Algemene Taalwetenschap

Inleiding

Jennifer Spenader, februari 2006
(Some slides: Petra Hendriks)

General Linguistics

Introduction

Jennifer Spenader, February 2006
(Most slides: Petra Hendriks)

Practicalities

- Course form:
  - 11 lectures, 1 group meeting (werkcollege), 1 "responsiecollege"
- Literature:
  - W. O'Grady, M. Dobrovolsky & F. Katamba (eds.), *Contemporary Linguistics: An Introduction*. Addison Wesley Longman, 1997.
  - Klapper: Verkrijgbaar in klapperwinkel GMW of online

- Examination:
  - 5 "opgavensets" + exam

Goal of the course

- Introduction to linguistics
  - Phenomena studied
  - Methods used
  - Results so far, current theories
- Preparation for other AI courses
  - E.g. Taal- en Spraaktechnologie
- Aids in understanding linguistic factors in other subjects

How to succeed in this class

1. Come to the lectures
   - Most important material covered here
   - TIP: Read through the opgaven and glance through the chapter BEFORE the lecture. This will help you focus!
2. Do the reading assignments
   - Pay attention to the reading guide of what is most important. Don't get lost! Some chapters are very long, and very detailed!
3. Do the homework on time.
4. Read through the exercises at the back of each chapter we read: this will also give you an idea of what is important!

Levels of language

• Text/Dialogue	⇨	Pragmatics	(lecture 11)
• Sentences	⇨	Syntax	(lectures 5 en 6)
		Sentence semantics	(lecture 10)
• Words	⇨	Morphology	(lecture 4)
		Lexical semantics	(lecture 9)
• Syllables	⇨	Phonology	(lecture 3)
• Sounds	⇨	Phonetics	(lecture 2)

We'll also look at…

- Language Acquisition (college 8)
- Natural Language Processing (college 12)
- Differences and similarities between langauges
- Computational applications
- Relation between language and other cognitive processes

What is language?

What is language?

- Language is:
  - ⇨ Form of knowledge
  - ⇨Formal system
  - ⇨A code?
  - ⇨A kind of behavior?
  - – Tool for communication
  - – Social activity
  - – …

How did we learn language?

- Is language special?
- Is language just a communication system based on other cognitive abilities?
- Is learning language different from learning how to e.g. walk or recognize faces?
- Are the communicative systems of animals just a simpler form of language or are they completely different types used for a similar purpose?
- Are humans “programmed” for language?

Is language special?

- Is language special compared to other communication methods
  - E.g. Animal communication
- Are we “programmed for language” or is language just part of general cognitive skills common to all humans
- Is part of language genetic?

Two perspectives

View One Language is based on general cognitive skills, and learned like any other learning.

View Two Language is a special type of knowledge, for which we are programmed. There is a "universal grammar" underlying all languages. Specific learning mechanisms are used by children when acquiring language.

Two perspectives

View One Language is based on general cognitive skills, and learned like any other learning.

- Pyschologists, computer scientists working with language

View Two Language is a special type of knowledge, for which we are programmed. There is a "universal grammar" underlying all languages. Specific learning mechanisms are used by children when acquiring language.

- (Generative) linguists

Arguments against **View 1**

- Poverty of stimulus
  - Language has a structure too complicated to be learned from data alone. Children reach levels of competence too quickly for all their learning to be the result of their experience.There must be an “LAD” (Language Acquisition Device)

- Lack of negative evidence
  - Children very rarely get negative evidence about what is not permitted in the language. But without negative examples learning is not possible.

Evidence why **View 2** must be correct: language variation not totally free

- **View 2** predicts that Languages should all be similar
  - Typological research seems to confirm this
  - Researchers in Generative Linguistics working on principles of UG (Universal Grammar" also believe that this means that by studying one language in depth, you can discover principle of all languages

Evidence why **View 2** must be correct: Creoles

- Pidgens: languages used as a “lingua franca” in situations with language contact between several mutually unintelligable languages.
- Children to pidgen speakers add morphosyntax distinctions to the language, turning it into a full-fledged language, a CREOLE
- Seems to indicate that there are programmed distinctions

Evidence why **View 2** must be correct: Language skills and General Intelligence don't correlate

