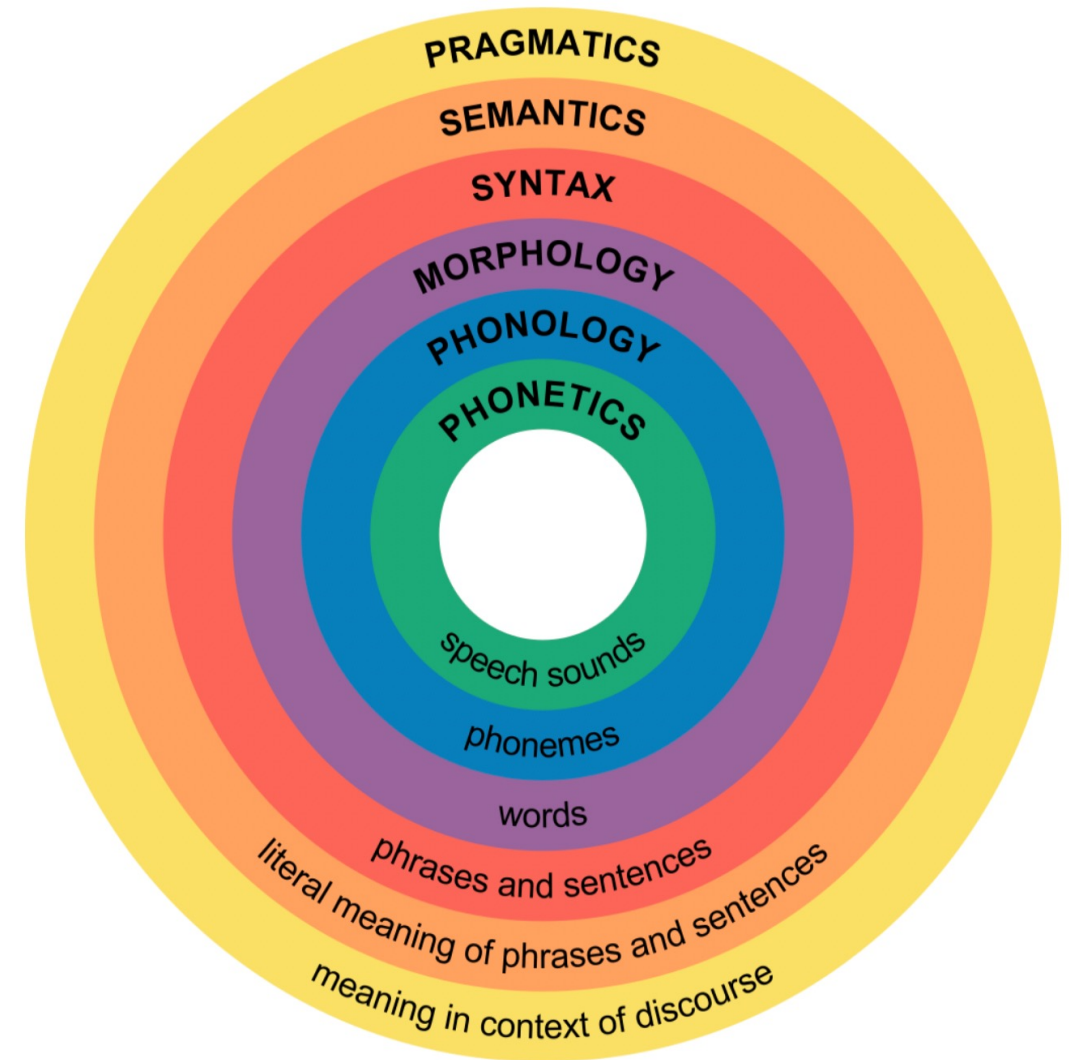
"Reglene som styrer hvordan setninger og fraser kan bygges opp i et språk." [5]

## 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]



# Morfologi - Hva er et ord?

*Når vi skal telle ord...*

Enkeltord, engelsk "token"

Tell alle ordene og tegnsetting.

Ordtyper

Tell alle unike ord og ev. tegnsetting.

Leksem

Tell alle unike leksikalske oppslag og ev. tegnsetting.

# Hva er et ord?

## Innholdsord

Objekter, handlinger, egenskaper, ideer o.l. med semantisk innhold

Åpen klasse: får stadig nye ord, f.eks. "*elsparkesykkel*"

Om "*elsparkesykkel*" for de nysgjerrige: <https://www.sprakradet.no/svardatabase/sporsmal-og-svar/elsparkesykkel/>

## Funksjonsord

Konjunksjoner, preposisjoner, artikler, pronomen o.l. uten betydelig semantisk innhold

