Foundations of Language Science and Technology
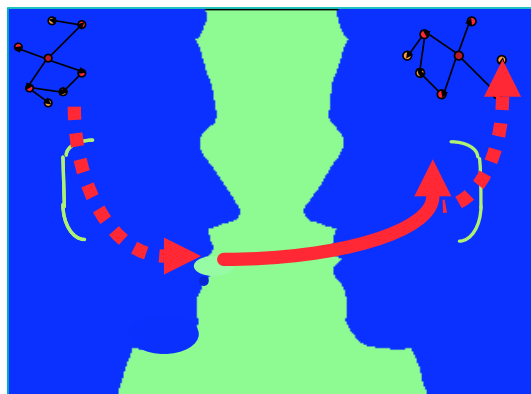
Introduction
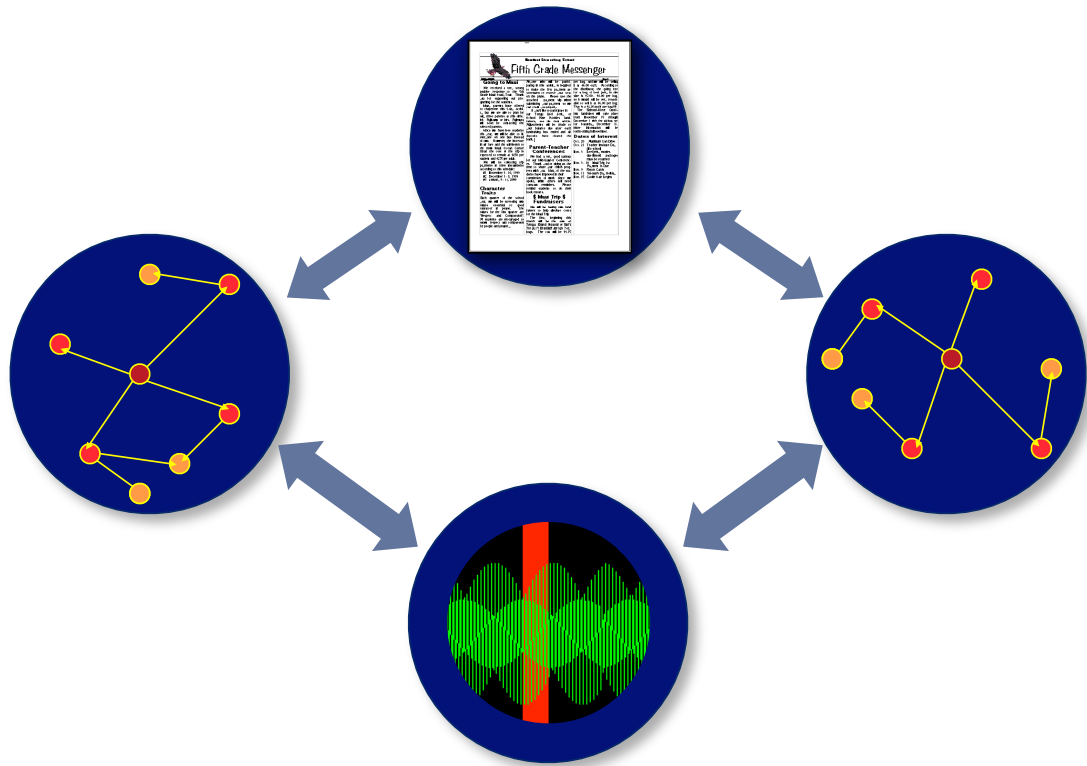
Alexander Koller
October 22, 2008

