MA Language Sciences
Profilmodul / Ringvorlesung

Computational Linguistics

The aims and methods of computational linguistics

- The relation between Informatics and Linguistics
- Computationally ‘responsible’ linguistic approaches: what does that mean?
- Application areas

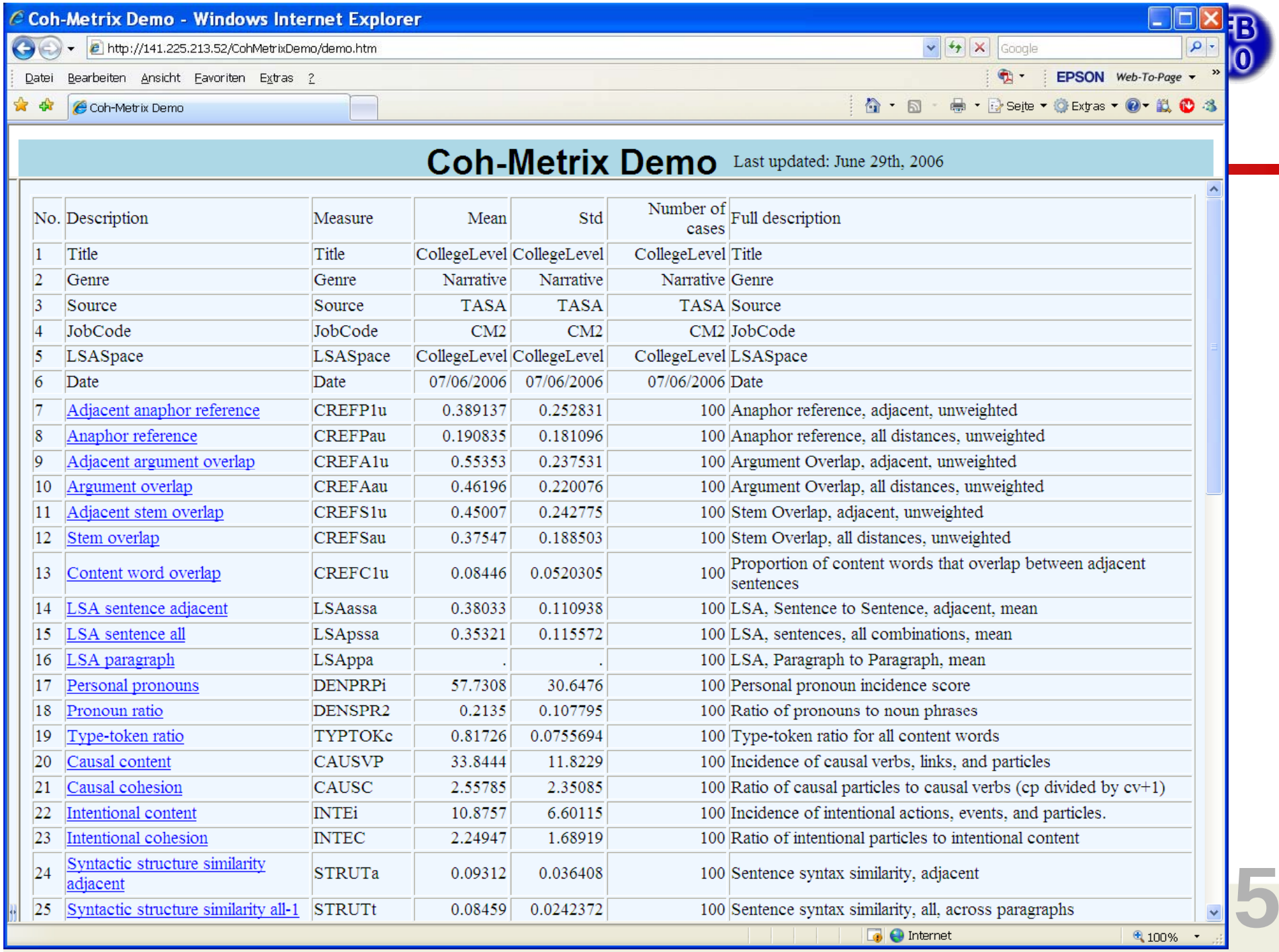
Orientation

- Examples
  - already many computational applications of ‘knowledge’ of language: i.e., where some linguistics appears in a computational context

- Starting question:
  - What kinds of linguistic knowledge are involved?

Many computational applications of ‘knowledge’ of language already

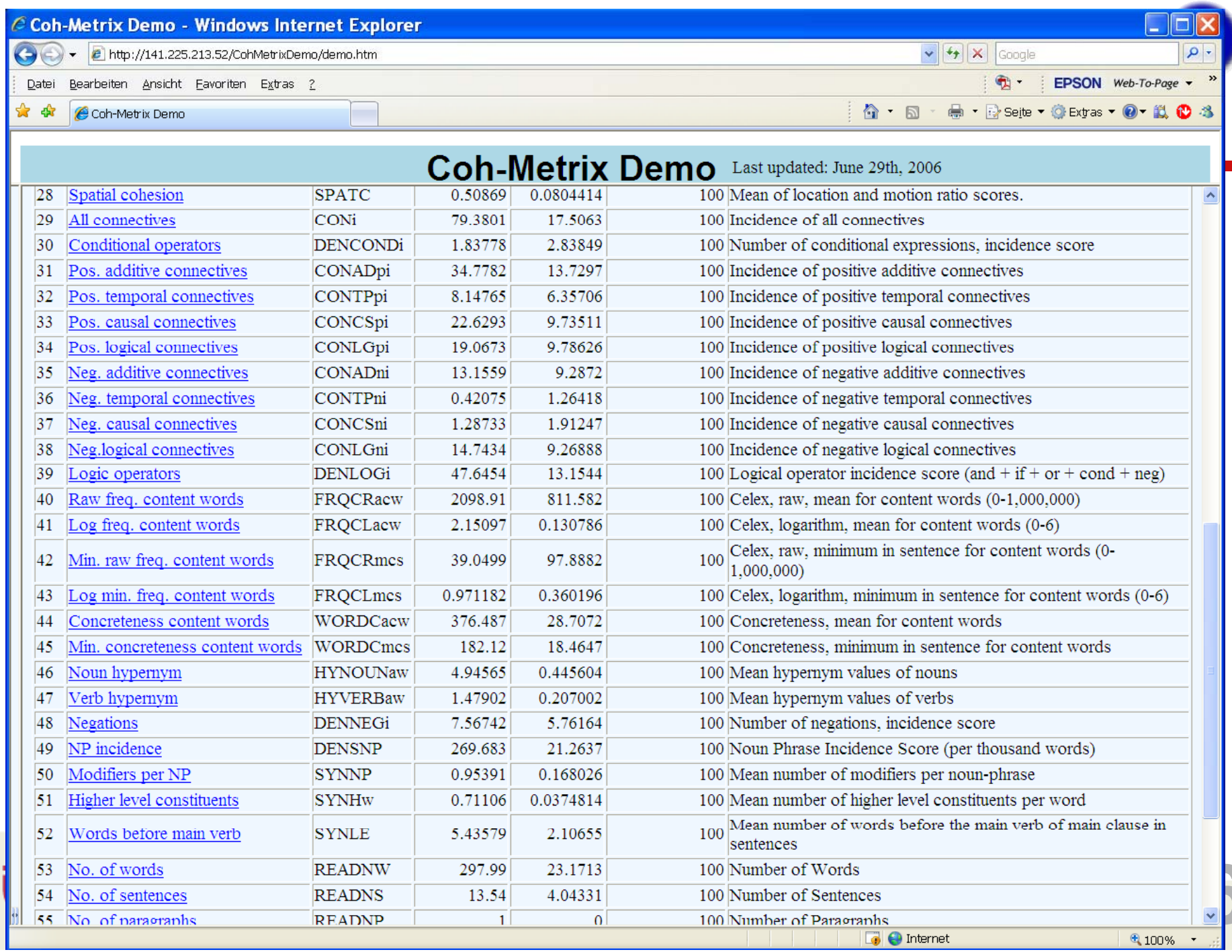
- speech recognition / speech synthesis
  - word recognition
  - announcement systems
  - telephone information systems
- indexing documents for retrieval (stemming, tagging)
- establishing various metrics for evaluating or describing texts
- dictionaries / lexical databases
- machine translation
- report generation
- automatic dialogue systems
- …



Coh-Metrix Demo

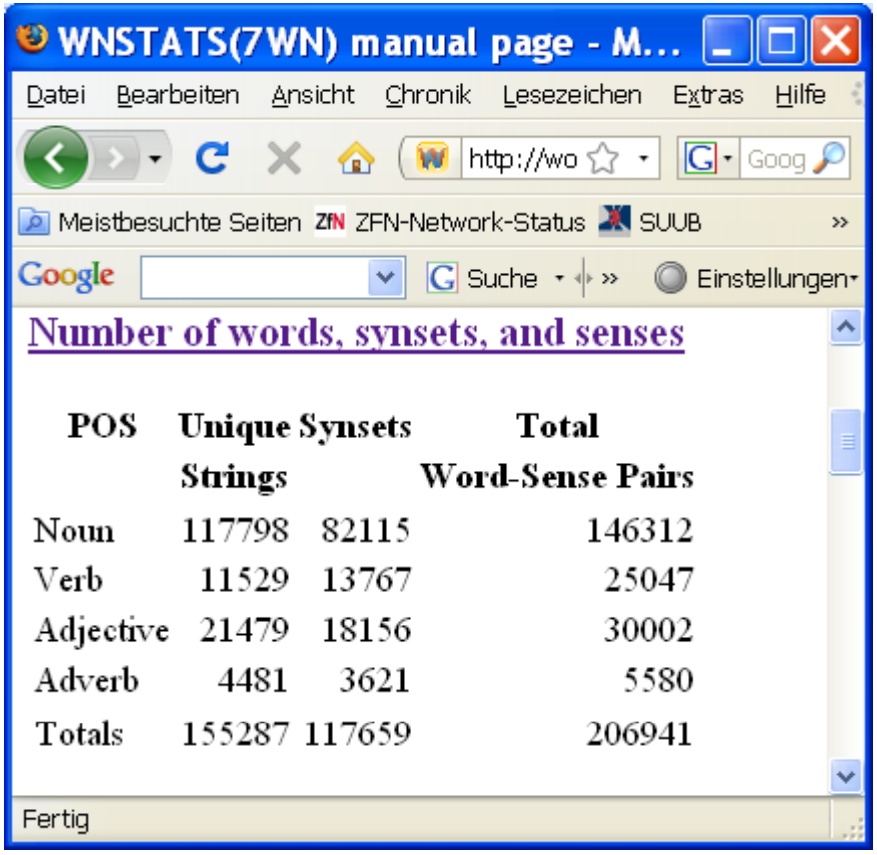
Last updated: June 29th, 2006

No.	Description	Measure	Mean	Std	Number of cases	Full description
1	Title	Title	CollegeLevel	CollegeLevel	CollegeLevel	Title
2	Genre	Genre	Narrative	Narrative	Narrative	Genre
3	Source	Source	TASA	TASA	TASA	Source
4	JobCode	JobCode	CM2	CM2	CM2	JobCode
5	LSASpace	LSASpace	CollegeLevel	CollegeLevel	CollegeLevel	LSASpace
6	Date	Date	07/06/2006	07/06/2006	07/06/2006	Date
7	Adjacent anaphor reference	CREFP1u	0.389137	0.252831	100	Anaphor reference, adjacent, unweighted
8	Anaphor reference	CREFPau	0.190835	0.181096	100	Anaphor reference, all distances, unweighted
9	Adjacent argument overlap	CREFA1u	0.55353	0.237531	100	Argument Overlap, adjacent, unweighted
10	Argument overlap	CREFAau	0.46196	0.220076	100	Argument Overlap, all distances, unweighted
11	Adjacent stem overlap	CREFS1u	0.45007	0.242775	100	Stem Overlap, adjacent, unweighted
12	Stem overlap	CREFSau	0.37547	0.188503	100	Stem Overlap, all distances, unweighted
13	Content word overlap	CREFC1u	0.08446	0.0520305	100	Proportion of content words that overlap between adjacent sentences
14	LSA sentence adjacent	LSAassa	0.38033	0.110938	100	LSA, Sentence to Sentence, adjacent, mean
15	LSA sentence all	LSApssa	0.35321	0.115572	100	LSA, sentences, all combinations, mean
16	LSA paragraph	LSAppa	.	.	100	LSA, Paragraph to Paragraph, mean
17	Personal pronouns	DENPRPi	57.7308	30.6476	100	Personal pronoun incidence score
18	Pronoun ratio	DENSPR2	0.2135	0.107795	100	Ratio of pronouns to noun phrases
19	Type-token ratio	TYPTOKc	0.81726	0.0755694	100	Type-token ratio for all content words
20	Causal content	CAUSVP	33.8444	11.8229	100	Incidence of causal verbs, links, and particles
21	Causal cohesion	CAUSC	2.55785	2.35085	100	Ratio of causal particles to causal verbs (cp divided by cv+1)
22	Intentional content	INTEi	10.8757	6.60115	100	Incidence of intentional actions, events, and particles.
23	Intentional cohesion	INTEC	2.24947	1.68919	100	Ratio of intentional particles to intentional content
24	Syntactic structure similarity adjacent	STRUTa	0.09312	0.036408	100	Sentence syntax similarity, adjacent
25	Syntactic structure similarity all-1	STRUTt	0.08459	0.0242372	100	Sentence syntax similarity, all, across paragraphs

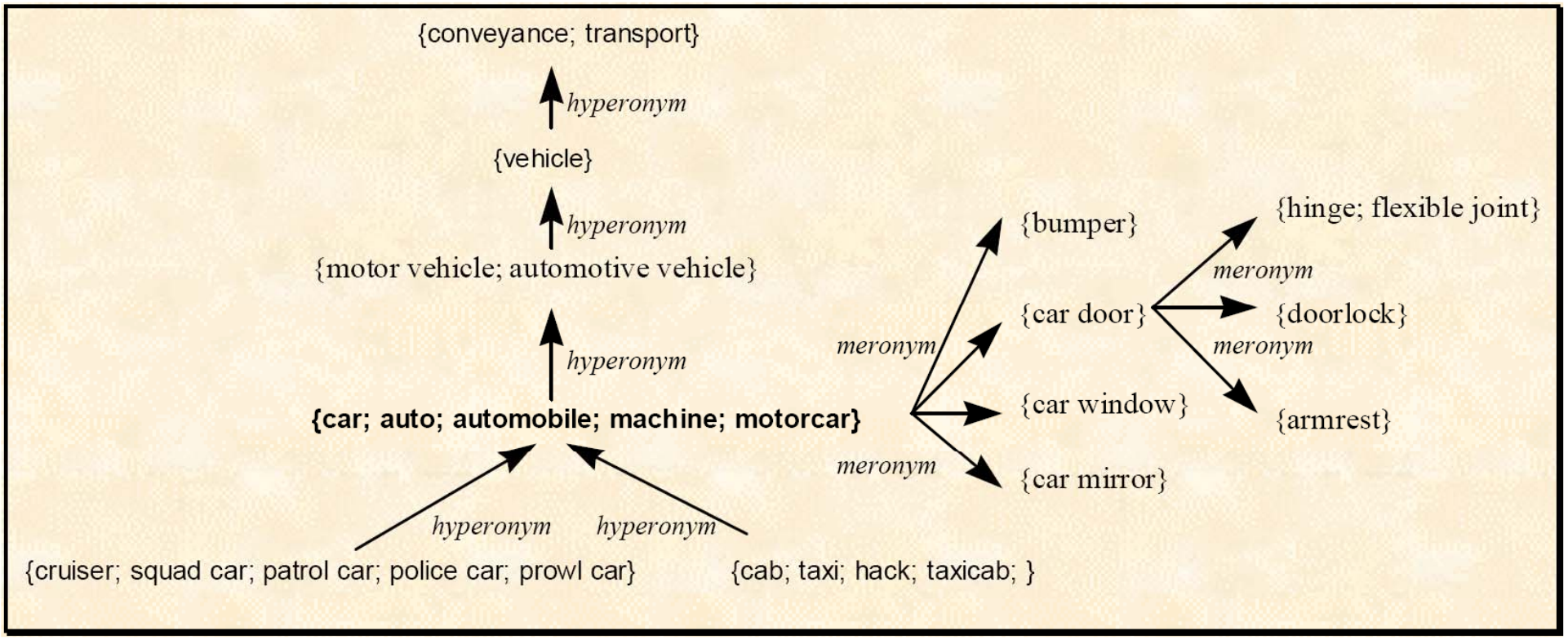

Coh-Metrix Demo - Windows Internet Explorer
http://141.225.213.52/CohMetrixDemo/demo.htm
Coh-Metrix Demo
Coh-Metrix Demo Last updated: June 29th, 2006
28 Spatial cohesion SPATC 0.50869 0.0804414 100 Mean of location and motion ratio scores.
29 All connectives CONi 79.3801 17.5063 100 Incidence of all connectives
30 Conditional operators DENCONDi 1.83778 2.83849 100 Number of conditional expressions, incidence score
31 Pos. additive connectives CONADpi 34.7782 13.7297 100 Incidence of positive additive connectives
32 Pos. temporal connectives CONTPpi 8.14765 6.35706 100 Incidence of positive temporal connectives
33 Pos. causal connectives CONCSpi 22.6293 9.73511 100 Incidence of positive causal connectives
34 Pos. logical connectives CONLGpi 19.0673 9.78626 100 Incidence of positive logical connectives
35 Neg. additive connectives CONADni 13.1559 9.2872 100 Incidence of negative additive connectives
36 Neg. temporal connectives CONTPni 0.42075 1.26418 100 Incidence of negative temporal connectives
37 Neg. causal connectives CONCSni 1.28733 1.91247 100 Incidence of negative causal connectives
38 Neg.logical connectives CONLGni 14.7434 9.26888 100 Incidence of negative logical connectives
39 Logic operators DENLOGi 47.6454 13.1544 100 Logical operator incidence score (and + if + or + cond + neg)
40 Raw freq. content words FRQCRacw 2098.91 811.582 100 Celex, raw, mean for content words (0-1,000,000)
41 Log freq. content words FRQCLacw 2.15097 0.130786 100 Celex, logarithm, mean for content words (0-6)
42 Min. raw freq. content words FRQCRmcs 39.0499 97.8882 100 Celex, raw, minimum in sentence for content words (0-1,000,000)
43 Log min. freq. content words FRQCLmcs 0.971182 0.360196 100 Celex, logarithm, minimum in sentence for content words (0-6)
44 Concreteness content words WORDCacw 376.487 28.7072 100 Concreteness, mean for content words
45 Min. concreteness content words WORDCmcs 182.12 18.4647 100 Concreteness, minimum in sentence for content words
46 Noun hypernym HYNOUNaw 4.94565 0.445604 100 Mean hypernym values of nouns
47 Verb hypernym HYVERBaw 1.47902 0.207002 100 Mean hypernym values of verbs
48 Negations DENNEGi 7.56742 5.76164 100 Number of negations, incidence score
49 NP incidence DENSNP 269.683 21.2637 100 Noun Phrase Incidence Score (per thousand words)
50 Modifiers per NP SYNNP 0.95391 0.168026 100 Mean number of modifiers per noun-phrase
51 Higher level constituents SYNHw 0.71106 0.0374814 100 Mean number of higher level constituents per word
52 Words before main verb SYNLE 5.43579 2.10655 100 Mean number of words before the main verb of main clause in sentences
53 No. of words READNW 297.99 23.1713 100 Number of Words
54 No. of sentences READNS 13.54 4.04331 100 Number of Sentences
55 No. of paragraphs READNP 1 0 100 Number of Paragraphs
Internet
100%

Lexical Databases

Lexical Databases: e.g., WordNet



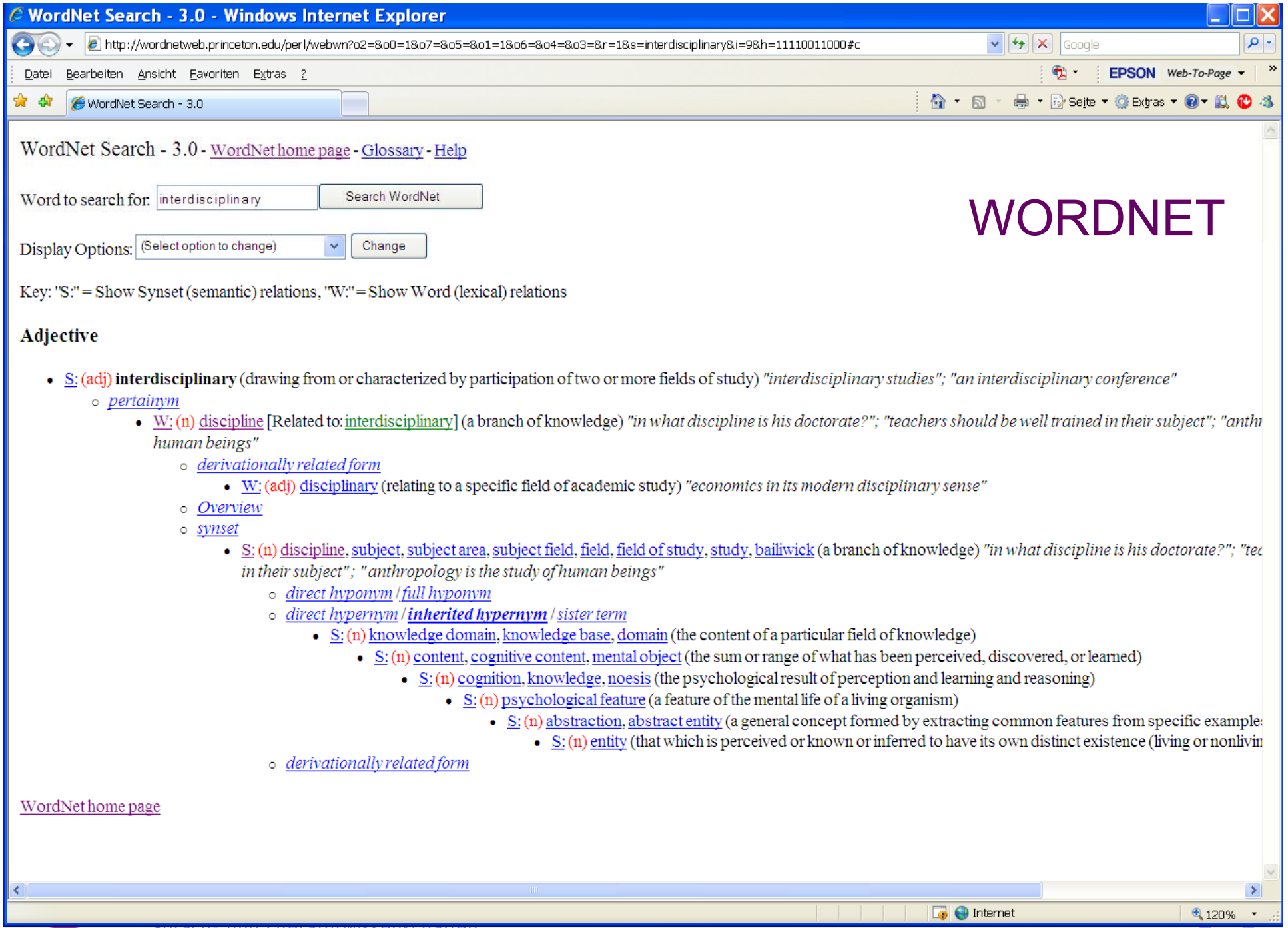
WordNet organisation



The WordNet approach to distinguishing meanings: ‘synsets’