- **Normal intelligence with impaired language:**
  - Broca's aphasia
    - can usually understand what words mean,
    - have trouble performing the motor or output aspects of speech
    - E.g. Yes … ah … Monday … er Dad and Peter H … (his own name), and Dad … er hospital
  - Wernicke's aphasia
    - speak extremely fluent nonsense
    - E.g. Well this is …. mother is away here working her work out o'here to get her better, but when she's looking, the two boys looking in other part.
    - Have also lost comprehension ability

Evidence why **View 2** must be correct: Normal language with impaired intelligence

- Williams syndrome

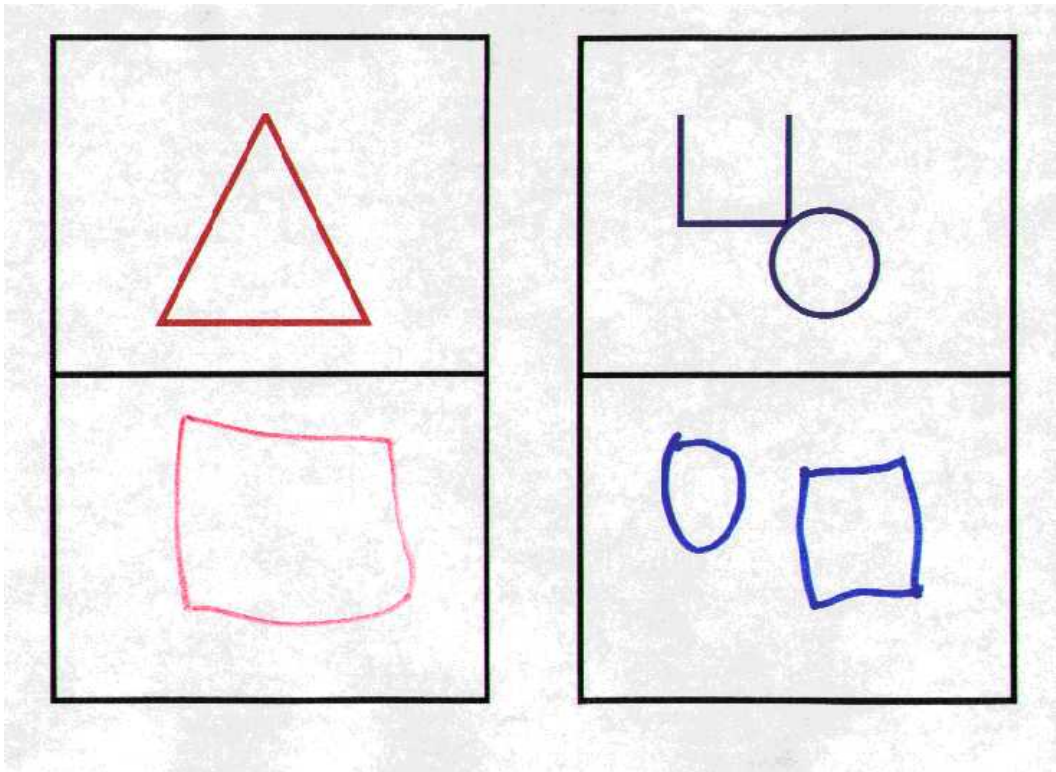
William's Syndrome

Characteristics of individuals with William's syndrome

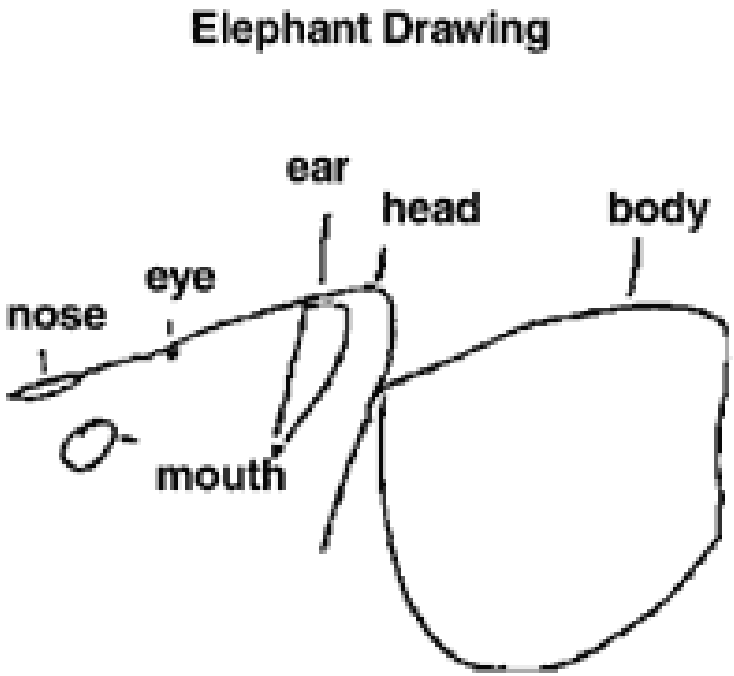
- Elf-like appearance
  - Some believe legends about Elves may have come from William's syndrome
- Mild mental retardation + spatial understanding problems
- Very verbal, talkative, expressive
  - Also called "Cocktail party syndrome
- Exceptional music ability

Williams syndrome

- http://www.cog.jhu.edu/faculty/landau/lang-cog-lab/SPACE.html



Elephant drawing



- http://www.wsf.org/behavior/research/scientist.htm

Williams syndrome: genetic correlates to language

- the region from the center of the 7 chromosome to the q11 band is missing in individuals with William's syndrome

- Normal individuals have two copies of 7q10-q11

Supernumerary ring 7

- a condition where a person has the usual 46 chromosomes as well as an extra, tiny ring chromosomes containing the region from the center of the 7 chromosome to the q11 band.
- the person has three copies of 7q10-q11, instead of just two.
- Opposite of Williams Syndrome
  - Remember: in Williams Syndrome 7q10-q11 is missing

Characteristics of individuals with Supernumerary ring 7

- Severe retardation of expressive speech development
  - K.D. Lichtenbelt et al. (2005), Supernumerary ring chromosome 7 mosaicism: Case report, investigation of the gene content, and delineation of the phenotype, American Journal of Medical Genetics Part A, 1552-4825
