

729G49 Språk och datorer (VT2025)

Introduktion till språkvetenskap

Lingvistik 1

Marcel Bollmann

Institutionen för datavetenskap (IDA)



Baserat på material från
Mikael Svensson & Evelina Rennes

Vem är jag?

- Ursprungligen från Tyskland 🇩🇪
- Bor och jobbar i Sverige sedan 2021
- Studerade **lingvistik** för min Bachelor och Master
- Doktorerade i **datalingvistik**
(*computational linguistics*)



Upplägg av "lingvistikdelen"

- Föreläsningar (6×)
- Seminarier (5×)
 - Övningsuppgifter
 - Två grupper (grupp 1 ~ labbgrupp Jenny, grupp 2 ~ labbgrupp Kättriin)
- Digital salstentamen (ny!)
 - 📅 2025-05-28 (omtenta: 2025-08-13)
 - 🕒 4 timmar

Mål

- Att ha en överblick över språkvetenskapens olika forskningsområden
- Att ha en kunskap om centrala språkvetenskapliga frågor
- Att kunna centrala språkvetenskapliga begrepp och termer
- Att kunna utföra grundläggande språkliga analyser

Lingvistikföreläsningar

F1: Introduktion till språkvetenskap Överblick

F2: Introduktion till morfologi och syntax Ord, ordklasser, lexikon

F3: Morfologi Morfem, ordens struktur

F4: Syntax I Konstituenten, fraser, frasstruktur

F5: Syntax II Satsdelar, huvud- och bisatser

F6: Semantik Lexikal- och satssemantik

Vad är lingvistik?



Varför lingvistik?

Lingvistik (*eller språkvetenskap*)
handlar om människans **förmåga**
att kommunicera.

Hur är språket uppbyggt?

Hur förstår vi språk?

Hur lär sig ett barn språk?

Hur förändras språk över tid?

Hur kan vi lära AI att använda språk?

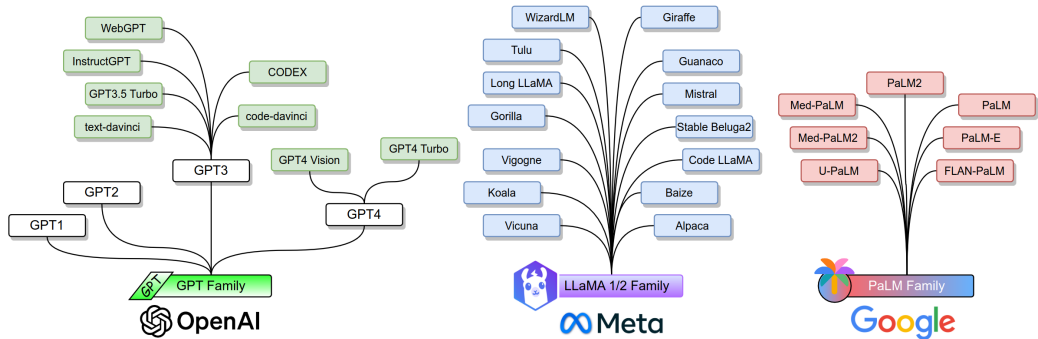
Hur skiljer sig språk åt i olika dialekter?

Hur uppfattas språk av öron och hjärna?

Vad spelar språk för roll för individens identitet?

Stora språkmodeller (large language models, LLMs)

LLMs: ~ statistiska modeller av mänskligt språk, tränad på stoooooongaa mängder data



Källa: Minaee et al. (2024)

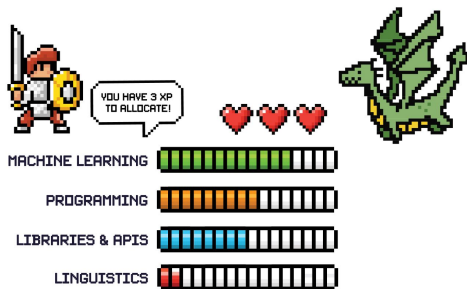
Stora språkmodeller (large language models, LLMs)

- **GPT-SW3** blev känt som “Sveriges svar på ChatGPT”
 - Tränad på BerzeLiUs!
- **TrustLLM** är ett europeiskt forskningsprojekt om “tillitsfulla” språkmodeller för de germanska språken
 - Koordineras av LiU!



Källa & mer information: [GPT-SW3](#) och [TrustLLM](#)

Men varför lingvistik?