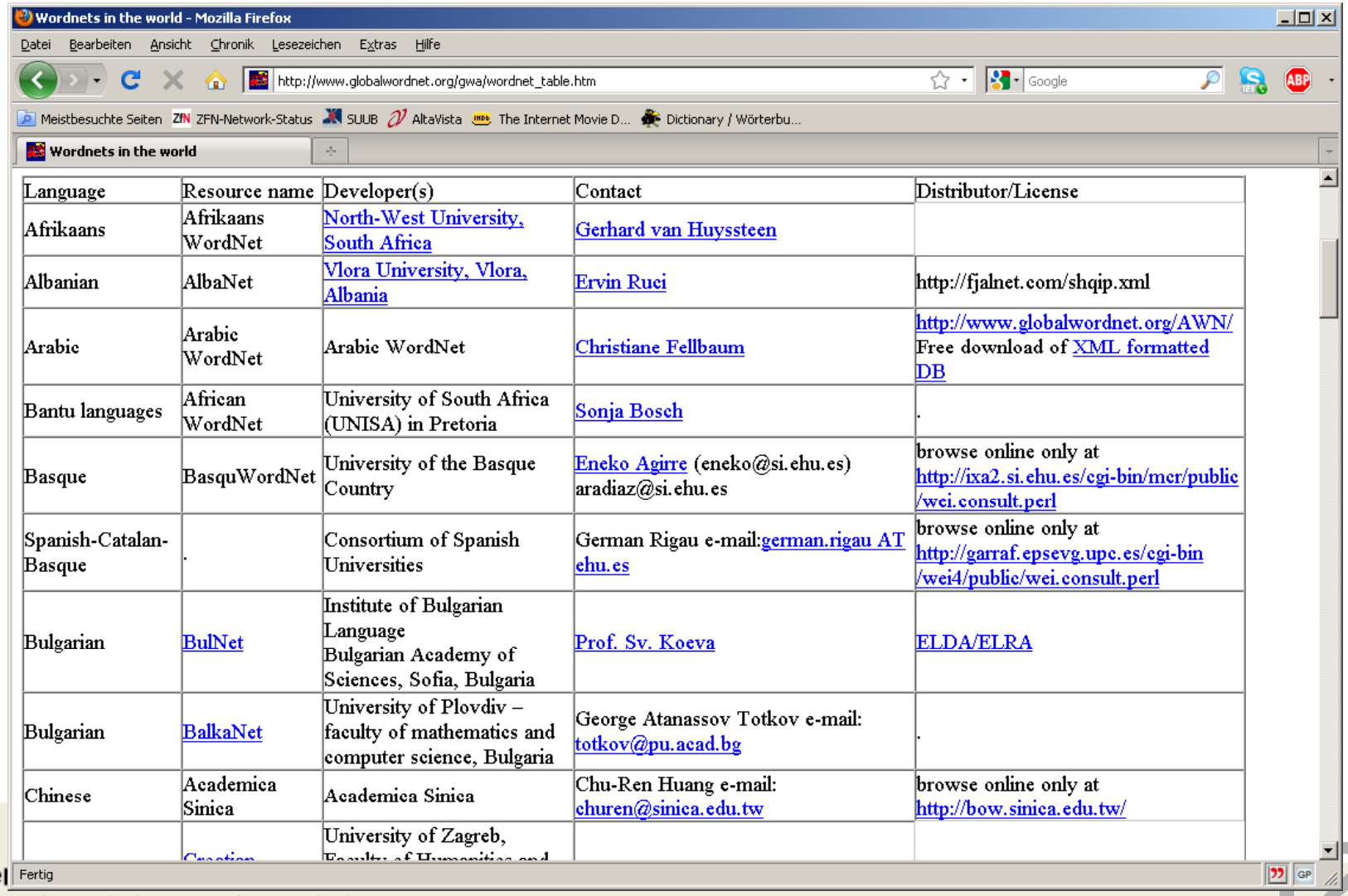
artefact: case

- {*carton, case0, box*}
- {*case1, bag*}
- {*case2, pillowcase, pillowslip, slip2, bed linen*}
- {*bag1, case3, grip, suitcase, traveling bag*}
- {*cabinet, case4, console, cupboard*}
- {*case5, container*}
- {*shell, shell plating, case6, casing1, outside surface*}
- {*casing, case7, framework*}


WordNet Search - 3.0 - Windows Internet Explorer
http://wordnetweb.princeton.edu/perl/webwn?o2=&o0=1&o7=&o5=&o1=1&o6=&o4=&o3=&r=1&s=interdisciplinary&i=9&h=11110011000#c
Datei Bearbeiten Ansicht Favoriten Extras ?
WordNet Search - 3.0
WordNet Search - 3.0 - WordNet home page - Glossary - Help
Word to search for: interdisciplinary
Search WordNet
Display Options: (Select option to change)
Change
Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations
Adjective
S: (adj) interdisciplinary (drawing from or characterized by participation of two or more fields of study) "interdisciplinary studies"; "an interdisciplinary conference"
pertainym
W: (n) discipline [Related to: interdisciplinary] (a branch of knowledge) "in what discipline is his doctorate?"; "teachers should be well trained in their subject"; "anthr
human beings"
derivationally related form
W: (adj) disciplinary (relating to a specific field of academic study) "economics in its modern disciplinary sense"
Overview
synset
S: (n) discipline, subject, subject area, subject field, field, field of study, study, bailiwick (a branch of knowledge) "in what discipline is his doctorate?"; "tec
in their subject"; "anthropology is the study of human beings"
direct hyponym / full hyponym
direct hypernym / inherited hypernym / sister term
S: (n) knowledge domain, knowledge base, domain (the content of a particular field of knowledge)
S: (n) content, cognitive content, mental object (the sum or range of what has been perceived, discovered, or learned)
S: (n) cognition, knowledge, noesis (the psychological result of perception and learning and reasoning)
S: (n) psychological feature (a feature of the mental life of a living organism)
S: (n) abstraction, abstract entity (a general concept formed by extracting common features from specific example
S: (n) entity (that which is perceived or known or inferred to have its own distinct existence (living or nonlivin
derivationally related form
WordNet home page
Internet
120%
WORDNET

Global WordNet Association

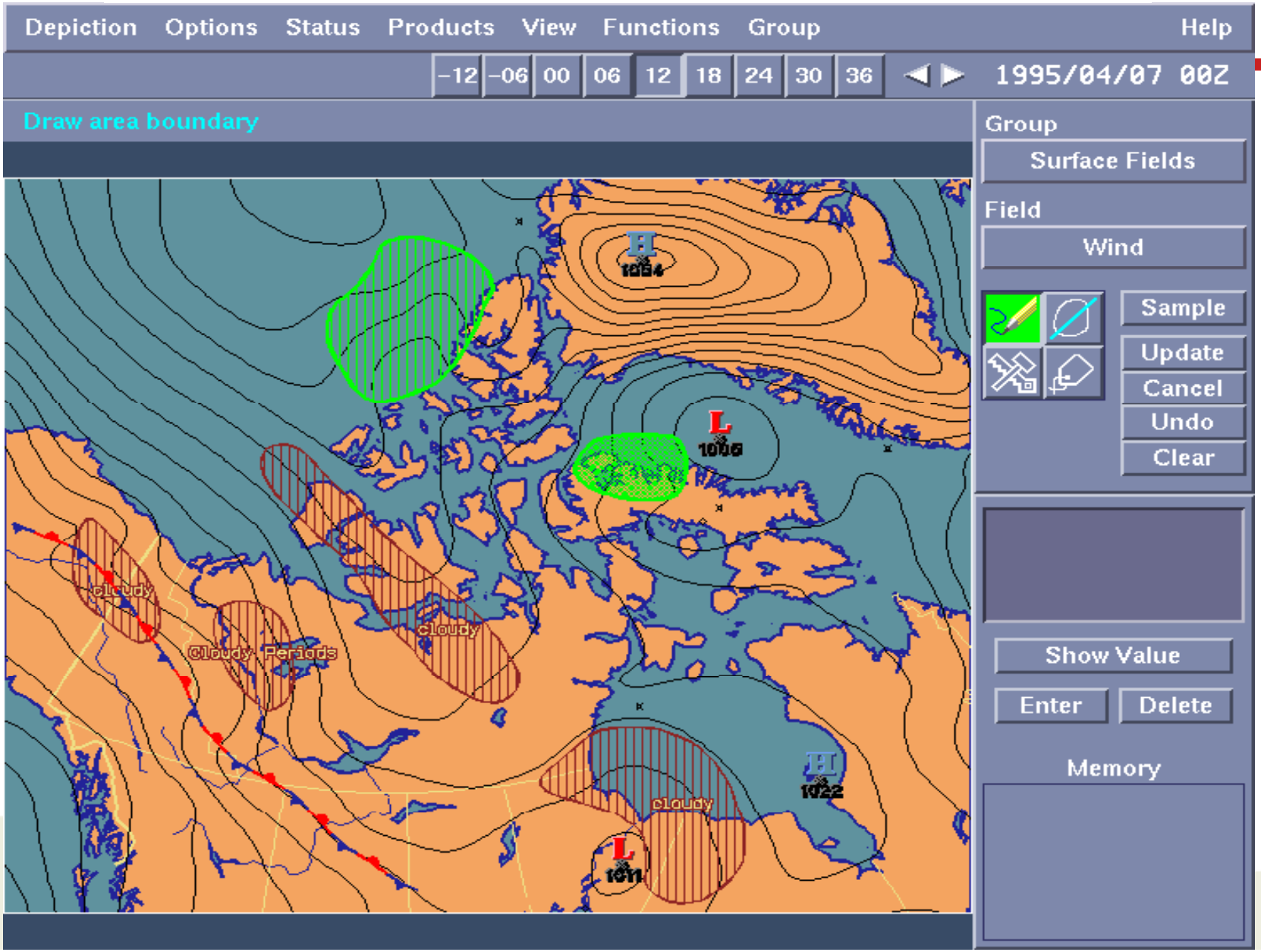
http://www.globalwordnet.org/gwa/wordnet_table.htm



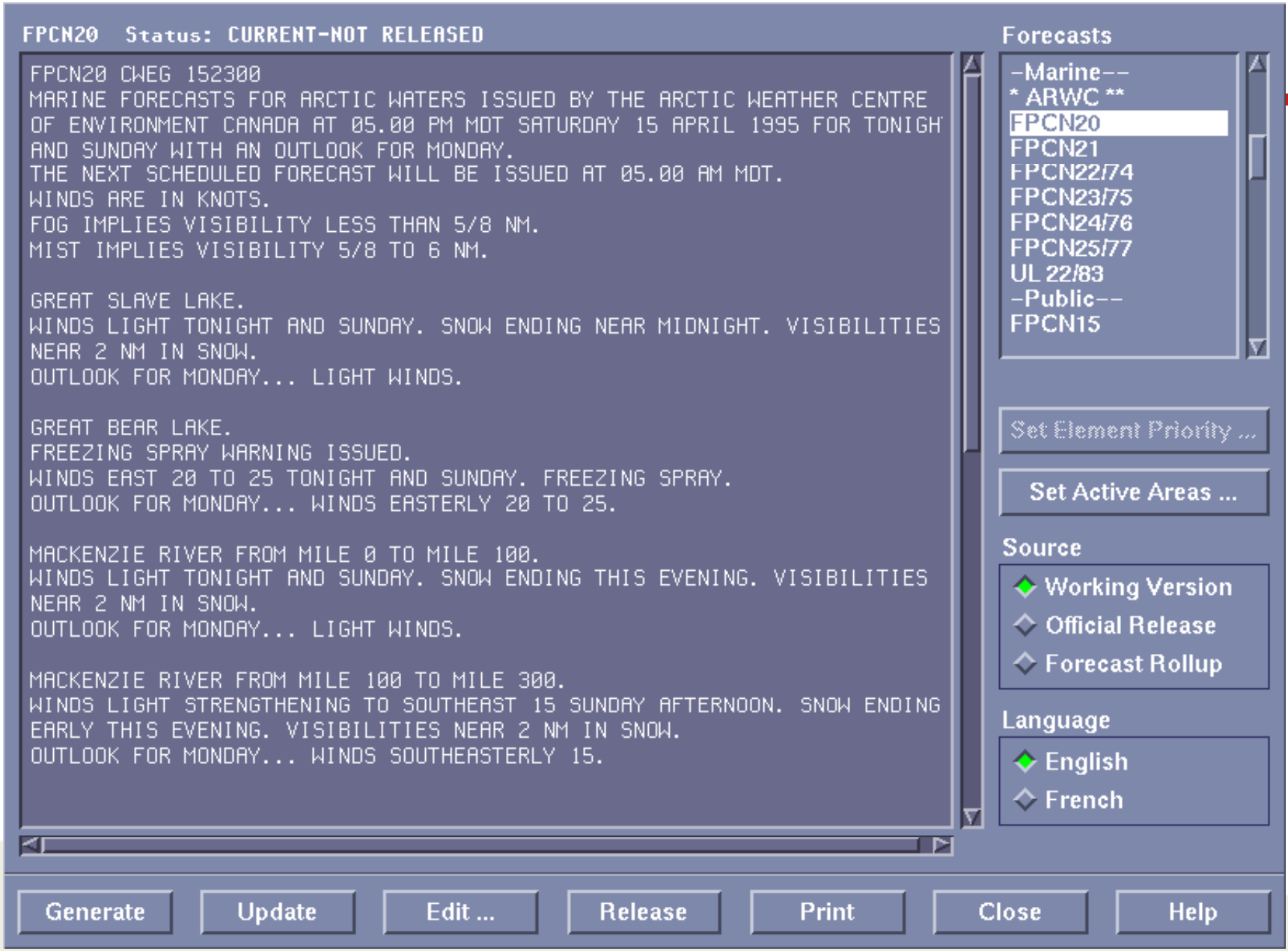
Language	Resource name	Developer(s)	Contact	Distributor/License
Afrikaans	Afrikaans WordNet	North-West University, South Africa	Gerhard van Huyssteen	
Albanian	AlbaNet	Vlora University, Vlora, Albania	Ervin Ruci	http://fjalnet.com/shqip.xml
Arabic	Arabic WordNet	Arabic WordNet	Christiane Fellbaum	http://www.globalwordnet.org/AWN/ Free download of XML formatted DB
Bantu languages	African WordNet	University of South Africa (UNISA) in Pretoria	Sonja Bosch	.
Basque	BasquWordNet	University of the Basque Country	Eneko Agirre (eneko@si.ehu.es) aradiaz@si.ehu.es	browse online only at http://ixa2.si.ehu.es/cgi-bin/mcr/public/wei.consult.perl
Spanish-Catalan-Basque	.	Consortium of Spanish Universities	German Rigau e-mail:german.rigau AT ehu.es	browse online only at http://garraf.epsevg.upc.es/cgi-bin/wei4/public/wei.consult.perl
Bulgarian	BulNet	Institute of Bulgarian Language Bulgarian Academy of Sciences, Sofia, Bulgaria	Prof. Sv. Koeva	ELDA/ELRA
Bulgarian	BalkaNet	University of Plovdiv – faculty of mathematics and computer science, Bulgaria	George Atanassov Totkov e-mail: totkov@pu.acad.bg	.
Chinese	Academica Sinica	Academica Sinica	Chu-Ren Huang e-mail: churen@sinica.edu.tw	browse online only at http://bow.sinica.edu.tw/
		University of Zagreb,		

Report Generation

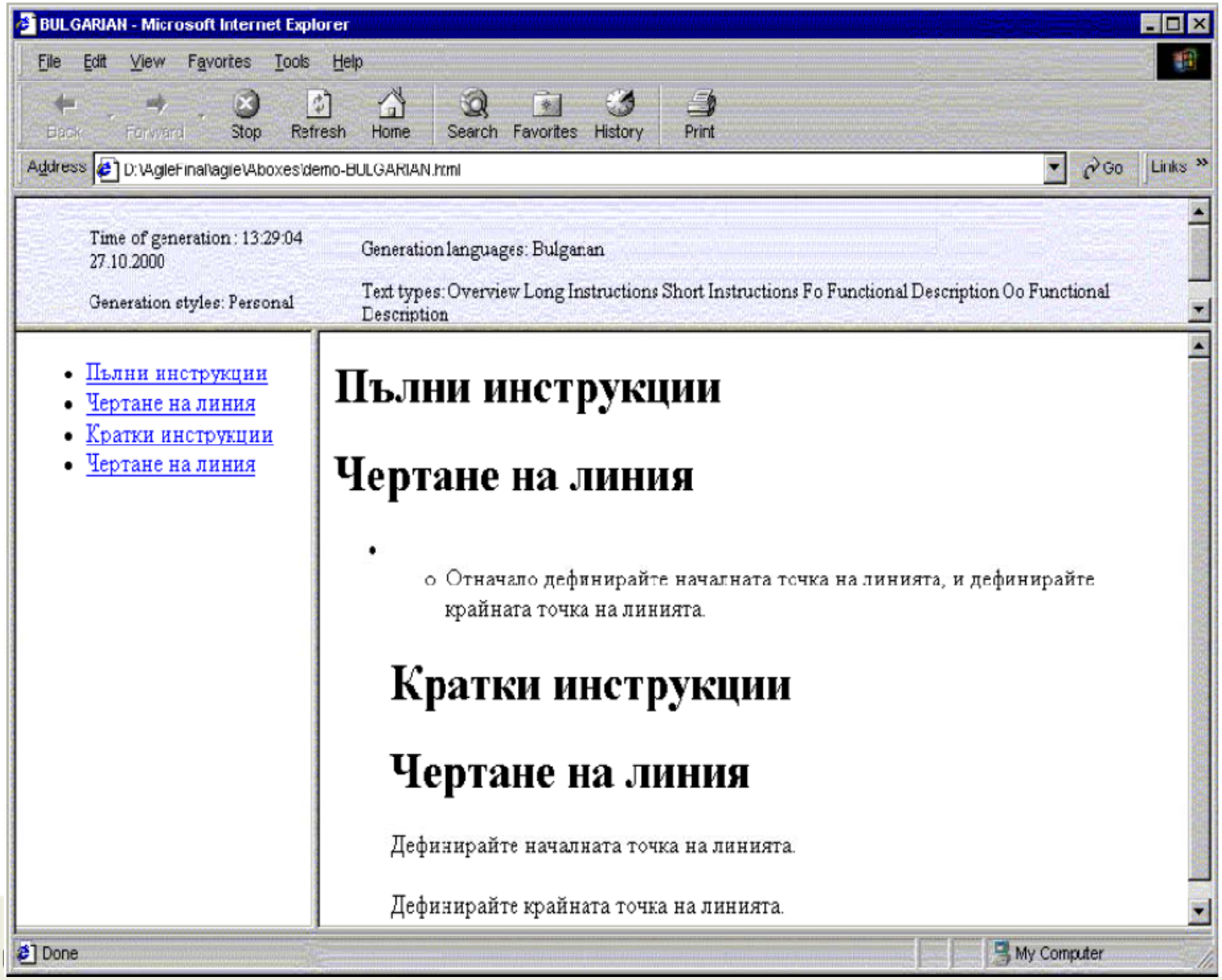
Report Generation: FoG: Input

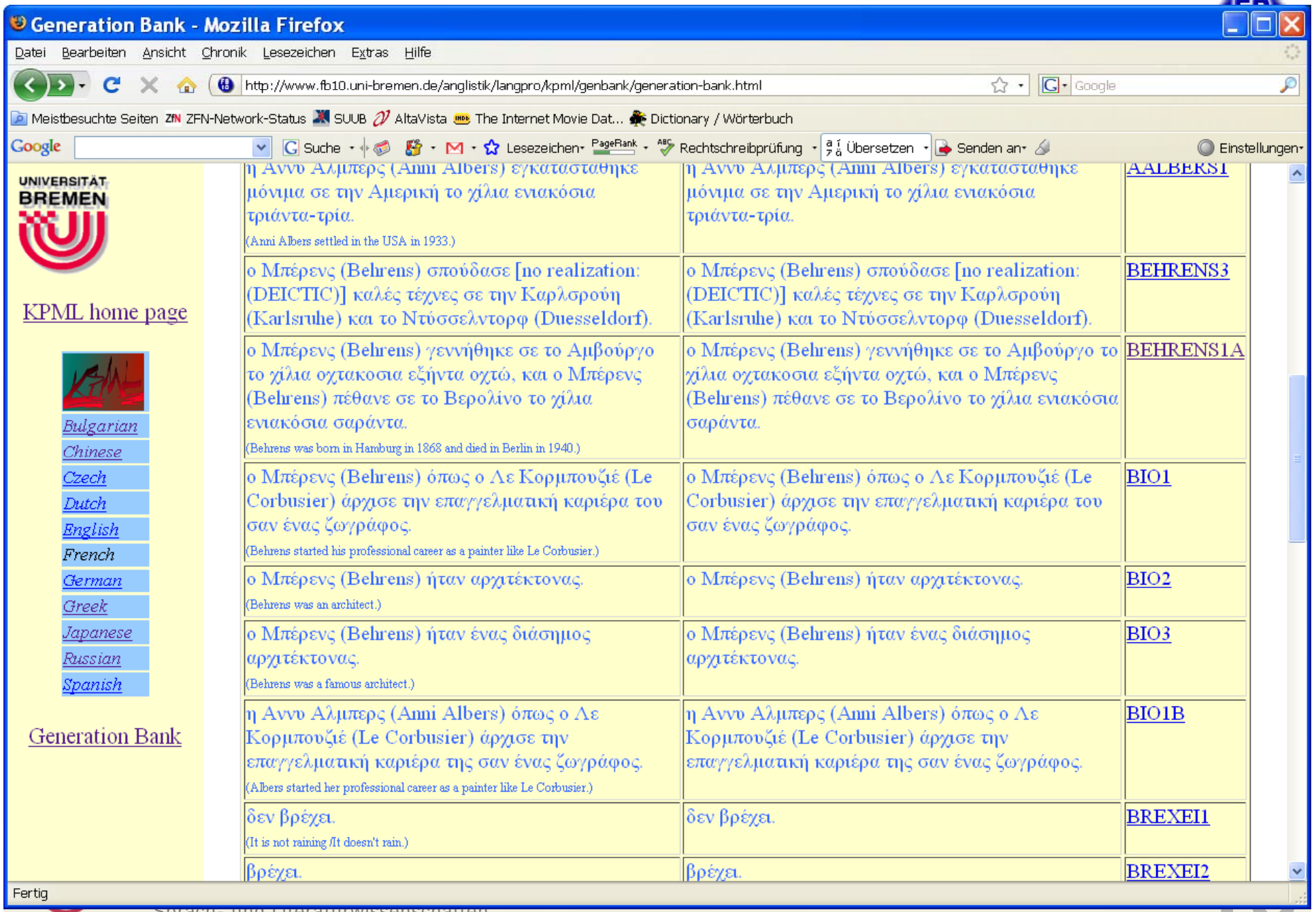


FoG: Output


FPCN20 Status: CURRENT-NOT RELEASED
FPCN20 CWEG 152300
MARINE FORECASTS FOR ARCTIC WATERS ISSUED BY THE ARCTIC WEATHER CENTRE
OF ENVIRONMENT CANADA AT 05.00 PM MDT SATURDAY 15 APRIL 1995 FOR TONIGH
AND SUNDAY WITH AN OUTLOOK FOR MONDAY.
THE NEXT SCHEDULED FORECAST WILL BE ISSUED AT 05.00 AM MDT.
WINDS ARE IN KNOTS.
FOG IMPLIES VISIBILITY LESS THAN 5/8 NM.
MIST IMPLIES VISIBILITY 5/8 TO 6 NM.
GREAT SLAVE LAKE.
WINDS LIGHT TONIGHT AND SUNDAY. SNOW ENDING NEAR MIDNIGHT. VISIBILITIES
NEAR 2 NM IN SNOW.
OUTLOOK FOR MONDAY... LIGHT WINDS.
GREAT BEAR LAKE.
FREEZING SPRAY WARNING ISSUED.
WINDS EAST 20 TO 25 TONIGHT AND SUNDAY. FREEZING SPRAY.
OUTLOOK FOR MONDAY... WINDS EASTERLY 20 TO 25.
MACKENZIE RIVER FROM MILE 0 TO MILE 100.
WINDS LIGHT TONIGHT AND SUNDAY. SNOW ENDING THIS EVENING. VISIBILITIES
NEAR 2 NM IN SNOW.
OUTLOOK FOR MONDAY... LIGHT WINDS.
MACKENZIE RIVER FROM MILE 100 TO MILE 300.
WINDS LIGHT STRENGTHENING TO SOUTHEAST 15 SUNDAY AFTERNOON. SNOW ENDING
EARLY THIS EVENING. VISIBILITIES NEAR 2 NM IN SNOW.
OUTLOOK FOR MONDAY... WINDS SOUTHEASTERLY 15.
Forecasts
-Marine--
* ARWC **
FPCN20
FPCN21
FPCN22/74
FPCN23/75
FPCN24/76
FPCN25/77
UL 22/83
-Public--
FPCN15
Set Element Priority ...
Set Active Areas ...
Source
Working Version
Official Release
Forecast Rollup
Language
English
French
Generate
Update
Edit ...
Release
Print
Close
Help

Multilingual Report Generation + Grammar Engineering


BULGARIAN - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Forward Stop Refresh Home Search Favorites History Print
Address D:\AgileFinal\agile\Aboxes\demo-BULGARIAN.html
Go Links
Time of generation: 13:29:04 27.10.2000
Generation languages: Bulgarian
Generation styles: Personal
Text types: Overview Long Instructions Short Instructions Fo Functional Description Oo Functional Description
Пълни инструкции
Чертане на линия
Кратки инструкции
Чертане на линия
Пълни инструкции
Чертане на линия
Отначало дефинирайте началната точка на линията, и дефинирайте крайната точка на линията.
Кратки инструкции
Чертане на линия
Дефинирайте началната точка на линията.
Дефинирайте крайната точка на линията.
Done
My Computer


Generation Bank - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
http://www.fb10.uni-bremen.de/anglistik/langpro/kpml/genbank/generation-bank.html
Meistbesuchte Seiten ZFN-Network-Status SUUB AltaVista The Internet Movie Dat... Dictionary / Wörterbuch
Suche Lesezeichen PageRank Rechtschreibprüfung Übersetzen Senden an Einstellungen
UNIVERSITÄT BREMEN
KPML home page
Bulgarian
Chinese
Czech
Dutch
English
French
German
Greek
Japanese
Russian
Spanish
Generation Bank
η Αννυ Αλμπερς (Anni Albers) εγκαταστάθηκε μόνιμα σε την Αμερική το χίλια ενιακόσια τριάντα-τρία.
(Anni Albers settled in the USA in 1933.)
AALBERS1
ο Μπέρενς (Behrens) σπούδασε [no realization: (DEICTIC)] καλές τέχνες σε την Καρλσρούη (Karlsruhe) και το Ντύσσελντορφ (Duesseldorf).
BEHRENS3
ο Μπέρενς (Behrens) γεννήθηκε σε το Αμβούργο το χίλια οχτακοσια εξήντα οχτώ, και ο Μπέρενς (Behrens) πέθανε σε το Βερολίνο το χίλια ενιακόσια σαράντα.
(Behrens was born in Hamburg in 1868 and died in Berlin in 1940.)
BEHRENS1A
ο Μπέρενς (Behrens) όπως ο Λε Κορμπουζιέ (Le Corbusier) άρχισε την επαγγελματική καριέρα του σαν ένας ζωγράφος.
(Behrens started his professional career as a painter like Le Corbusier.)
BIO1
ο Μπέρενς (Behrens) ήταν αρχιτέκτονας.
(Behrens was an architect.)
BIO2
ο Μπέρενς (Behrens) ήταν ένας διάσημος αρχιτέκτονας.
(Behrens was a famous architect.)
BIO3
η Αννυ Αλμπερς (Anni Albers) όπως ο Λε Κορμπουζιέ (Le Corbusier) άρχισε την επαγγελματική καριέρα της σαν ένας ζωγράφος.
(Albers started her professional career as a painter like Le Corbusier.)
BIO1B
δεν βρέχει.
(It is not raining /It doesn't rain.)
BREXEI1
βρέχει.
BREXEI2
Fertig

