

Hybrid NLP

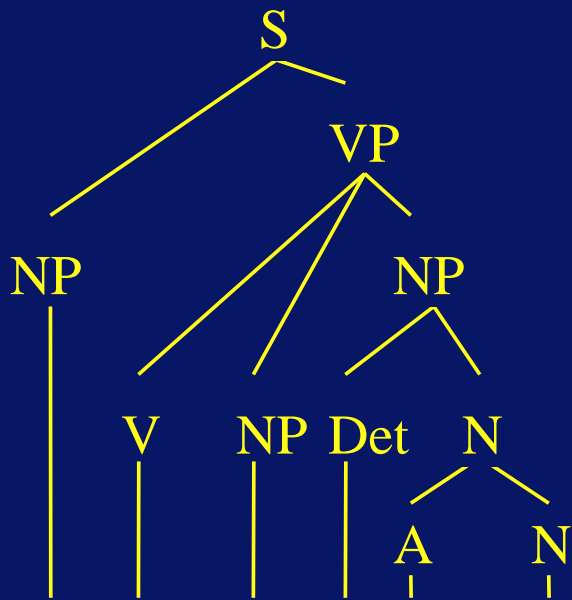
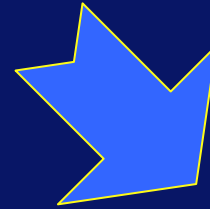
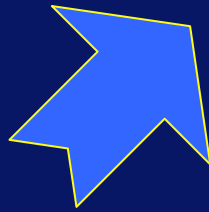
OUTLINE

- **Problems of Deep and Shallow Processing**
- **Hybrid Architectures**
- **An Advanced Platform for Hybrid NLP: Deep Thought**
- **Applications for Hybrid Processing**
- **Conclusion and Outlook**

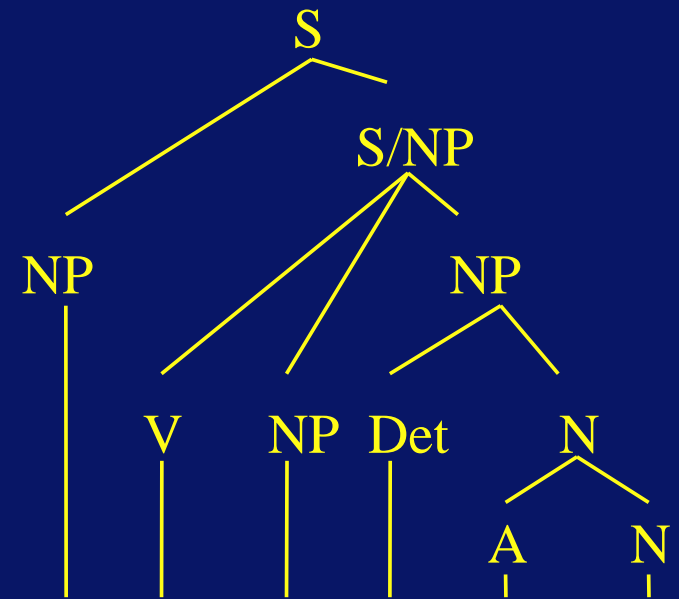
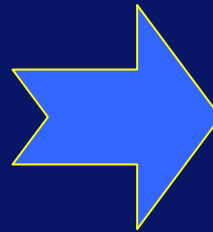
DEEP & SHALLOW PROCESSING

- **deep methods** for morphological - syntactic - semantic processing exploit our knowledge about the structure of human language
- as opposed to **shallow methods** such as pattern matching grammars, n-gram language models
- **deep methods** are needed for getting at the meaning of language input
- **shallow methods** perform a partial or heavily under-specified analysis sufficient for certain applications

$\exists x[(\text{old}'(\text{penny}'))(x) \wedge (\text{Past}(\text{give}'(\text{sue}', \text{paul}', x)))]$



Sue gave Paul an old penny.



Sue gab Paul einen alten Pfennig.

ONCE UPON A TIME

■ Broad industrial research in deep parsing

- Xerox - LFG
- Siemens - LFG
- IBM Germany - HPSG
- Hewlett Packard - GPSG and HPSG
- IBM USA - PLNLP and Slot Grammar

■ Very large projects

- EUROTRA
- LILOG
- LS-GRAM

GRAMMAR FRAMEWORKS

- **Head-Driven Phrase Structure Grammar (HPSG)**
- **Lexical Functional Grammar (LFG)**
- **Tree-Adjunction Grammar (TAG)**
- **Categorial Grammar (CG)**
- **Dependency Grammar (DG)**
- **GB-Minimalist Program**

HPSG

- **Head-Driven Phrase Structure Grammar by Pollard and Sag**
- **Uniform formalism: typed feature structures**
- **High degree of lexicalization: very few PS-rules, rich lexicon structure**
- **Ontological structure: Multiple inheritance type hierarchy**

