

Lingvistik I

Introduktion till språkvetenskap: områden och forskningsfrågor

(baserat på material från Mikael Svensson)

Upplägg

- Föreläsningar (6 st)
- Seminarier (5 st)
 - Övningsuppgifter
 - Halvklass (grupp 1 = ”Daniel”, grupp 2 = ”Julia”)
- Dugga
 - 2023-05-17
 - 4 timmars skrivtid
 - Omdugga 2023-08-15

Mål

Att ha en överblick över språkvetenskapens olika forskningsområden

Att ha en kunskap om centrala språkvetenskapliga frågor

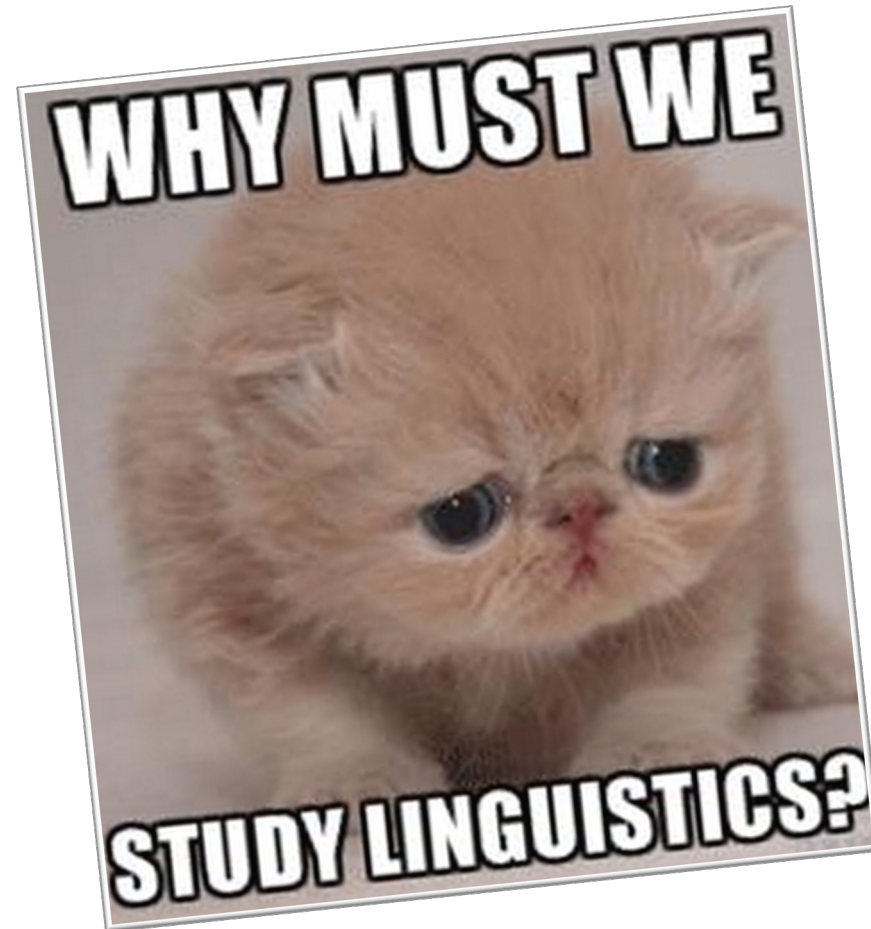
Att kunna centrala språkvetenskapliga begrepp och termer

Att kunna utföra grundläggande språkliga analyser

Att kunna tillgodogöra sig språkvetenskaplig forskning



Övningsuppgifter,
dugga



Varför?

Lingvistik handlar om den kanske mest fantastiska av människans alla förmågor: **förmågan att kommunicera**

Hur är språket uppbyggt?

Vad spelar språk för roll för individens identitet?

Hur förstår vi språk?

Hur lär sig ett barn språk?

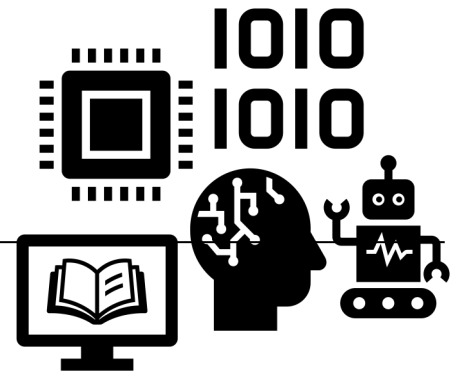
Hur förändras språk över tid?

Hur kan vi lära AI att använda språk?

Hur skiljer sig språk åt i olika dialekter?

Hur uppfattas språkljud av öron och hjärna?

Språkmodeller

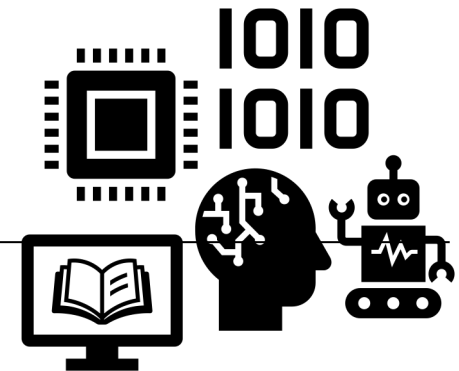


Språkmodeller

Stort framtidsområde inom AI (och framtiden är nu!)

Att träna stora* språkmodeller: att modellera mänskligt språk genom att träna en statistisk modell på stora* mängder data

ChatGPT, GPT-3 (GPT-4!), GPT-SW3 (tränad på BerzeLiUs!), LaMDA, T5, ...