Dialogue

One example application: spatial dialogues…

Drive Instructions

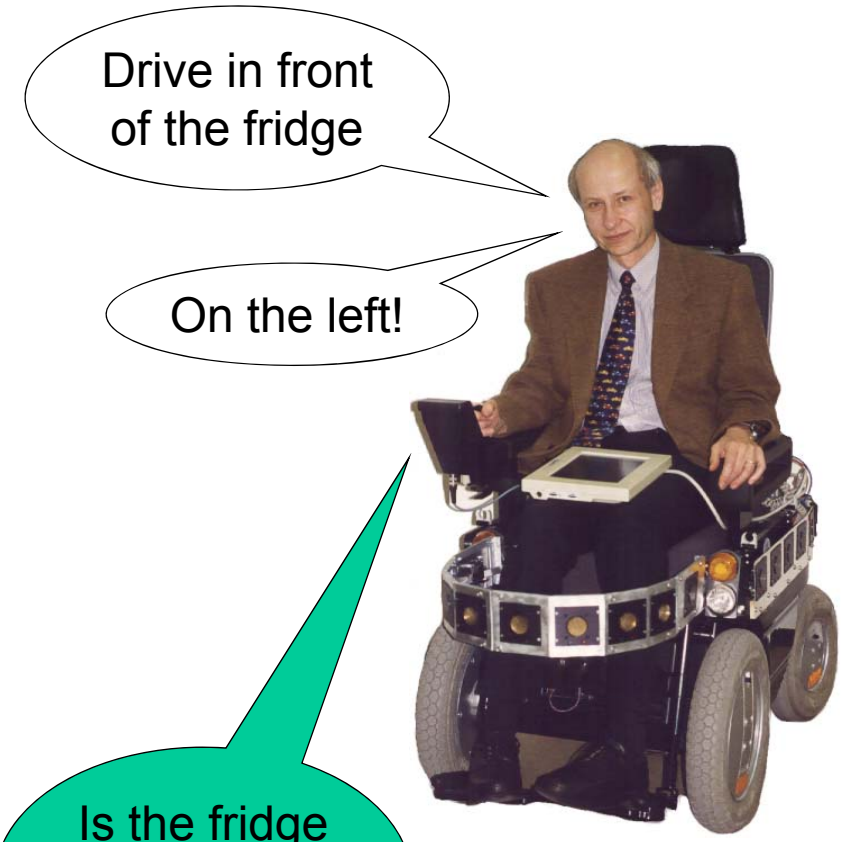
Drive to the kitchen

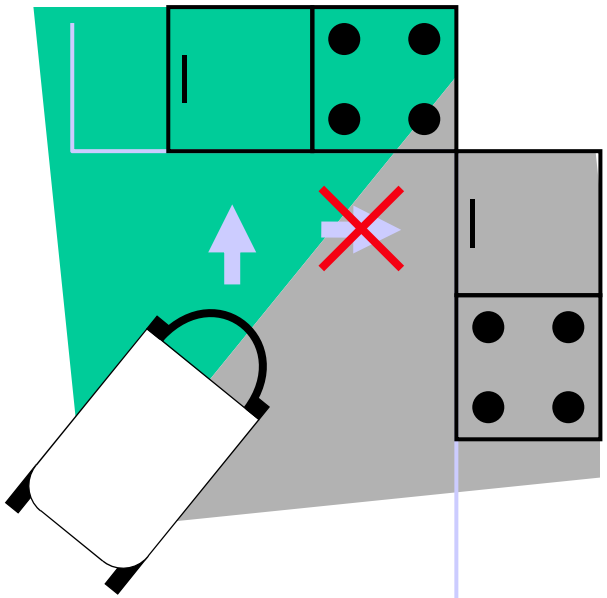
Drive in front of the fridge

Augmenting Maps

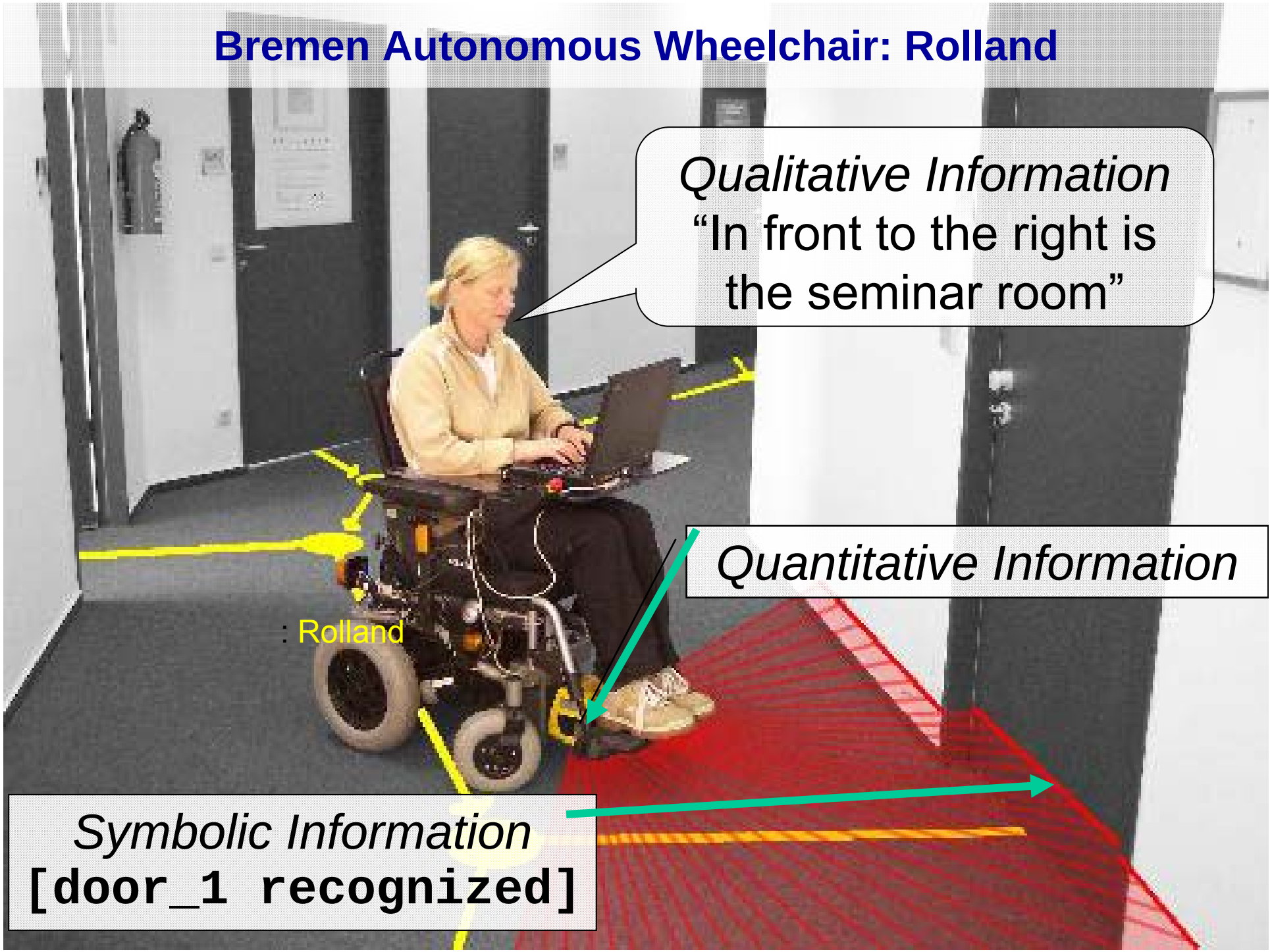
The fridge is on the right, and left of the stove

Situated Dialogs





Bremen Autonomous Wheelchair: Rolland



What kind of linguistic knowledge is involved in these kinds of systems?

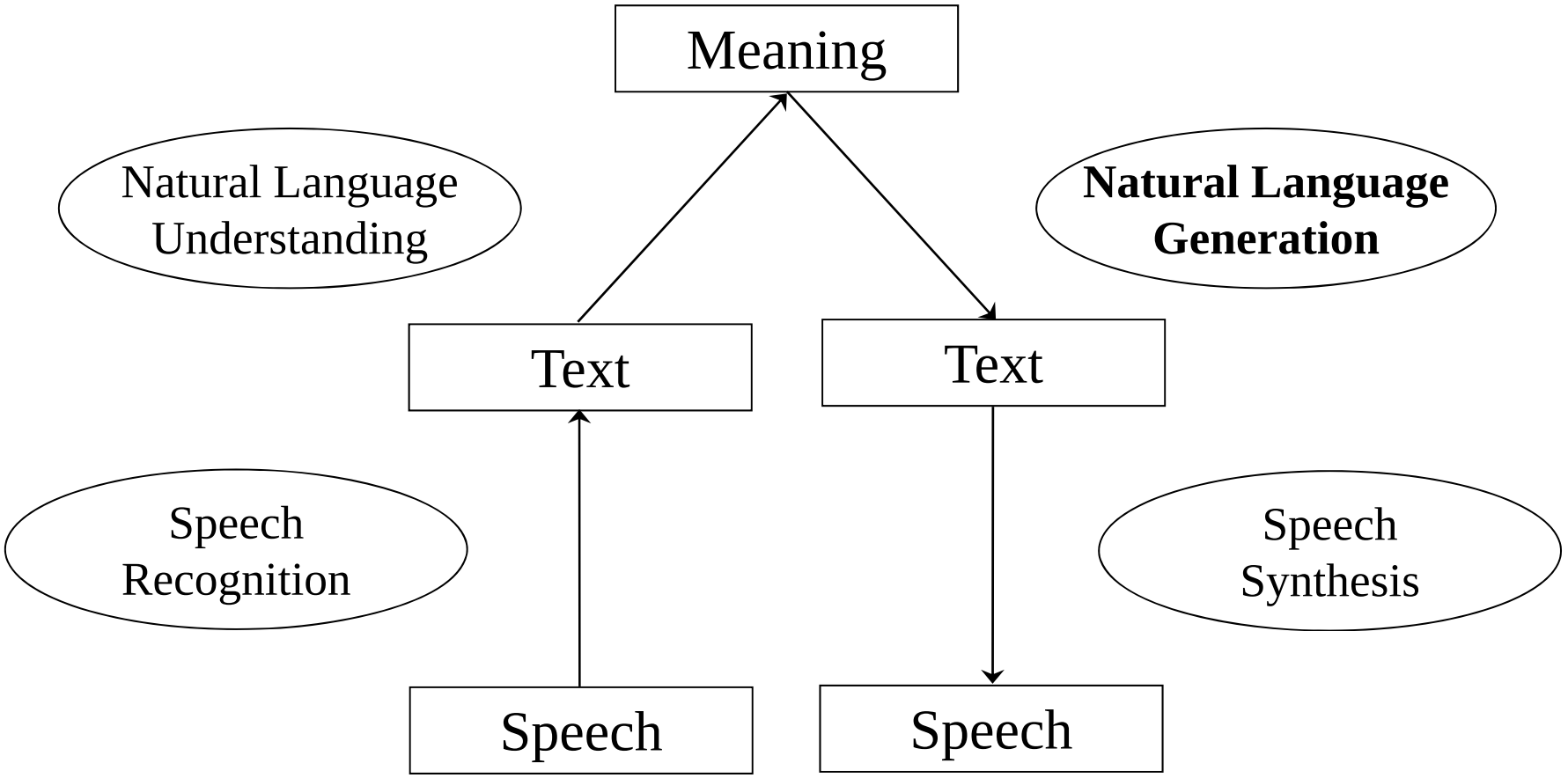
Levels of language technology

- Phonetic / Phonological → speech processing / speech synthesis
- Syntactic / Grammatical → tactical generation / parsing technology
- Semantics → reasoning, logic, knowledge representation, ontologies
- Pragmatics → dialogue acts, dialogue processing

Close relation between theory and practice

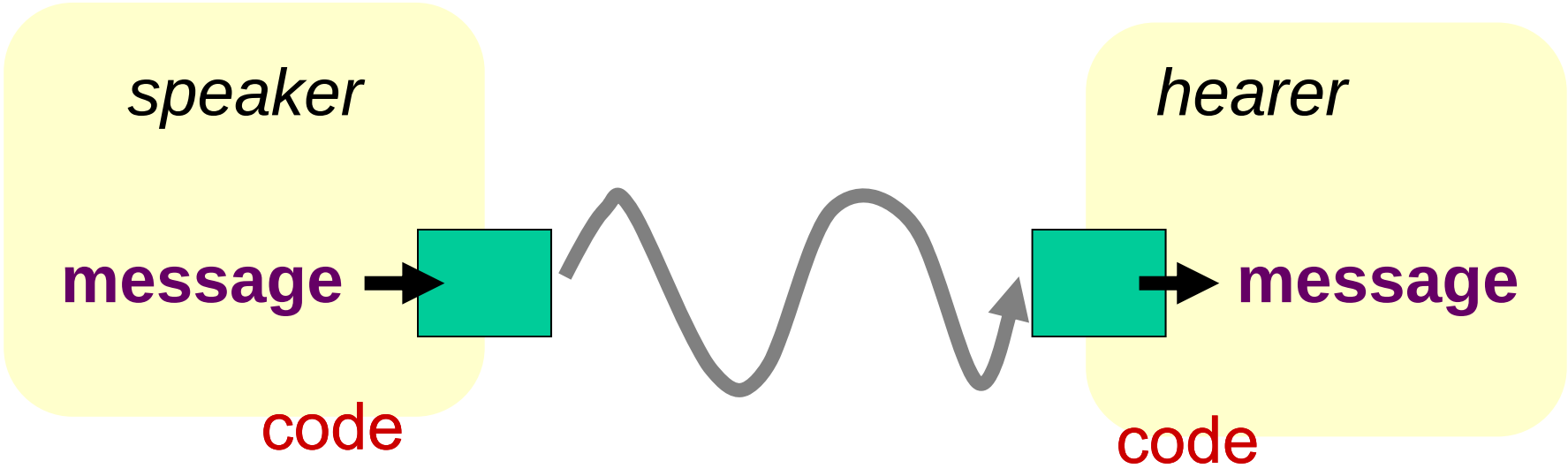
- Theory informs our modelling
- The performance of those models in real or test applications gives us input about the quality of the theory used

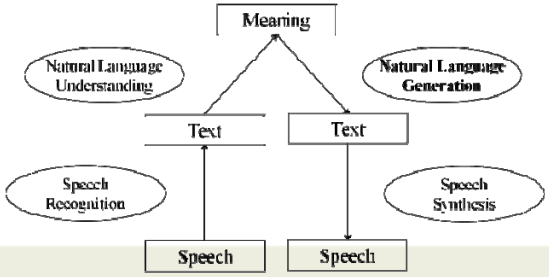
Language Processing Tasks

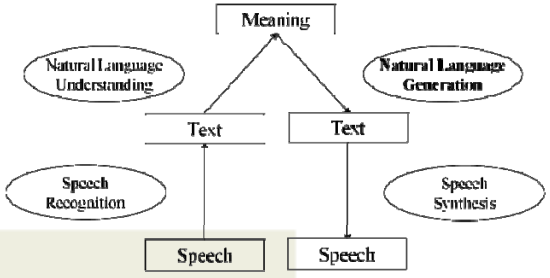


So, what kind of thing is language anyway?

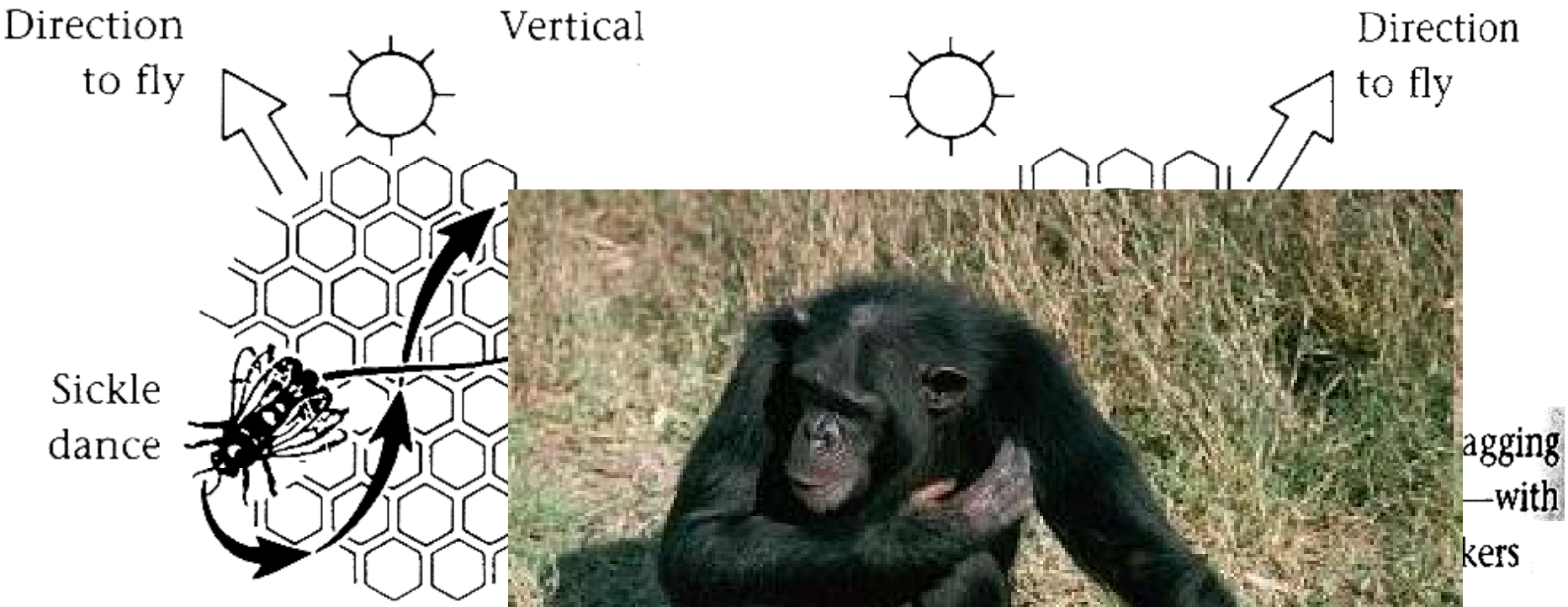
Communication Model







Bee communication

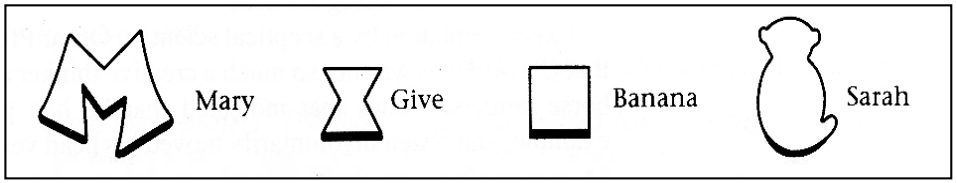


From: Michael Dobrovolsky (19
Dobrolovksy / Katamba *Contem*

Language acquisition by non-human primates

- 1930-1940s: **Gua + Viki**
  [baby chimpanzees, spoken language]

- 1960s
  - **Sarah**
  - **Lana**
  - **Kanzi**



[large buttons on keyboard]

- 1970s
  - **Washoe** [chimp, ASL]
  - **Koko** [Gorilla, ASL]
  - **Nim Chimpsky** …

Patterson (1978) “Language is no longer the exclusive preserve of man.”

The debate continues…

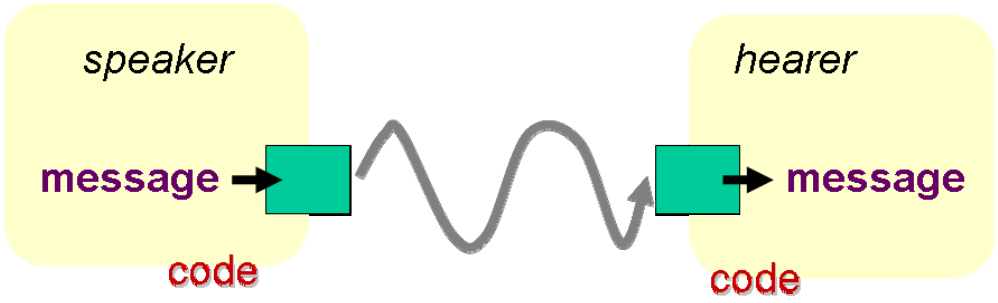
- **Seidenberg and Petitto:**

  "We believe that … there is no basis to conclude that signing apes acquired linguistic skills."

- **Fouts:**

  – "When these projects [Washoe, Lana, Sarah and Nim] are taken together, it can be seen that chimpanzees are within the range of language behaviour of humans and therefore have the capacity for language."

Basic Questions

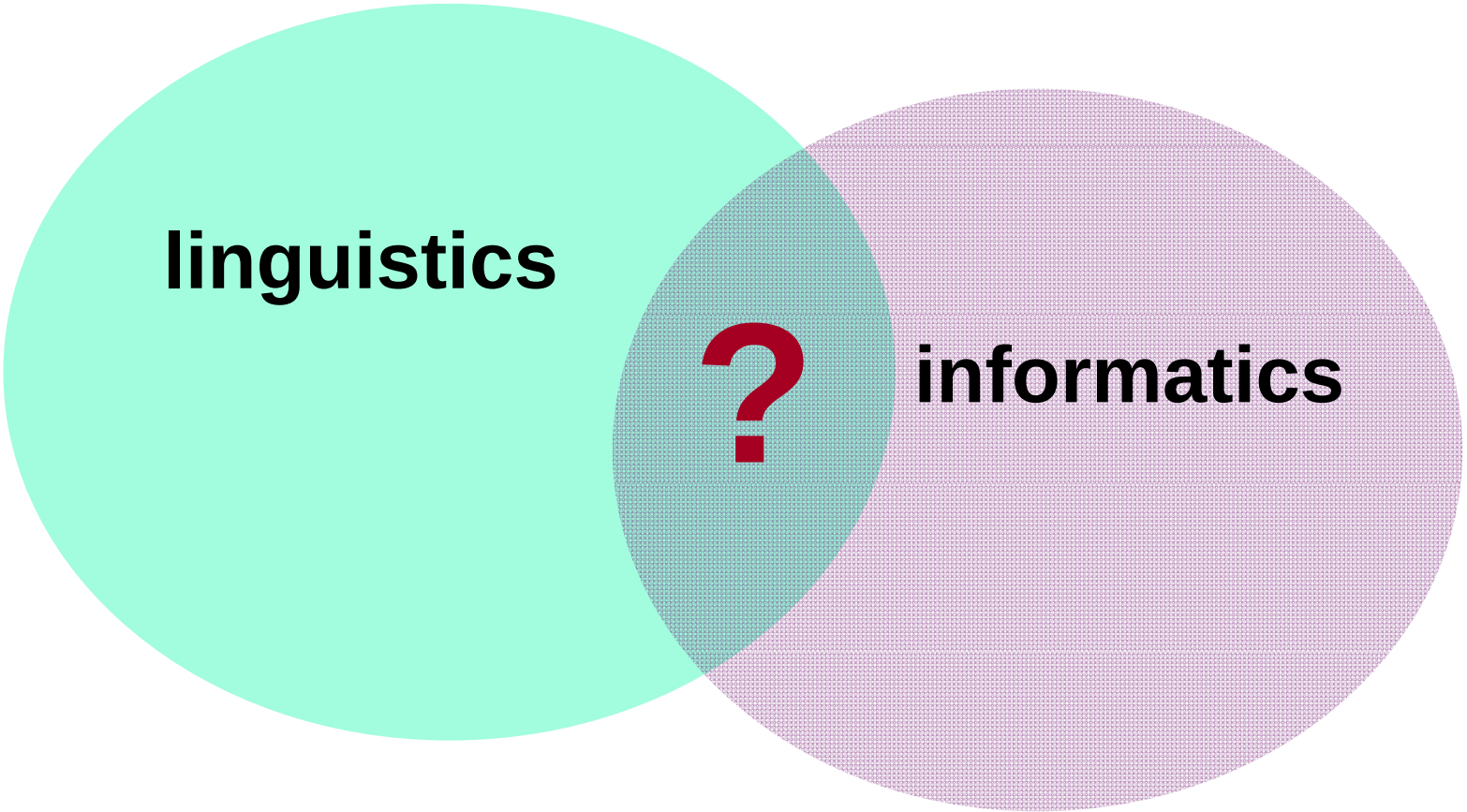


- Can we say something about the kind of code that human's employ?
- Something that brings out what is unique in this system of language in relation to others?
- Something that explains at the same time how it is that any child will learn the language(s) that it is first exposed to?