based on slides by Hans Uszkoreit
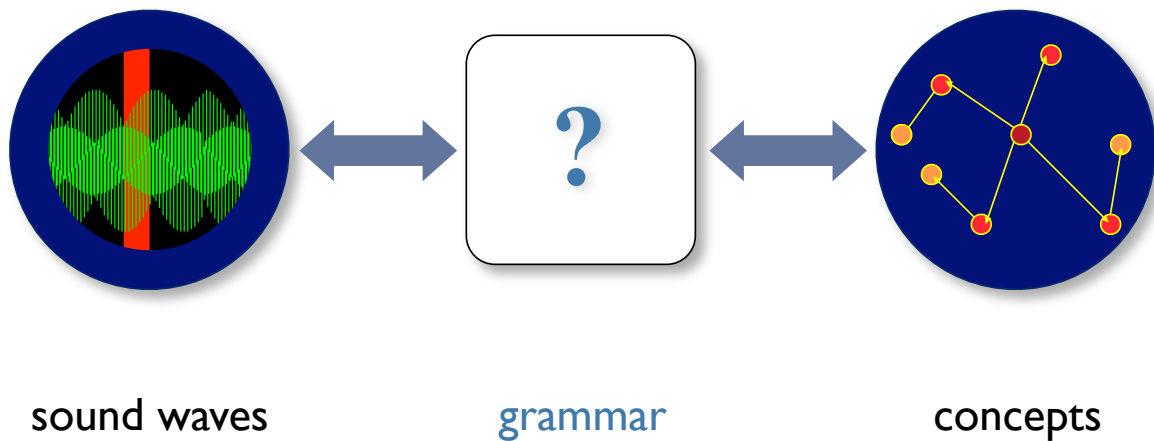

The Miracle
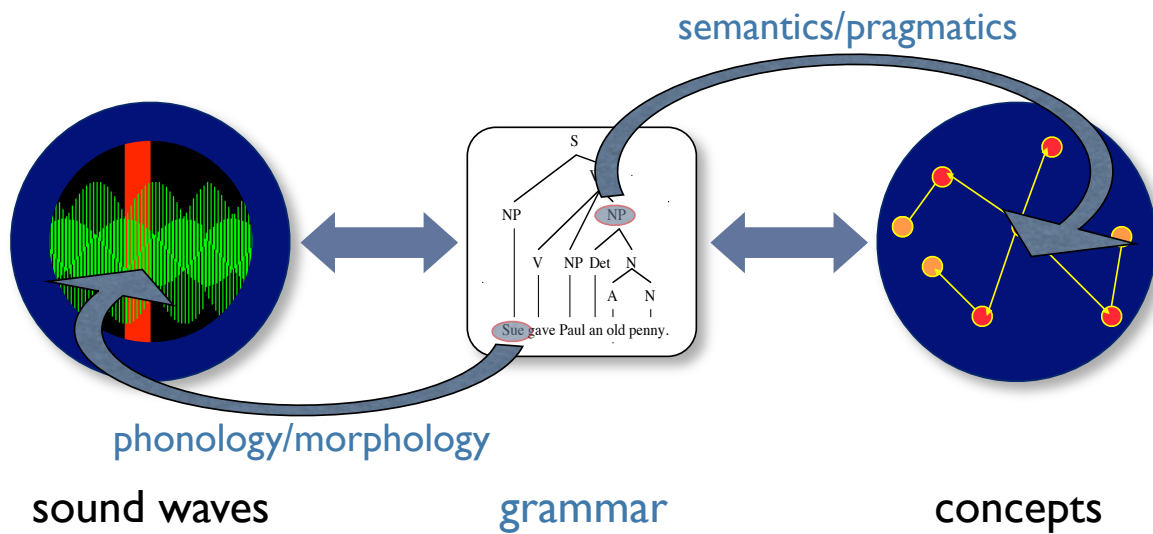
Language is the Medium

What happens in between?

sound waves   grammar   concepts

What happens in between?