- Few words, late development of more words
- poor articulation remains even in adulthood
- Mild mental retardation Gita Tan-Sindhunata, Sérgio Castedo, Beike Leegte, Irma Mulder, Anneke Y. vd Veen, Annemieke H. vd Hout, Titte J. Wiersma, Anthonie J. Van Essen (2000) Molecular cytogenetic characterization of a small, familial supernumerary ring chromosome 7 associated with mental retardation and an abnormal phenotype, American Journal of Medical Genetics 147-152.

Do all languages have the same level of complexity? Possible arguments against **View 2**

- If languages were found that were very different from all languages inventoried so far, might be evidence against believing that basics of language are "hard-wired"

- Could also be evidence of the lowest level of complexity found

The Pirahã

Dan Everett's photos

Piraha


SURINAME
FRENCH GUIANA
ATLANTIC OCEAN
BOA VISTA
MACAPÁ
AMAZON RIVER
MANAUS
BELÉM
TERESINA
AMAZONAS
RIO BRANCO
PORTO VELHO
Brazil
CUIABÁ
BRASÍLIA
GOIÂNA
THE PANTANAL
BOLIVIA
PARAGUAY
SÃO PAULO
IGUAÇU FALLS
CURITIBA
CHILE
ARGENTINA
URUGUAY
PORTO ALEGRE
OCEAN
Madeira River
Maici River
Humaitá
Trans-Amazon Highway
Detail
0
600 km

The **Pirahã**

- The **Pirahã** are an indigenous hunger gatherer tribe of Amazonian Indians in Brazil
  - Only about 200 members
- only surviving member of the Mura language family, all other members having become extinct in the last few centuries.
  - Language isolate: has no known connection to other languages

Characteristics of Pirahã

- Lack of number words
  - makes Pirahã a fascinating test case of **the Sapir-Whorf Hypothesis**, and more generally of the link between language and cognition.
- Claims that they have no quantification
- Smallest **phoneme inventories** of any known language
- An extremely limited clause structure
- No abstract color words other than terms for light and dark.
- Few specific kin one word covers both "father" and "mother".
- The personal pronouns (and seemingly no other words) may have been borrowed from an unrelated Tupian language.
- Pirahã can be whistled, hummed, or encoded in music.

