Introduction to Linguistics
Linguistics 1

Marcel Bollmann

Department of Computer and Information Science (IDA)

LINKÖPING UNIVERSITY

Structure of the "Linguistics" part

- Lectures (6×)

- Seminars (5×)
    - Exercises *(övningsuppgifter)* on the website
    - Two groups (group 1 ~ lab group Daniel, group 2 ~ lab group Jenny)

- Exam *(dugga)*
    - 📅 2024-05-15 (*omdugga:* 2024-08-21)
    - 🕐 4 hours

Goals

- Get an overview of different subareas of linguistics

- Know about central research questions in linguistics

- Know important linguistic terminology

- Be able to conduct basic linguistic analyses

Lectures on linguistics

L1: Introduction to Linguistics Overview

L2: Introduction to Syntax and Morphology Words, parts of speech, lexicon

L3: Morphology Morphemes, the structure of words

L4: Syntax I Constituents, phrases, sentence structure

L5: Syntax II Main and subordinate clauses

L6: Introduction to Semantics The meaning of words and sentences

What is linguistics?

①②③④⑤⑥⑦

Why study linguistics?

Linguistics *(lingvistik, språkvetenskap)* studies one of the most fascinating human abilities: the ability to **communicate with language**

How is language structured?

How do we understand language?

How do children learn language?

How does language change over time?

How can we make AI "understand" language?

How do dialects of a language differ?

How do we perceive language in the brain?

How is language connected to one's identity?

Large language models (LLMs)

LLMs: ~ statistic models of human language, trained on huuuuuuuuuuuge amounts of data
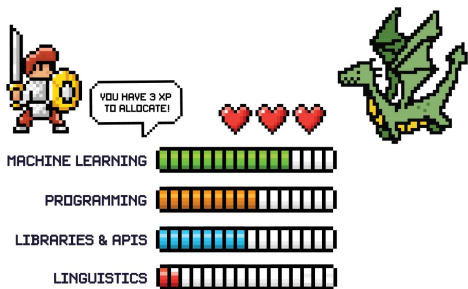
Large language models (LLMs)

- **GPT-SW3** gained traction as "Sweden's answer to ChatGPT"
  - Trained on BerzeLiUs!

- **TrustLLM** is a large, EU-funded research project on trustworthy LLMs for European languages
  - Led by LiU!

DANIEL GILLBLAD forskningsledare AI Sweden

IDAG SLÄPPS SVERIGES SVAR PÅ CHATGPT

SENASTE NYTT PÅ TV4.SE Storbanken sänker bolåneräntor

More information on ⬈ **GPT-SW3** and ⬈ **TrustLLM**

But why linguistics?

> *For applied **natural language processing (NLP)**, a little bit of linguistics knowledge can go a long way and prevent some expensive mistakes.*
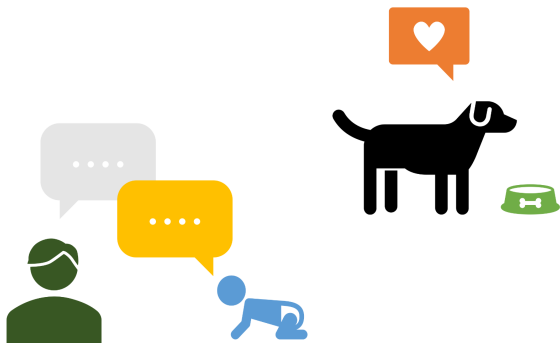
What is "natural" language?

no formal
languages, e.g.
programming
languages

Source: Unsplash

Language ≠ communication

- All living beings communicate, but not all have language

Properties of human language

- **Generative**: a limited set of words can express an unlimited set of meanings

- **Recursive**: a structure can contain itself (e.g. sentence within a sentence)
    - *Hunden jagade katten som jagade råttan som jagade...*

- **Symbolic**: the connection between words and meaning is purely a convention
    - 🐕: *dog, hund, chien, perro, ...*

- **Displacement**: we can refer to things that are not spatially or temporally present
    - *Where is my wallet? — It was here yesterday!*

Language ≠ thought

- Small children can think before they can speak

- Animals without language can think

- Many types of cognitive activity do not require language

- We store the meaning of linguistic communication in a more abstract form

- Some types of brain damage result in aphasia *(afasi)* without any other cognitive problems

Written vs. spoken language

Language can be expressed in different **modalities** (*speech, signs, writing*).