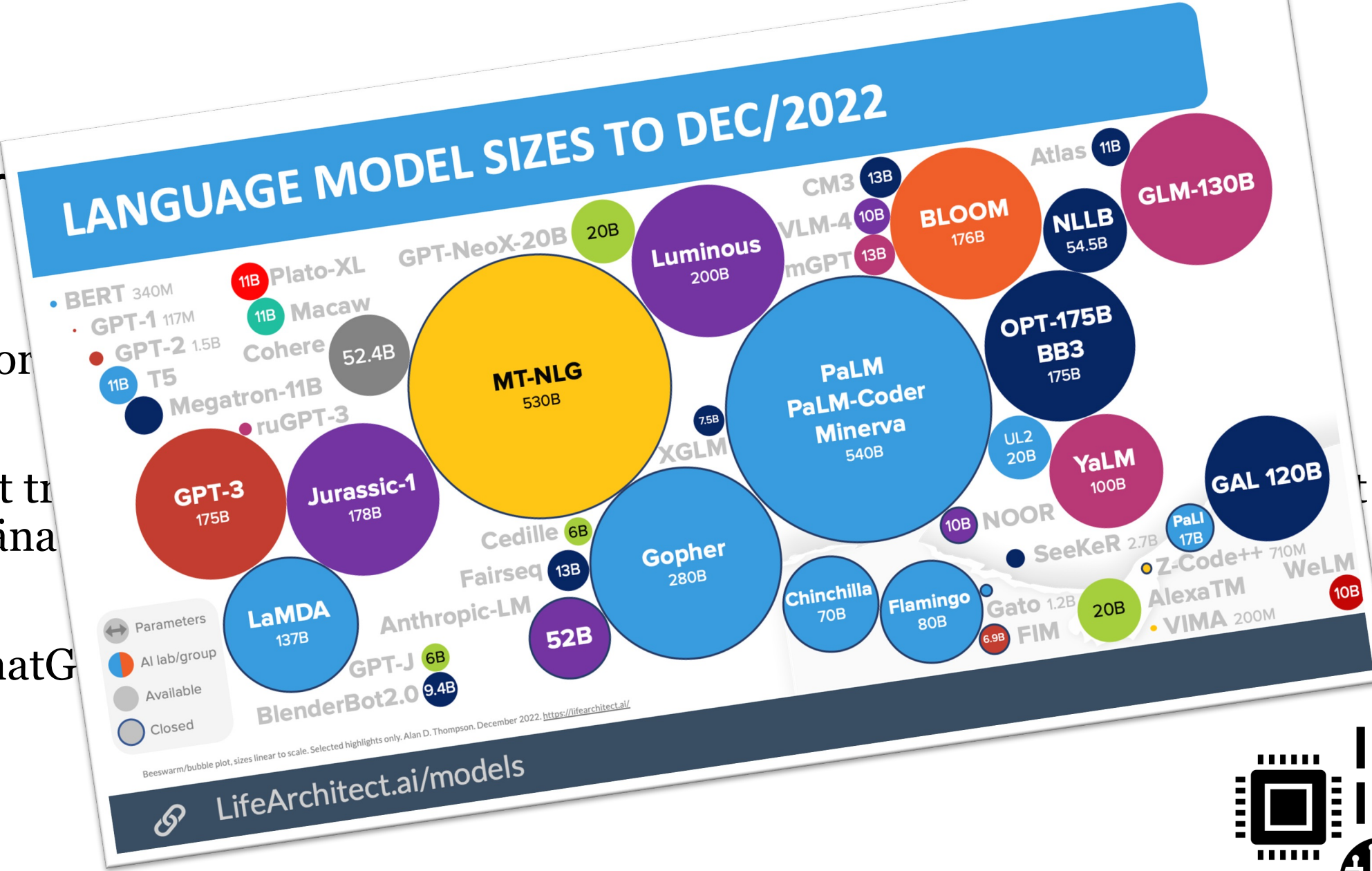


Spr

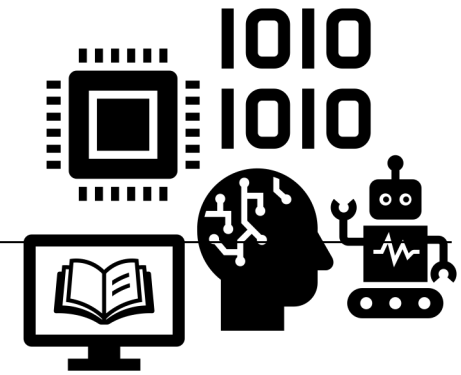
• Stor

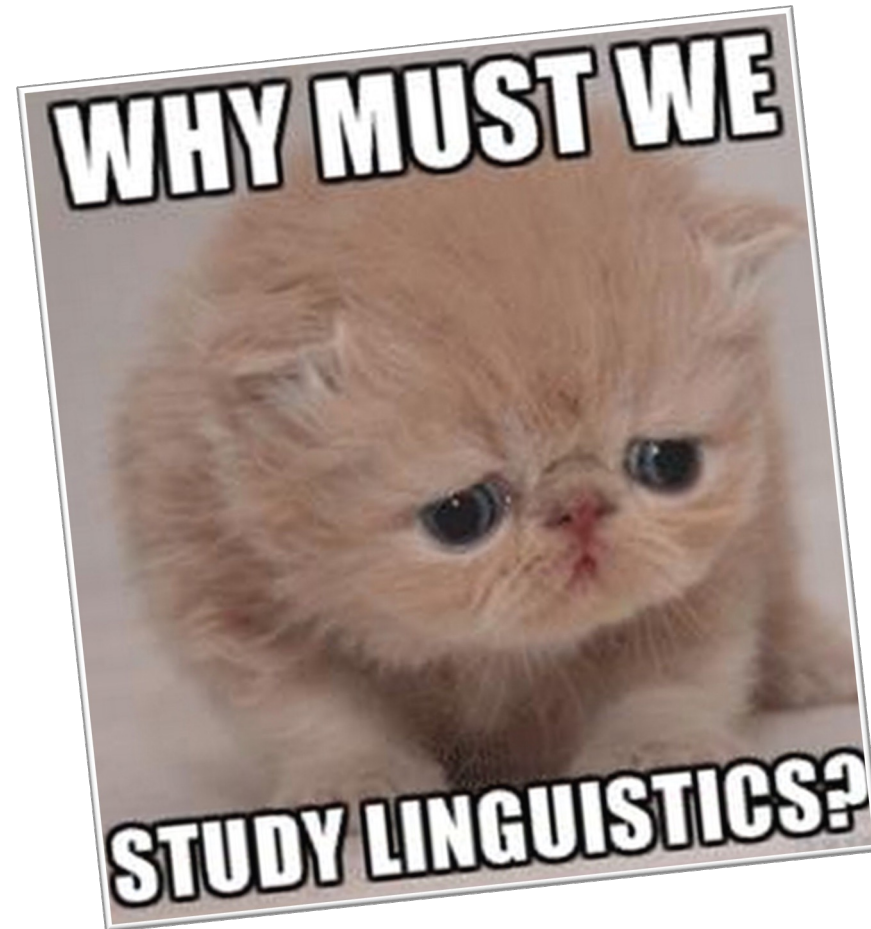
• Att tr
träna

• ChatG



5, ...







Do I need to learn linguistics



Linguistics is not a strict re
be helpful. Linguistics is th
deeper understanding of
NLP. Understanding the
you design and improve
backgrounds in comput



Can you convince me that I need to know linguistics in order to work with natural language processing?



Linguistics can be a valuable asset when working with natural language processing (NLP) for several reasons: 👍 👎

1. Understanding the structure and grammatical rules of different languages can help you design and improve NLP models. This can be especially important when working with less-resourced languages.
2. Linguistics can help you identify and understand language phenomena that are not immediately obvious, such as idiomatic expressions, figurative language, and cultural references.
3. It provides a deeper understanding of the meaning of words and sentences, which is crucial when working on tasks such as sentiment analysis, text summarization, and machine translation.
4. Linguistics can also give you a better understanding of how humans use language in different contexts and how to create natural-sounding text generation models.

That being said, you don't have to have a background in linguistics to work in NLP. There are many people who work in NLP who have backgrounds in computer science, mathematics, or other related fields. However, having knowledge in linguistics can give you a deeper understanding of the field of NLP and can help you make better models.