semantics/pragmatics

phonology/morphology

sound waves   grammar   concepts

Different grammar traditions

Phrase-structure Grammar

Dependency Grammar

Categorial Grammar

A grammar

Phrase-structure Grammar

S → NP VP

A grammar

Phrase-structure Grammar
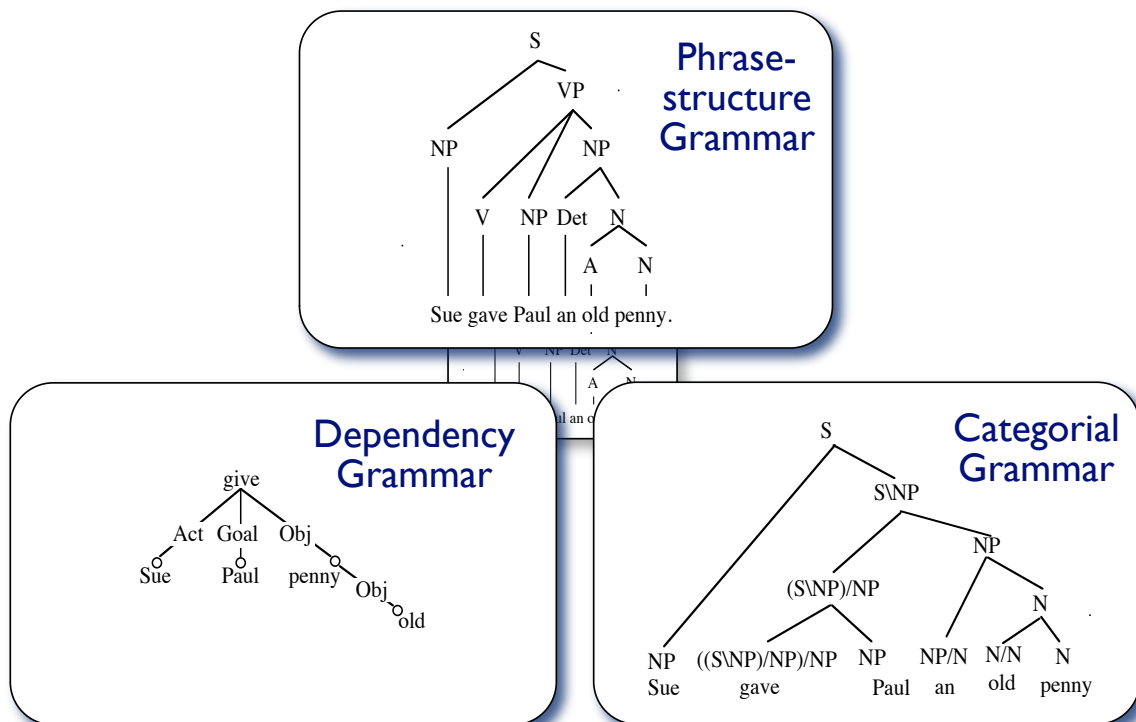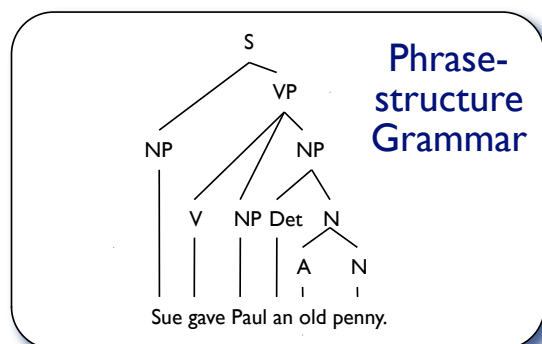
S → NP VP