“ For applied *natural language processing (NLP)*, a little bit of linguistics knowledge can go a long way and prevent some expensive mistakes. ”

Källa: Ines Montani (2021)

Vad är "naturligt" språk?

inte formella
språk, t.ex.
programmer-
ingsspråk

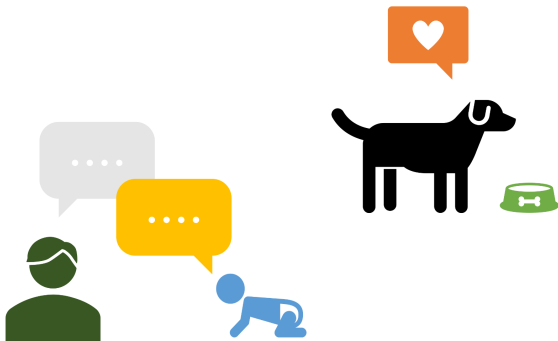


```
attachEvent("onreadystatechange",H),e.attachE
boolean Number String Function Array Date RegE
_={};function F(e){var t=_[e]={};return b.ea
t[1])===!1&&e.stopOnFalse){r=!1;break}n=!1,u&
?o=u.length:r&&(s=t,c(r))}return this},remove
nction(){return u=[],this},disable:function()
re:function(){return p.fireWith(this,argument
ending",r={state:function(){return n},always:
romise)?e.promise().done(n.resolve).fail(n.re
id(function(){n=s},t[1^e][2].disable,t[2][2].
=0,n=h.call(arguments),r=n.length,i=1!==r|e&
(r),l=Array(r);r>t;t++)n[t]&&b.isFunction(n[t
/><table></table><a href='/a'>a</a><input typ
/TagName("input")[0],r.style.cssText="top:1px
test(r.getAttribute("style")),hrefNormalized:
```

Källa: Unsplash

Språk $\neq$ kommunikation

- Alla varelser *kommunicerar* men bara människor har *språk*



Egenskaper hos mänskligt språk

- **Generativt:** begränsat antal ord kan uttrycka ett oändligt antal betydelser
- **Rekursivt:** en typ av struktur kan vara del av samma struktur
 - *Hunden jagade katten som jagade råttan som jagade...*
- **Symboliskt:** kopplingen mellan ett ord och dess betydelse är konventionaliserat
 - 🐕: *dog, hund, chien, perro, ...* inget ord är “mer naturligt” än ett annat
- **Displaced:** vi kan referera till saker utanför den omedelbara kontexten
 - *Var är min mobiltelefon? Den var här igår!*

Språk $\neq$ tänkande

- Små barn tänker innan de har språk.
- Varelser utan mänskligt språk tänker.
- Det finns många typer av kognitiv aktivitet som inte innefattar språk.
- Vi lagrar betydelsen av språklig kommunikation i mer abstrakt form.
- Skador i vissa regioner i hjärnan resulterar i afatiska problem utan konsekvenser för övrig kognition.

Skrivspråk och talspråk

Språk kan uttryckas på olika sätt: **tal, tecken, skrift.**

Talspråk ses oftast som **primärt** eftersom:

- Alla kulturer har talspråk men inte alla har skriftspråk.
- Barn tillägnar sig talspråk intuitivt men måste lära sig att skriva genom instruktioner.
- Vi antar att talat språk har existerat långt före (50.000–100.000 år sedan) skrivet språk (5.000 år sedan).

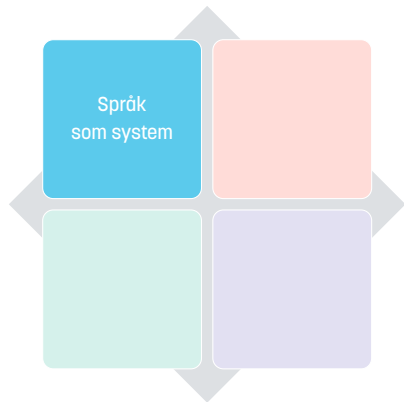


Språk som system



Språk som system

- **Fonologi:** läran om språkljud och hur de är organiserade (Yule, § 3–4)
- **Morfologi:** läran om hur ord byggs upp av de “minsta betydelsebärande enheter” (Yule, § 5–6)
- **Syntax:** läran om hur ord kombineras för att forma fraser, satser och meningar (Yule, § 7–8)
- **Semantik:** läran om språkliga uttrycks betydelse (Yule, § 9)



/dehɛ:rka:f:əteʊdrɪkba:t/

det här kaffet är odrickbart

Fonetik och fonologi

Fonetik handlar om vad som utgör möjliga språkljud.

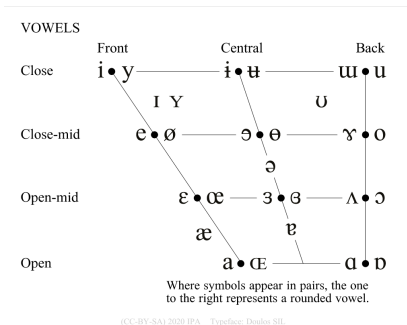
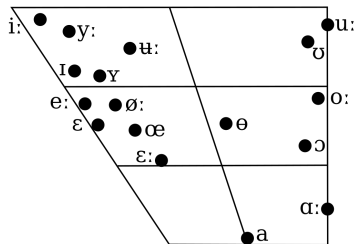


Diagram över vokalljud

Fonologi handlar om hur språkljuden är organiserade och kombineras.