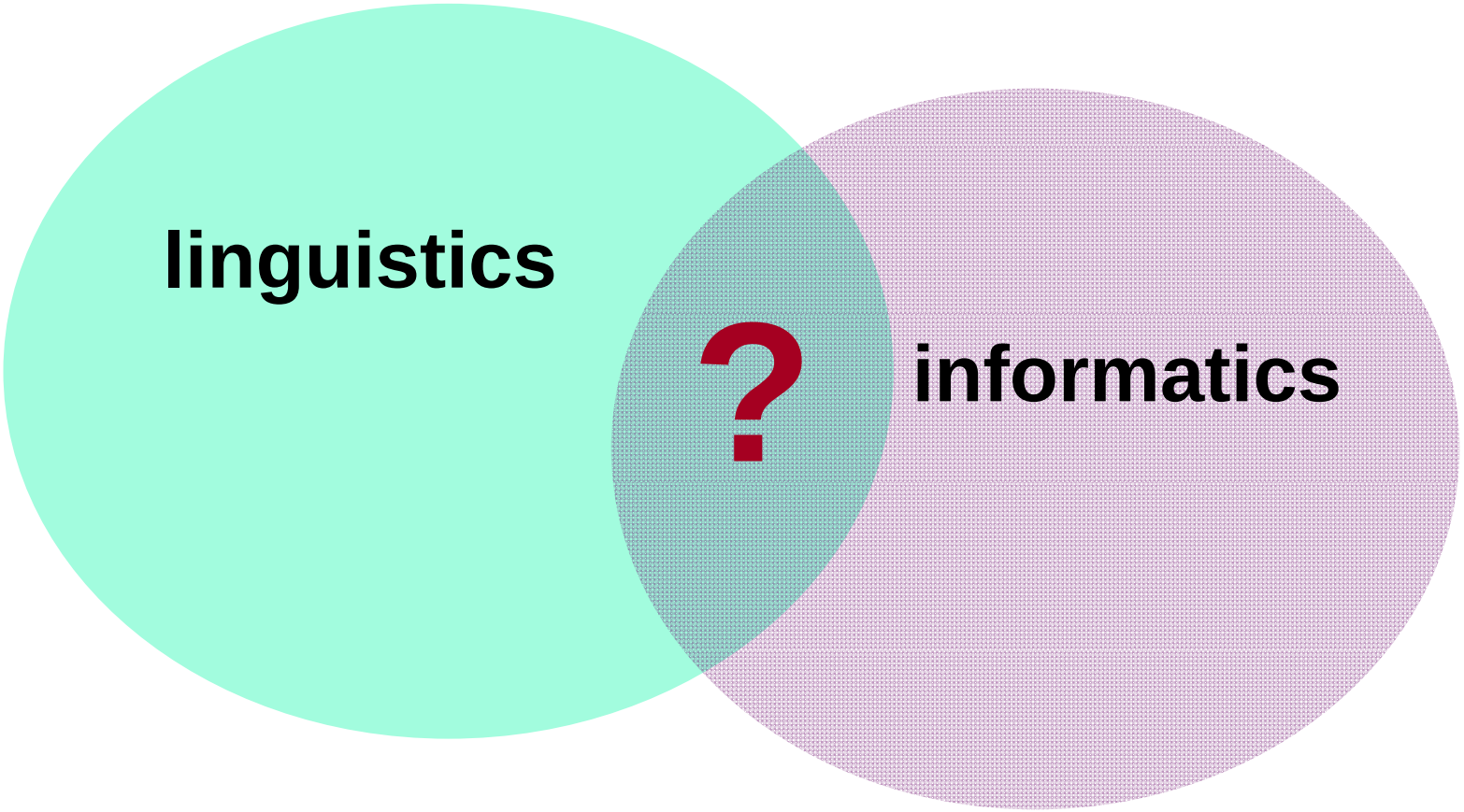
History of Computational Linguistics

- History
  - what is computational linguistics?
  - where did it come from?
  - what does it have to do with linguistics?

Computational Linguistics ?



Computational Linguistics ?



Computer Science

- Formal Language Theory (mathematics: 1930s + 1940s)
  - a language is a set of strings
  - grammars can describe sets of strings
  - we can divide grammars into distinct classes according to how ‘powerful’ they are

Example

- A ‘language’ might be the set of strings:
  - {ab, aabb, aaabbb, aaaabbbb, …}

- This can be produced by a grammar of the following kind:

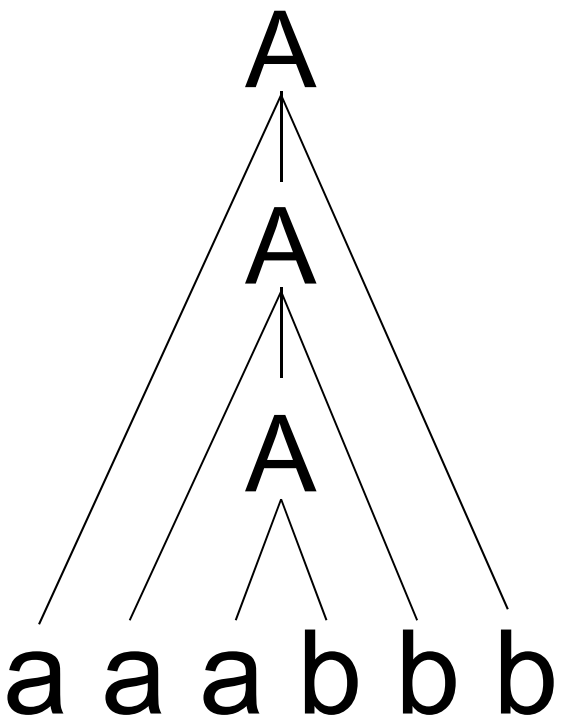
$$A \rightarrow ab$$

$$A \rightarrow aAb$$

(Questions of ‘meaning’ left out for the time being)

Example

- The grammar ‘assigns’ structures to strings



A → aAb

A → aAb

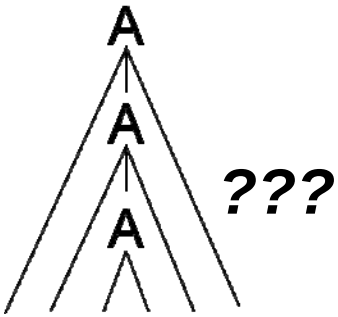
A → ab

Example

- It was noted that some strings just can't be produced by some kinds of grammars...

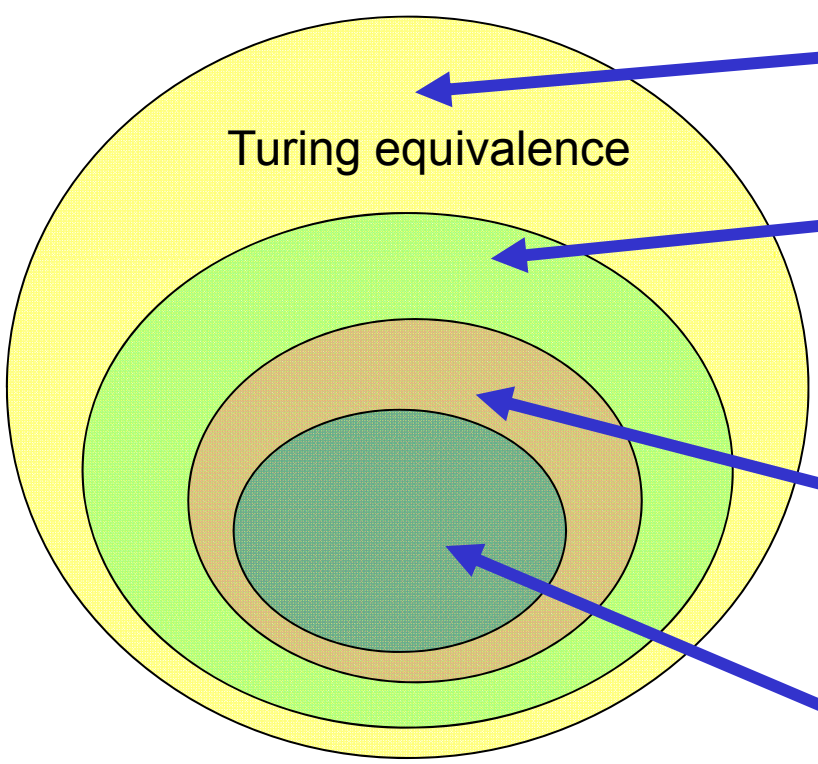
a a b b c c

a a a b b b c c c



$a^n b^n c^n$

‘Chomsky’ Hierarchy



Type 0 : unrestricted

$$\alpha \rightarrow \beta$$

Type 1 : context-sensitive

$$\alpha \, A \, \beta \rightarrow \alpha \, \gamma \, \beta$$

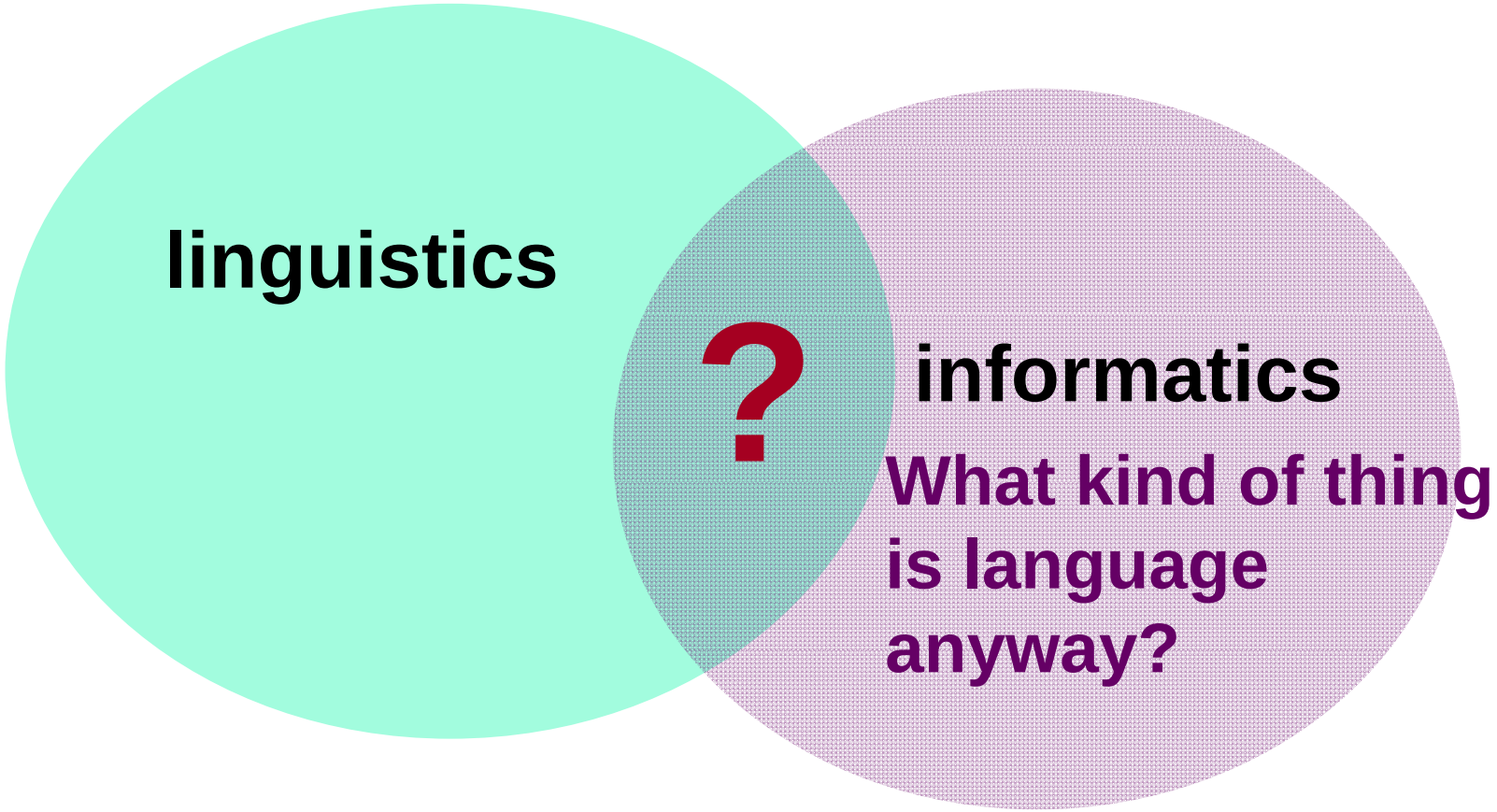
Type 2 : context-free

$$A \rightarrow \gamma$$

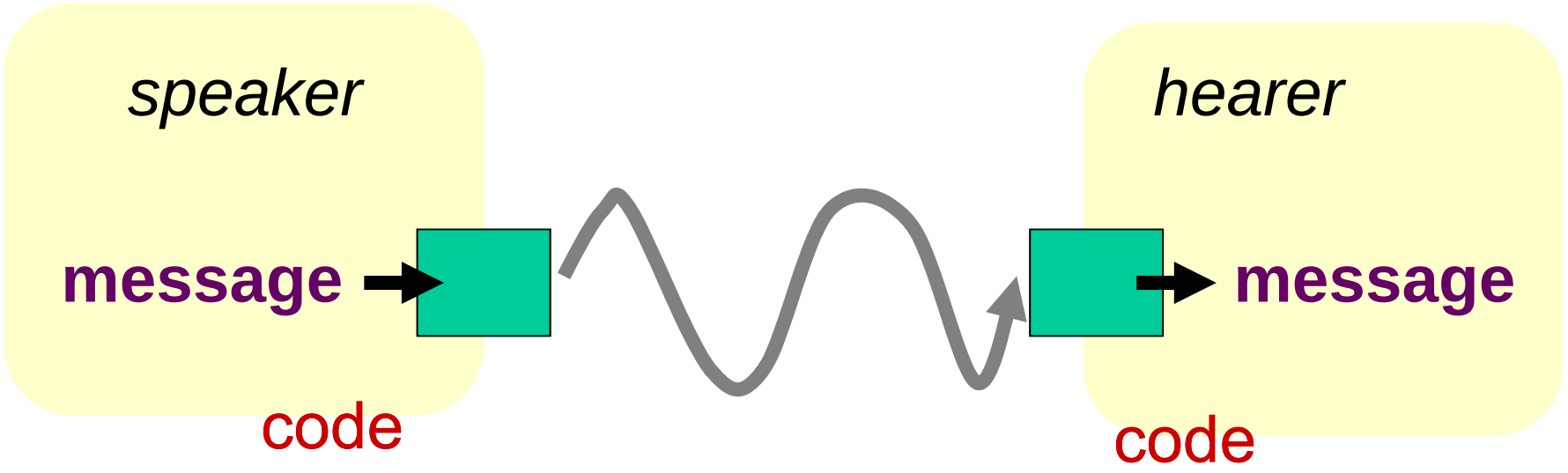
Type 3 : regular

$$A \rightarrow a$$
$$A \rightarrow aB$$

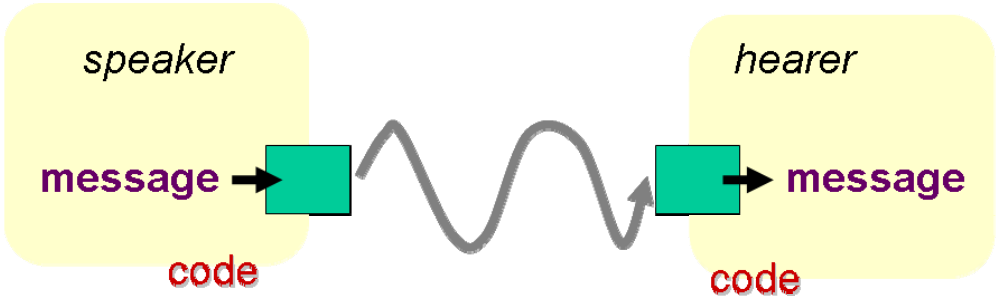
Computational Linguistics ?



Communication Model

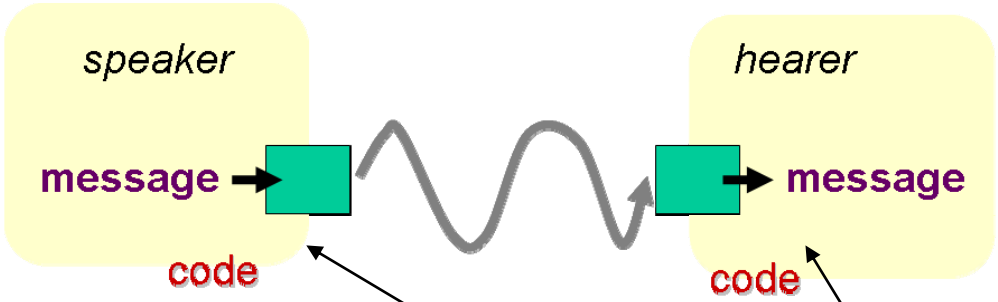


Basic Questions



- Can we say something about the kind of code that human's employ?
- Something that brings out what is unique in this system of language in relation to others?
- Something that explains at the same time how it is that any child will learn the language(s) that it is first exposed to?

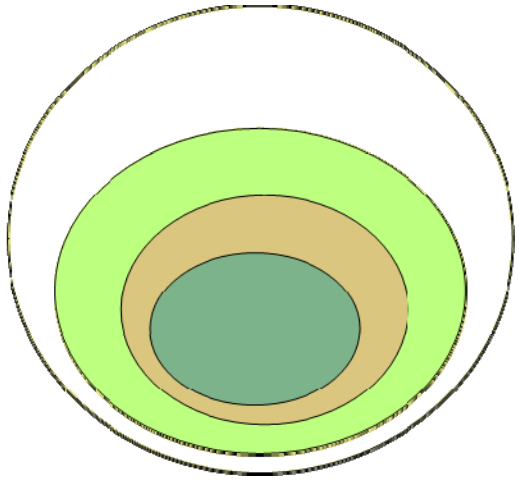
Chomsky's 'revolution'



- In short:
  - just what kind of code is this?

Chomsky's 'revolution'

- Fairly easy to show that the kinds of structures observable in natural human language do not fit in the simpler classes of grammatical system
  - not regular
  - not context-free
  - …?



Chomsky's 'revolution'

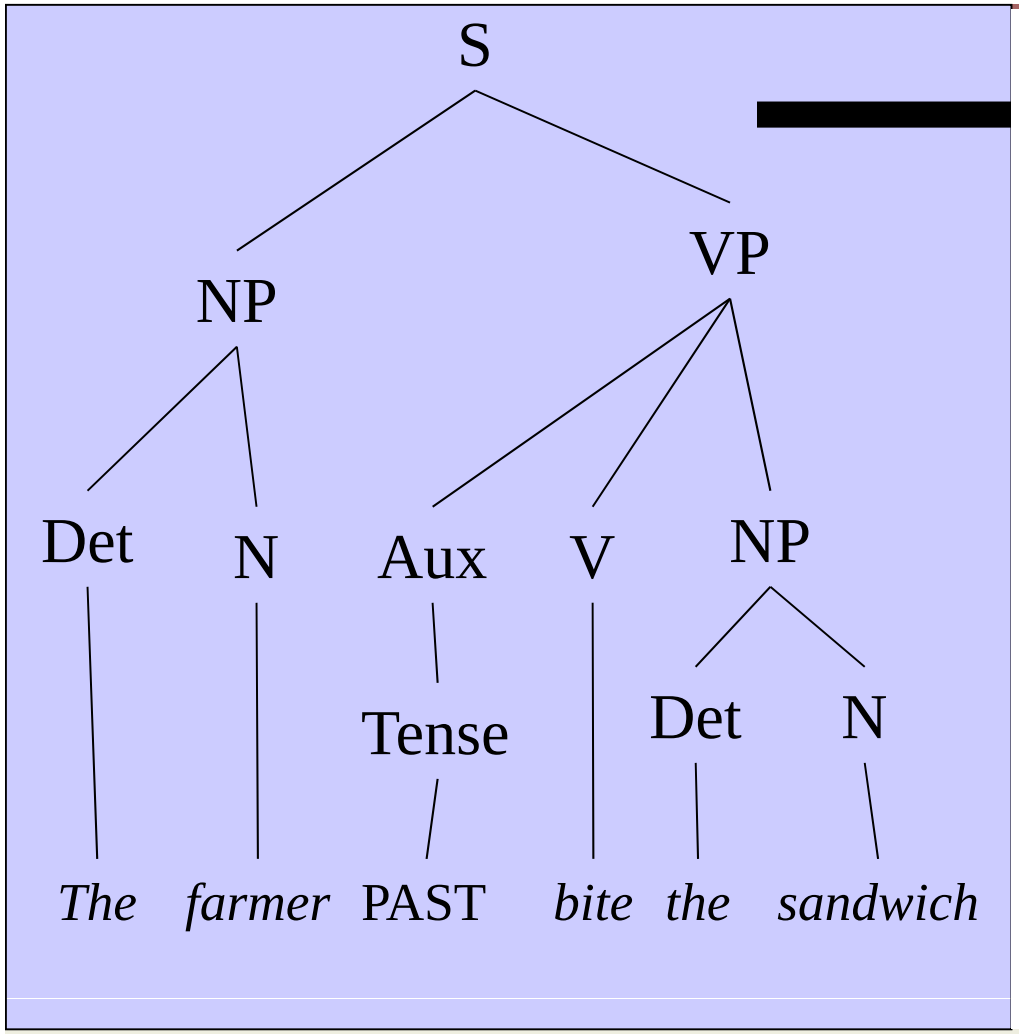
Chomsky 1956

Grammar much simplified
(particularly with the interface with morphology)
if a rule such as the following were possible:

have^past → had
be^en → been
take^ing → taking
will^past → would
can^past → could
M^present → M
walk^past → walked
take^past → took
etc.

Affix-hopping

Af V ⇒ V Af



Chomsky's 'revolution'

A rule of the form

(25) $Z \hat{} X \hat{} W \rightarrow Z \hat{} Y \hat{} W$

indicates that X can be rewritten as Y only in the context Z--W. It can easily be shown that the grammar will be much simplified if we permit such rules.

Chomsky 1956:118

Chomsky's$_{(1957)}$ position

- Previous and other current accounts of syntax not powerful enough to deal with the basic facts of language
  - not precise or restrictive enough: how can children learn an unstructured set of methods?
  - finite-state models do not match human language
  - meaning was not addressed and therefore something is seriously wrong (Behaviourism)

Chomsky: definitions (1957)

- Language:
  - language is a set (finite or infinite) of sentences
- Grammar:
  - a device that generates all the grammatical sequences [of that language] and none of the ungrammatical ones.