Lingvistik, föreläsningar

F1: Introduktion till språkvetenskap, översikt

F2: Inledning till syntax och morfologi, ord, ordklasser, lexikon

F3: Morfem och morfologi (ordens struktur)

F4: Syntax: Konstituenten. Fraser och satser (strukturer som bildas av ord)

F5: Satsdelar, huvud- och bisats

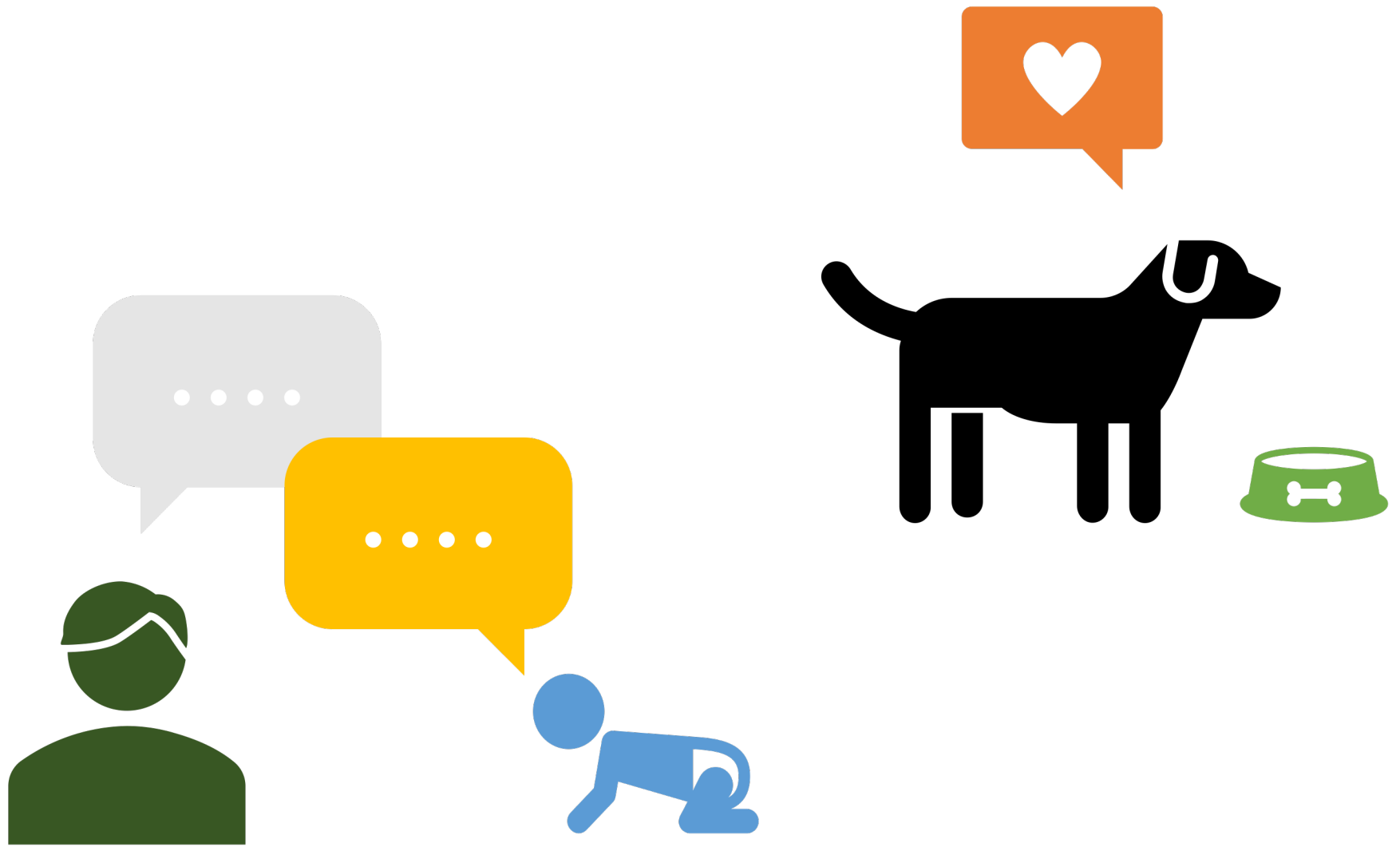
F6: Introduktion till semantik, lexikal och satssemantik

Dagens föreläsning

- Vad är mänskligt språk?
- Språkvetenskapliga områden
 - Språk som system
 - Språk och hjärnan
 - Språkets sociala och kommunikativa funktion
 - Språkutveckling
- Metoder inom lingvistik



Vad är mänskligt språk?



Språk \neq kommunikation

Alla varelser kommunicerar men bara människor har ett språk som är generativt, rekursivt, symboliskt och *displaced*

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 d.v.s. ordet *hund* är inte mer naturligt för djuret än *chien* eller *perro*

Displaced: Vi kan referera till saker utanför den omedelbara kontexten

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 d.v.s. ordet *hund* är inte mer naturligt för djuret än *chien* eller *perro*

Displaced: Vi kan referera till saker utanför den omedelbara kontexten

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 d.v.s. ordet *hund* är inte mer naturligt för djuret än *chien* eller *perro*

Displaced: Vi kan referera till saker utanför den omedelbara kontexten

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 d.v.s. ordet *hund* är inte mer naturligt för djuret än *chien* eller *perro*