Tests of Pirahã

- Without numerals, the Pirahã do not count. They use only approximate measures, and in tests were unable to consistently distinguish between a group of four objects and a similarly-arranged group of five objects. When asked to duplicate groups of objects, they duplicate the number correctly on average, but almost never get the number exactly in a single trial.

- As of 2004, most of the remaining Pirahã speakers were monolingual, knowing only a few words of Portuguese. It is the belief of the Pirahã people that their language is the best one to speak, so there seems to be no immediate danger of Pirahã dying out.

Animal communication

- Do animals communicate like we do?

- Vervet monkey:
  - Have signals for
    - Snake
    - Leopard
    - eagle

Nim Chimpsky

- Play on words with name "Noam Chomsky"
- Taught sign language from birth
- Most frequent 2-3 word signs
  - Play me
  - Me Nim
  - Tickle me
  - Play me Nim
  - Eat me Nim
  - Eat Nim eat

“Real language”

- Separated from events and time
  - I.e. you can tell about things that have happened or could happen
- Syntactic structure which allows more combinations
  - Meaning not just in symbols, but also in how they are combined
- Maybe language is key to why we are so cognitively advanced

What did she do yesterday?

What will she do tomorrow?

Can she tell us?

Language affect cognition? Does language make use different from our genetic "cousins"?

- (Sapir)-Whorf hypothesis
  - The language that we speak influences how we understand the world
    - HOPI = Whorf believed that they had no concept of time, because they had no tense system
    - ESKIMOS = Whorf believed that because snow was so important to them, they had a larger number of lexical items for snow
      - THIS IS HOWEVER A MYTH!

Weaker Whorfian Hypothesis

- Language has a slight affect on cognition
  - Spatial perception

- Language is a prerequisite for certain types of cognition
  - Children don't understand that others can have false beliefs until they acquire sentential complements
    - E.g. John said that he was leaving.

Turing Machines vs. Language

- At some level language must be a kind of computation
- Since Turing machines can “model” all computations, thus a TM must be able to model langauge

Language can be described as a formal system

- Comparable to a logical system or a programming language
- Order is important
- Recursive
- Can be described as having a formal level of complexity, placed on the Chomsky hierarchy scale (Chomsky, 1956).

Chomsky hierarchy

In order of increasing generative complexity:

Finite languages

- Type-3 Languages: Regular A => a B en A => a
- Type-2 Languages: Context-free A => γ
- Type-1 Languages: Context sensitive α A β => α γ β
- Type-0 Recursively enumerable languages: No restrictions on the form that rules can have, equivalent with Turing Machines

Complexity of language

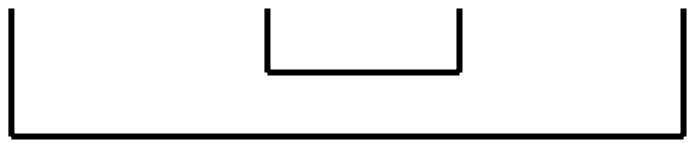
Where does natural language fit on the Chomsky hierarchy?

- If…then constructions
- Subject-V-agreement

➔ Language is at least context free

Context-sensitive

- English uses nested dependencies
  - The man the boy saw laughed.



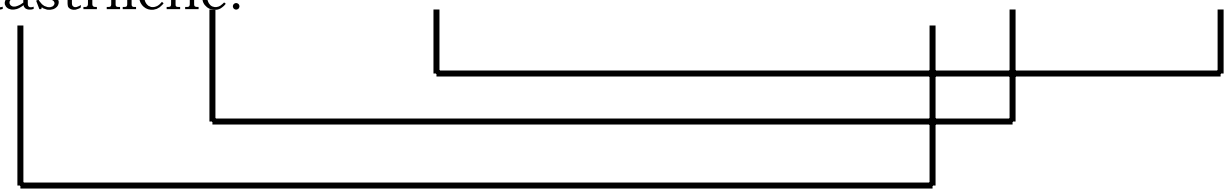
Context-gevoeligheid

- Dutch uses crossed dependencies:
  - dat we Hans het huis helpen verven.

