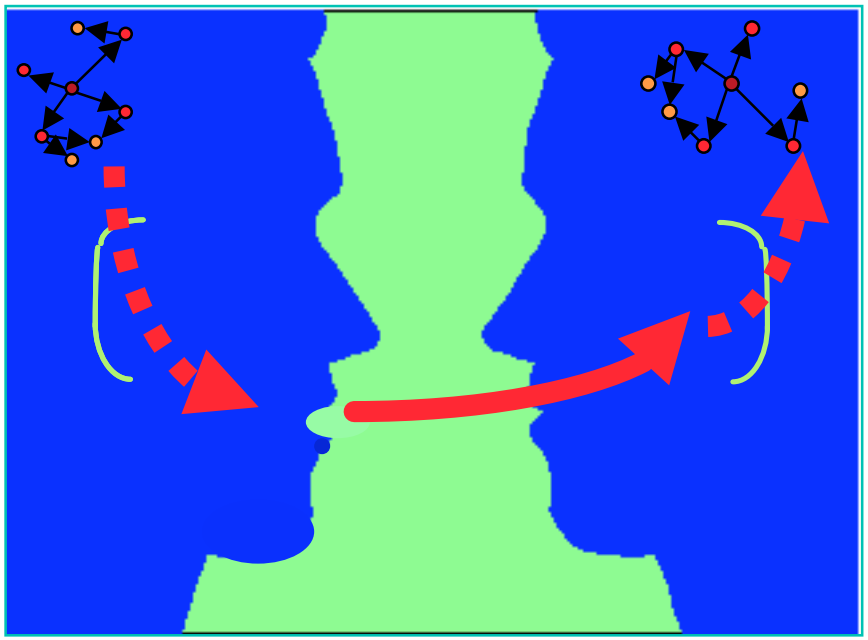
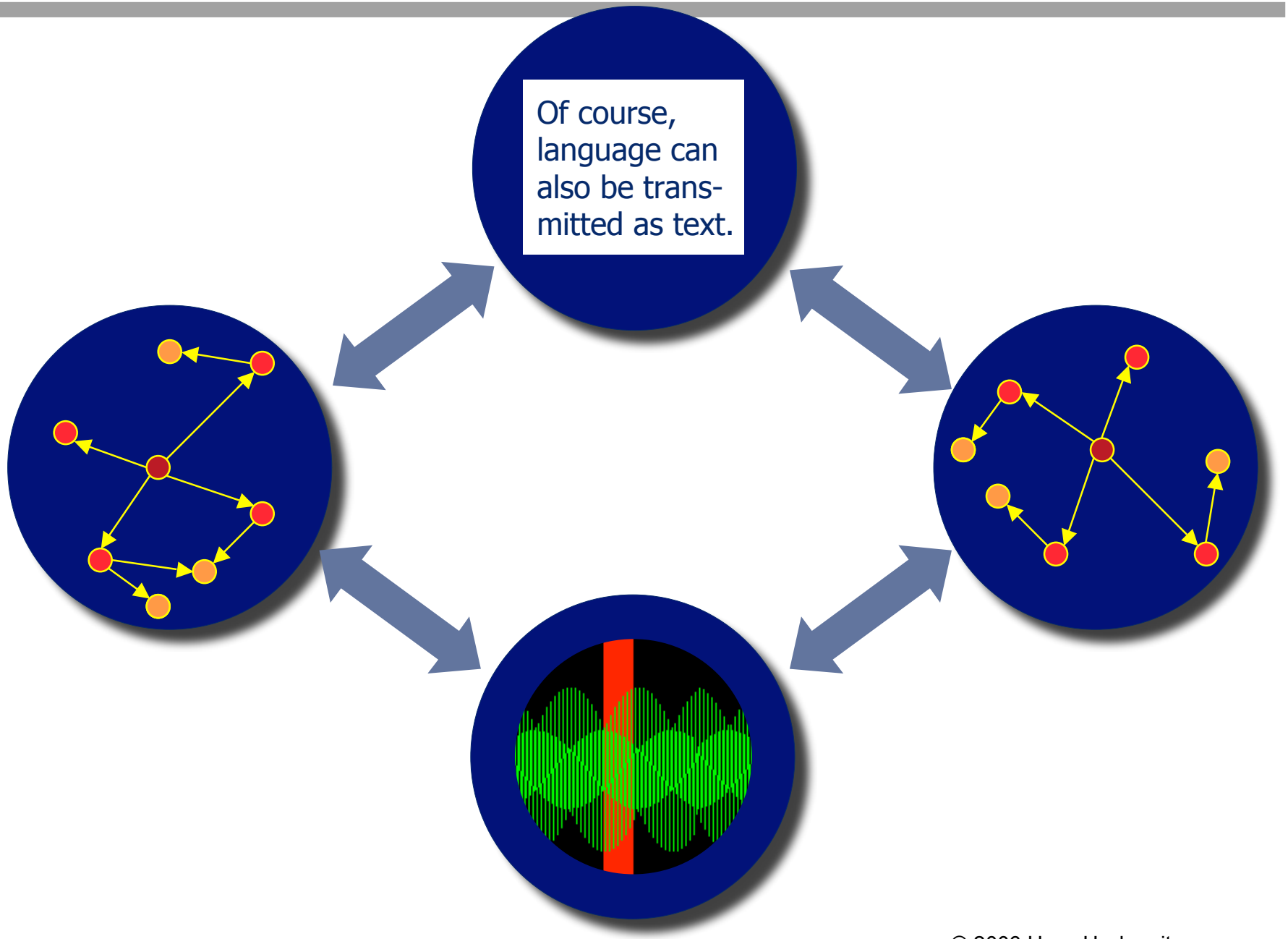
FOUNDATIONS OF LANGUAGE SCIENCE AND TECHNOLOGY

Introduction

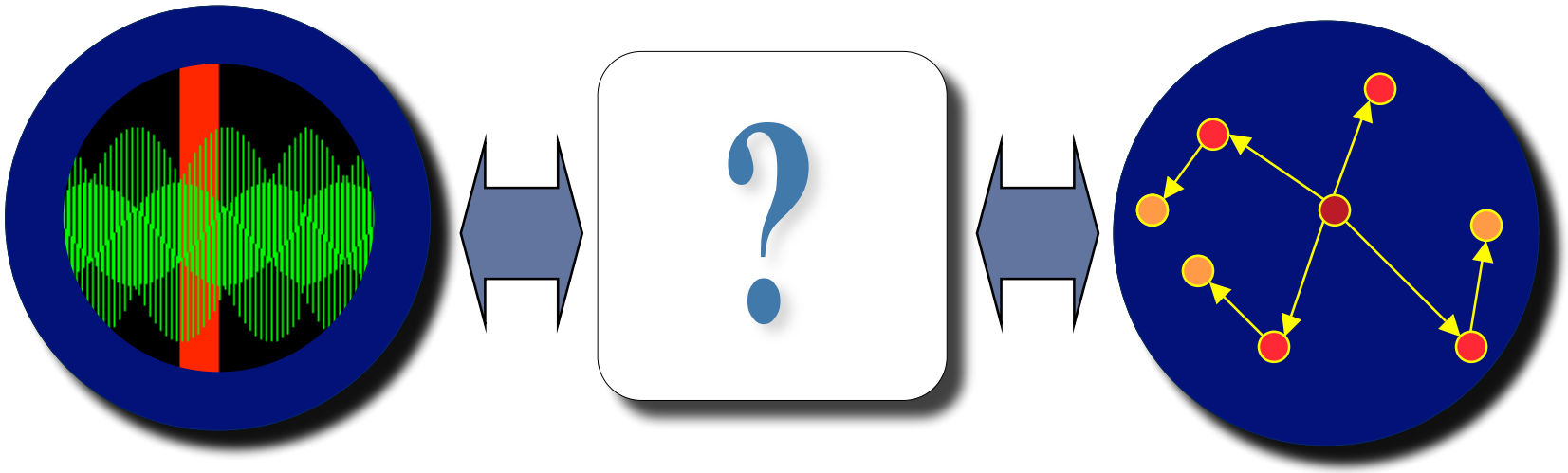
THE MIRACLE



Language is the Medium



WHAT HAPPENS IN BETWEEN?