Displaced: Vi kan referera till saker utanför den omedelbara kontexten

Språk \neq tänkande

Små barn tänker innan de har ett språk

Varelser utan mänskligt språk tänker

Det finns många typer av kognitiv aktivitet som inte innefattar språk

Språk \neq tänkande

Vi använder icke-lingvistisk information för att förstå språklig betydelse

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

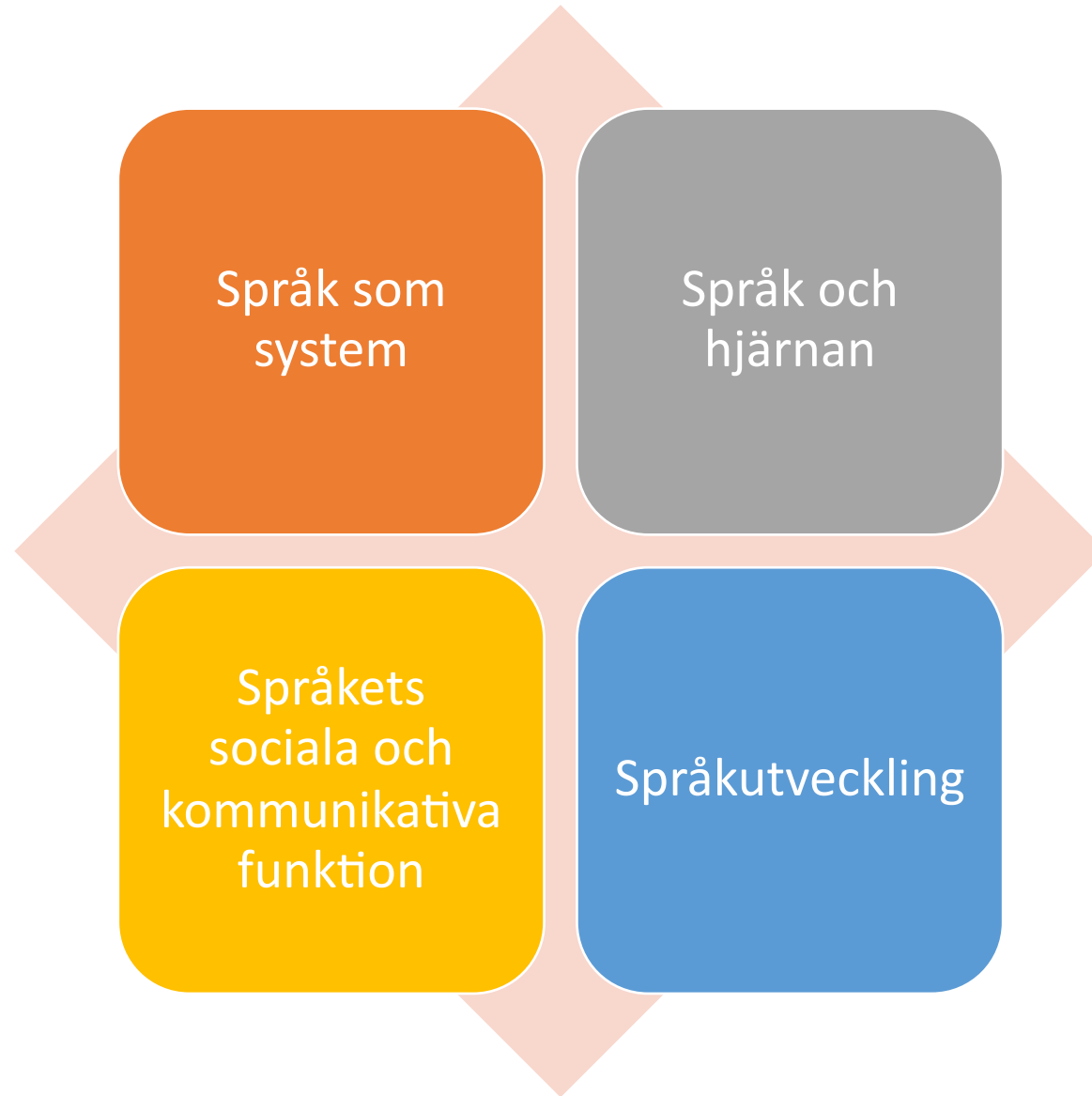
Skrivspråk och språk

Språk är en kognitiv förmåga som kan uttryckas på olika sätt (tal, tecken, skrift).

Tal ses oftast som primärt eftersom

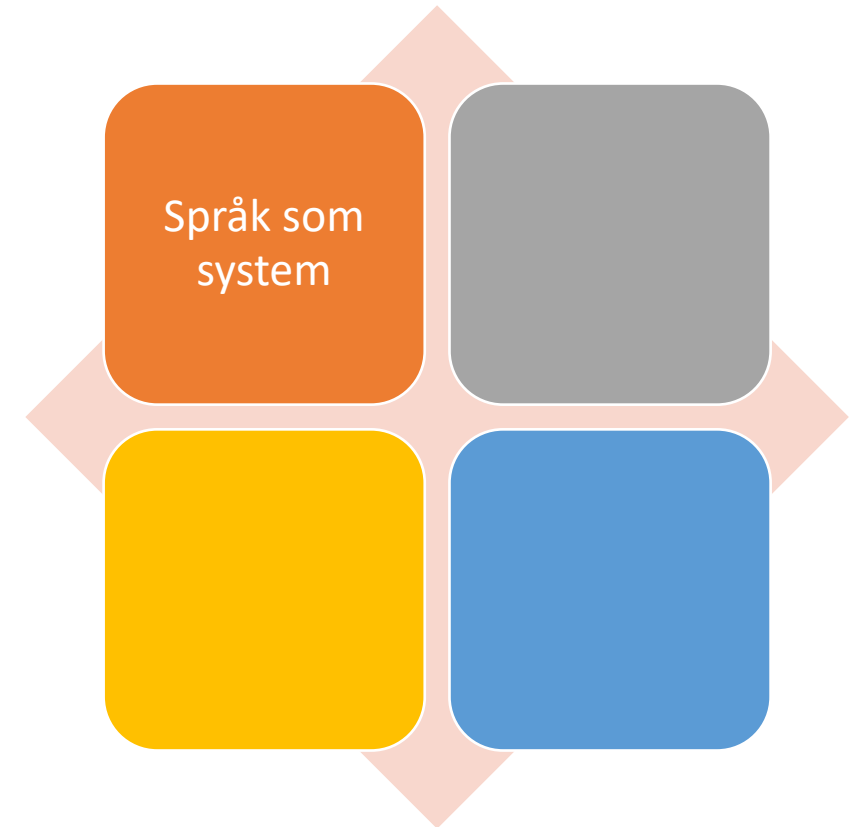
- alla kulturer har talspråk men inte alla har skriftspråk
- barn tillägnar sig talspråk intuitivt men måste lära sig att skriva genom instruktioner
- det finns argument för att anta att talat språk existerat långt före (50.000 till 100.000 år sedan) skrivet språk (5.000 år sedan)

Språkvetenskapliga områden



Språk som system

- **Fonetik/fonologi:** läran om språkljud och hur de är organiserade (Yule, kap. 3–4)
- **Morfologi:** läran om hur ord byggs upp av mindre betydelsebärande enheter som kallas morfem (Yule, kap. 5–6)
- **Syntax:** läran om hur ord kombineras för att forma fraser, satser och meningar (Yule, kap. 7–8)
- **Semantik:** läran om språkliga uttrycks betydelse (Yule, kap. 9)



/dehɛ:^rka:fəteʊdrikbɑ:t/

dethärkaffetärodrickbart
/dehɛ:^rka:fəteʊdrɪkba:t/

det här kaffet är odrickbart
/de hɛ:r kɑ:fət e ʊdriktɑ:t/

Språkvetenskapliga områden

- Vad som utgör möjliga språkljud: **fonetik**
- Hur språkljuden är organiserade och kombineras: **fonologi**
- Hur språkljuden motsvarar meningsbärande enheter (morfem/ord): **morfologi/lexikon**

dethärkaffetärodrickbart
/dehɛ:rka:fəteʊdrɪkba:t/

/de hɛ:r/	det här	pronomen
/kɑ:fət/	kaffet	substantiv
/e/	är	verb
/ʊdrikbɑ:t/	odrickbart	adjektiv

Vilka typer/klasser av ord som enheterna utgör: **morfologi/lexikon**

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

Hur ord kombineras till större enheter och hur dessa enheter relaterar till varandra: **syntax**

'det här kaffet är odrickbart' = [[Det som refereras till som 'kaffet' och som kan identifieras i talkontexten som framträdande pekas ut med 'det här' har egenskapen 'är' av att vara 'odrickbart'.]]

Vad enskilda morfem/ord betyder och vilken betydelse som skapas genom att de relateras till varandra: **semantik**

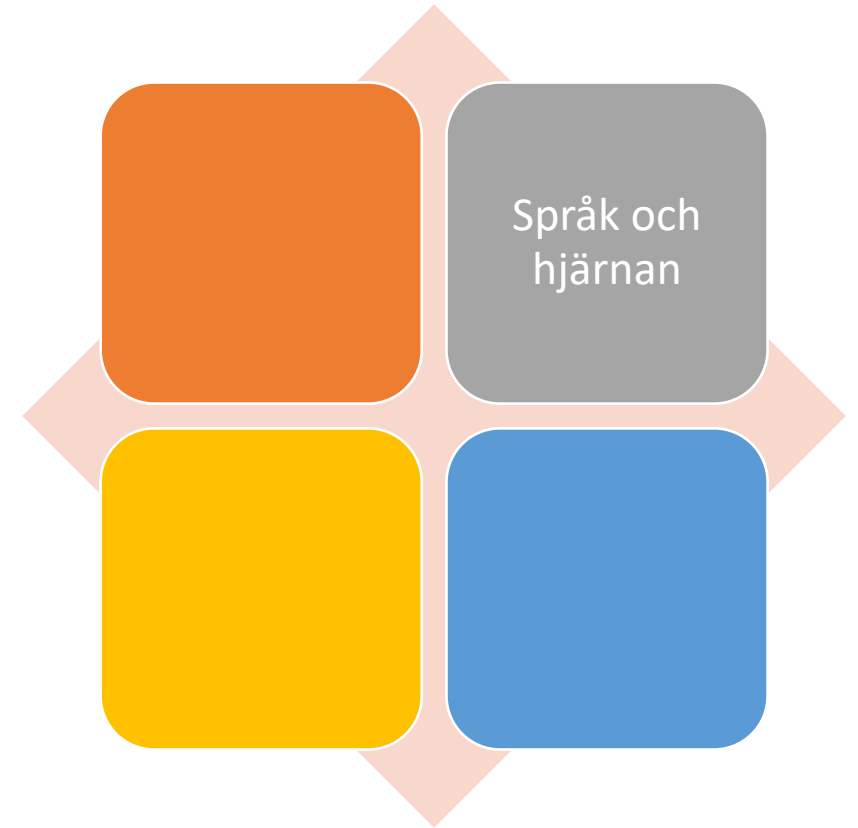
”Det här kaffet är odrickbart!”