Spoken language is often seen as the **primary form** of language:

- All cultures have spoken language, but not all have writing.
- Children learn to speak intuitively, but must be taught how to write.
- Spoken language has probably existed much longer (50.000–100.000 years) than written language (5.000 years).

Language as a system

Language and the brain

Social and communicative function of language

Language development
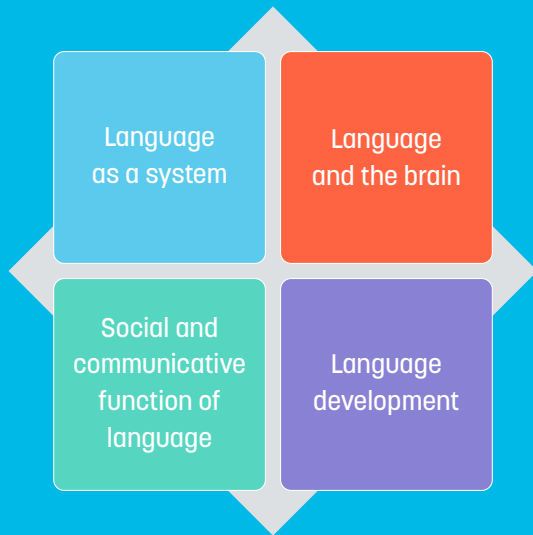
Language as a system

Language as a system

- **Phonology**: the study of sounds in a language and how they are organized (Yule, § 3–4)

- **Morphology**: the study of how words are built up from "smallest meaningful units" (Yule, § 5–6)

- **Syntax**: the study of how words combine to form phrases and sentences (Yule, § 7–8)

- **Semantics**: the study of the meaning of linguistic expressions (Yule, § 9)
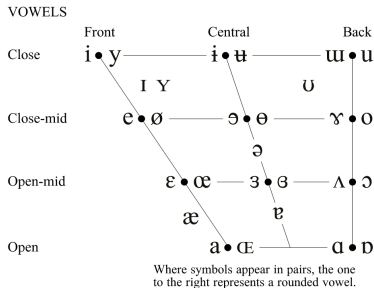
Language as a system

/dehɛːrkafːəteʊdrɪkbaːʈ/
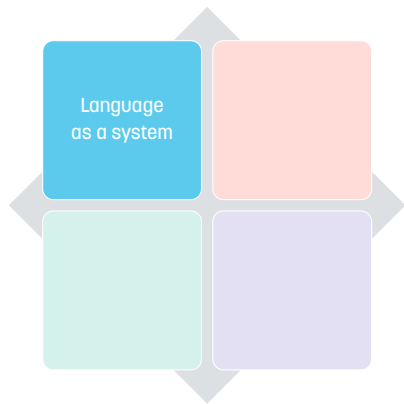
det här kaffet är odrickbart

Phonetics and phonology

Phonetics *(fonetik)* studies the speech sounds we produce.

Chart of vowel sounds

Phonology *(fonologi)* studies how sounds are organized and used in a language.
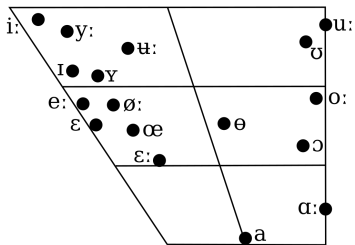
Vowels in Central Standard Swedish

Sources: IPA and Jeffrey Connell & Moxfyre

Phonology

$$d\underline{ö}r - d\underline{ö}d$$

- A **phoneme** *(fonem)* is an abstract unit of sound that can distinguish one word from another.

- Different sounds that belong to the same phoneme are called **allophones**.
  - *d<u>ö</u>r* – phoneme /ø/ – phone/sound [œ]
  - *d<u>ö</u>d* – phoneme /ø/ – phone/sound [ø]

- Two words that differ in only one phoneme are called a **minimal pair**.
  - *<u>sk</u>ena* /ɧeːna/ — *<u>tj</u>ena* /ɕeːna/

Morphology

Morphology *(morfologi)* studies the structure of words.