A grammar
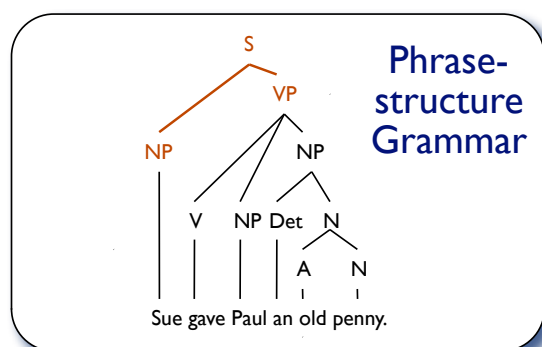
Phrase-structure Grammar

S → NP VP

VP → V NP NP
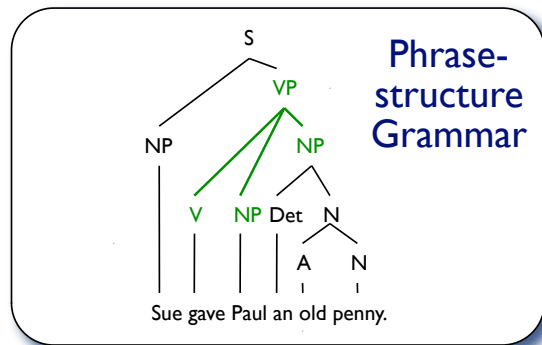
A grammar

Phrase-structure Grammar
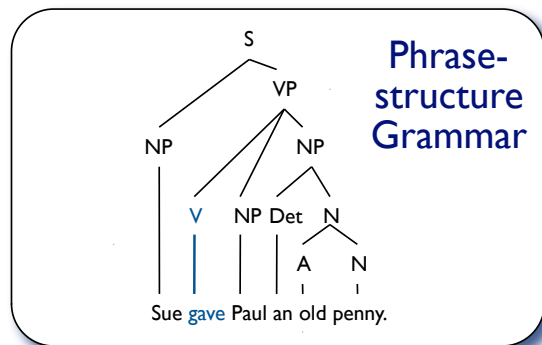
S → NP VP

VP → V NP NP

V → gave

Transformation Grammar
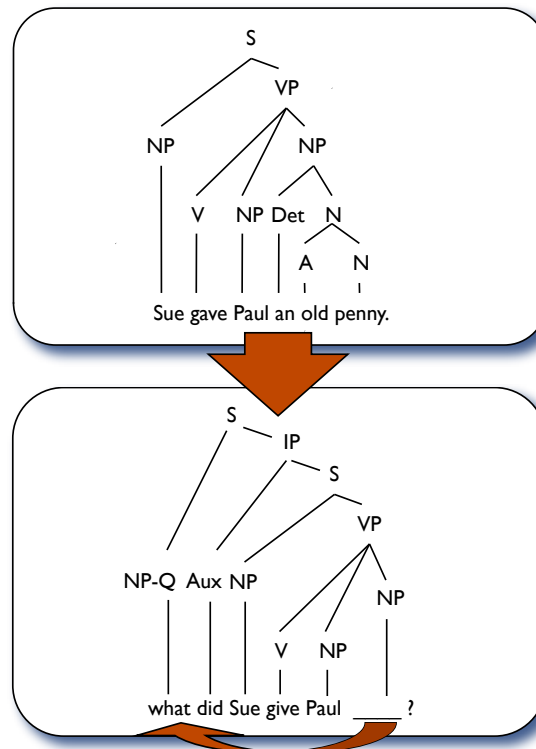
Unification Grammar
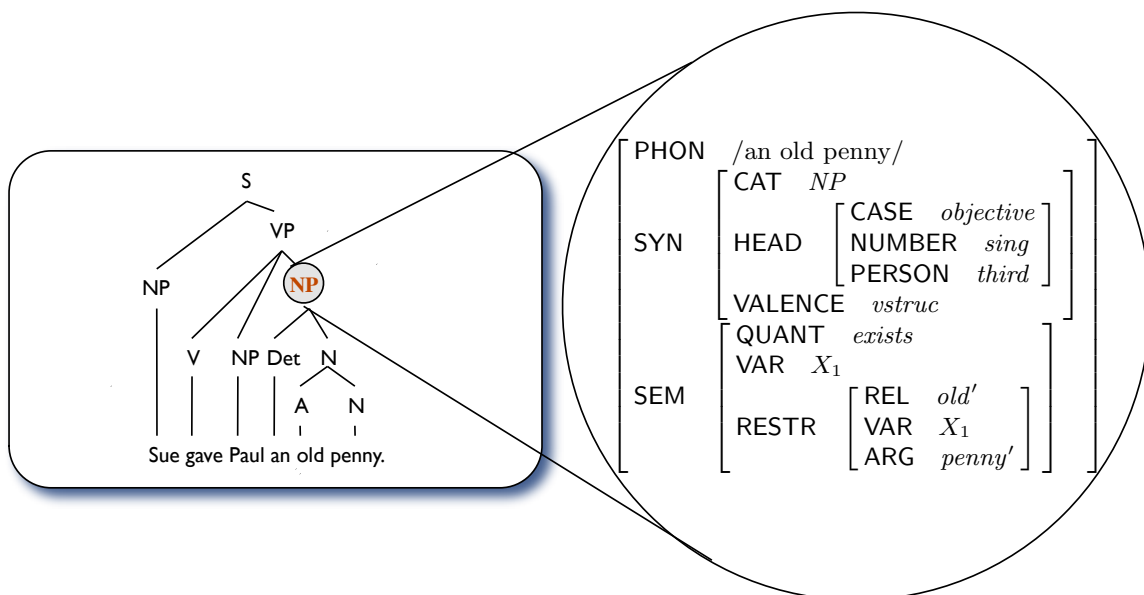
Grammar Size

- Estimates for lexicon size (English):
  - Shakespeare: 29.000 word forms, 25.000 head words
  - common estimates of the vocabulary of a college graduate: 20.000 words active / 25.000 passive
  - a million-word corpus of American English: about 38.000 distinct head words
  - total size of English vocabulary: 1 million general vocabulary, plus 1 million scientific and technical terms.