- Klaga på kvaliteten på kaffet
- Avråda någon från att dricka kaffet
- Begära en ny kopp kaffe

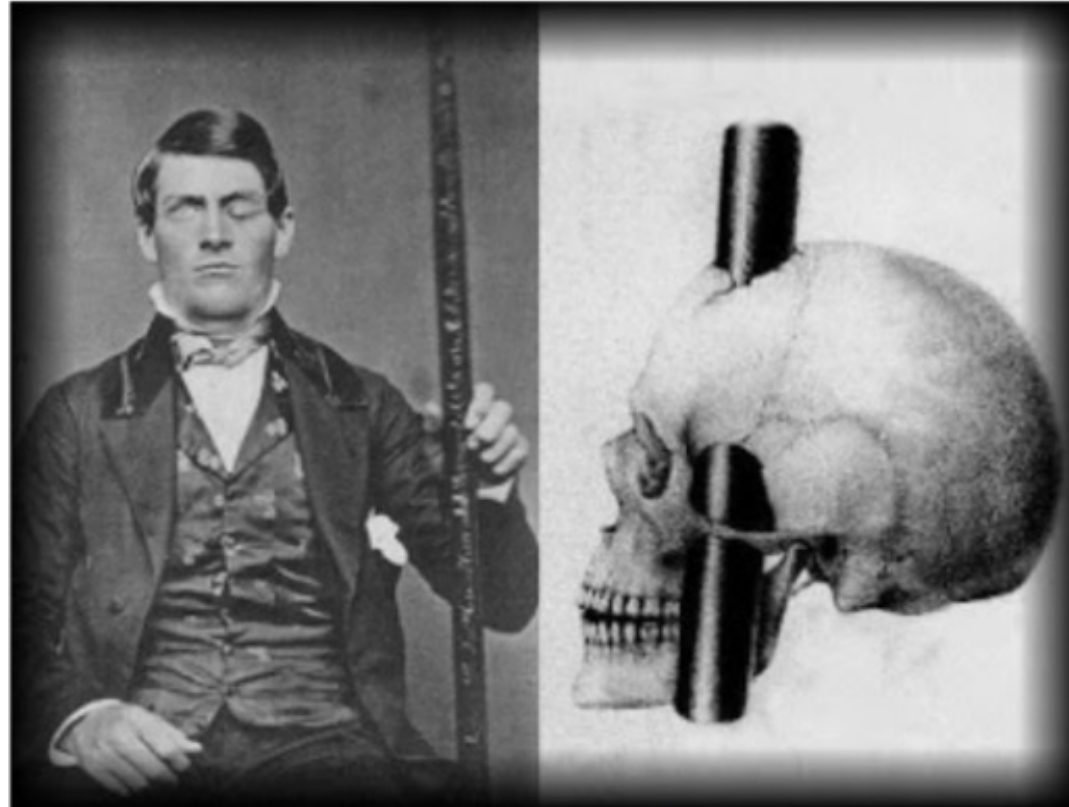
Hur vi tolkar vad talare/skrivare vill kommunicera genom att relatera yttranden till kontext: **pragmatik**

Språk och hjärnan

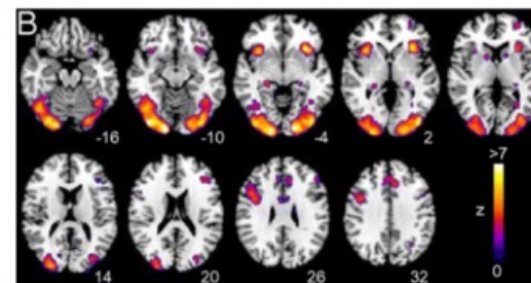
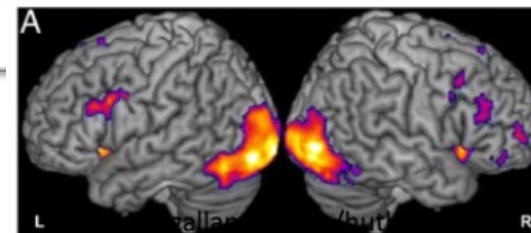
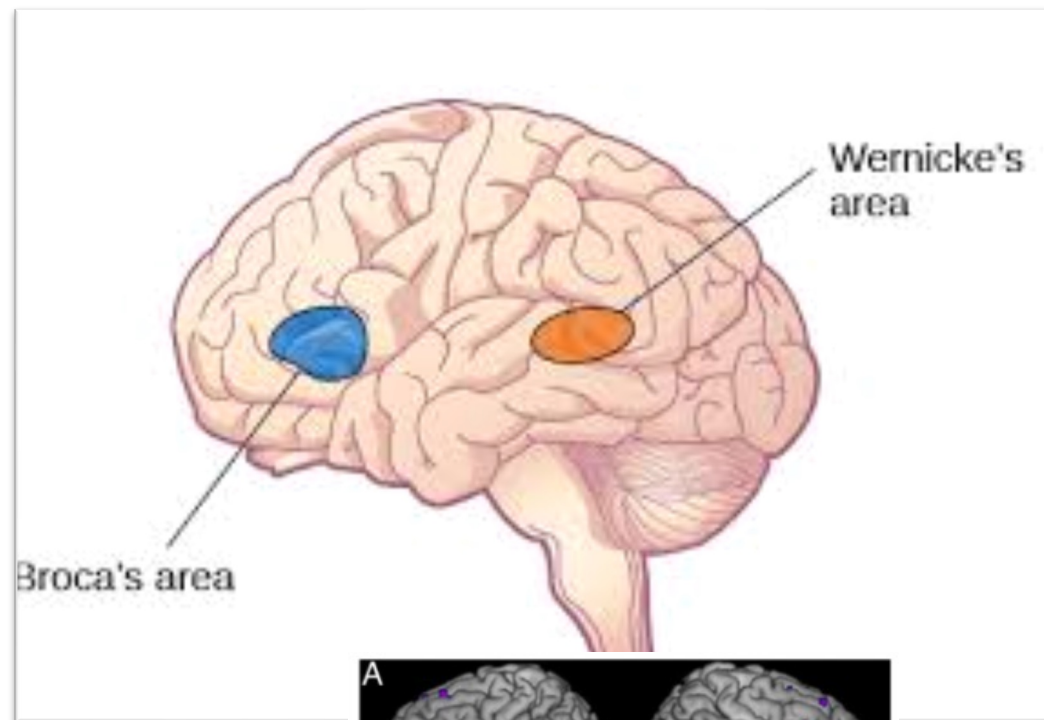
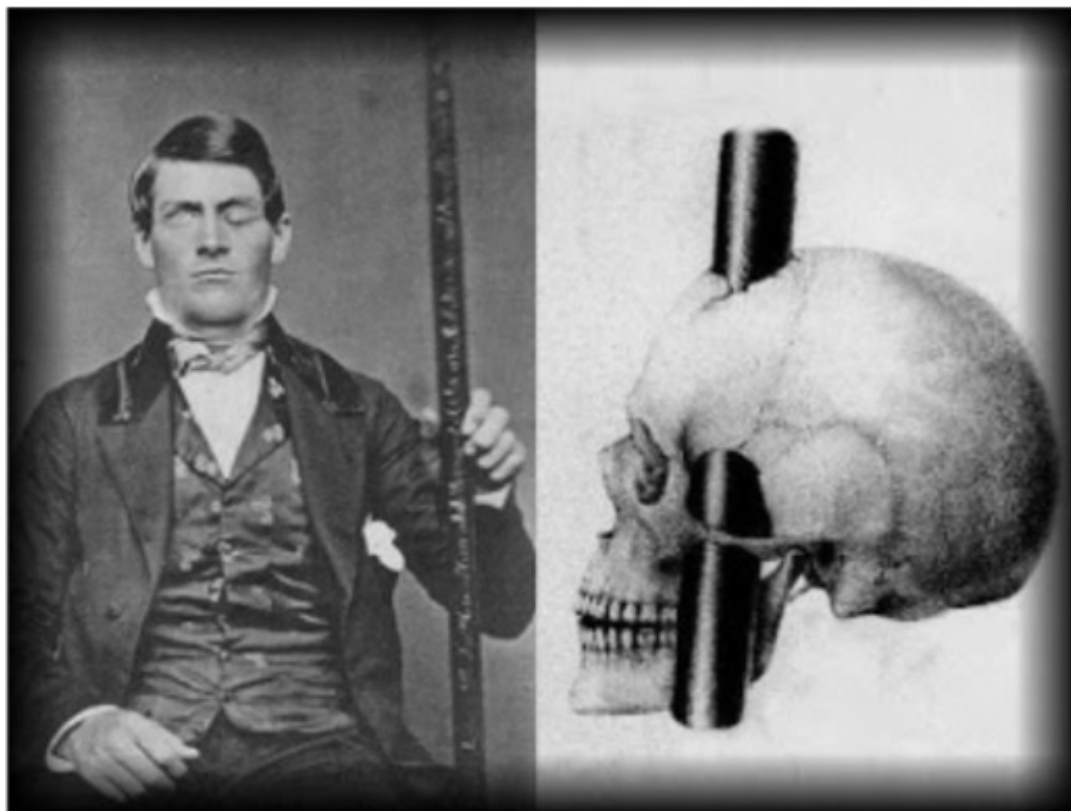
- **Psykolingvistik:** läran om språkets mentala processer och hur språklig information är representerad, t.ex. taligenkänning, läsförmåga, minnesanvändning (Yule, kap. 12)
- **Neurolingvistik:** läran om kopplingen mellan neurofysiologi och språkkunskap, t.ex. hjärnaktivitet under olika språkprocesser, språkutveckling och afasi (Yule, kap 12)