Vokaler i standardsvenskan

Källor: IPA och Jeffrey Connell & Moxfyre

Fonologi och fonem

dör – död

- Ett **fonem** är en abstrakt enhet av ljud som är **betydelseskiljande**.
- Olika ljud som hör till samma fonem kallas **allofoner**.
 - dör – fonem /ø/ – fon/ljud [œ]
 - död – fonem /ø/ – fon/ljud [ø]
- Två ord som endast skiljer sig på en fon kallas ett **minimalt ordpar**.
 - skena /ʃe:na/ – tjena /tʃe:na/

Morfologi

Morfologi handlar om ordens struktur.

- Ett **morfem** är ett språks minsta meningsbärande enhet.

kaffe - *t*
ROT def.

o - *drick* - *bar* - *t*
neg. ROT adj. kasus

Ordklasser

- En **ordklass** är en grupp av ord med liknande grammatisk funktion.

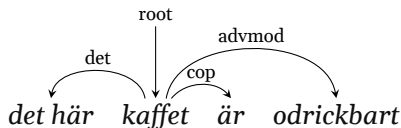
/dehɛ:r/	<i>det här</i>	pronomen
/kaf:ət/	<i>kaffet</i>	substantiv
/e/	<i>är</i>	verb
/ʊdrɪkba:t/	<i>odrickbart</i>	adjektiv

- **Lexikonet** avgör vilka ordklasser ett ord kan tillhöra.
 - Ordklass kan ändras genom morfologiska processer.

Syntax

Syntax handlar om hur ord kombineras till större enheter (och hur dessa enheter relaterar till varandra).

$[_S [_{NF} [_{NF} \text{det här}] \text{kaffet}] [_{VF} \text{är} [_{ADJF} \text{odrickbart}]]]$



Semantik

Semantik handlar om vad enskilda morfem, ord, fraser och meningar betyder.



“milk drinkers turn to powder”

- **Polysemi:** ord och fraser kan ha flera betydelser
 - *turn to* ~ att vända sig till
 - *turn to* ~ att övergå till
- Exempel på andra semantiska relationer...
 - **Synonymi:** köpa – inhandla
 - **Antonymi:** köpa – sälja
 - **Hyponymi:** titta – stirra



Satssemantik

- **Tematiska roller** handlar om ordens semantiska funktioner inom en mening.

Kalle sparkade *bollen*
AGENT TEMA

Bollen sparkades av *Kalle*
TEMA AGENT

- **“Textual entailment”** handlar om den semantiska relationen mellan meningar.
 - **Entailment** (*‘logiskt medförande’*): *Kalle slutade röka* $\longrightarrow$ *Kalle rökte*
 - **Motsägelse**: *Kalle klarade duggan* $\longleftrightarrow$ *Kalle blev underkänd*

Andra språkvetenskapliga områden

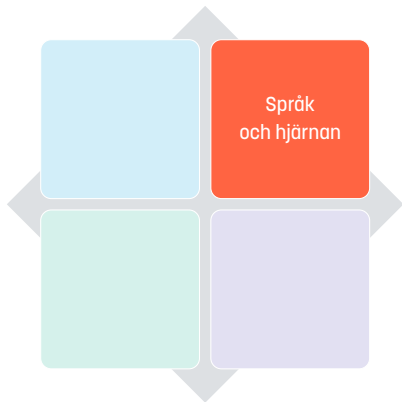


Andra språkvetenskapliga områden



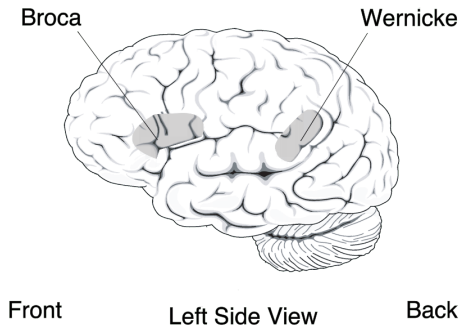
Språk och hjärnan

- **Psykolingvistik:** läran om språkets mentala processer och hur språklig information är representerad (Yule, § 12)
 - Hur tillägnar sig barn språk?
 - Hur använder vi vårt minne när vi talar?
 - Hur lär vi oss ett nytt språk?
- **Neurolingvistik:** läran om hjärnaktivitet under olika språkprocesser (Yule, § 12)
 - Hur uppstår tal- och språkstörningar?
 - Vilka hjärndelar är aktiva när vi talar, läser eller lyssnar på språk?