Grammar Size

- English Resource Grammar (LinGO): unification grammar with 60% coverage of newspaper text
  - 100.000 lines of code
  - average feature structure > 300 nodes

- Language has structure, but it's really complex!

Lots of languages

(Wikipedia)
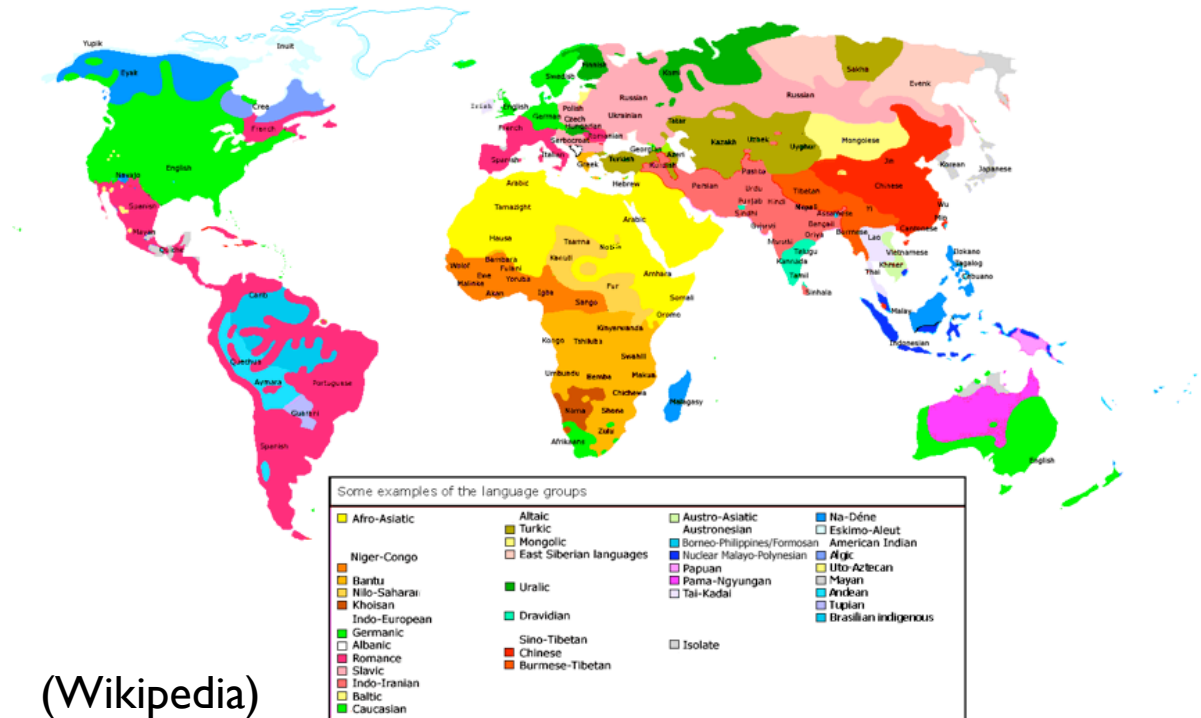
How many languages?

- According to Ethnologue: 6.912.

- 230 in Europe, 2.197 in Asia
  (of which 832 in Papua New Guinea)

- Bible translations exist for 2.200 languages.

- 250 language families (such as the Indoeuropean language family).

The Tower of Babel
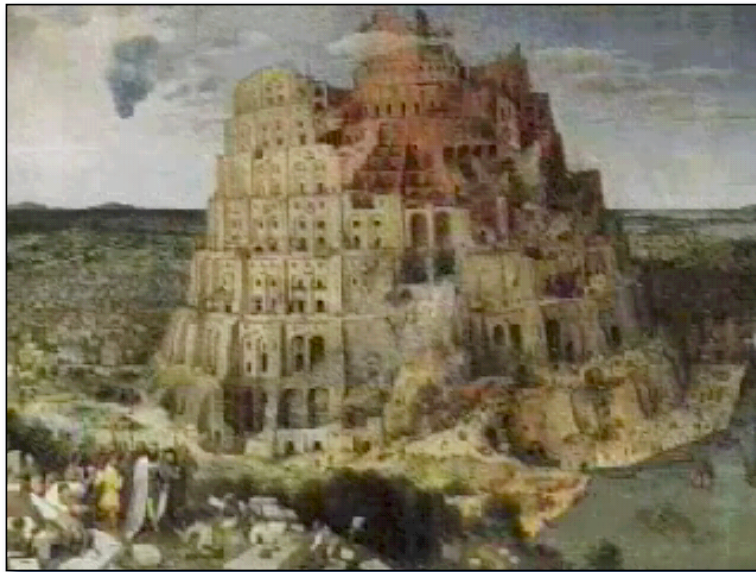
Interdisciplinary Landscape