- A **morpheme** *(morfem)* is the smallest meaningful unit of a language.

$$kaffe - \quad t$$
Root   def.

$$o \quad - drick - bar - \quad t$$
neg.   Root   adj.   case

Part of speech

- A **part of speech** *(ordklass)* is a group of words with similar functions in a sentence.

/dehɛːr/	*det här*	**pronoun** *(pronomen)*
/kɑfːət/	*kaffet*	**noun** *(substantiv)*
/e/	*är*	**verb** *(verb)*
/ʊdrɪkbɑːʈ/	*odrickbart*	**adjective** *(adjektiv)*

- The **lexicon** determines what parts of speech a word can have; morphological processes can change it.

Syntax

Syntax *(syntax)* studies how words combine to form phrases and sentences.

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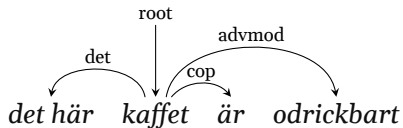
Semantics

Semantics *(semantik)* studies what words, morphemes, phrases, and sentences mean.

det här kaffet är odrickbart

the object is being prominently
pointed out in the speech context

there is a property being
assigned to the object

the object referred to as "kaffet",
which is prominently pointed out in the speech context,
is described as having the property "odrickbart".

Lexical semantics

"milk drinkers turn to powder"

- **Polysemy** *(polysemi)*: words or phrases can have multiple meanings
    - *turn to* ~ directing one's activity
    - *turn to* ~ become changed or transformed

- Examples for other semantic relations...
    - **synonymy**: *köpa – inhandla*
    - **antonymy**: *köpa – sälja*
    - **hyponymy**: *titta – stirra*

Phrasal semantics

- **Thematic roles** describe the function of a word in sentence.

  Kalle sparkade bollen
  AGENT   THEME

 *Bollen sparkades av Kalle*
  THEME   AGENT

- **Textual entailment** describes the semantic relation between sentences.

  – **Entailment** *(innebörd)*: *Kalle slutade röka* ⟶ *Kalle rökte*

  – **Contradiction** *(motsättning)*: *Kalle klarade duggan* ⟷ *Kalle blev underkänd*

Language and the brain

3

Language and the brain

- **Psycholinguistics** *(psykolingvistik)*: the study of how language is represented and processed in the mind and brain (Yule, §12)

    – How do children acquire language?

    – How do we use our memory when we speak?

    – How do we learn a second language?

- **Neurolinguistics** *(neurolingvistik)*: the study of brain activity with regard to language (Yule, §12)

    – How do speech or language disorders occur?

    – What parts of the brain are active when we speak or try to understand language?

Language and the brain
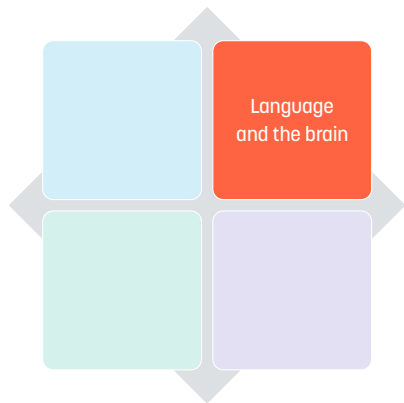
Broca and Wernicke areas

- Damage to certain brain areas is clearly linked to specific types of **aphasia** *(afasi)*.

Social and communicative function of language

◯◯◯**4**◯◯◯
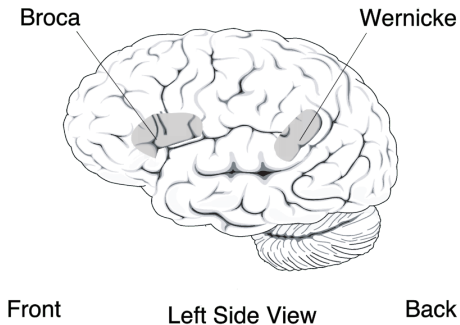
Social and communication function of language

- **Pragmatics**: the study of how we use language to communicate in certain contexts (Yule, § 10)

- **Discourse analysis**: the study of structure beyond individual utterances, i.e. entire conversations (Yule, § 11)