A Phrase Structure Grammar

(a) S $\rightarrow$NP + VP

(b) NP $\rightarrow$ Det + N

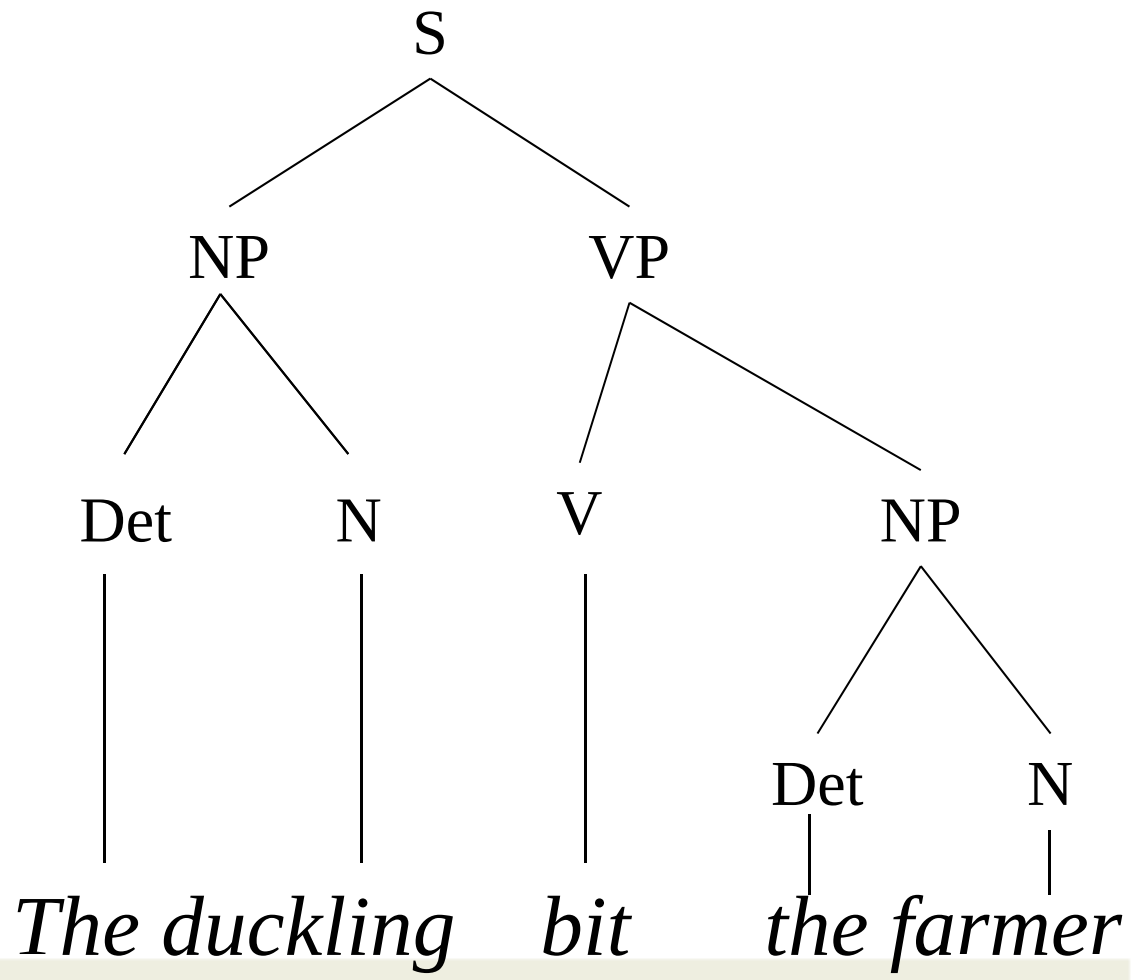
(c) VP $\rightarrow$ V + NP

(d) Det $\rightarrow$ *the*

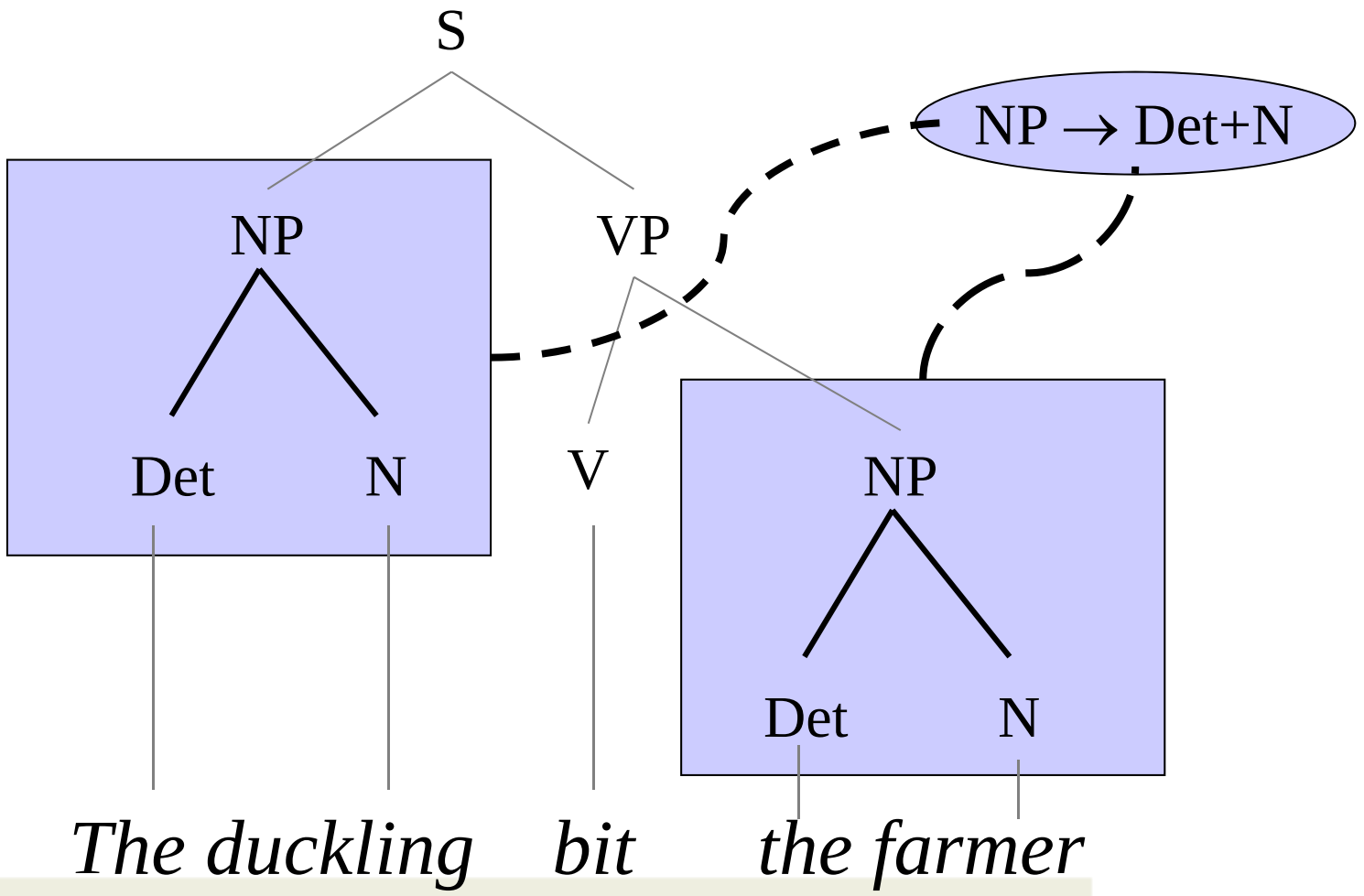
(e) N $\rightarrow$ {*dog, duckling, sandwich, ...*}

(f) V $\rightarrow$ {*bite, chase, hop, kill, ...*}

Phrase Structure Markers



Phrase Structure Markers: reuse of structural knowledge



Capturing relations (or failing to)

- The duckling bit the farmer
- The farmer was bitten by the duckling

- S →NP+V+NP
- S →NP+be+V+by+NP

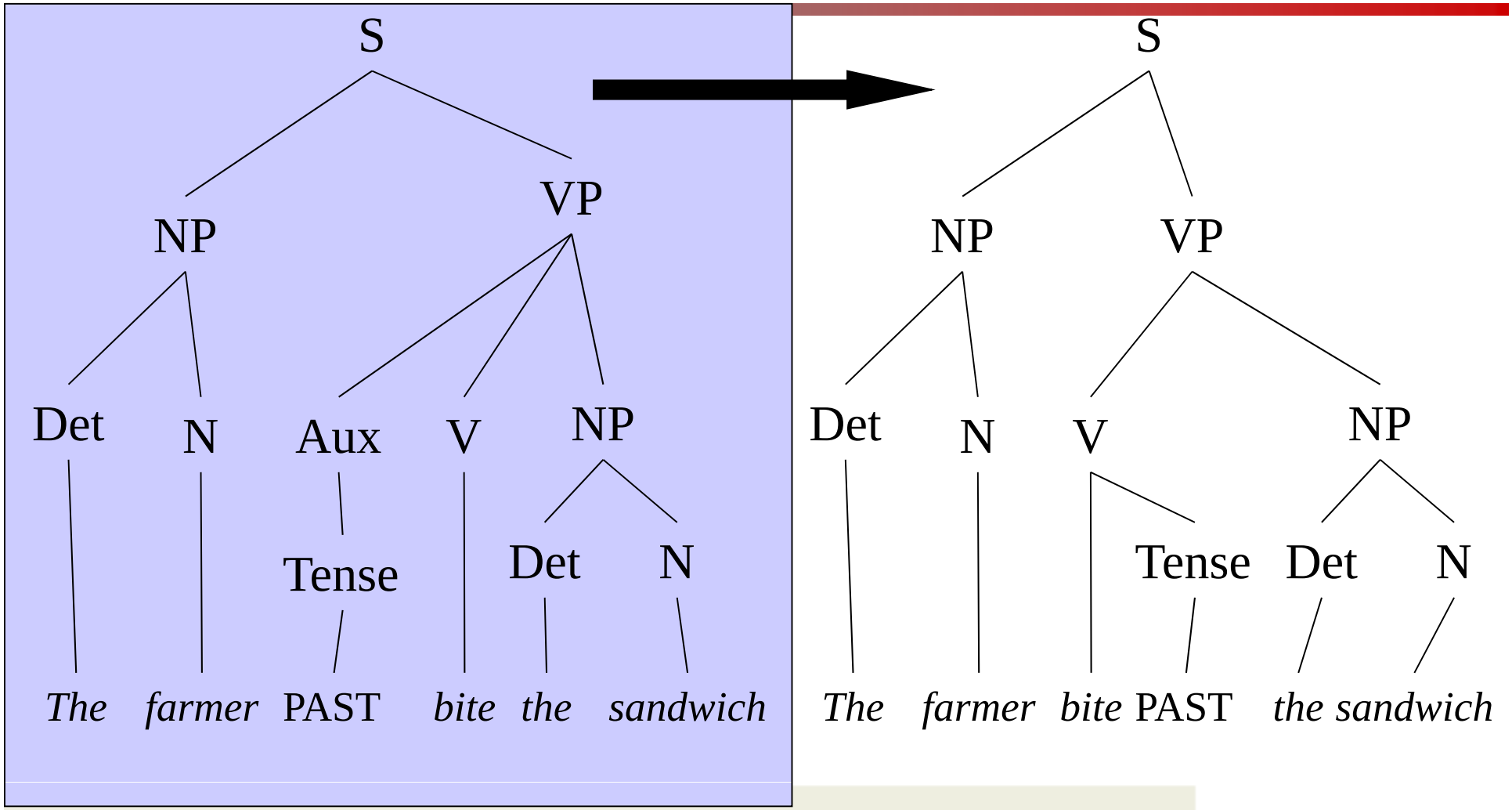
- *The sandwich bit the farmer
- *The farmer was bitten by the sandwich

The Passive Transformation (Chomsky, 1957)

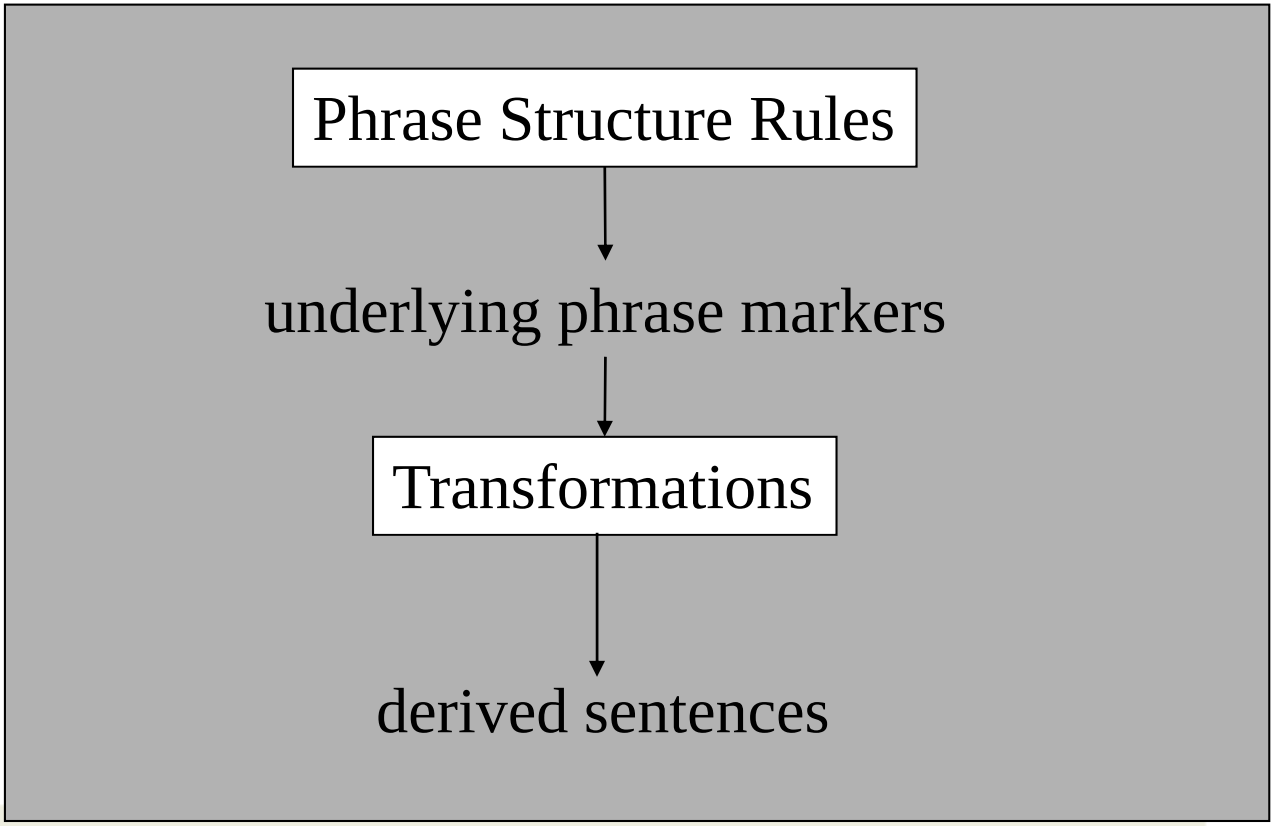
$NP_1\ V\ NP_2$

Affix-hopping

Af V ⇒ V Af



Transformational Grammar in *Syntactic Structures* (1957)



Chomsky and the first stabs at meaning

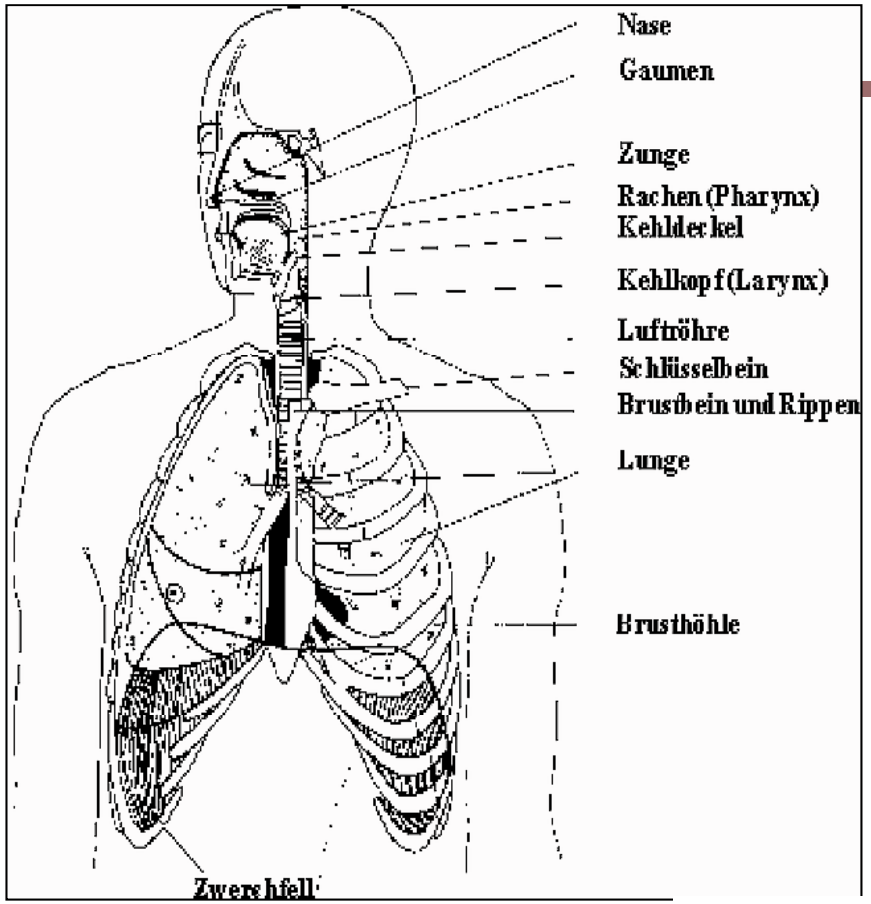
"In order to understand a sentence it is necessary to know the kernel sentences from which it originates (more precisely, the terminal strings underlying these kernel sentences) and the phrase structure of each of these elementary components, as well as the transformational [or derivational] history of development of the given sentence. The general problem of analyzing the process of 'understanding' is thus reduced...to the problem of explaining how kernel sentences are understood..."

Chomsky (1957:92) *Syntactic Structures.*

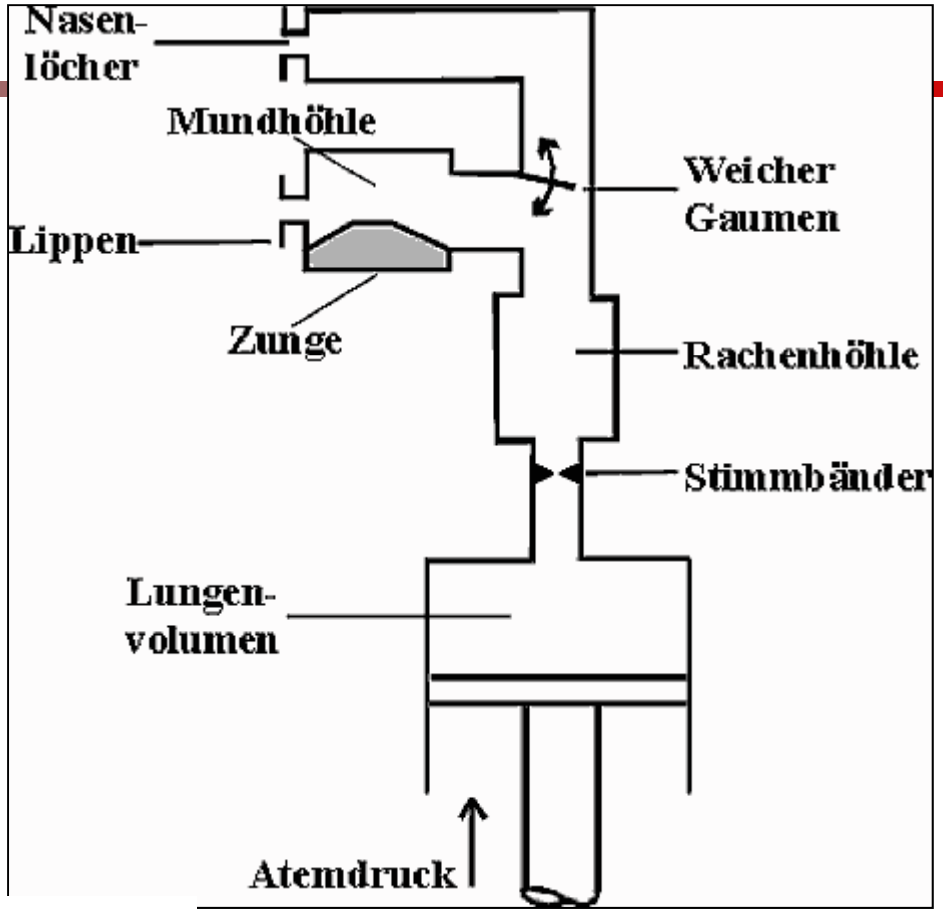
Phrase Structure Rules

make possible a very different kind of model !!

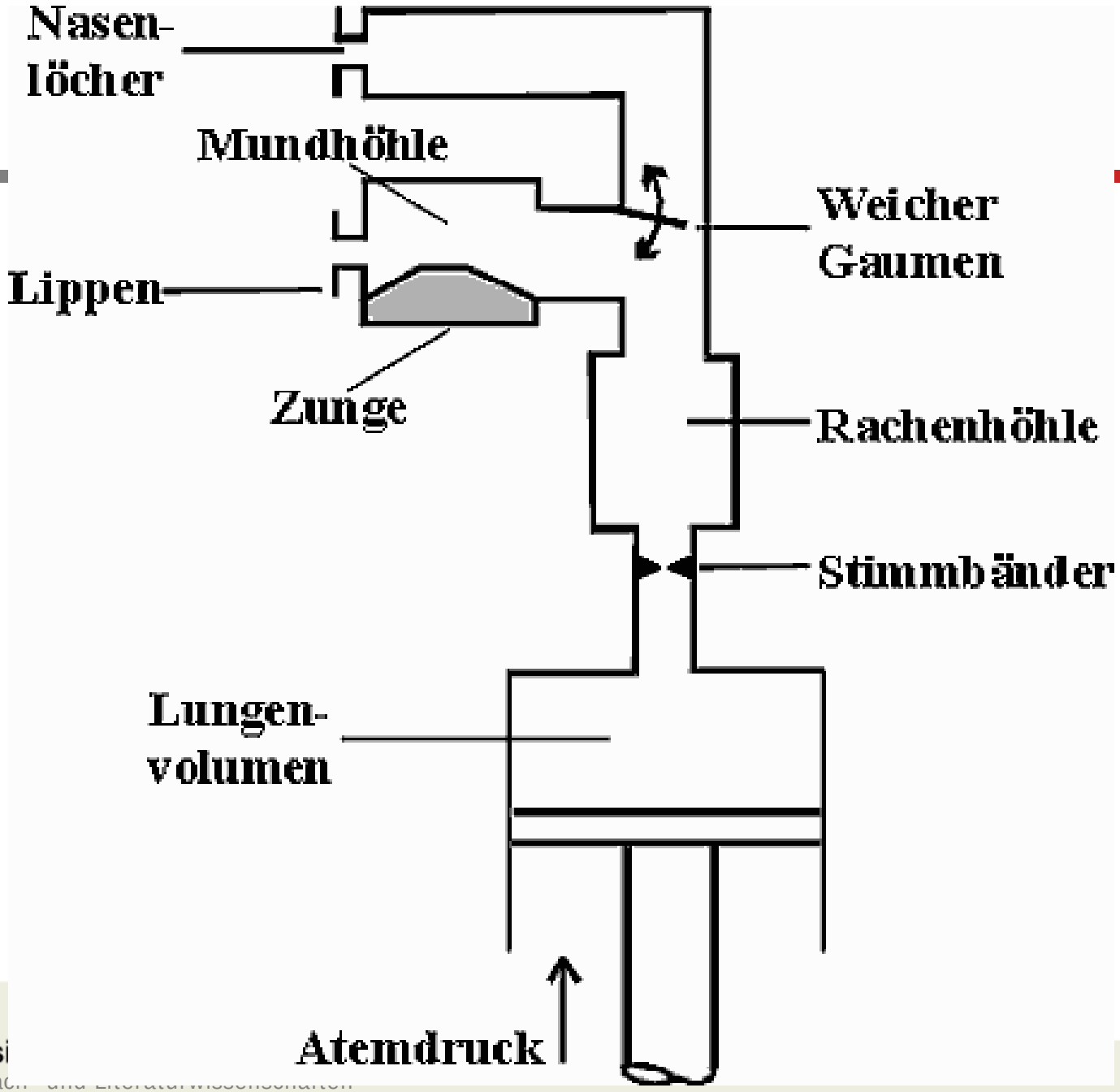
FORM



FUNCTION



models


Nasen-
löcher
Mundhöhle
Weicher
Gaumen
Lippen
Zunge
Rachenhöhle
Stimmbänder
Lungen-
volumen
Atemdruck

New Model of Syntax / Grammar

- A generative model
  - where we can follow the rules and build sentences
  - or even better, some ‘machine’ can follow the rules and builld sentences

- Noam Chomsky (1957)
  - **Autonomous Syntax**

New Model of Syntax / Grammar

- and some ‘machine’ can also do the analysis, building the structure on the basis of the rules…
  - that ‘machine’ can be a computer
  - or a brain…

Some consequences...

- **Psychology**: excited about the possibility of getting at the brain's operations via transformations
- **English Studies**: excited about the fact that meaning and grammar were now acceptable again
- **Philosophy**: excited about the mental modelling suggested

Some cracks...

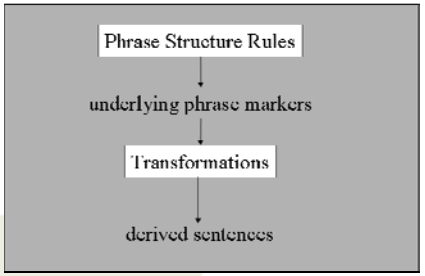
Katz and Fodor: transformations

"It would be theoretically most satisfying, if we could take the position that transformations never changed meaning."