Brocas och Wernickes områden

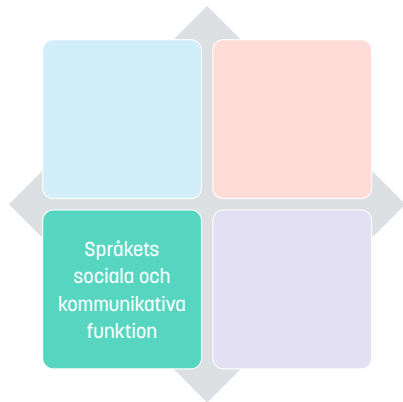
- Skador på vissa områden i hjärnan är kopplade till specifika typer av **afasi**.



Källa: [Wikimedia Commons](#)

Språkets sociala och kommunikativa funktion

- **Pragmatik**: läran om språkanvändning och interaktion mellan språk och kontext (Yule, § 10)
- **Diskursanalys**: läran om strukturer och språkliga fenomen över yttrandegränserna (Yule, § 11)
- **Sociolingvistik**: läran om språkets sociala funktioner (Yule, § 18–20)
 - Både inom och mellan språkgrupper



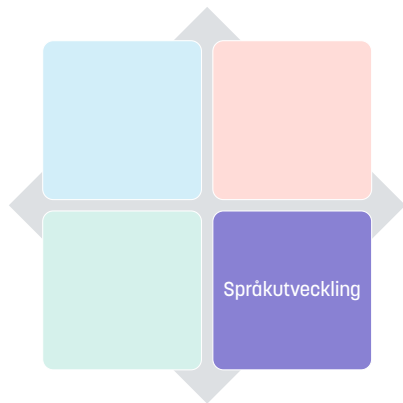
Pragmatik

Pragmatik handlar om betydelsen som talaren vill förmedla i en viss kontext.



Språkutveckling

- **Språkinläring:** läran om språkets utveckling i individen (Yule, § 13–14)
 - Både hur barn tillägnar sig språk och hur vuxna lär sig språk (*andraspråksinläring*)
- **Historisk lingvistik:** läran om språkets utveckling och hur det förändras över tiden mellan generationer av talare (Yule, § 1–2)



Barnspråksutveckling

tidsperiod	utvecklingsstadium
0–6 månader	vokaljoller
6–12 månader	joller
12–18 månader	ettordsfas, 8–10 ord
18–24 månader	tvåordfas, enkla meningar



“ *Samson tog jomfrwne j sin fampn oc bar henne nidh oc satte henne vpa sin hæst oc en annan hæst hade han . ther førde han gwll oc kostelighe haffuor vpa .* ”

“ *Samson tog jungfrun i sin famn och bar ned henne och satte henne på sin häst, och han hade en annan häst. Den lade han upp guld och dyra gåvor uppå.* ”

Källa: [Wikipedia](#)

Metoder inom lingvistik



Preskriptiv vs. deskriptiv lingvistik

- 👍 Inom forskningen följer vi vanligtvis en **deskriptiv** (*beskrivande*) metod.
 - Deskriptiv lingvistik: att analysera och beskriva hur språk faktiskt används
- 👎 Det står i motsats till en **preskriptiv** metod som ibland används inom utbildning.
 - Preskriptiv lingvistik: att fastställa regler för vilket språkbruk som är rätt och “sant”

Hur får man språkliga data?

Vi behöver observera språkliga data för att kunna analysera dem.

- **Inspelningar** eller naturligt förekommande **textdata** (*korpusdata*)
 - “naturligt förekommande” ~ inte skrevs specifikt för att analyseras
- **Tester och utfrågning** av infödda talare
 - “Vilken av dessa meningar tycker *du* låter mest naturlig?”
 - “Uppfattar *du* denna mening som ogrammatisk?”
- **Introspektion**
 - “Uppfattar *jag* det här som naturligt/fel/etc.?”

Välformade uttryck

- En **asterisk (*)** framför ett uttryck anger vanligtvis **ogrammatiska exempel**.
 - *Bilen är röd.*
 - *Bilarna är röda.*
 - **Bilen är röda.*
 - **Bilarna är röd.*
 - *Huset är rött.*
 - *Husen är röda.*
- Om vi har sådana bedömningar kan vi dra **slutsatser**:
 - *röd* används med singular + utrum
 - *rött* används med singular + neutrum
 - *röda* används med plural och är genusneutralt

Språkteknologi

Språkteknologi handlar om analys av språk med hjälp av datorer.

- Kan behandla *stora mängder data* inom kort tid
- Kan *automatisera analysen* av språkliga strukturer

Mer om detta i Jennys föreläsningar och kursen **729G86 Language Technology!**

Nästa gång

F2: Introduktion till morfologi och syntax Ord, ordklasser, lexikon

Att läsa

- Yule, kapitel 5 (*Word formation*)
- Josefsson, kapitel 4 (*Ordklasser*)