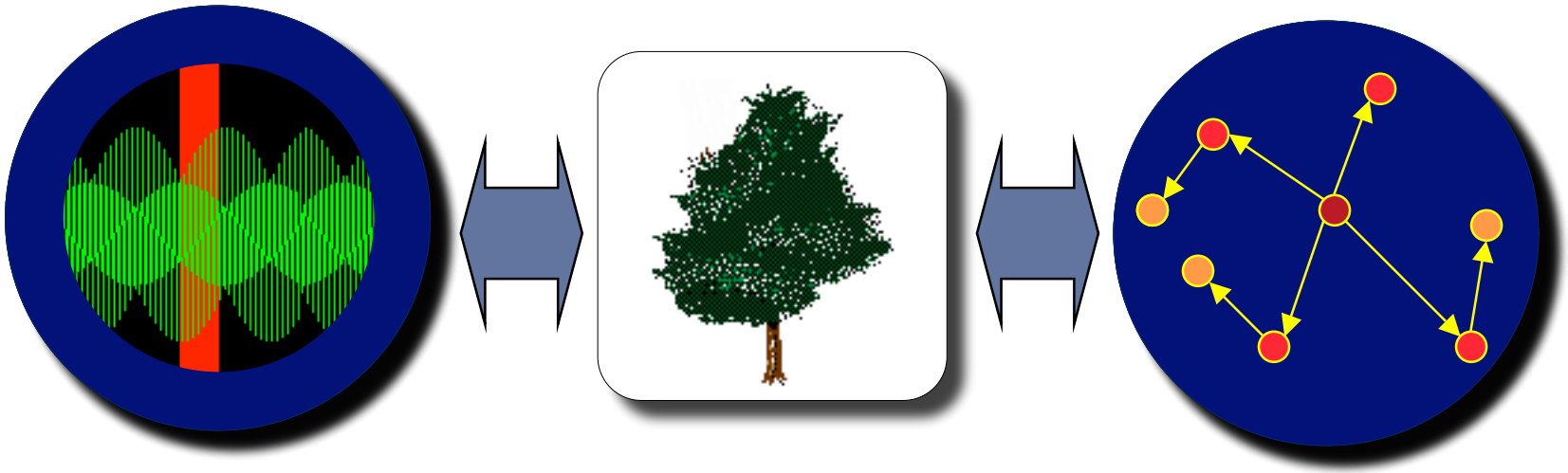


sound waves

Grammar

activation of concepts

WHAT HAPPENS IN BETWEEN?

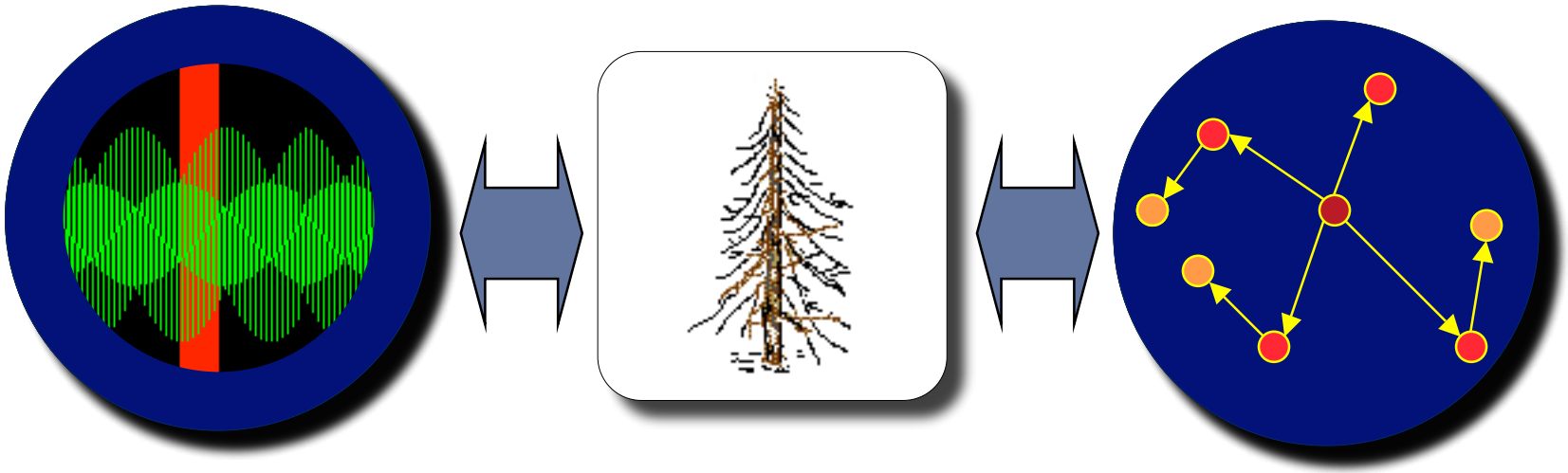


sound waves

Grammar

activation of concepts

WHAT HAPPENS IN BETWEEN?