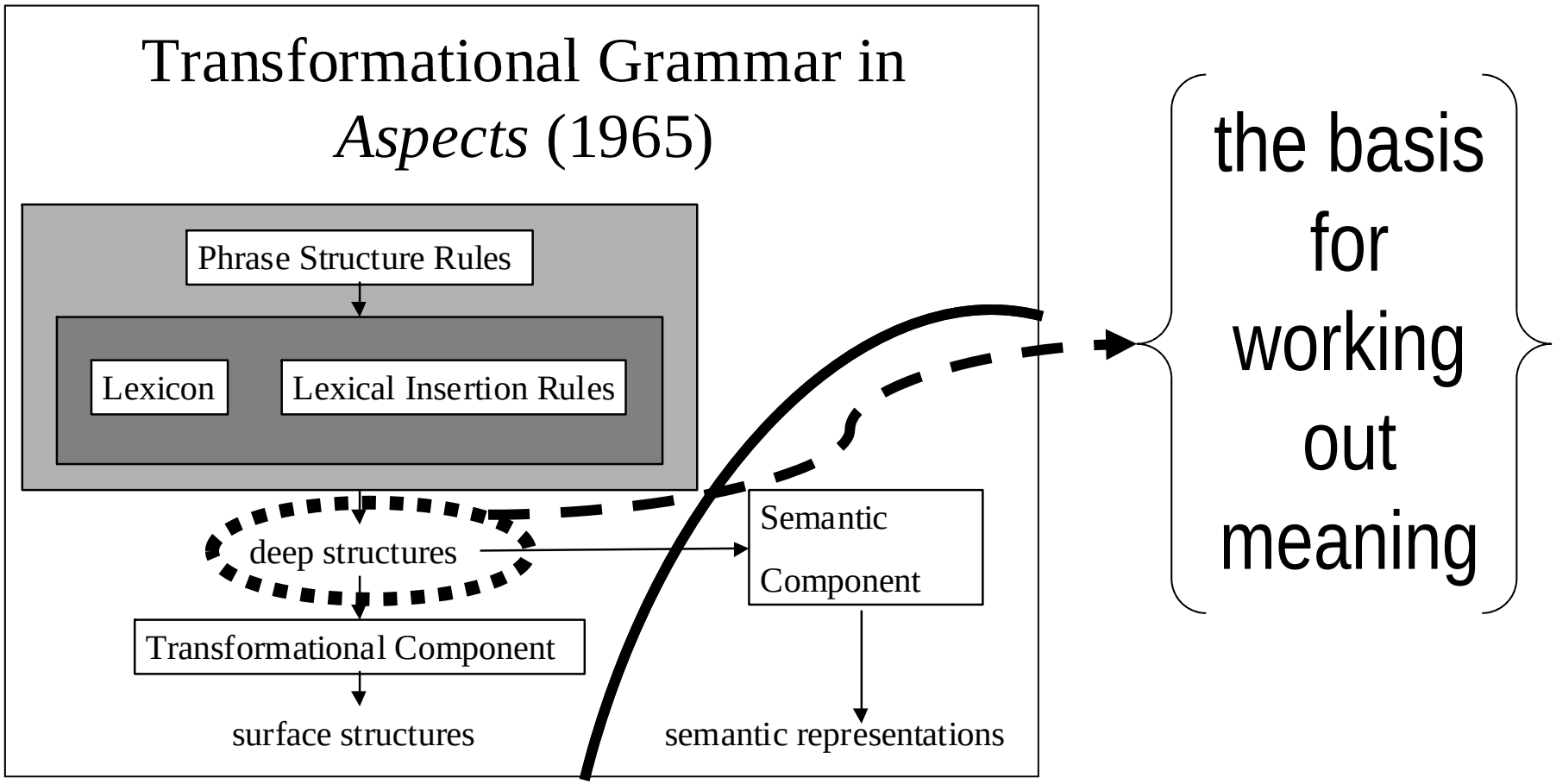
Katz and Fodor (1964: 515)

Katz and Postal: *An Integrated Theory of Linguistic Descriptions* (1964)

•the "Katz-Postal Hypothesis"



Interpretive Semantics



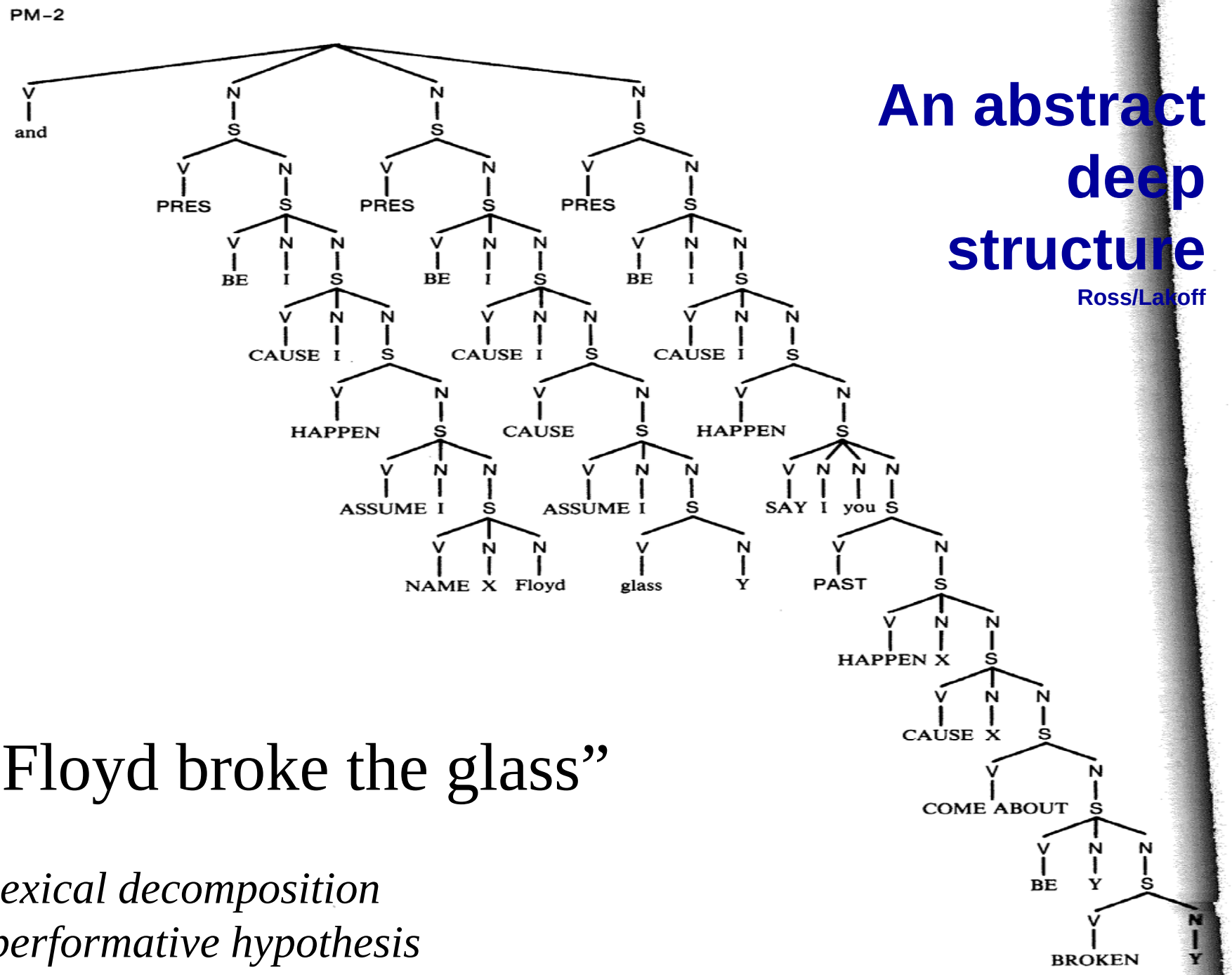
Changes in perspective from *Syntactic Structures* (1957) to *Aspects* (1965)

- Kernel to deep structure: deep structure becoming increasingly ‘abstract’

- Transformations becoming increasingly general and *nonspecific*

An abstract deep structure

Ross/Lakoff

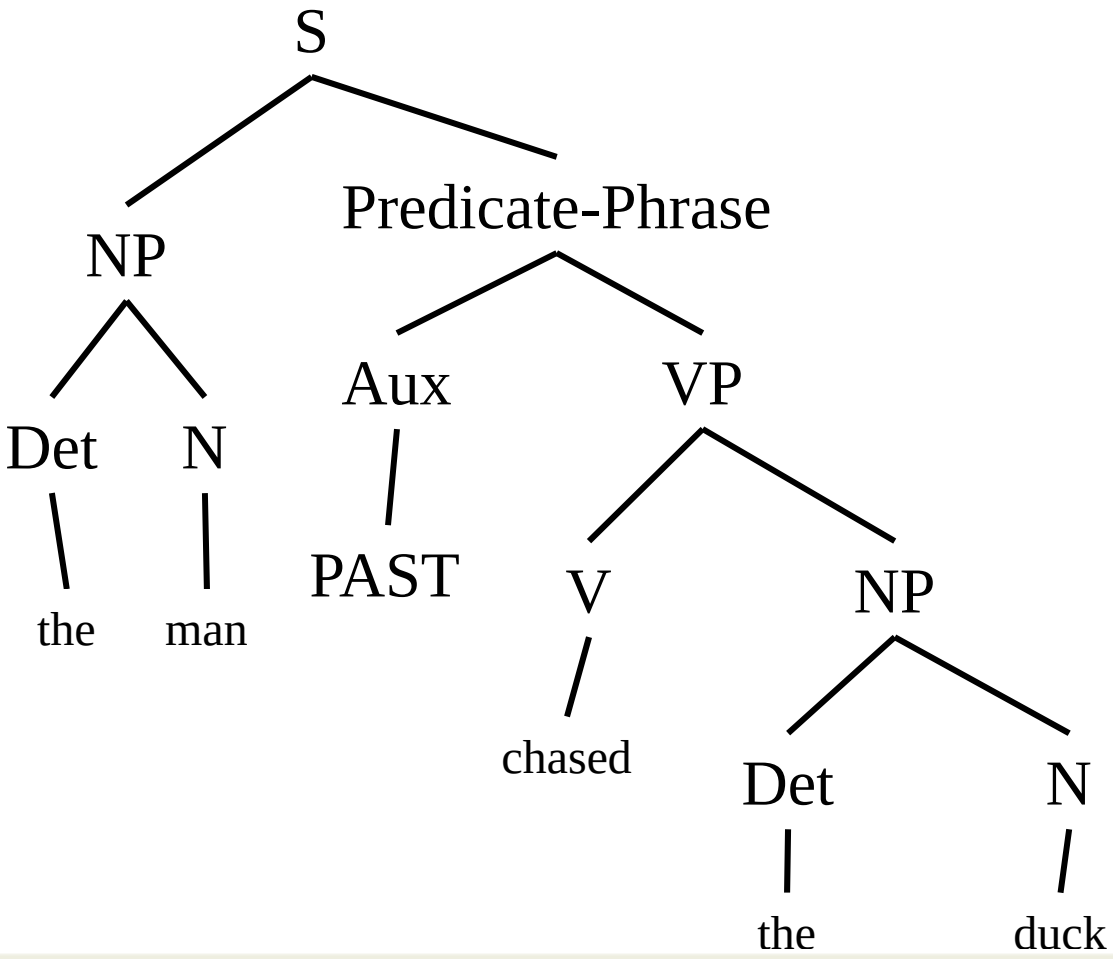


"Floyd broke the glass"

- *lexical decomposition*
- *performative hypothesis*

Meeting Logic:1

The man chased the duck

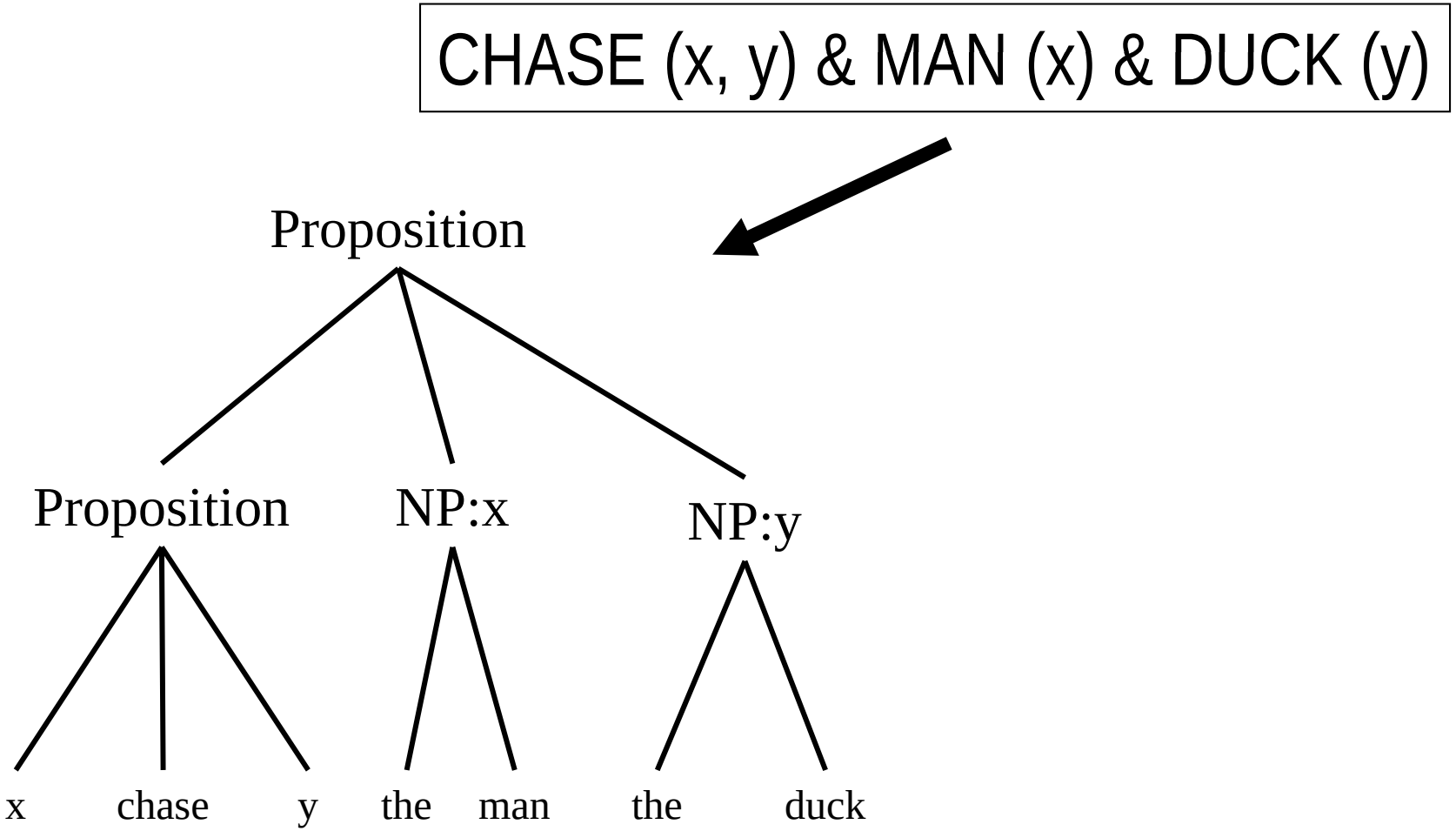


Meeting Logic:1

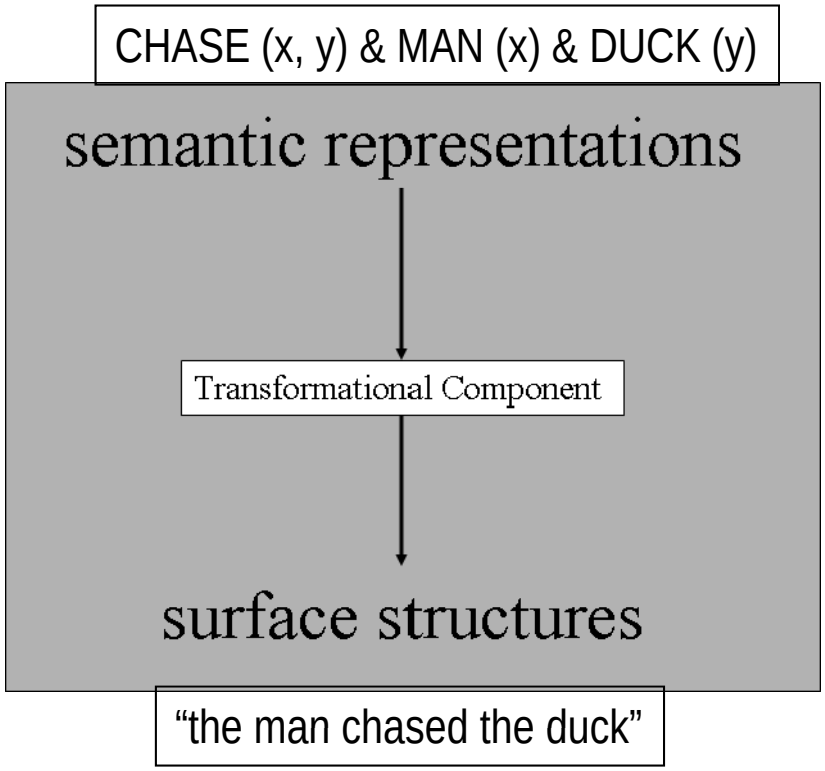
The man chased the duck

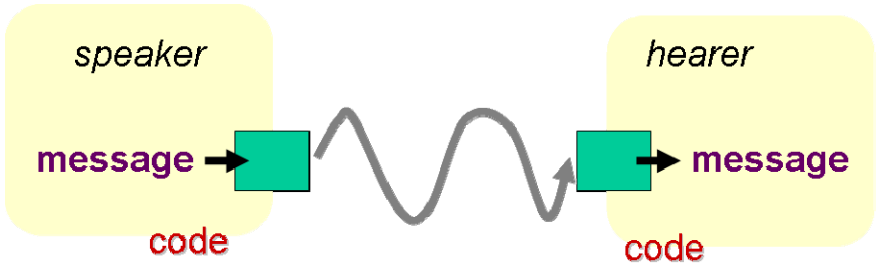
$((\text{The man})_{NP}\ (\text{PAST}\ (\text{chased}\ (\text{the duck})_{NP})_{VP})_{\text{predicate-phrase}}$

Meeting Logic:2



Generative Semantics



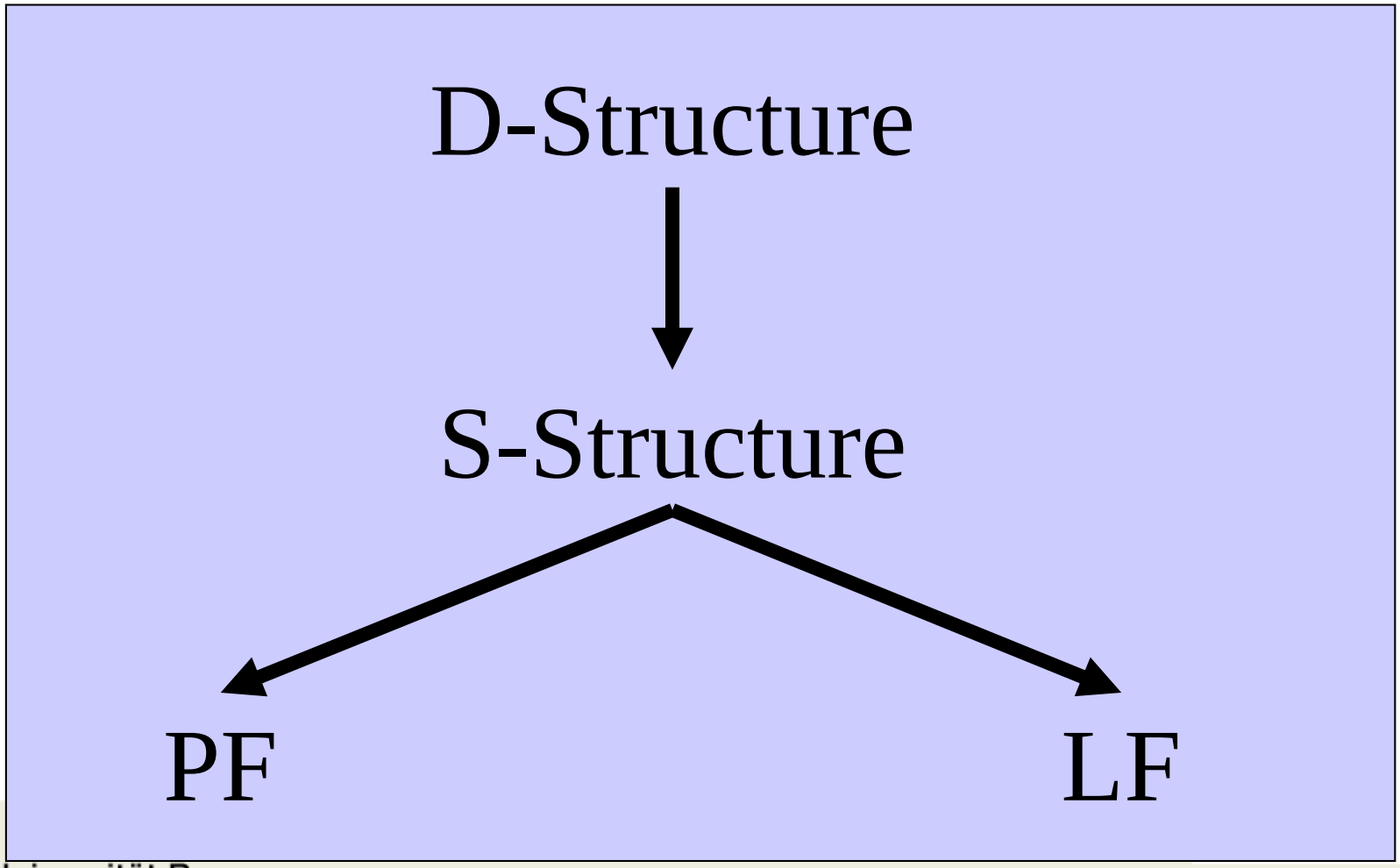


isn’t that just what we need?

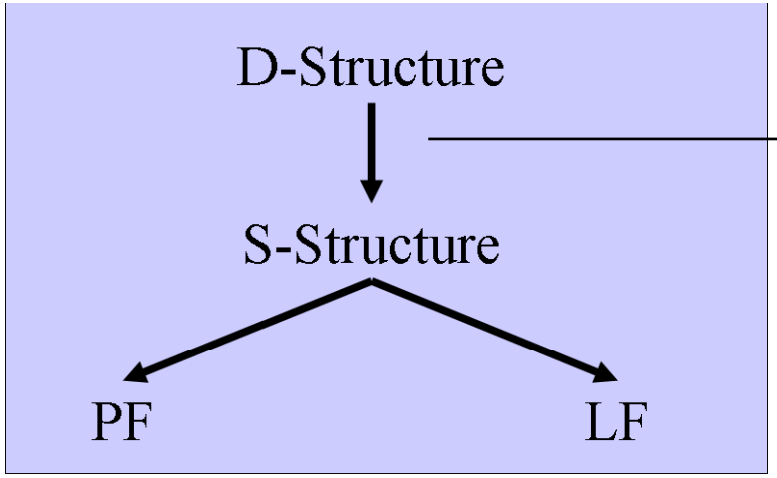
Early Problem

- Transformations clearly **could** often change meaning….
  - negation
  - passive
    - “Two languages are spoken by everyone in this room”
    - “Everyone in this room speaks two languages”

Chomsky: Extended Standard Theory



Deeper Problems and the beginnings of computational linguistics



the early accounts of these formal relationships turned out to be extremely unpleasant **computationally**

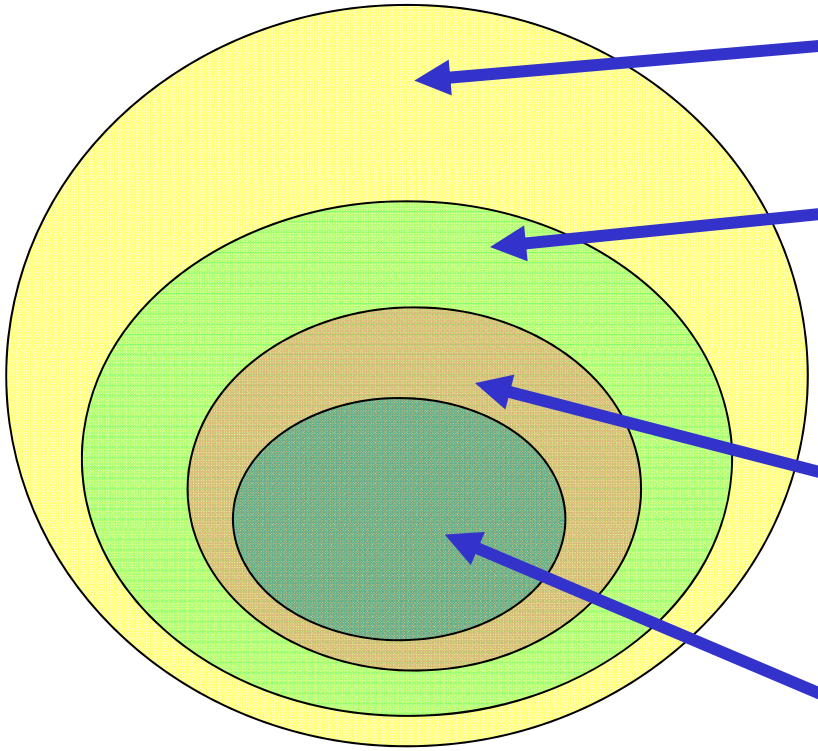
A rule of the form

(25) $Z \hat{} X \hat{} W \rightarrow Z \hat{} Y \hat{} W$

indicates that X can be rewritten as Y only in the context Z--W. It can easily be shown that the grammar will be much simplified if we permit

Peters / Ritchie (1973)
"On the generative power of transformational grammar"

‘Chomsky’ Hierarchy



Type 0 ***TRANSFORMATIONAL GRAMMAR***

Type 1 : context-sensitive

$$\alpha A \beta \rightarrow \alpha \gamma \beta$$

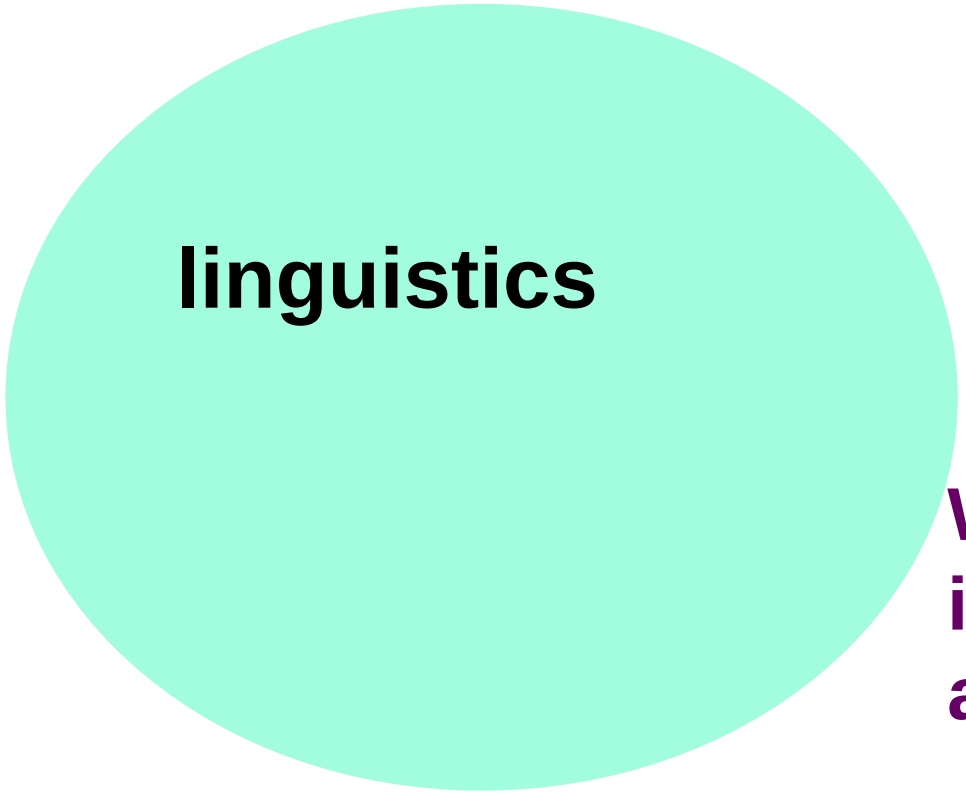
Type 2 : context-free

$$A \rightarrow \gamma$$

Type 3 : regular

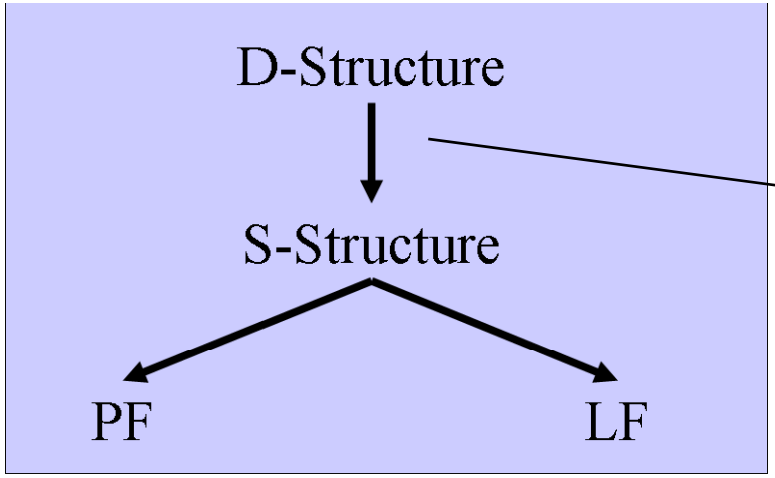
$$A \rightarrow a$$
$$A \rightarrow aB$$

Computational Linguistics ?