Motivations

models of grammar

language technology applications

models of human language processing
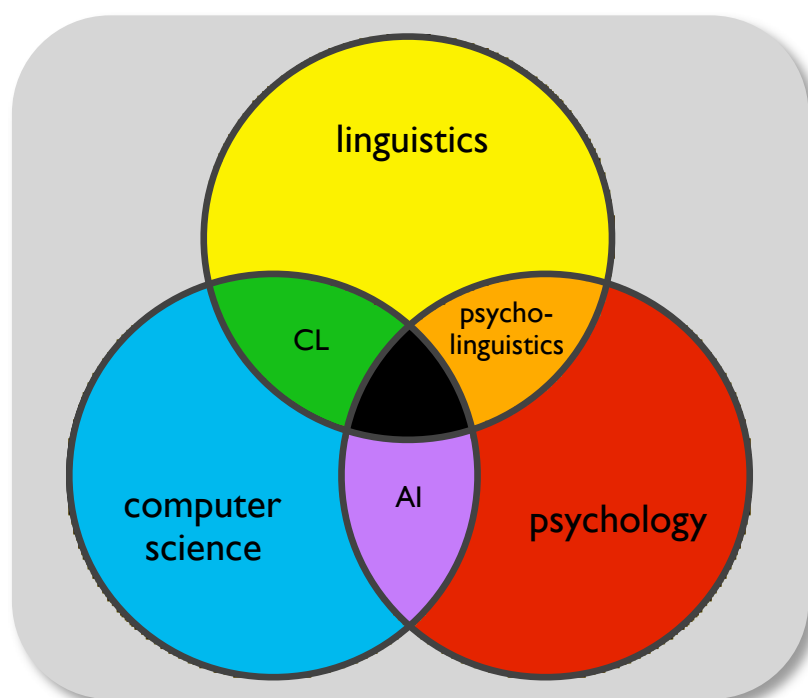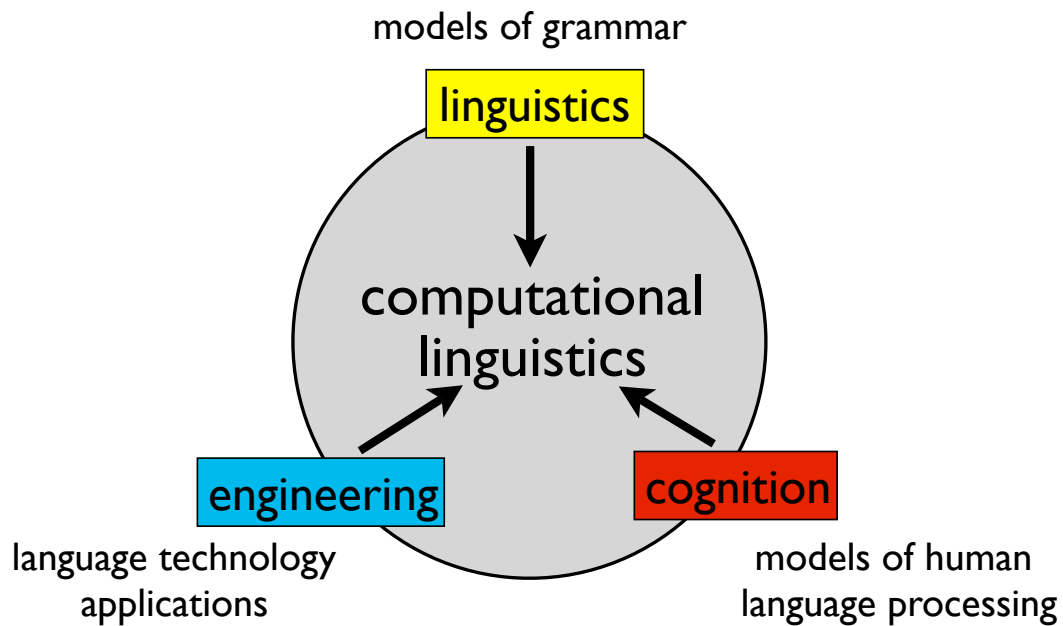
Central Questions

- Linguistic knowledge: What are the contents and structures of this knowledge?