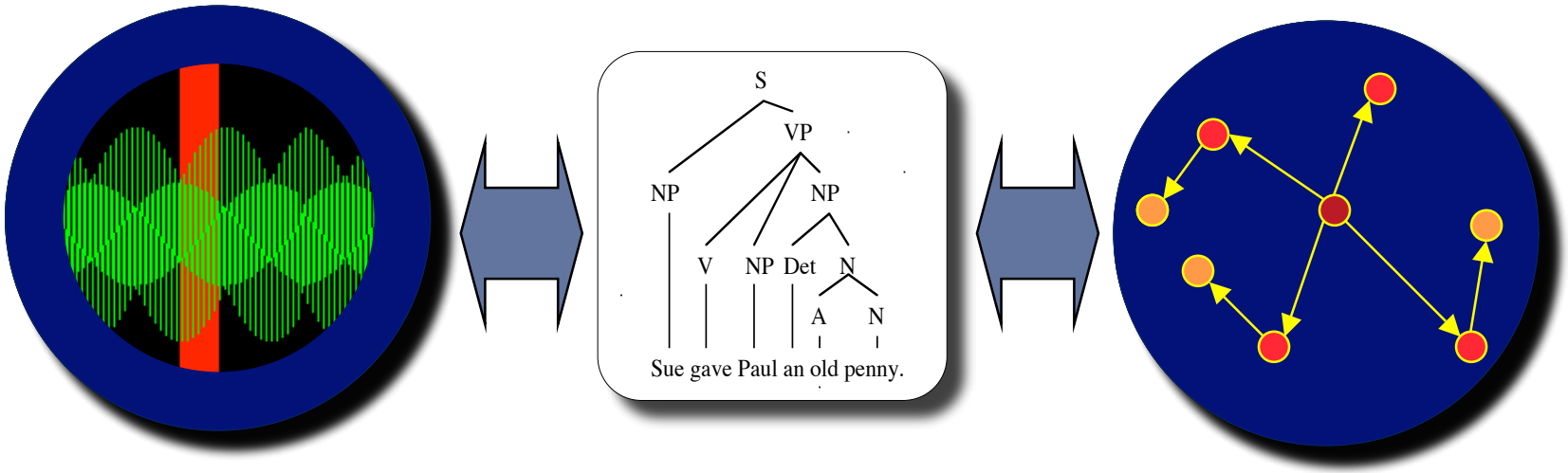


sound waves

Grammar

activation of concepts

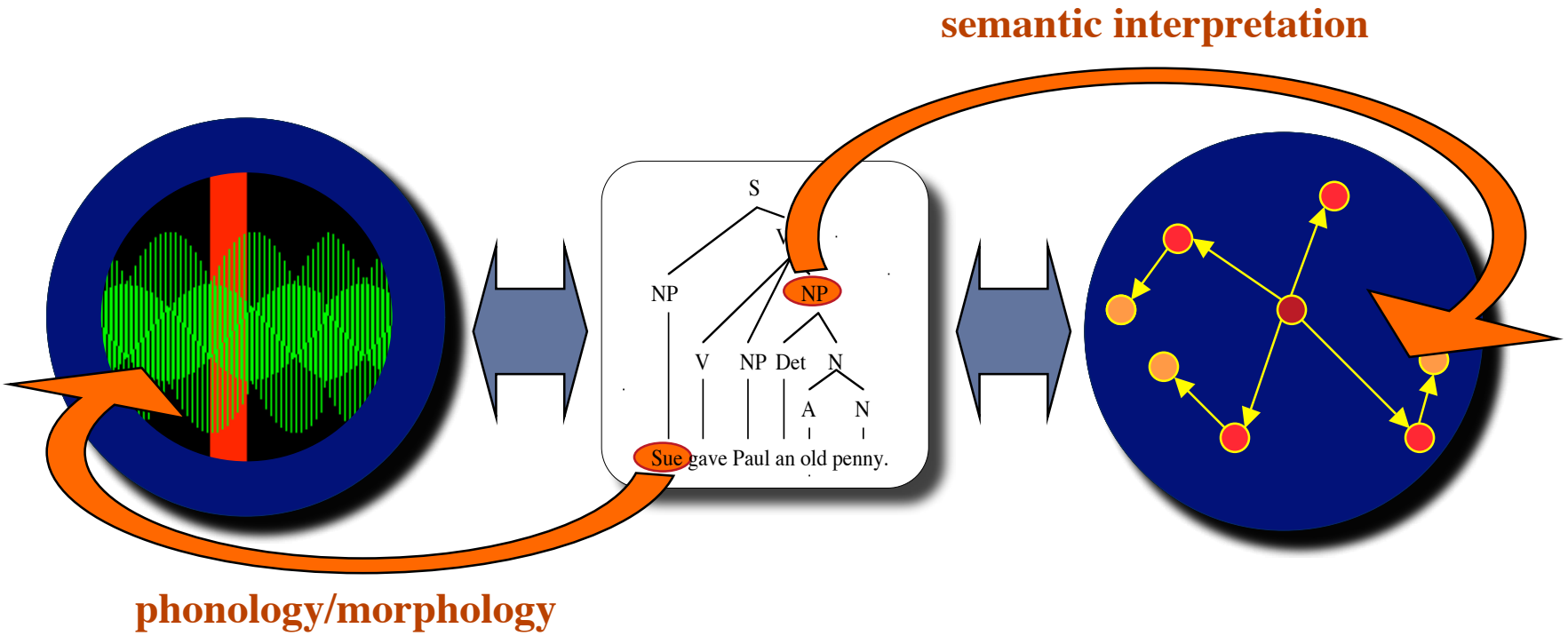
WHAT HAPPENS IN BETWEEN?