- **Sociolinguistics**: the study of the social function of language (Yule, § 18–20)
  - cultural norms, speech communities

Social and communicative function of language
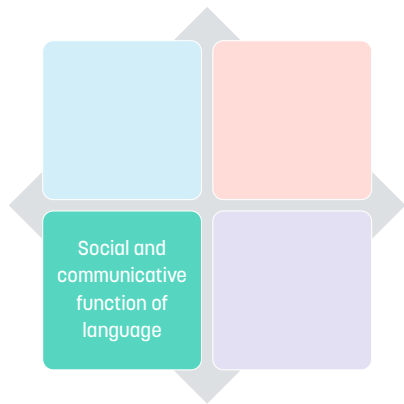
Pragmatics

Pragmatics *(pragmatik)* studies what we want to communicate by saying/writing something in a given context.

I want to complain about this coffee

det här kaffet är odrickbart

I want to tell others not to drink this coffee

*I want someone to bring me a new cup of coffee*

Language development

Language development

- **Language acquisition** *(språkinlärning)*:
  the study of how we acquire a new language
  (Yule, § 13–14)

  – either as a child or as an adult
    (second-language acquisition)

- **Historical linguistics** *(historisk lingvistik)*:
  the study of how a language changes and
  develops over time (Yule, § 1–2)

Language
development
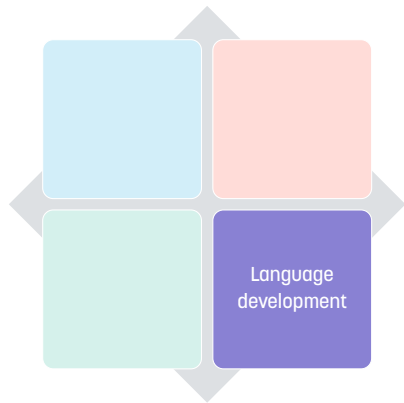
Language acquisition

timeframe	stage of development
0–6 months	babbling *(joller)* with vowels
6–12 months	babbling with vowels and consonants
12–18 months	speaks 8–10 words
18–24 months	speaks simple sentences

där boll

dockan borta
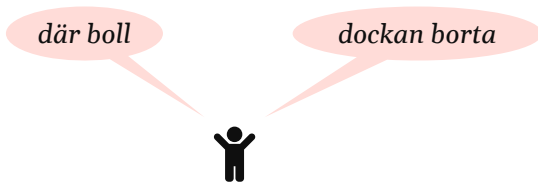
Historical linguistics

> *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.*

Methods used in linguistics

How do you "do" linguistics?

- Linguistics is the **scientific study** of human language.

👍 In academia, we take a **descriptive** approach to language.

    – Descriptivism: *objectively* analysing and describing how language is *actually used*

👎 This contrasts with **prescriptivism** sometimes found in education.

    – Prescriptivism: establishing *rules* that tell others how to use language *"correctly"*

How can we obtain language data?

To **analyse and describe** how language is used, we can use:

- Recordings or naturally occurring text data *(corpus data)*
  - "naturally occuring" ~ not written specifically for this analysis

- Tests and questioning of native speakers
  - "Which of these sentences sounds more natural to you?"
  - "Do *you* perceive this sentence as ungrammatical?"

- Introspection
  - "Do *I* think this sounds natural/wrong/etc.?"

Linguistic well-formedness

- An **asterisk (*)** before an expression usually indicates **ungrammatical examples**.

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

- With such grammaticality judgments, we can draw conclusions:

  - *röd* is used with singular + common gender *(utrum)*
  - *rött* is used with singular + neuter gender *(neutrum)*
  - *röda* is used with plural + any gender

Language technology

Language technology *(språkteknologi)* is the analysis and interpretation of language with computational methods.

- Advantage: can analyse **large amounts of data** quickly!

- Often performs **automatic annotation** of linguistic structure

- More in the *språkteknologi* lectures and in 729G17 Language Technology!

Wrapping Up

Wrapping up

Before the seminar

Do the following exercises:

- Josefsson övningsbok chapter 2 *(ordklasser och morfologi)*

- *Övningsuppgifter* from the website

Next time: Introduction to syntax and morphology

Reading:

- Yule chapter 5, Josefsson chapter 4 *(ordklasser)*