Neurolingvistik



Neurolingvistik



Språkets sociala och kommunikativa funktion

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



Språkets sociala och kommunikativa funktioner

Pragmatik

Semantik är betydelsen av det språkliga uttrycket. Pragmatik är betydelsen som talaren vill förmedla i en viss kontext.

- Det är väldigt kallt här inne
- Kan du skicka saltet?

Språkutveckling

- **Språkinlärning:** Språkets utveckling i individen (ontogenesis). Hur barn tillägnar sig språk och hur vuxna lär sig språk (Yule, kap. 13–14)
- **Språkets uppkomst/språkhistoria:** Språkets utveckling i språkgruppen (phylogensis). Hur förändras språket över tiden mellan generationer av talare? (Yule, kap. 1–2)



Språkutveckling

- Barn som har språklig interaktion med andra språkanvändare lär sig språk, oavsett vilket språk som talas
- Språkinlärning kräver social interaktion. Språklig exponering räcker inte.
- Anpassad input 'parentese'
- Barn följer samma språkutveckling, oavsett språk



Språkutveckling

Barnspråksutveckling

- 0–6 veckor
- 6–8 månader
- 12–18 månader
- 18–20 månader

vokaljoller, velarer

joller

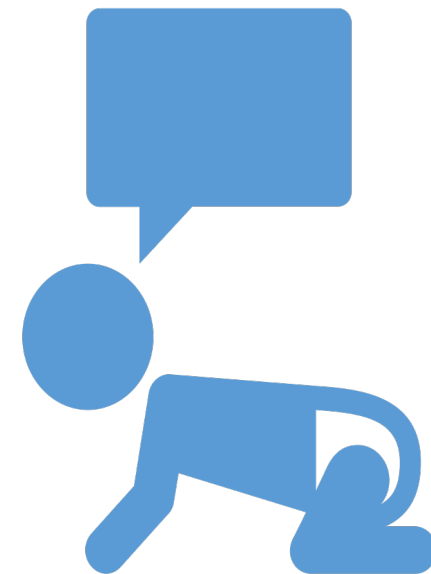
ettordsfas

tvåordsfas

Barn ca två år

Där boll.

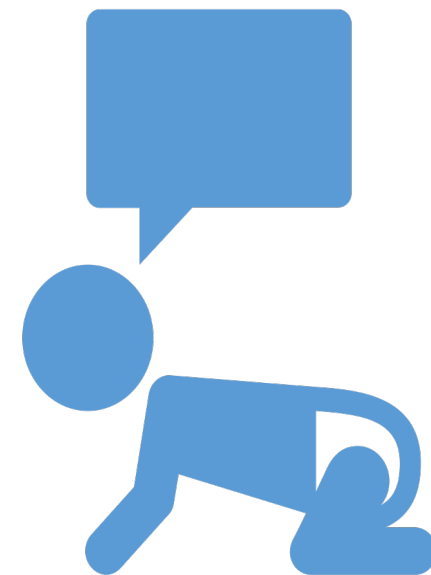
Dockan borta.



Poverty of stimulus: Language Acquisition Device

Jag gedde en docka till Pelle.

Du är den bästaste pappan i världen.



Fornsvenska ca 1450

Her effter bõriæs eth høffuist æuintyr aff Nampnos och Falantin, aff all theres mandom, som the bedriffuo j theres daga

Metoder inom lingvistik

Metoder

Observera språkliga data

- Inspelningar eller naturligt förekommande textdata (korpusdata)
- Experiment
- Tester och utfrågning av infödda talare
- Introspektion

Beskriva språkliga data

- Universellt applicerbara kategorier, begrepp och termer
- Jämföra fenomen inom språk
- Jämföra fenomen mellan språk

Välformade uttryck

- Bilen är röd.
- Bilarna är röda.
- *Bilen är röda.
- *Bilarna är röd.
- Huset är rött.
- Husen är röda.