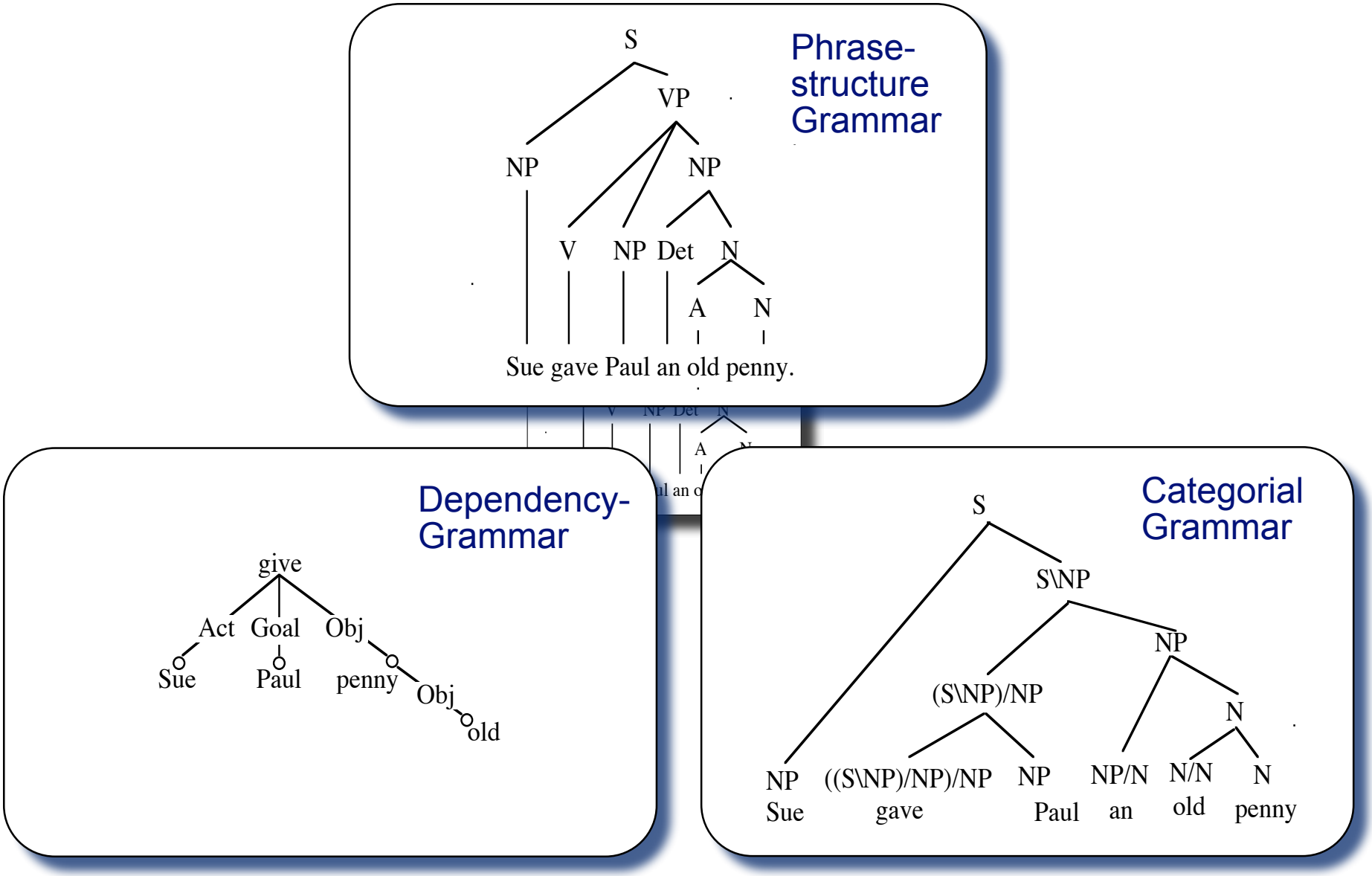


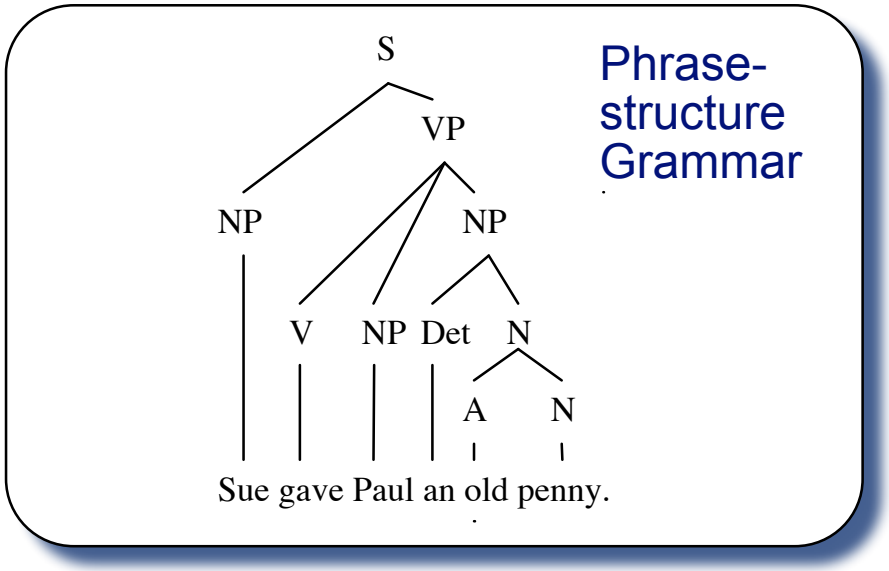
sound waves

Grammar

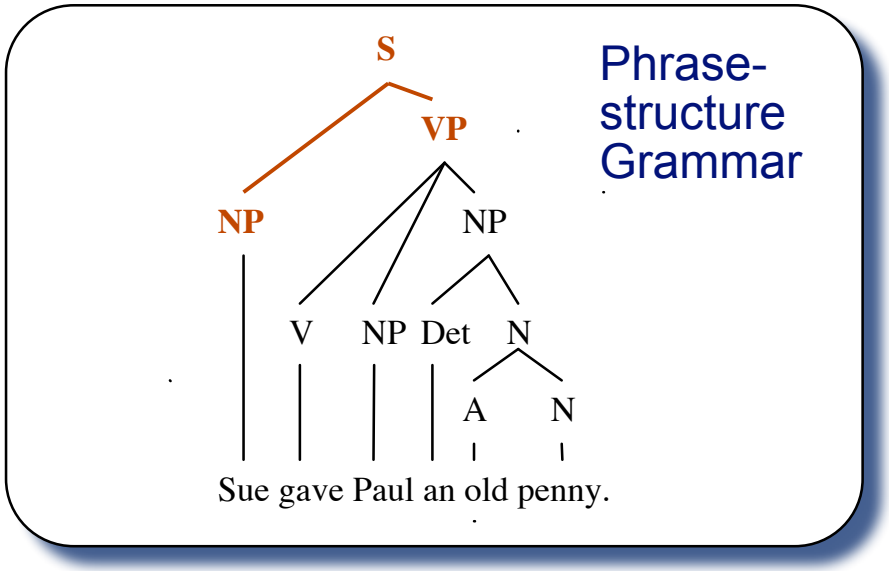
activation of concepts



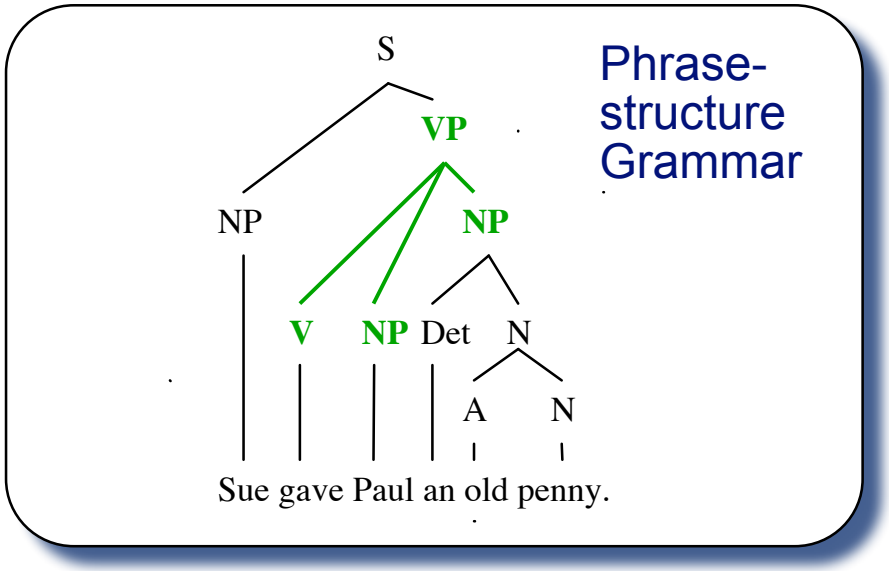




S → NP VP

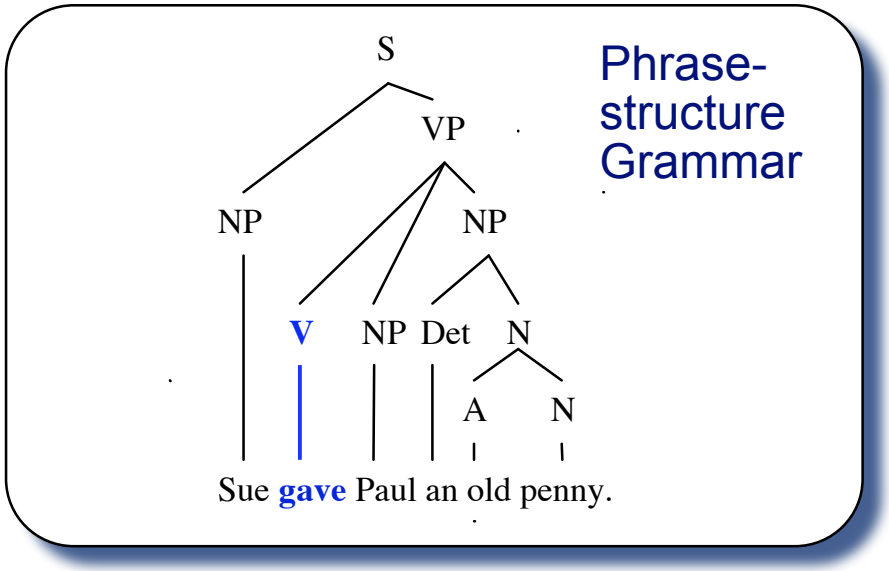


S → NP VP



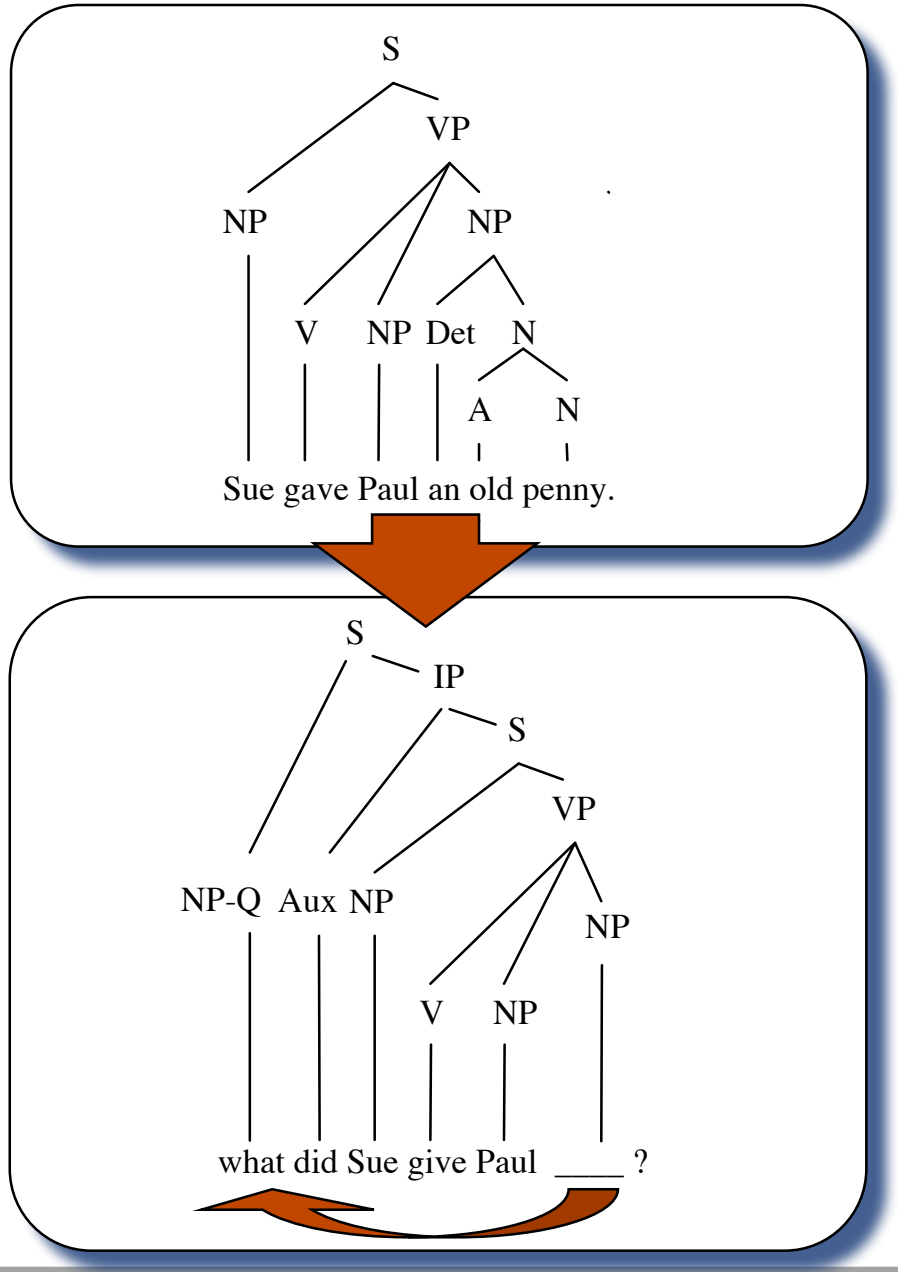
S → NP VP

VP → V NP NP



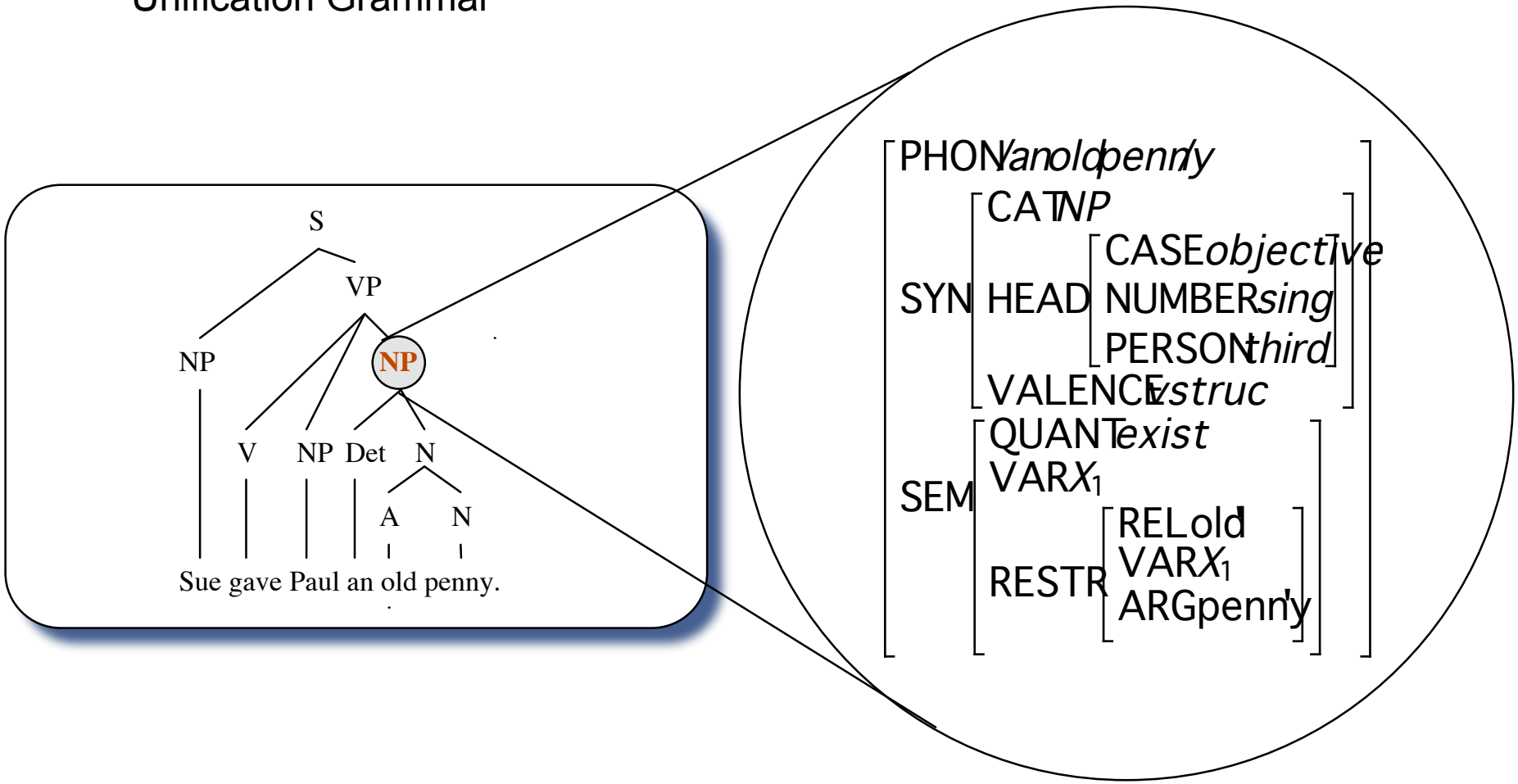
S → NP VP
VP → V NP NP

V → gave



Transformation Grammar

Unification Grammar



Size

- How large is the grammar.
- Let's start with the lexicon.

Estimates for English

- Shakespeare actively used 29.000 word forms mapping to about 25.000 head words