Context-gevoeligheid

- Zwitsers-Duits:
  - (Jan saït das mer)

d'chind em Hans es huus haend wele laa hälfe aastriiche.



"Jan zei dat we de kinderen (ACC) Hans (DAT) het huis (ACC) hadden willen laten helpen verven"

➔$xA^kB^lyC^kD^lz$ (resultaat van context-gevoelige regel)

Patterns

- Robot behavior
  - Patterns, making new patterns, recognizing patterns
- OCR
  - Recognizing patterns, recognizing symbols, their variations and their associated meaning

Linguistic Patterns

- Language is also a code, that can be used to communicate messages
- It has certain patterns
- Moreover, languages vary in certain ways, so we can even have expectations about the form of languages we know nothing about
  - However we still don't know exactly how to model the patterns of language
  - Linguistic Research: tries to MODEL language patterns

Summary

- Language can be seen as a code, following certain constraints
- It seems to be pre-programmed in some ways
- To what degree it is pre-programmed is
  - Controversial
  - Heavily researched
- Concentration in this course: Basics of linguistic knowledge/research

Language as a type of knowledge

- Implicit, unconscious knowledge
  - Fun about linguistics: you can "discover" things about language just by introspection
    - Making up own examples that illustrate points
    - Negative: armchair linguistics
- People have intuitions about language
- Competence vs. performance
  - Performance includes all mistakes people make
  - Asking about intuitions gets at competence
    - Negative: Linguistics have been shown to have very different intuitions from "naïve" speakers
    - Negative: claim here is then that you can learn nothing from e.g. corpus data!

Perspective on language

- All our unconscious knowledge about language = grammar
  - Aim of linguistics is to model this grammar

- Descriptive study

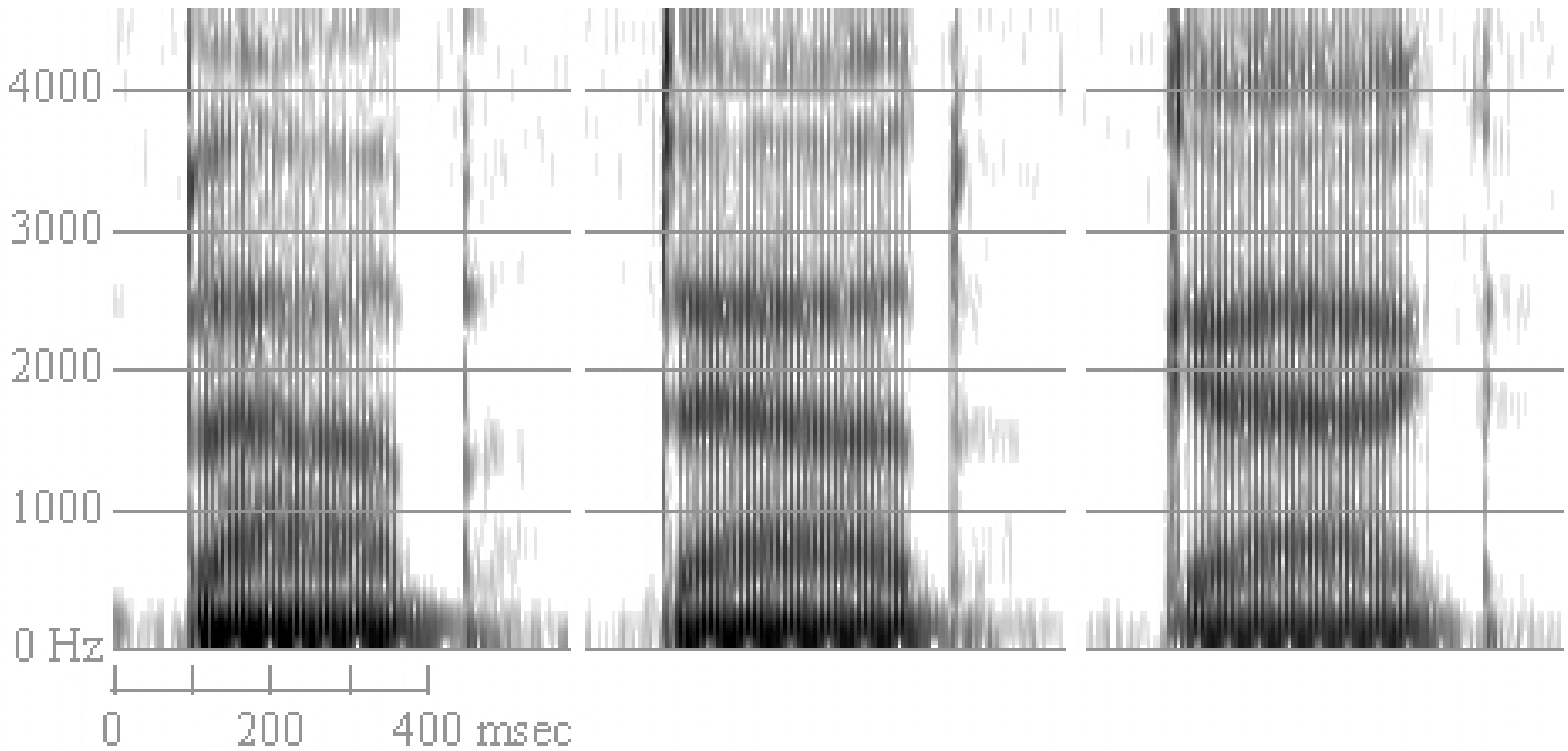
(c.f. prescriptive grammars)