Slutsatser:

”röd” + singular + utrum

”rött” + singular + neutrum

”röda” + plural, genusneutralt

”är” numerus och genusneutralt

Språk och datorer

Språkteknologi för behandling av stora mängder data:

- Tokenisering (dela upp text i ord och meningar)
- Lemmatisering (koppla böjningsformer till samma grundord)
- Taggning (ordklassbestämning)
- Parsning (satsdelsfunktioner)

- Tokenisering

- "det här" "kaffet" "är" "odrickbart"

Notera att "det här" fungerar som ett ord men skrivs som två

- Lemmatisering

- "det här=det här" "kaffet=kaffe" "är=vara" "odrickbart=odrickbar"

Knyt ord till sin grundform

- Taggning

- ”det här” = pronomen
- ”kaffet” = substantiv
- ”är” = verb
- ”odrickbart” = adjektiv

- Parsning

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

Lite mer detaljer...

Fonetik och fonologi

Fonem: individuella språkljud, minsta betydelseskiljande språkenheten

Fonetik: fonemens egenskaper

Fonologi: fonemens organisation

Fonem

bil – pil

- /b/ och /p/

sky – sy

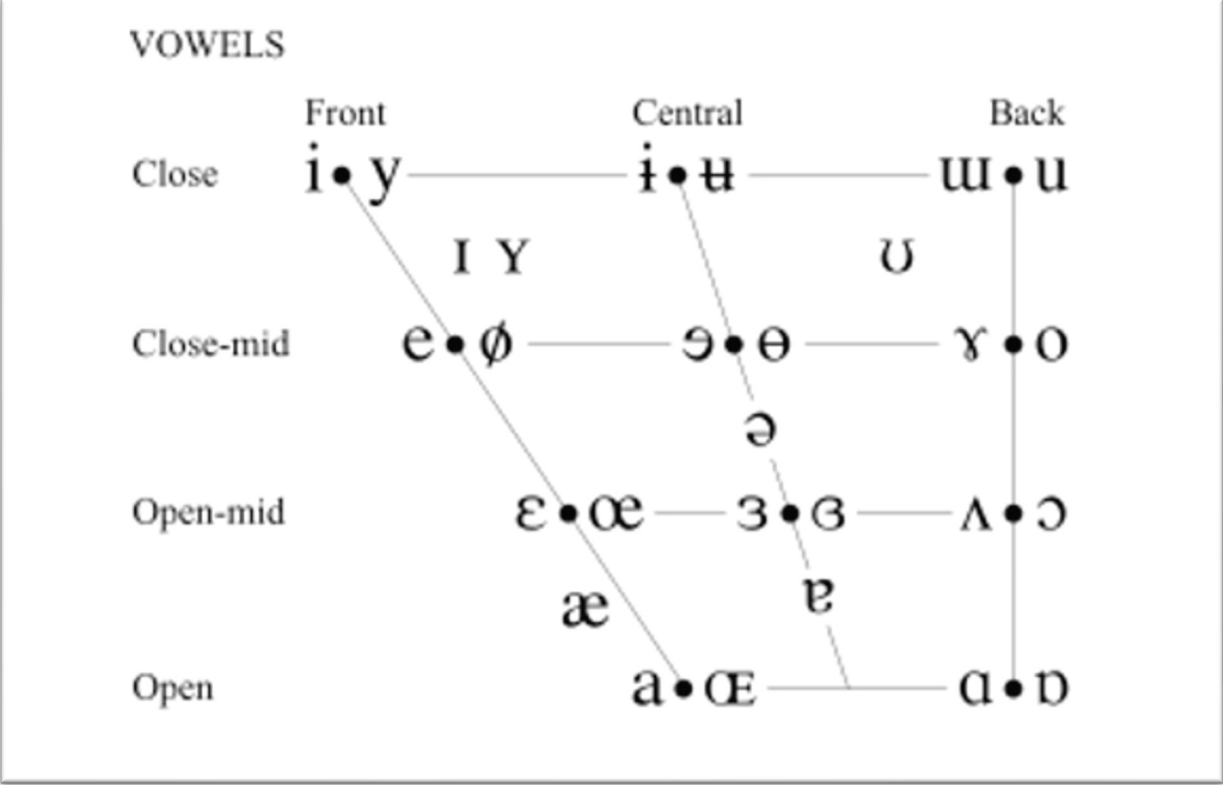
- /ŋy/ och /s/

Engelska

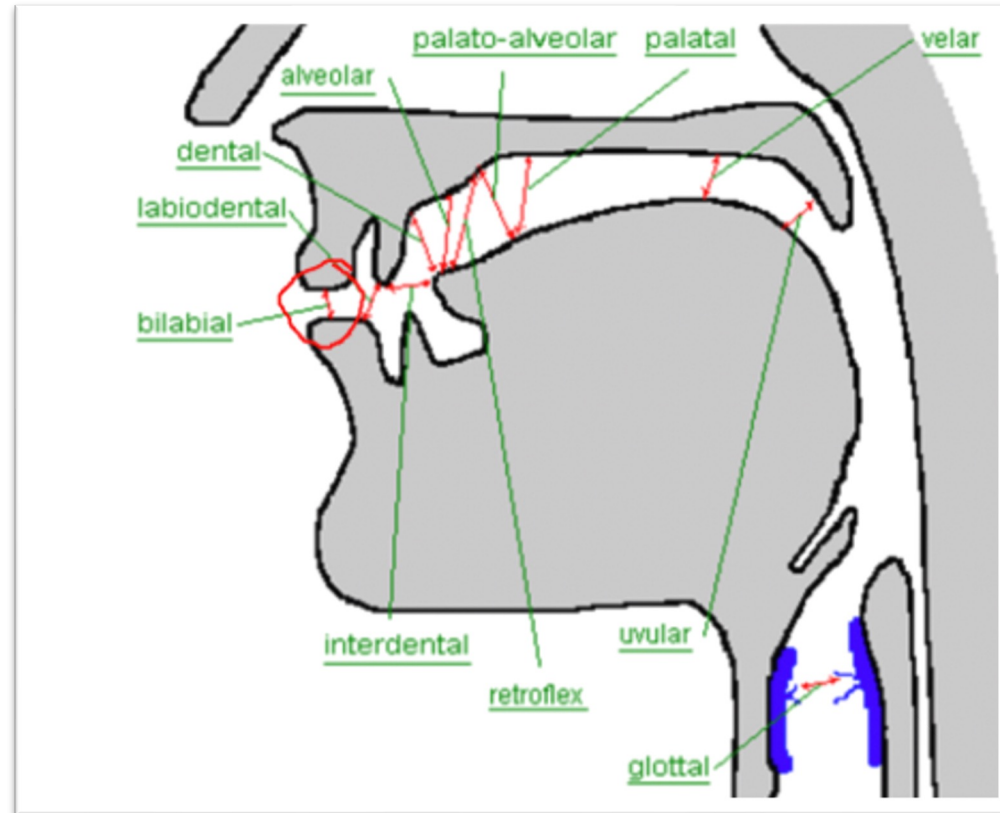
ice – eyes

- /aɪs/ och /aɪz/

Vokalfyrkant

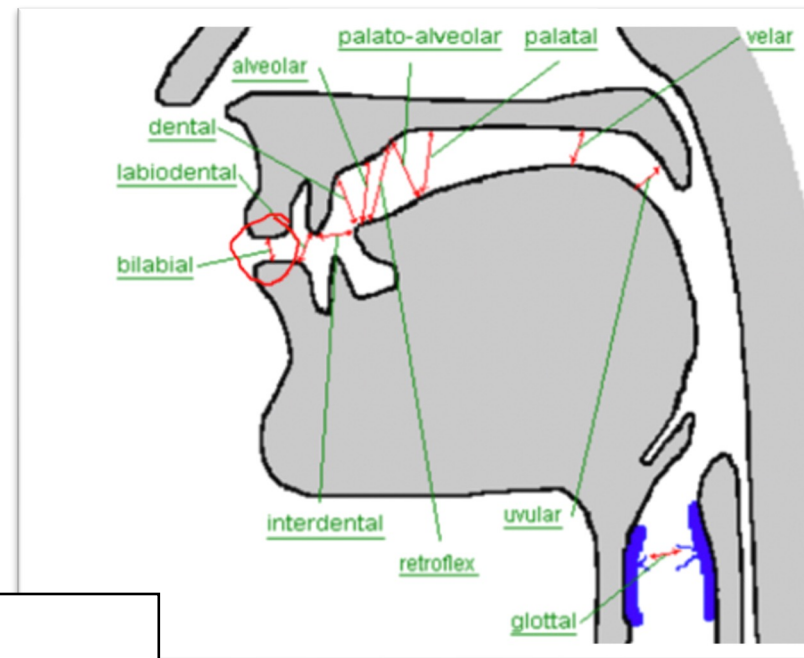


Artikulationsställen



Svenska konsonanter

artikulationsätt	artikulationsställe											
	labial		labio-dental		dental		alveolar/ palatal		palatal/ velar		velar	glottal
klusil	p	b			t	d	ʈ	ɖ	k	g		
frikativ	(f)		f	v	s		ʃ		ç	j	ħ	h
nasal/klusil		m			n		ɳ		ŋ			
lateral					l		ɭ					
tremulant							r					
	-ton	+ton	-ton	+ton	-ton	+ton	-ton	+ton	-ton	+ton	-ton	-ton
	<i>fonation</i>											



Klusil	Explosionsljud	/p/ /b/ /t/ /d/ /k/ /g/
Frikativa	Väsljud	/f/ /v/ /s/ /ʃ/ /ç/ /j/ /ħ/ /h/