- common estimates of the vocabluary of a college graduate:
  20.000 words active -- 25.000 words passive

- David Crystal's estimate
  60.000 words active -- 75.000 words passive

- Total Size of English Vocabulary

  1 million words without special scientific and technical terms
  2 million words including all scientific and technical terms

A million-word-corpus of American English exhibits about 38.000 head words.

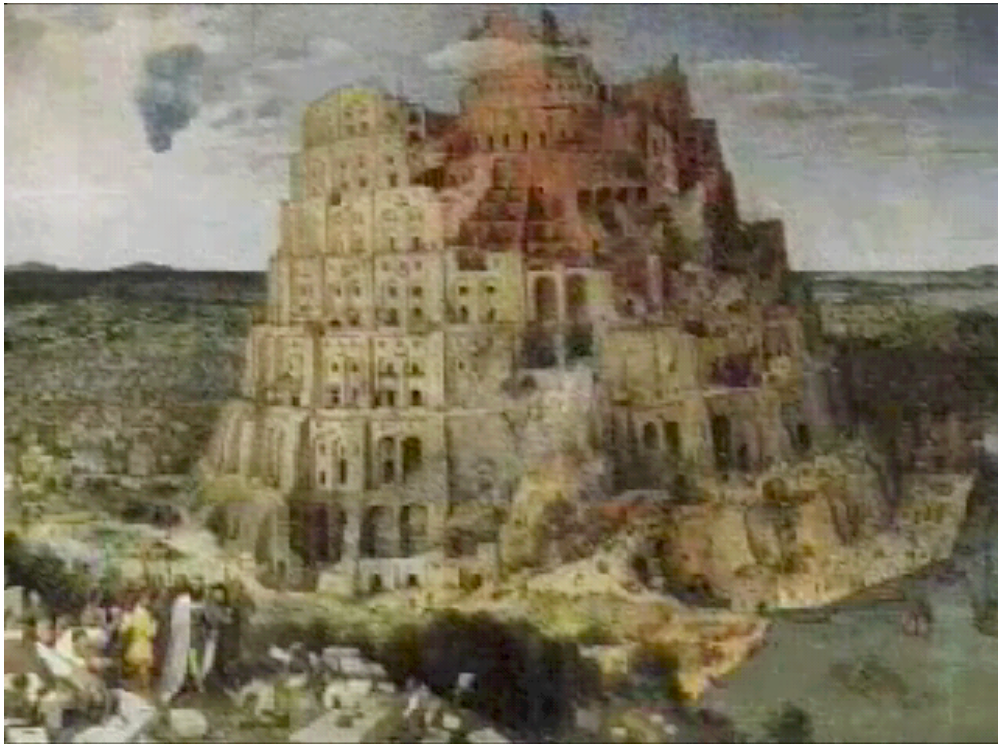
Size of a Grammar

- LinGO - English Resource Grammar

  (60% coverage of newspaper texts)

  - 8.000 types
  - 100.000 lines of code
  - average feature structure > 300 nodes