- Language processing: How do we produce and comprehend utterances?

- Language acquisition: How does a child learn their mother tongue?

- Language change: How do languages (dialects, etc.) emerge, change, evolve?

Areas of linguistics

- According to levels of linguistic description:
  - ‣ phonetics
  - ‣ phonology
  - ‣ morphology
  - ‣ syntax
  - ‣ semantics
  - ‣ discourse
  - ‣ pragmatics

http://www.give-challenge.org/

Areas of linguistics

- According to aspects of human language:
  - ‣ Psycholinguistics
  - ‣ Neurolinguistics
  - ‣ Historical Linguistics
  - ‣ Sociolinguistics
  - ‣ Ethnolinguistics
  - ‣ Dialectology
  - ‣ Applied Linguistics
  - ‣ Mathematical Linguistics

Derek Bentley (1952)

Computational Linguistics

- Levels:
  - comp. phonetics
  - comp. phonology
  - comp. morphology
  - comp. syntax
  - comp. semantics
  - comp. pragmatics

- Aspects:
  - comp. psycholinguistics
  - comp. neurolinguistics
  - comp. historical linguistics
  - comp. dialectology
  - comp. applied linguistics / applied comp. linguistics
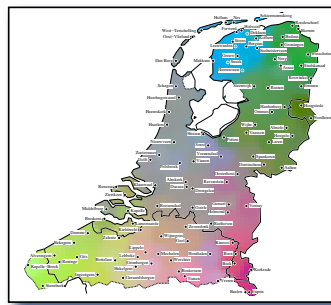
Nerbonne et al., 1999

Ambiguity 1

- phonetic ambiguities (homophony):
  their/there   tow/toe

- orthographic ambiguities (homography):
  read/read

- lexical ambiguities:
  bank/bank   ball/ball

- morphological ambiguities:
  undoable/undoable

Ambiguity 2

- Syntactic ambiguity:
  - (With the naked eye she couldn't see much.)
    So she watched the man with a telescope.
  - (She couldn't watch all suspects.)
    So she watched the man with a telescope.

- Semantic ambiguity:
  - The three selected special agents speak two foreign languages nearly accent-free. (Namely French & Russian.)
  - The three selected special agents speak two foreign languages nearly accent-free. (But only two of them master Russian.)

Ambiguity 3

- Pragmatic ambiguity:
  - Could you translate this text?
    (I need it tomorrow.)
  - Could you translate this text?
    (I doubt that anyone could do it.)

- Discourse ambiguity (here, anaphoric reference):
  - John sued Peter because *he* had thrown *him* down the stairs.

Preferences

- Certain readings are preferred over others:
  - Where is a bank?
  - Do you like plants?

- Preferences can be influenced by context:
  - "The mayor opened the ball" vs.
    "The goalkeeper opened the ball."
  - "The movie director married a star" vs.
    "The astronomer married a star."

Uszkoreit's Island Ambiguity

„Früher   stellten   die Frauen der Inseln am Wochenende Kopftücher mit
 in the past produced the women of the islands on the weekends   scarves   with
Blumenmotiven her, die ihre Männer an den folgenden Montagen auf dem
floral patterns   that their husbands on the following   Mondays   on   the
Markt im   Zentrum der   Hauptinsel verkauften."
market in the center   of the main island sold.   (Hans Uszkoreit)

The sentence exhibits a total of 13 lexical, syntactic, and referential ambiguities.

$2 \times 2 \times 2 \times 3 \times 3 \times 2 \times 4 \times 2 \times 4 \times 2 \times 2 \times 7 \times 2$
  = **258,048 readings**

Summary

- Miracle of language: How can sound waves move concepts from my head to yours?

- Levels of linguistic description.

- Computational linguistics and interdisciplinary perspectives.

- Ambiguity.

Homework until Friday

- Find or construct a highly ambiguous sentence in your native language.

- Present and explain it in class on Friday.

- Hand in sentence and brief explanation to me on Friday.

- You should manage at least six readings.