Phonetic knowledge

- What are the acoustic characteristics of the sounds of language, and how are they formed and recognized?

- Example 1: Production of [m]
- Example 2: acoustic characteristics of speech sounds
- How does ability to hear a sound relate to ability to produce it?

Spectrograms of ‘bab’, ‘dad’ and ‘gag’



Phonological knowledge

- What restrictions are there about what speech sounds can follow each other, and how are sounds influenced by other sounds nearby
- Example 1: *prak* vs. **rpak*
- Example 2: /e/ in *beer* vs. /e/ in *beet*
- Some sound differences carry meaning in a language, and some do not.

Morphological knowledge

- How are words formed and how are new words formed?
- Example: *jagers* = *jaag* + *-er* + *-s*
- Why can you *Theeleuten*, *koffieleuten* but not *Chocomelleuten*?
- Differences between languages: some languages have a rich morphology, other language have a much poorer morphology

Poor morphology

- Mandarin Chinese:
  - Ta chi fan le.

    zij eet maaltijd v.t.

Rich morphology

- Inuktitut:
  - **Qasuiirsarvigssarsingitluinarnarpuq.**

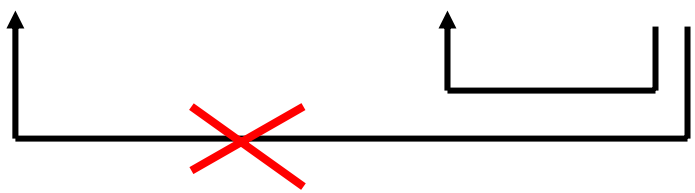
-Moe niet veroorzaken plaats-voor geschikt vind niet volledig iemand 3e-pers.e.v.

“Someone failed to find a completely appropriate resting place”

(Iemand vond niet een volledig geschikte rustplaats)

Syntactic knowledge

- How are words combined to make sentences?
- Example 1:
  - Marie zag Jan. SVO
  - Piet gelooft dat Marie Jan zag. SOV
- Example 2:
  - Piet gelooft dat Marie zichzelf zag.



Semantic knowledge

- How is the meaning of a sentence determined by the form of the sentence?
- Example 1:
  - Marie zag Jan ⇔Jan zag Marie.
    - Irene Kramer: Dutch children confused about subject but in production:
      - Marie zag.
      - Zag Jan.
- Example 2:
  - Marie zag de man met de verrekijker. (ambiguïteit)

Most

- Most ships load at night.

- Most linguists drink at night.

Pragmatic knowledge

- In what way is the meaning of a sentence dependent on the context in which it is spoken?
- Example1: Context-dependency of words like *hij*, *zij*, *gisteren*, *hier*, etc.
- Example 2: Information structure:
  - De hond beet Jan.
  - Jan werd gebeten door de hond.

Next time

- Phonetics!

- Acknowledgements:
  - Thanks to Philippe Schlenker's intro to linguistics notes for inspiration and information!