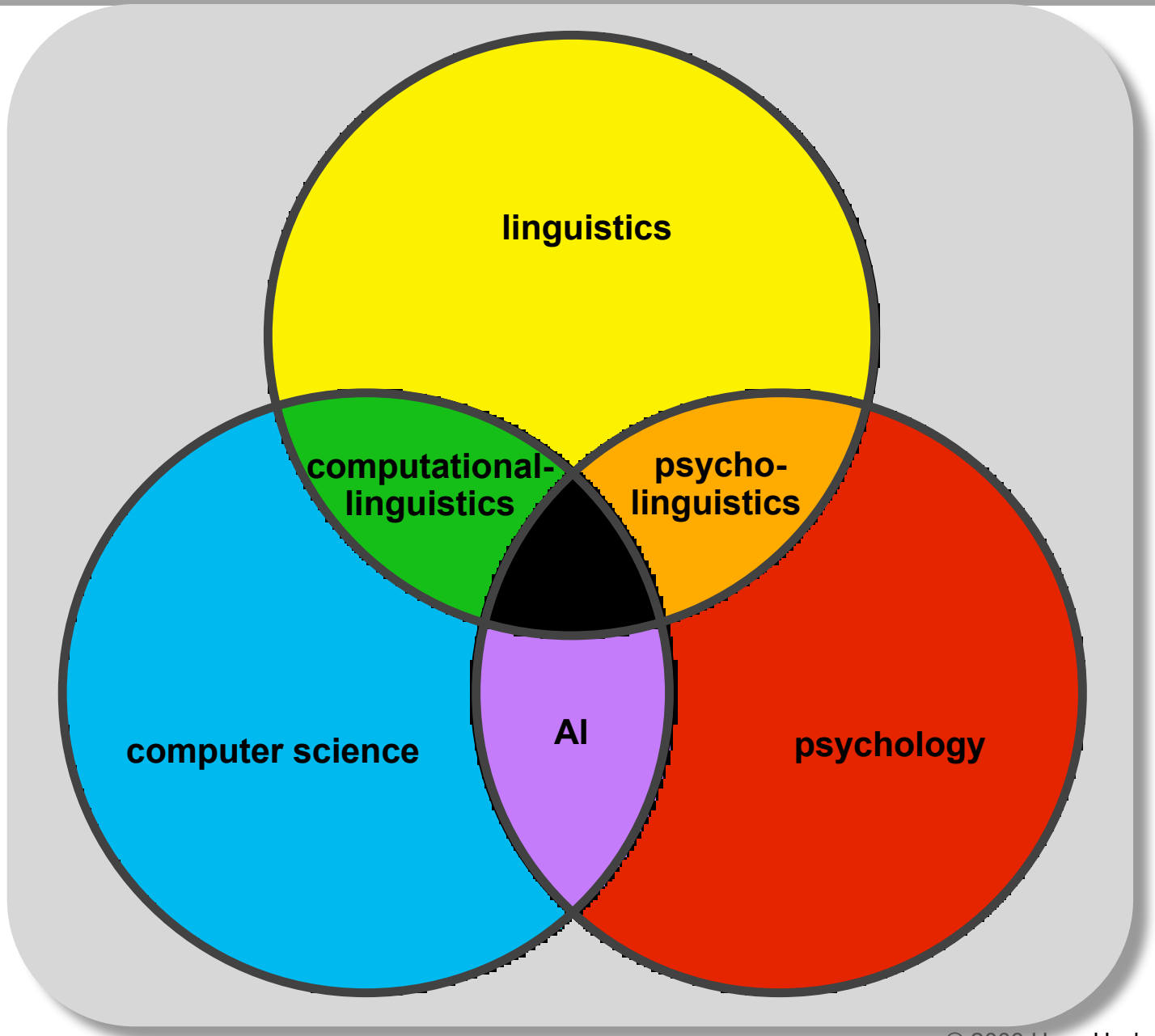
The Tower of Babel



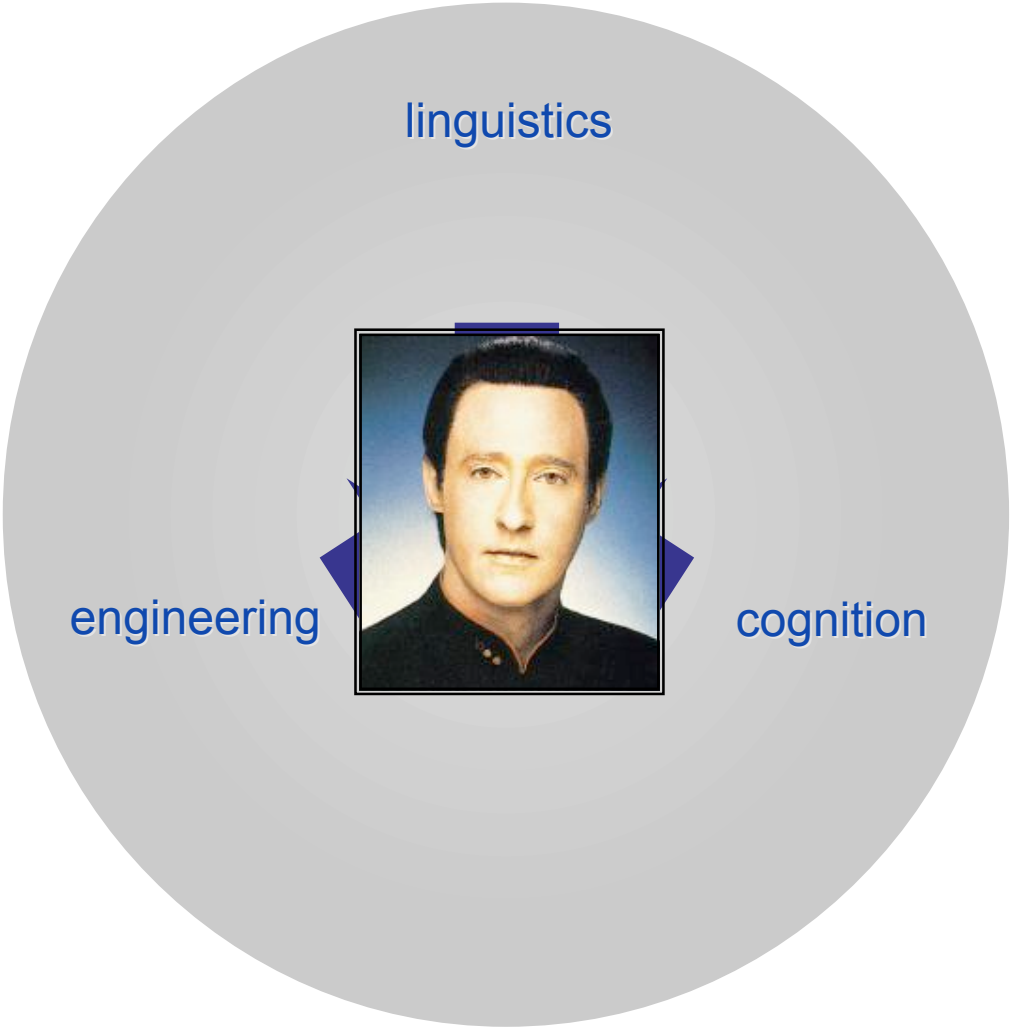
How Many Languages ?

- According to Ethnologue 6,809 languages
- 230 in Europe, 2197 in Asia (832 in Papua-New Guinea)
- Bible translations exist for 2.200 languages
- 250 families of languages (such as Indoeuropean Languages)