Lateral	Sidan av tungan	/l/
Nasal	Luftströmmen genom näsan	/m/ /n/ /ɳ/ /ŋ/
Tremulant	Vibrationsljud	/r/

Assimilation

- 'have' isolerat uttalas /hæv/
- 'have to' uttalas /hæftə/

Morfologi

- Beskriver ordens inre struktur
- Ord består av ett eller flera morfem
 - Bil (ett morfem)
 - Lastbil (två morfem "last" + "bil")
 - Lastbilar (tre morfem "last" + "bil" + "ar")
- Morfem är minsta betydelsebärande enheten i ett språk
- Centrala morfemtyper är **rot**, **avledning** och **böjning**

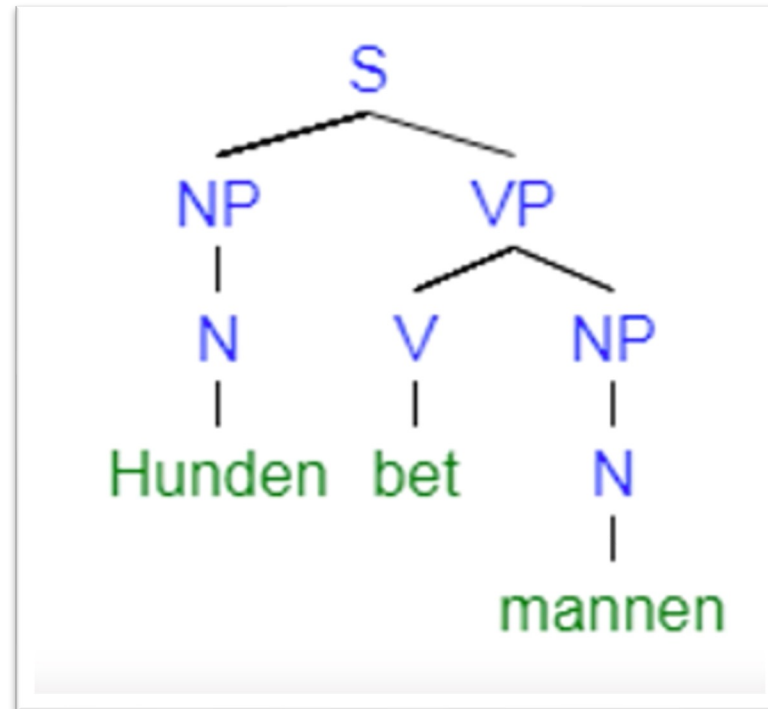
Syntax

- Beskriver hur ord kombineras till större konstituenten: fraser och satser
- Konstituenten beskrivs både utifrån form (fraser, satser) och funktion (satsdelar)

Satsanalys – form och funktion

Funktion	Subjekt	Predikat	Direkt objekt
	Hunden	bet	mannen
Form	Nominalfras	Verbfras	Nominalfras

Syntaxträd



Semantik

- Språkliga uttrycks betydelse
- Hur språkliga uttryck relaterar till varandra

- Lexikal semantik
- Satssemantik

Lexikal semantik

- köpa– inhandla (synonym)
- köpa – sälja (antonym)
- titta – stirra (hyponym)

Satssemantik

- Semantiska roller
 - Kalle (agent) sparkade bollen (tema)
 - Bollen (tema) sparkades av Kalle (agent)
- Satsrelationer ”förutsätter”, dvs. A förutsätter B
 - ”Kalle slutade röka” förutsätter ”Kalle rökte”

Nästa gång

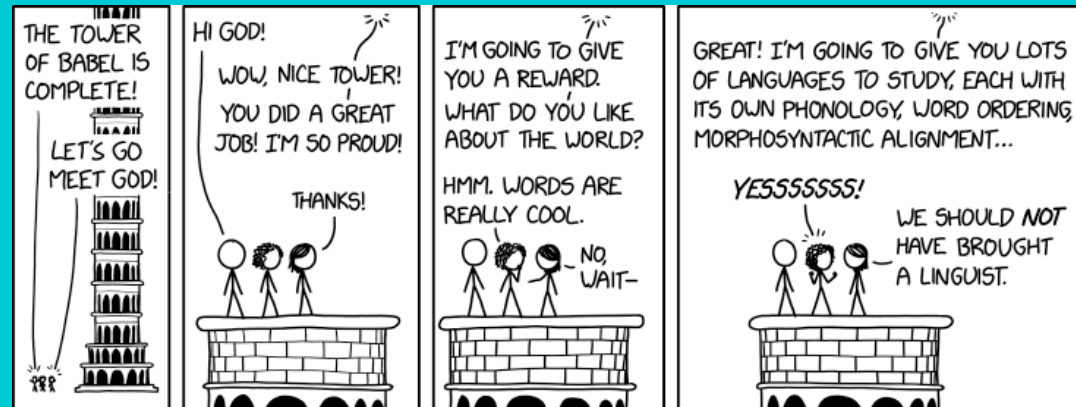
- Inledning till morfologi och syntax, lexikon och ordklasser
- Yule kapitel 5, Josefsson kapitel 4 (ordklasser)

Inför seminarium

- Gör övningar:
 - Josefsson övningsbok kapitel 2, ordklasser och morfologi
 - Utdelade övningsuppgifter

Om ni har frågor, skriv på Teams eller skicka ett mejl (evelina.rennes@liu.se)!

Tack för idag!



<https://xkcd.com/2421/>