Lukket klasse: sjeldent nye ord

*Nyttig verktøy til å finne ordklassen til et ord: NAOB*



# Morfemer

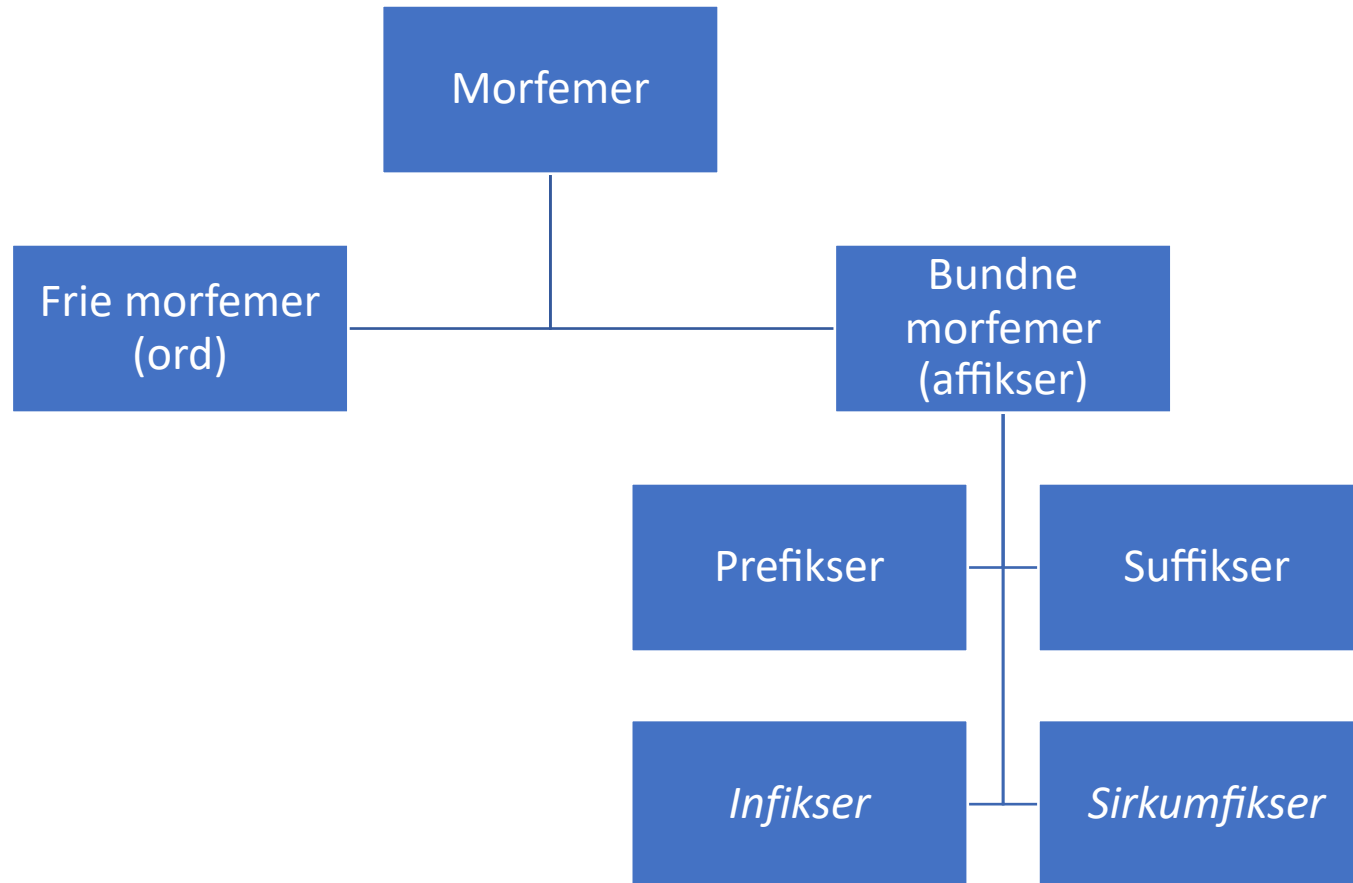
De minste meningsbærende delene av ord.

**Ett morfem:** *gutt, bil, hus*

**To morfemer:** *bil-er, gå-gate, u-venn*

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

# Oversikt over ulike typer morfemer



# Eksempler på affikser

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

En oversikt for de nysgjerrige: <https://snl.no/.taxonomy/3959>

## Prefiks

**u-**venn, **opp-**tre

## Suffiks

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

## Sirkumfiks (tysk)

**ge-**lieb-**t**, **ge-**schrieb-**en**

## Infiks (russisk)

откры-**ва**-ть

# Orddannelse

rot      bøyningssuffiks  
båt-ut-leie-r-e  
stamme

# Orddannelse

## Bøyning

*Bevarer ordklasse*

*Bevarer stammen*

gå → går

grønn → grønnere

jente → jenter

## Avledning

*Kan endre ordklasse*

*Endrer stammen*

barn → barnslig

oppnå → oppnåelig

mørk → mørkne

## Sammensetning

løkke + skrift → løkkeskrift

sparke + sykkel → sparkesykkel

under + vurdere → undervurdere

# Referanser

- [1] <https://www.nrk.no/vestland/regjeringen-apner-for-ti-nye-nasjonalparker-1.15639792> (7. sep. 2021)
- [2] <https://snl.no/fonetikk> (8. sep. 2021)
- [3] <https://snl.no/fonologi> (8. sep. 2021)
- [4] [https://snl.no/morfologi\\_-\\_grammatikk](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](https://snl.no/pragmatikk_-_spr%C3%A5kvitenskap) (8. sep. 2021)