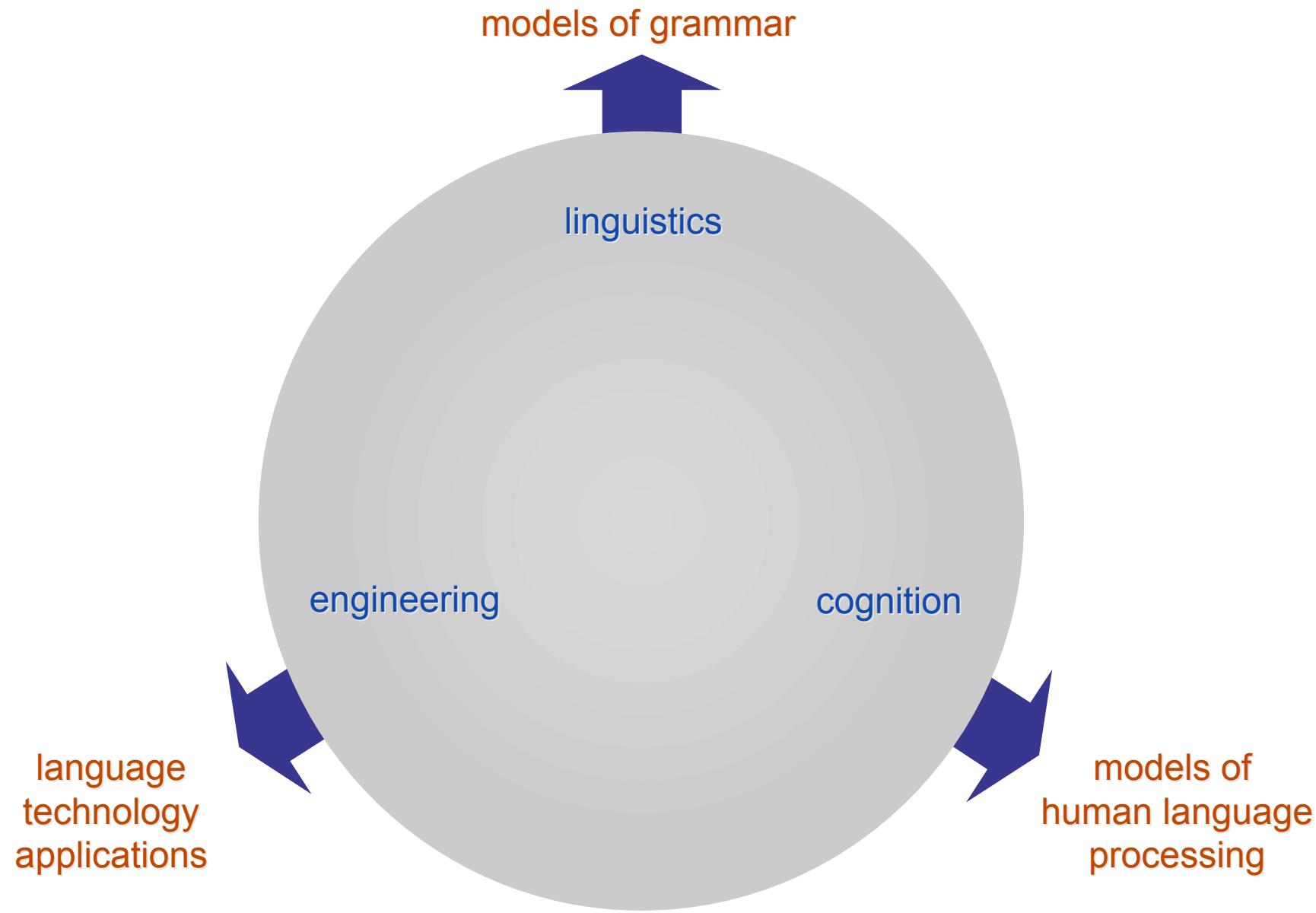
Transdisciplinary Interests



Motivations



Motivationen



Central Questions of Language Research

- LINGUISTIC KNOWLEDGE

  What are the contents and structures of this knowledge

- LANGUAGE PROCESSING

  How do we produce and comprehend linguistic utterances?

- LANGUAGE ACQUISITION

  How does the child learn his mother tongue?

- LANGUAGE CHANGE

  How do languages (dialects, sociolects) emerge, change, evolve?

Areas of Linguistics

- According to levels of linguistic description
  - Phonetics
  - Phonology
  - Morphology
  - Syntax
  - Semantics
  - Pragmatics/Discourse
- According to aspects of human language
  - Psycholinguistics
  - Neurolinguistics
  - Historical Linguistics
  - Sociolinguistics
  - Ethnolinguistics
  - Dialectology
  - Applied Linguistics
  - Mathematical Linguistics
  - Computational Linguistics

Combinations

- According to levels of linguistic description
  - Computational Phonetics
  - Computational Phonology
  - Computational Computational Morphology
  - Computational Syntax
  - Computational Semantics
  - Computational Pragmatics

- According to aspects of human language
  - Computational Psycholinguistics
  - Computational Neurolinguistics
  - Computational Historical Linguistics
  - Computational Sociolinguistics
  - Computational Ethnolinguistics ???
  - Computational Dialectology
  - Computational Applied Linguistics / Applied Computational Linguistics
  - Computational Mathematical Linguistics (funny)

Levels of Description

acoustic form

written form

phonetic or graphemic representation

morpho-phonological processing

syntactic representation

semantic representation

representation of the full meaning

Levels of Processing

acoustic form **written form**

phonetic processing | orthographic processing

phonetic or graphemic representation

morpho-phonological processing

morpho-phonological processing

syntactic processing

syntactic representation

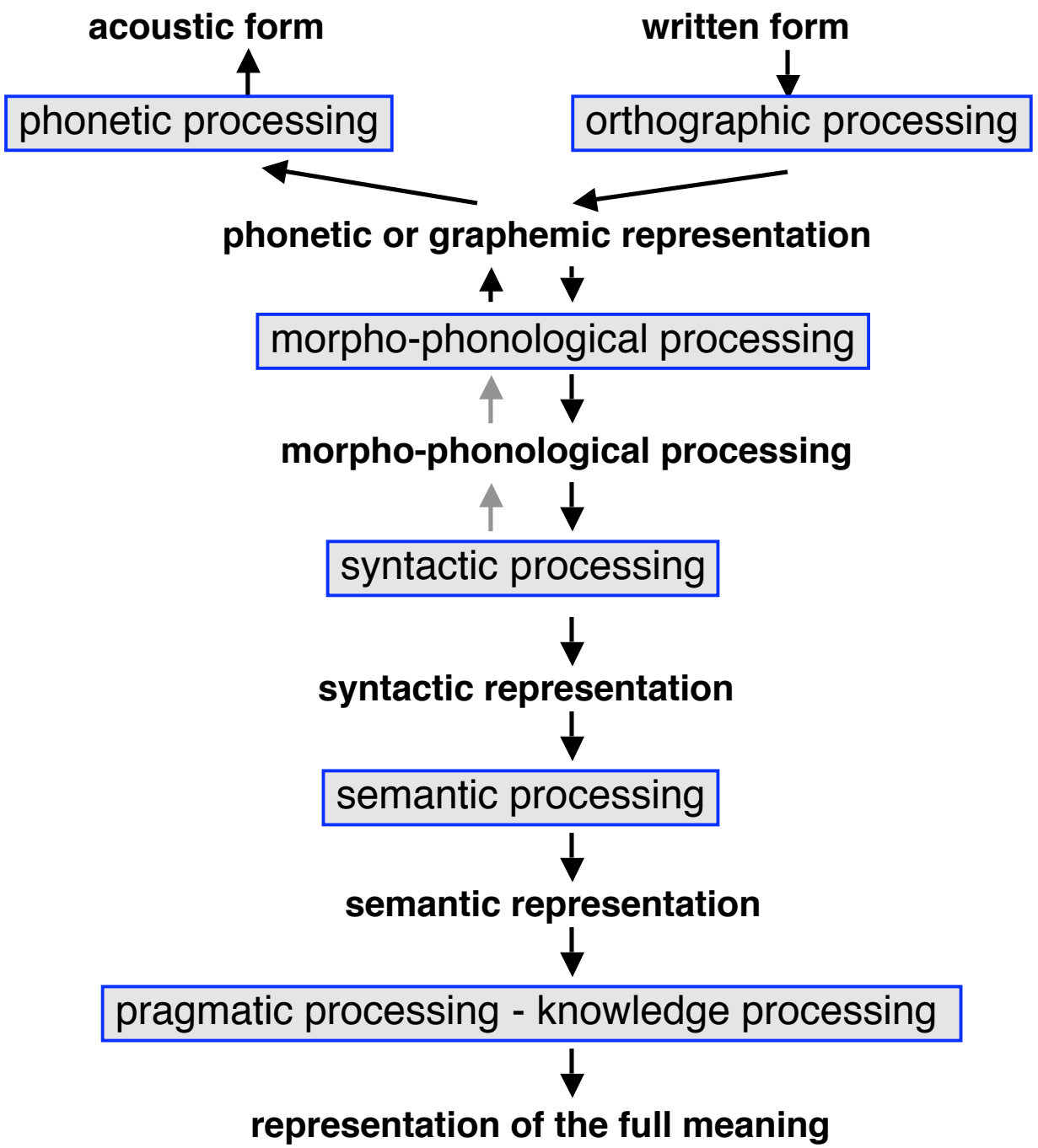
semantic processing

semantic representation

pragmatic processing - knowledge processing

representation of the full meaning

Text-to-Speech System



Why do we need deep processing for simple text-to-speech conversion

(I) The student will read the paper. (/riːd/)

(2) The students have read the paper. (/rɛd/)

(3) Will the students read the paper? (/riːd/)

(4) Have any citizens of good will read the paper? (/rɛd/)