What kind of thing is language anyway?

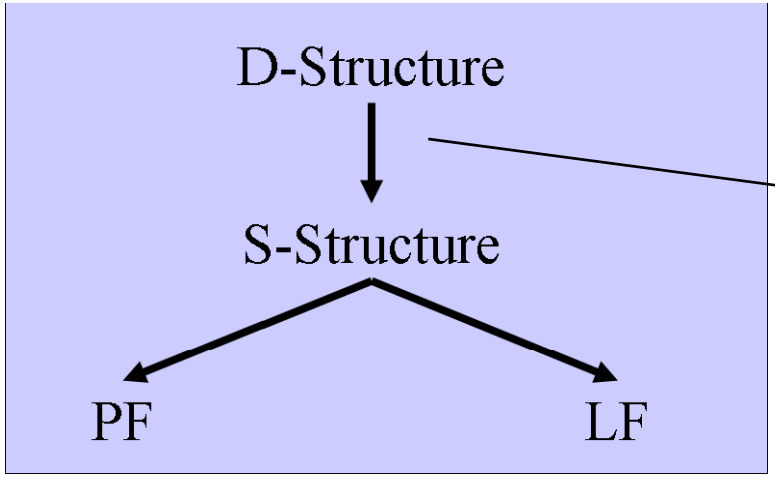
Deeper Problems and the beginnings of computational linguistics



can we specify the necessary relationships ***without*** requiring a formalism that dives off the deep end into full Turing equivalence?

LFG, TAG, CCG, …

Deeper Problems and the beginnings of computational linguistics



can we perhaps do **without these kind of grammars** altogether???

Genus: Homo

Years ago
3 million
1.7 million
700 thousand
130 thousand

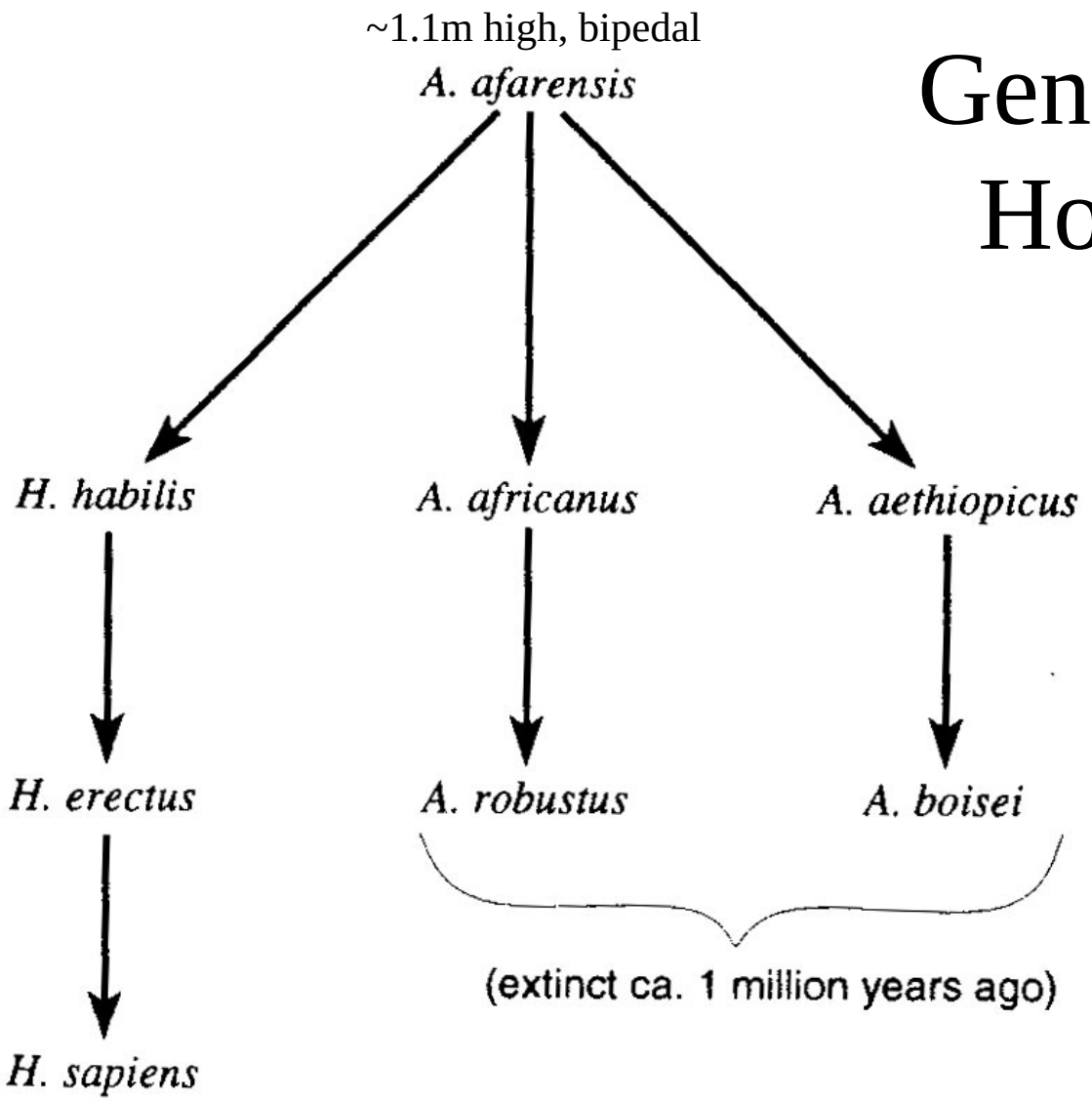
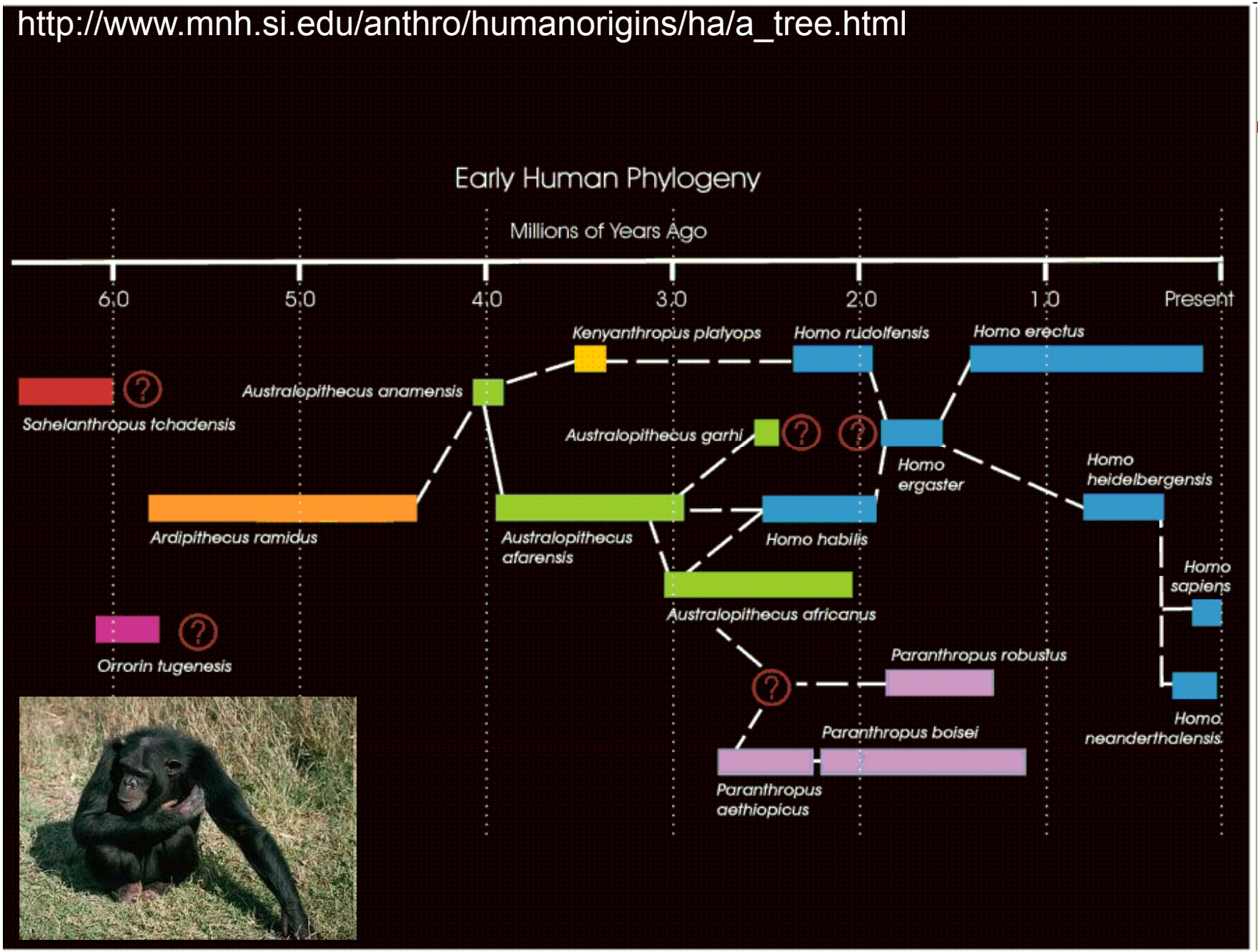


Fig. 2.1.1 Common part of four phylogenies proposed for the descent of humans from Australopithecines (after Johanson 1989).

Smithsonian

http://www.mnh.si.edu/anthro/humanorigins/ha/a_tree.html



More modern studies

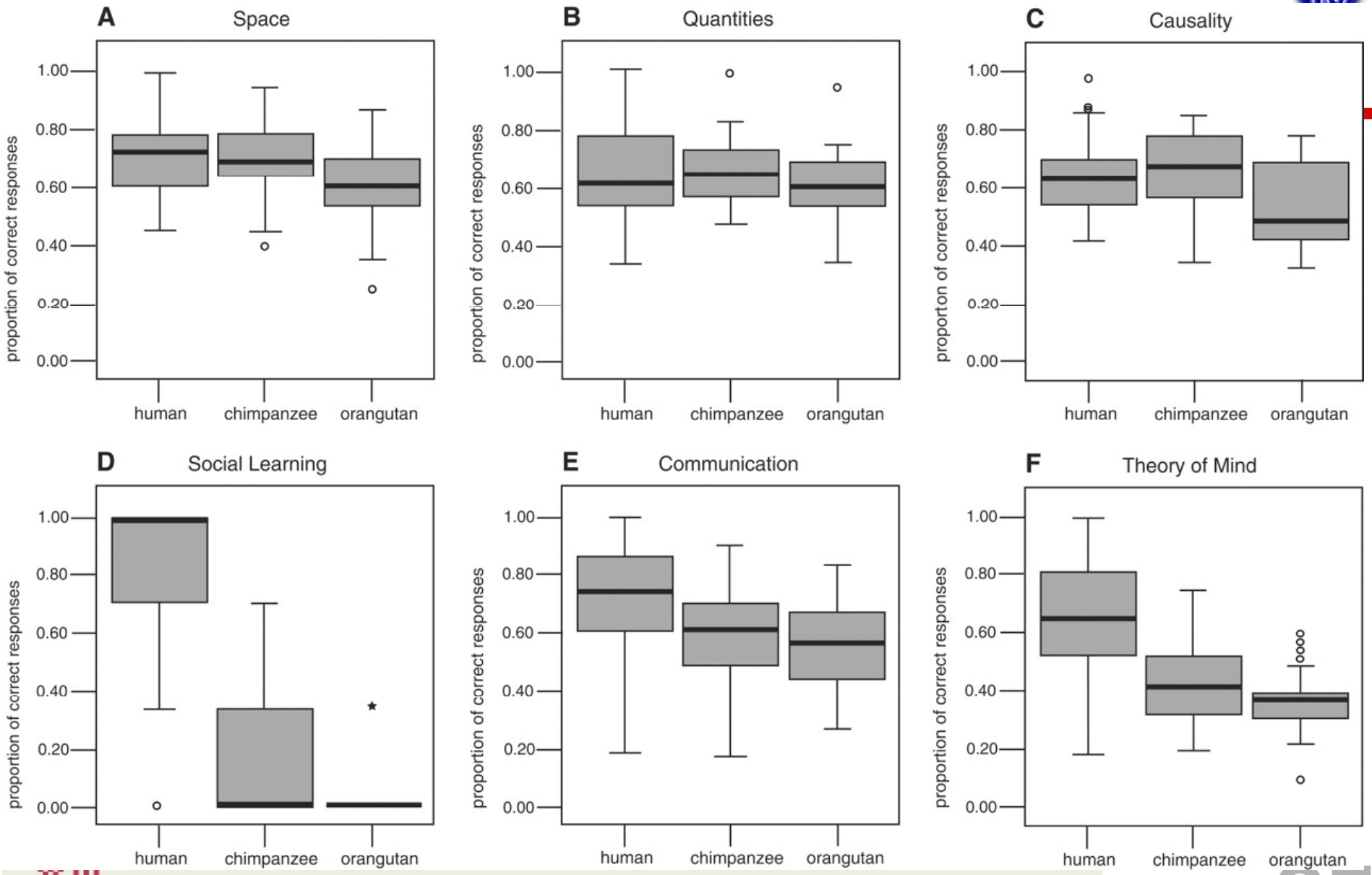
- Max Planck Institute for evolutionary anthropology in Leipzig:
  Tomasello and colleagues

  - trying to discover more finely just what skills go into language behaviour…
  - comparative experiments with young apes and children
  - just what are the abilities that are necessary for language acquisition? where are the differences?

MPI
evolutionary anthropology (Leipzig)

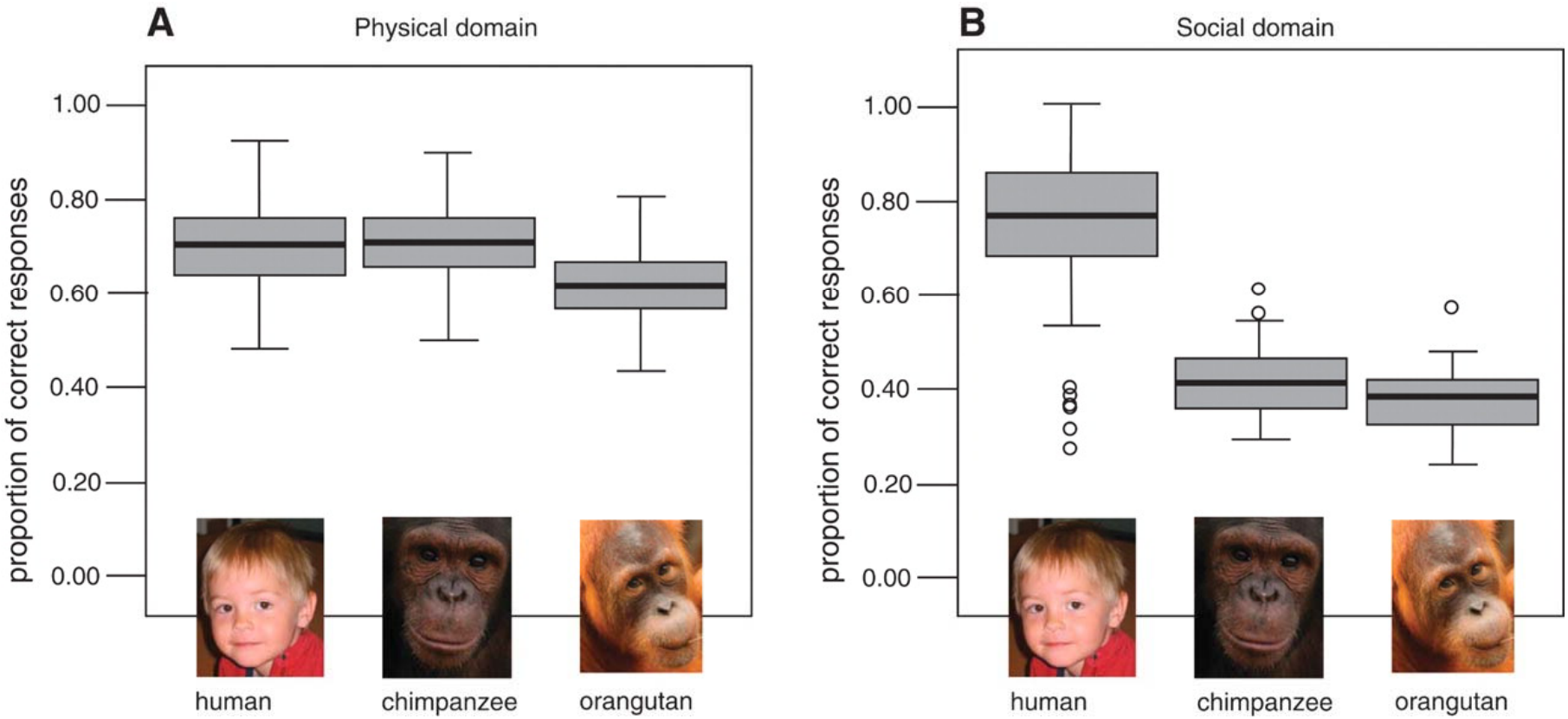
- tool use
- physical properties
- communication

- deception



Hermann et al. (2007) Humans Have Evolved Specialized Skills of Social Cognition: The Cultural Intelligence Hypothesis. *Science* 3

Comparisons



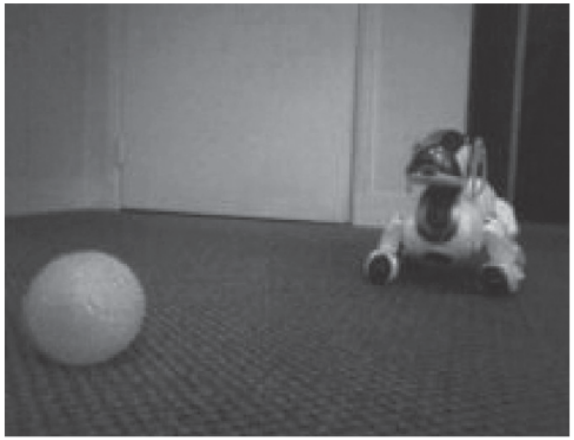
Hermann et al. (2007) Humans Have Evolved Specialized Skills of Social Cognition: The Cultural Intelligence Hypothesis. *Science* 3

Language as an Emergent Phenomenon

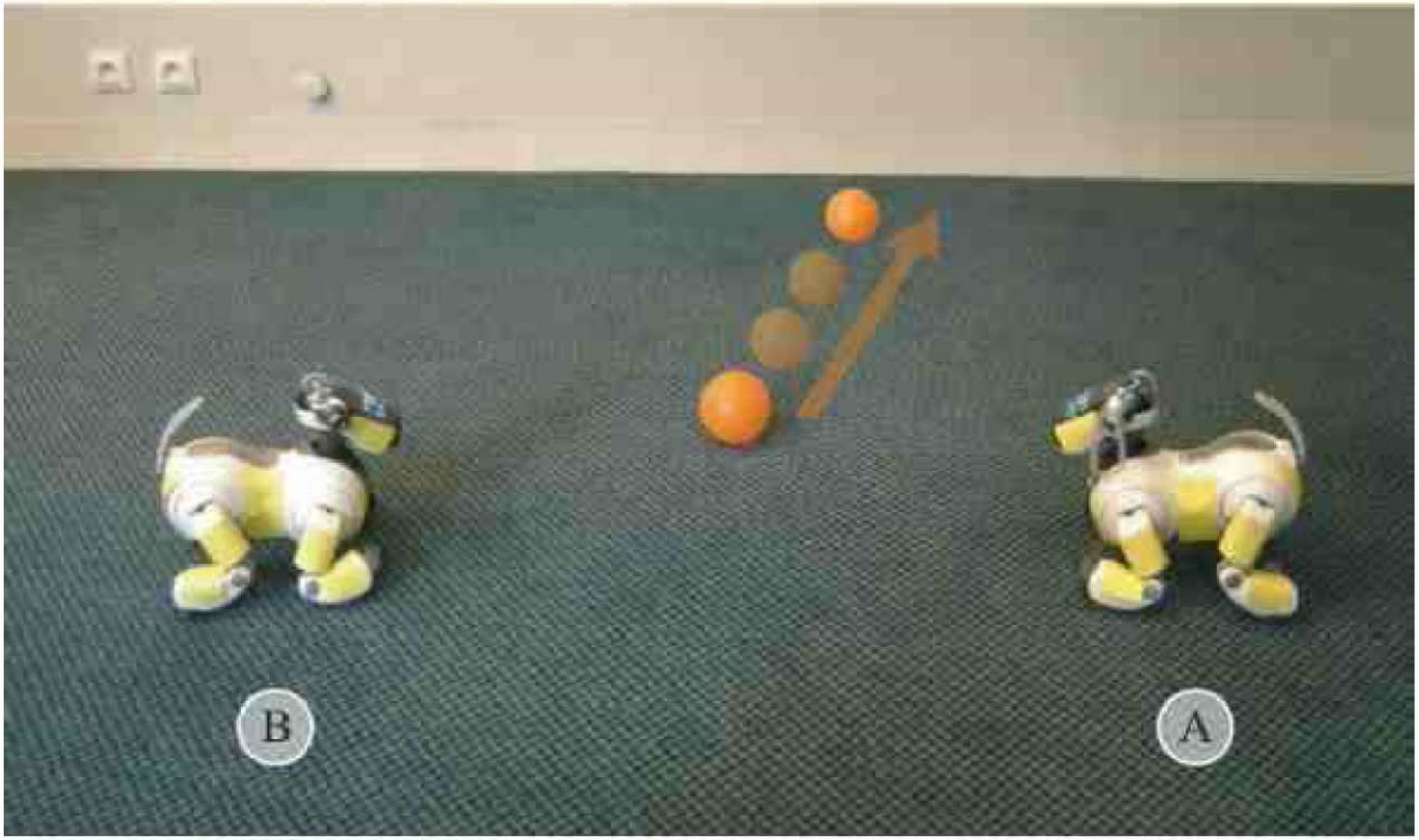
- Luc Steels and colleagues
  - How much can language systems establish themselves out of the constraints of goal-directed interaction?

(a)

(b)



Luc Steels: Emergence



L. Steels
M. Loetzsch

K. Coventry
T. Tenbrink
J. Bateman, (eds.)
(2009) *Spatial Language in Dialogue.*
Oxford University Press, Oxford

1: Agents embodied in physical robots. The speaker (in robot A) and
ırer (in robot B) together observe ball movement events and then play
ıage game to describe the scene to each other.

Convergence of Lexicons (Steels)

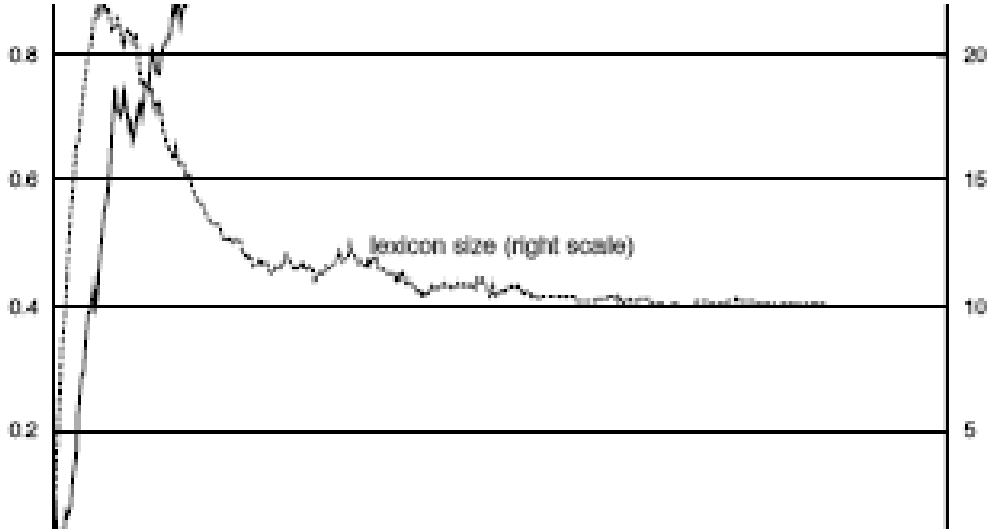


Fig. 1. A population of 5 agents plays a sequence of 5000 naming games, naming 10 objects. We see that the lexicon climbs up at first, reaching a peak of almost 25 after about 200 games. It then settles quickly to an optimum of 10 names for 10 objects thanks to the lateral inhibition dynamics.