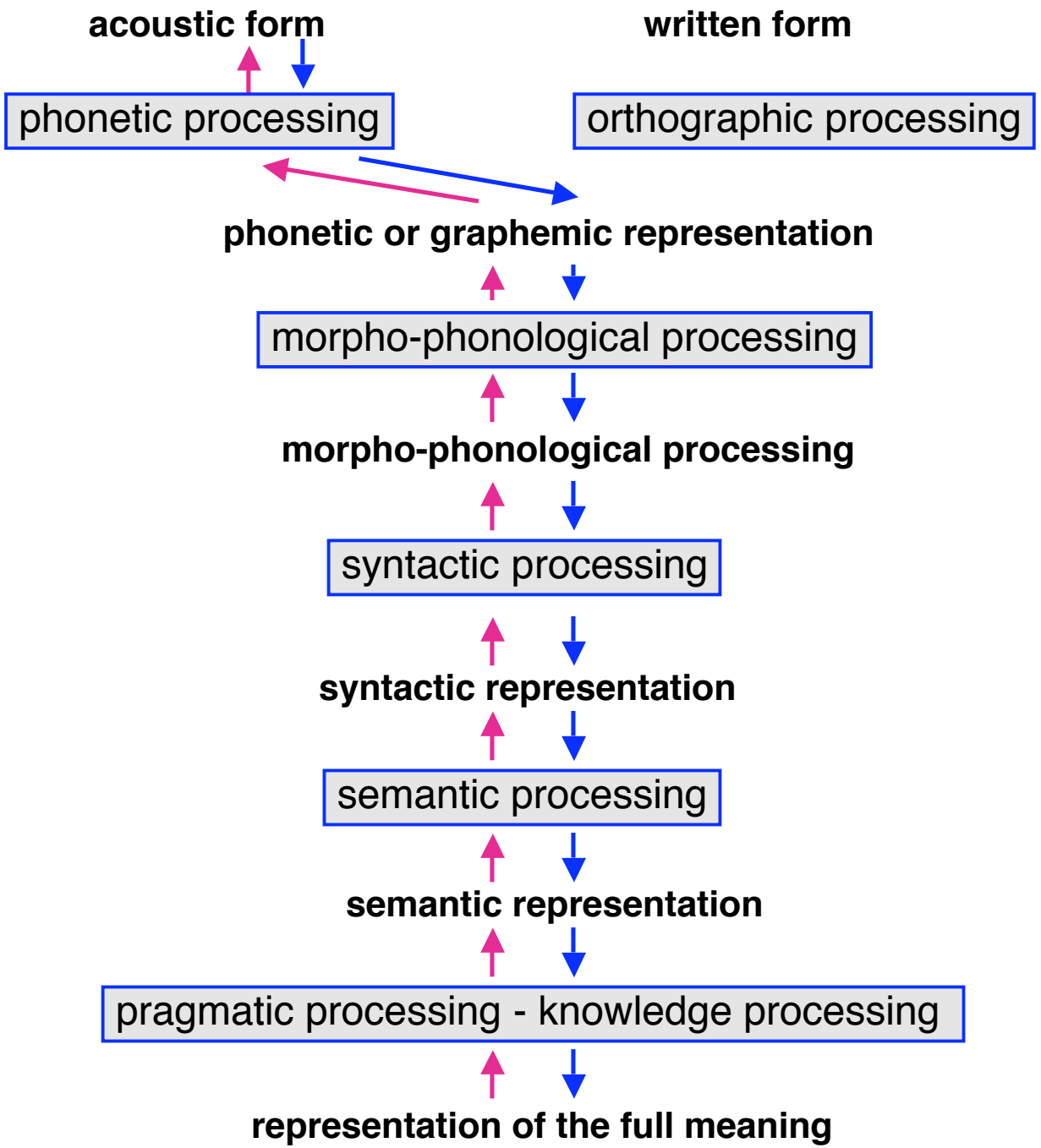
(5) Have the executors of the will read the paper? (/rɛd/)

(6) Have the students who will arrive next week read the paper yet? (/rɛd/)

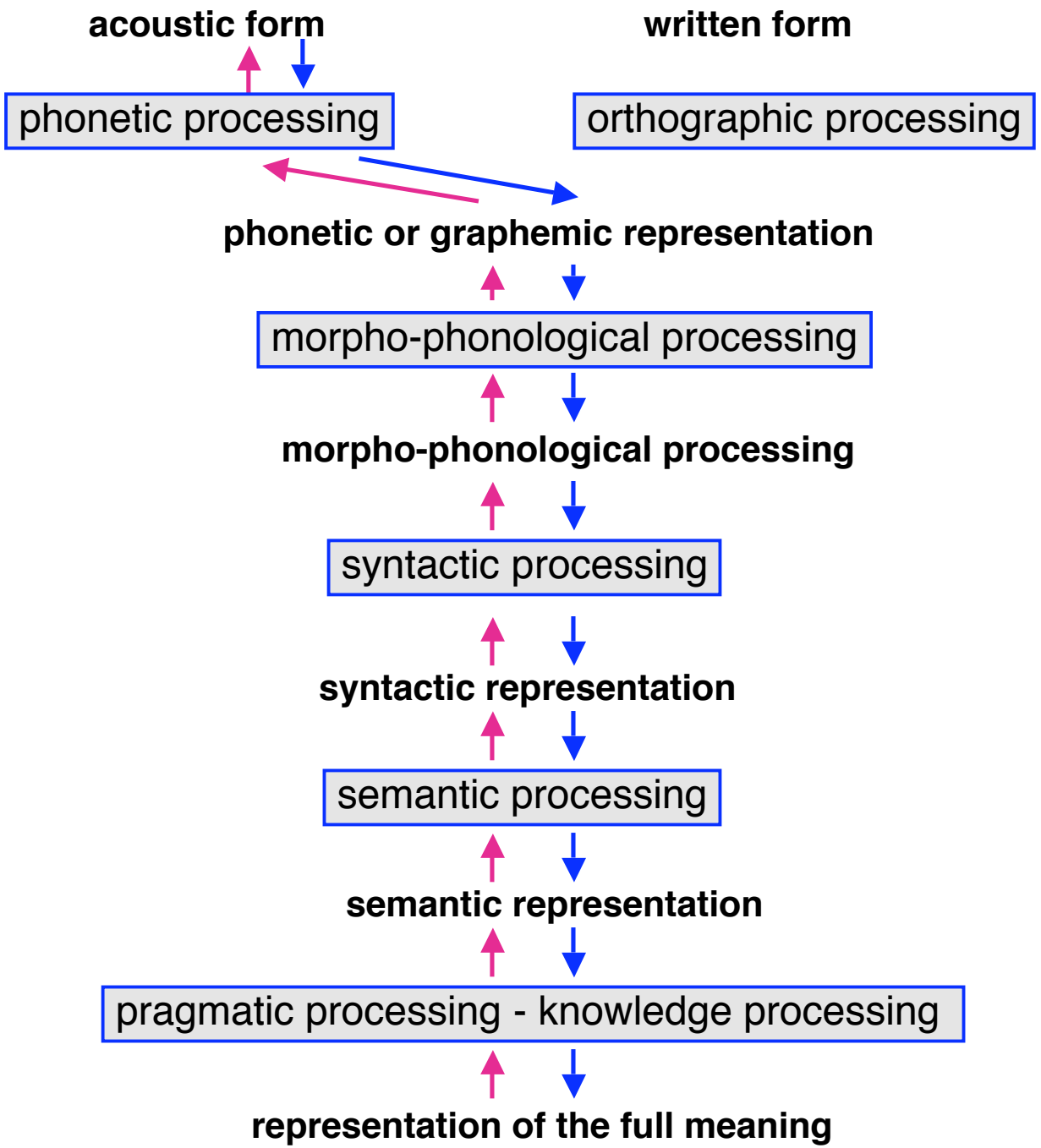
(7) Please have the students read the paper. (/riːd/)

(8) Have the students read the paper? (/rɛd/)

Speech Translation



Speech Translation



phonetic (homophony):

their	there
toe	tow

orthographic (homography):

read	read
undoable	undoable

lexical (homonymy):

bank	bank
ball	ball

syntactic

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

The three selected special agents
speak two foreign languages
nearly without an accent.
Namely French and Russian.

The three selected special agents
speak two foreign languages
nearly without an accent.
But only two of them master
Russian.

pragmatic

Could you translate this text?
I need it tomorrow.

Could you translate this text?
I even wonder if anybody could do it.

Certain readings are less preferred than others:

Where is a bank?

Do you like plants?

The preference can be influenced by context.

The goal keeper opened the ball. **vs.** ***The Mayor opened the ball.***

The astronomer married a star. **vs.** ***The movie director married a star.***

Ambiguity (a pathological example)

„Früher stellten die Frauen der Inseln am Wochenende Kopftücher mit

in the past produced the women of the islands on the weekends scarfs with

Blumenmotiven her, die ihre Männer an den folgenden Montagen auf dem

flower 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.

In the past the women of the islands produced scarfs with flower patterns on the weekends that were sold by their husbands on the following Mondays on the market in the center of the main island.

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

2 x 2 x 2 x 3 x 3 x 2 x 4 x 2 x 4 x 2 x 2 x 7 x 2 = 258,048