Convergence of Ontologies (Steels)

L. Steels
M. Loetzsch

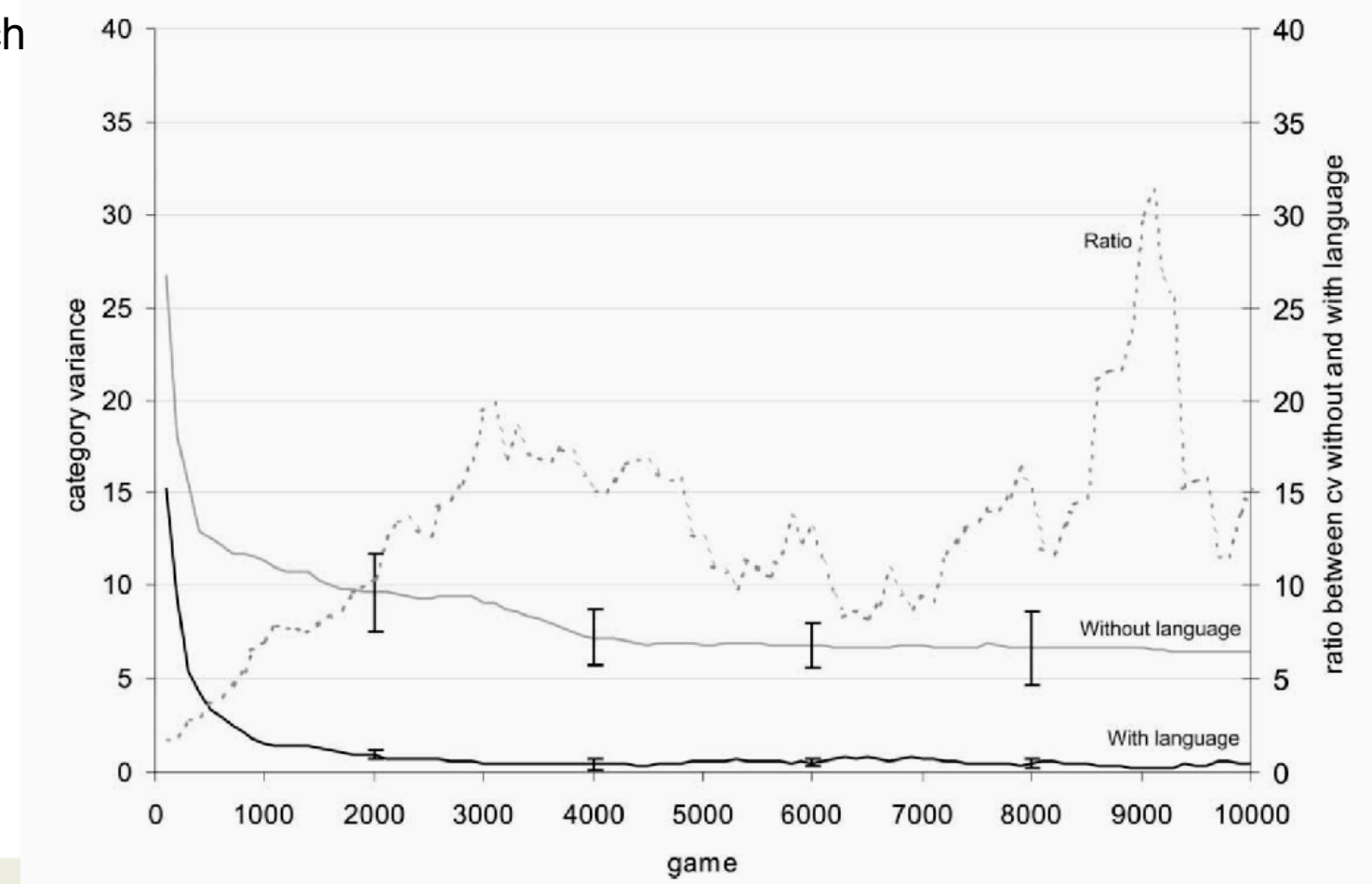


Figure 4. The graph plots the cumulative category variance between the ontologies in a population of agents with and without language. Agents play discrimination and interpretation games and construct an ontology as a side-effect. When there is no co-ordination through language, the ontologies do not converge, otherwise they do.

Perspective in Spatial Language

L. Steels
M. Loetzsch

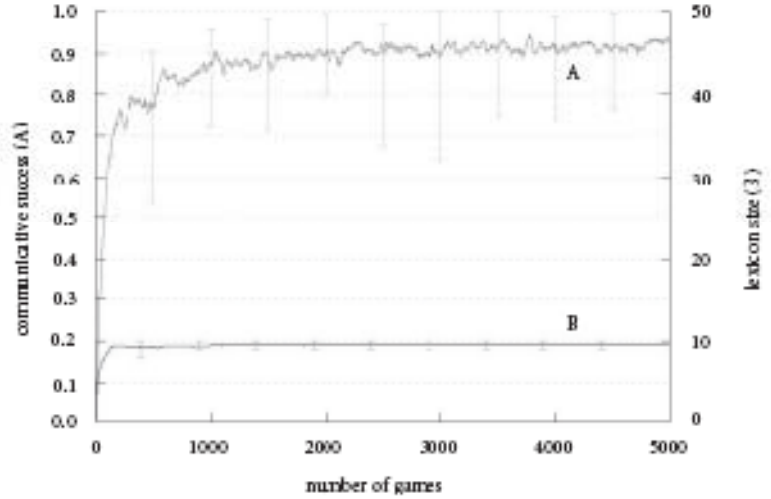


FIG. 6.9. Agents have the same sensory information and hence share their situation model. They quickly self-organize a lexicon and ontology.

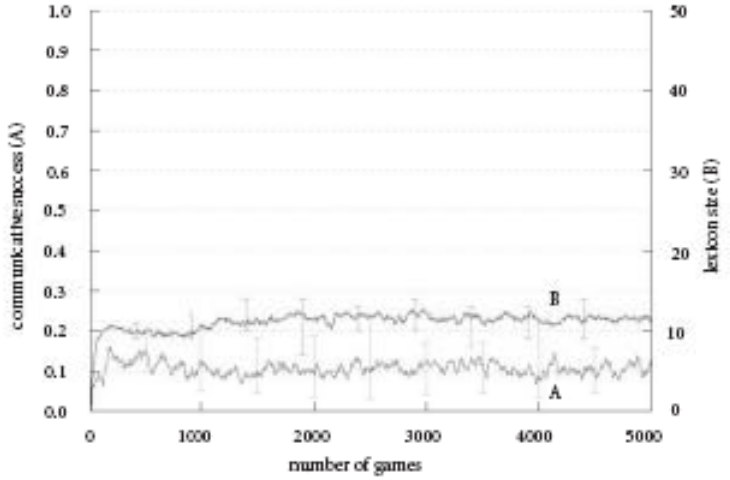


FIG. 6.10. Agents do not share sensory stimuli and do not consider perspective. The system does not come off the ground.

K. Coventry, T. Tenbrink, J. Bateman, (eds.) (2009) *Spatial Language in Dialogue.* Oxford University Press, Oxford

Remaining open problems

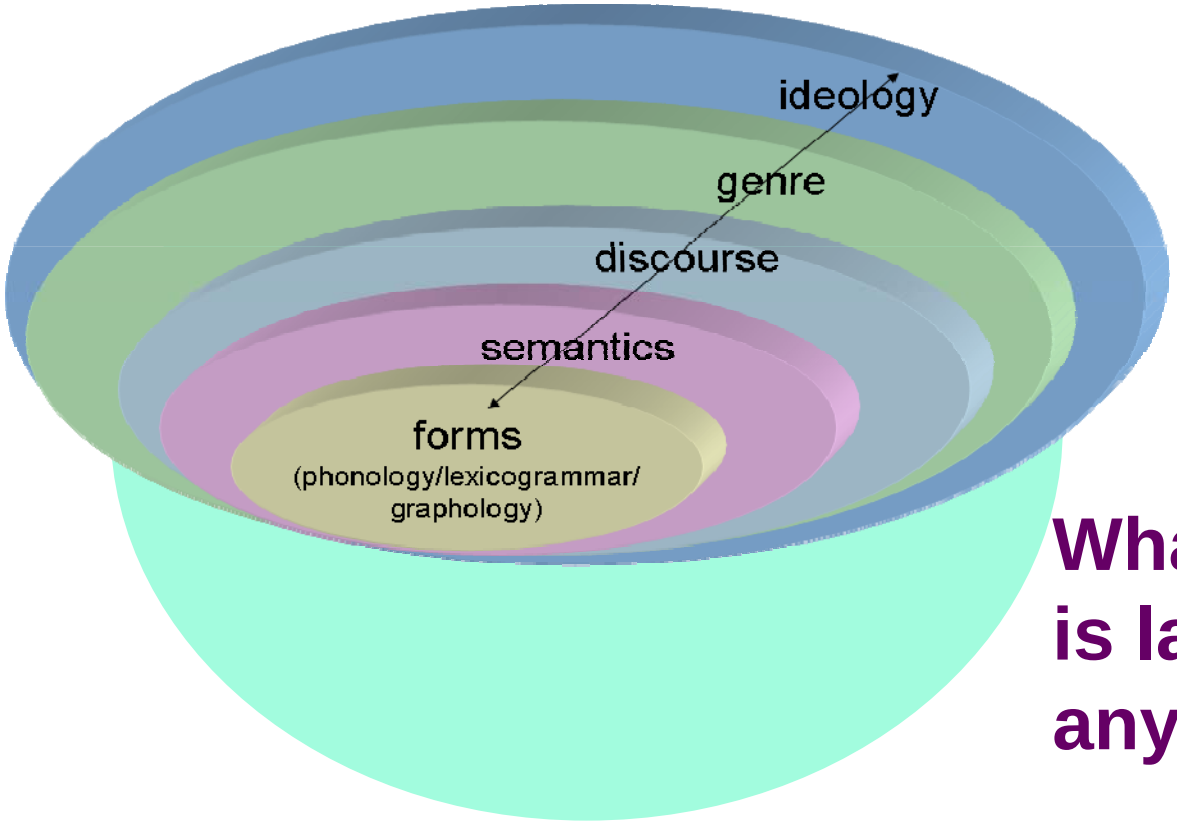
- Are the constraints of the communicative situation really sufficient to bring about the kind of grammatical organisations that we find?
  - e.g., probably mildly context sensitive: TAGs, CCGs, etc.

Steedman: Affordances, Plans …

- This problem is often rejected in usage-based accounts.
- Tomasello (2003: 301): *Constructing a language*

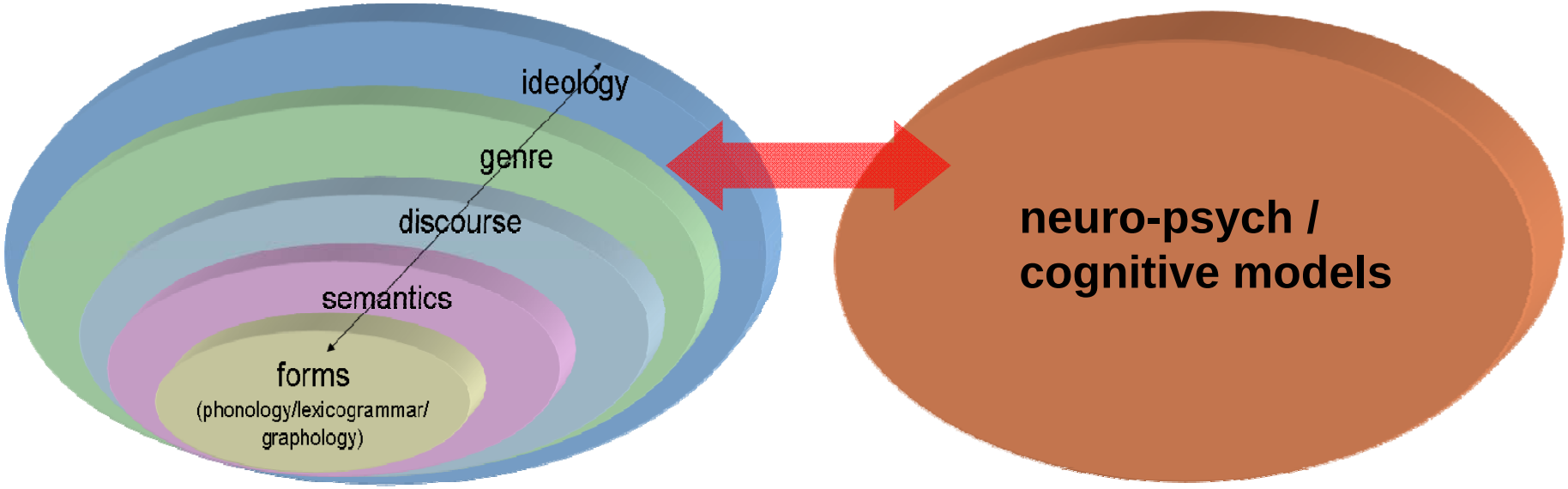
> "It is probably true that constraining linguistic generalizations appropriately is the aspect of the current usage-based theory that, at the moment, is its weakest link."

Computational Linguistics



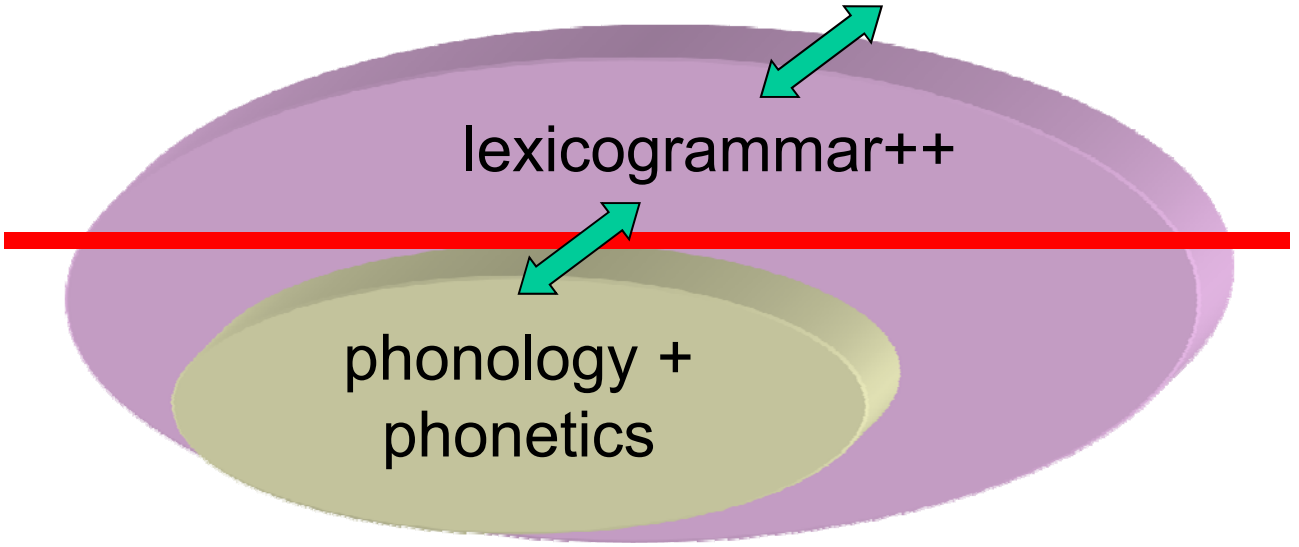
What kind of thing is language anyway?

Computational Linguistics



Language is ‘doubly articulated’ (Martinet)

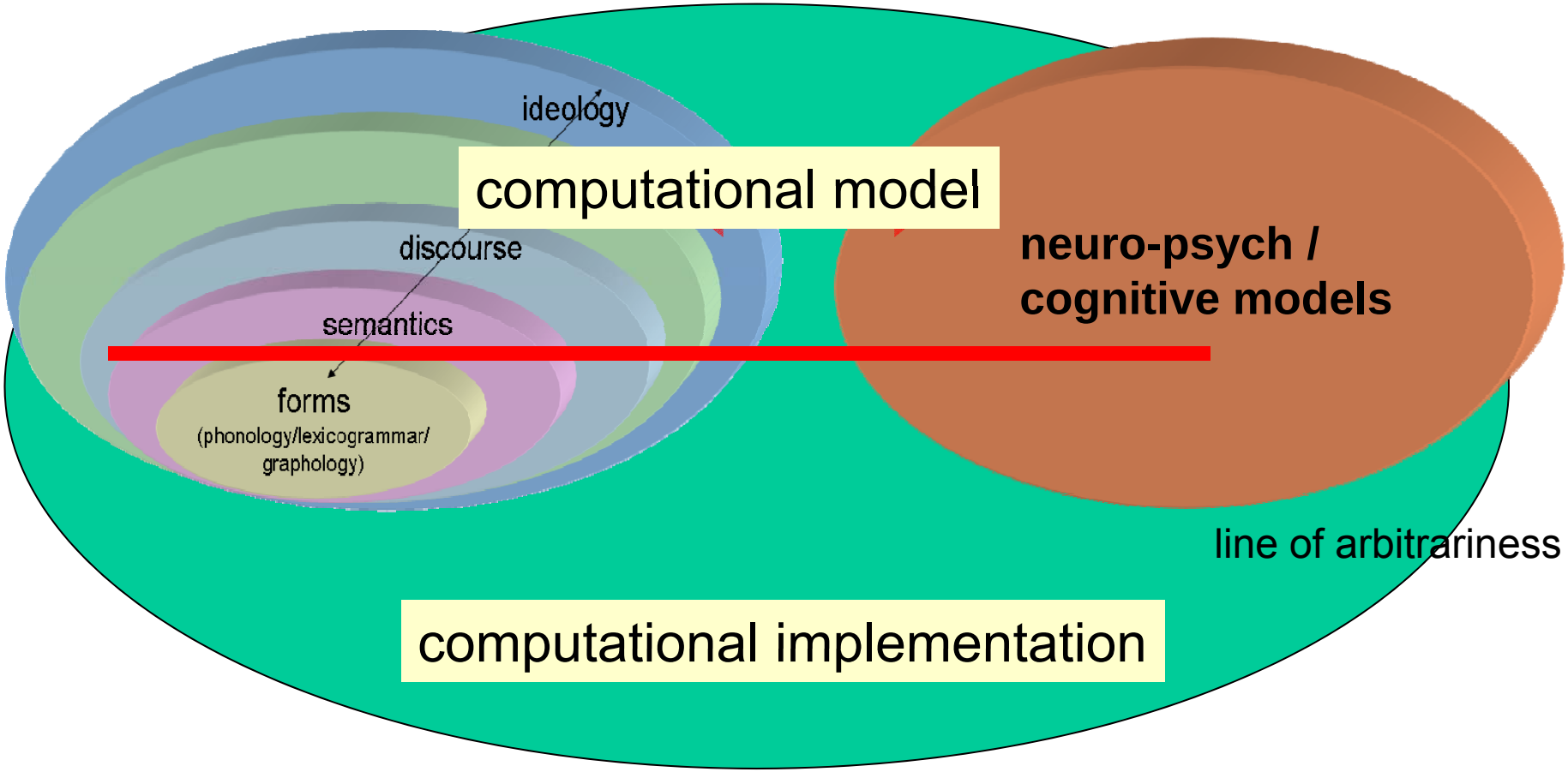
an analogy



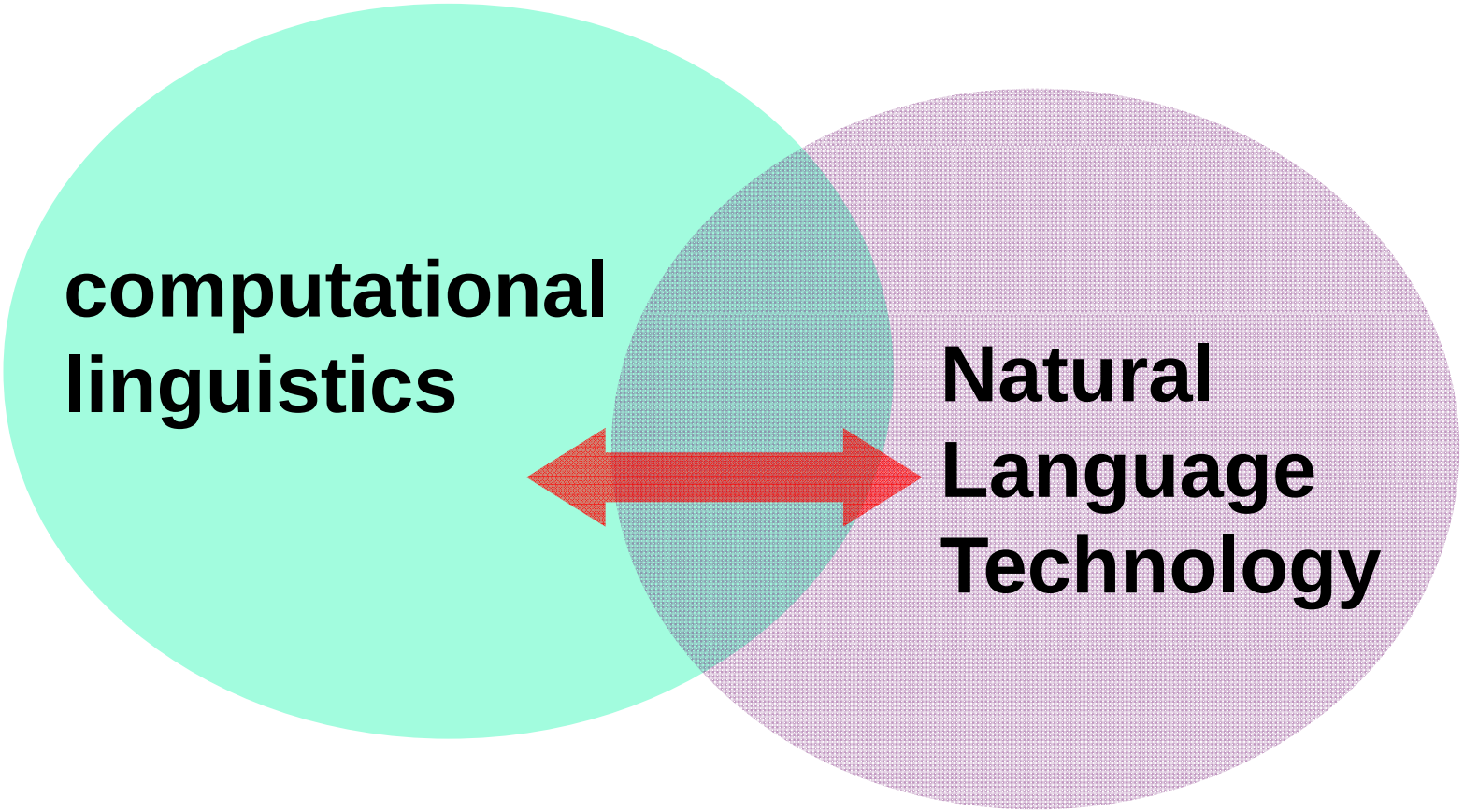
line of arbitrariness

…, /k/, /ɪ/, /t/, …

Computational Linguistics



Natural Language Technology



Profile: Computational Linguistics

- Essentially an integrative enterprise
  - diverse areas from **linguistics**:
    - syntax, semantics, phonetics, pragmatics, lexicon
  - diverse area from **informatics**:
    - logic, programming, knowledge representation, parsing methods, data structures,

Levels of language technology

- Phonetic / Phonological → speech processing / speech synthesis
- Syntactic / Grammatical → tactical generation / parsing technology
- Semantics → reasoning, logic, knowledge representation, ontologies
- Pragmatics → dialogue acts, dialogue processing

Jobs for linguists....

- Computational linguistics is in many respects a ‘booming’ field
  - many application areas
  - requiring:
    - detailed theoretical knowledge about distinct aspects of linguistics
    - detailed theoretical knowledge about distinct aspects of computation
    - practical knowledge about putting these together into working systems

Computational Linguistics

- History
  - what is computational linguistics? ✓
  - where did it come from? ✓
  - what does it have to do with linguistics? ✓
- To come: Theoretical + Practical Work
  - some frameworks and basic tools used in computational linguistics
  - grammars and semantics
  - some examples: analysis, generation, dialogue

Accompanying Reading

- Dale / Moisl / Somers (Hrgg)(2000) *A handbook of natural language processing: techniques and applications for the processing of language as text.* Dekker.

  Chapters:

  Preface, pp.iii-viii
  5. Semantic Analysis (Poesio)
  7. Natural Language Generation (McDonald)
  13. Machine Translation (Somers)
  14. Dialogue Systems: from theory to practice (Allen)

- Mitkov (Hrg)(2002) *Oxford Handbook of Computational Linguistics.* OUP.

  Chapter:

  25. Ontologies (Vossen)

- Meyer (2002) *Synchronic English Linguistics: An introduction.* GNV.

  Chapter:

  4.5 Formal Semantics (pp118—132)