Problems with Deep Analysis

- **Coverage (Development Time)**
- **Robustness (Coping with Out-of-Grammar Input)**
- **Efficiency (Runtime and Space Efficiency)**
- **Specificity (Selection among Readings)**

Problems with Shallow Analysis

■ Accuracy

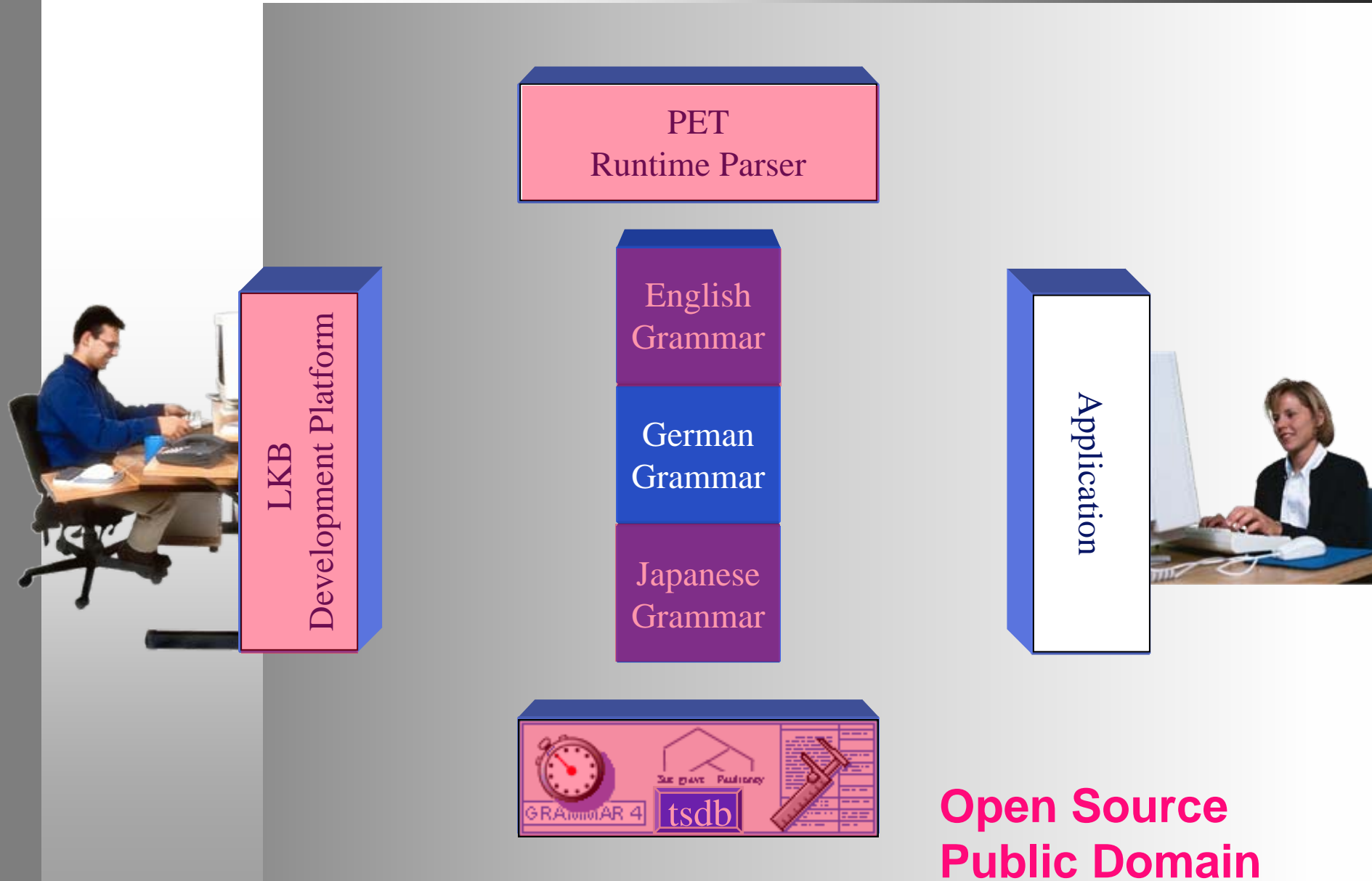
- Problems with embeddings, grammatical control, anaphora and modal as well as negative contexts.

According to SVP Raul Lopez, Slator expected him to be appointed CEO of Crawford Inc. at the upcoming share holders meeting.

After the retirement of Peter Smith, Mary Hopp was introduced by VP Brown as the new director of the marketing division.

After every former US based vicepresident except Lisa Ronell served as Chairman of the Board, the shareholders for the first time appointed a non-US Chairperson.

THE CORE MACHINERY



HOWEVER

- **Back to the problems of**
 - **robustness**
 - **coverage**
 - **specificity**

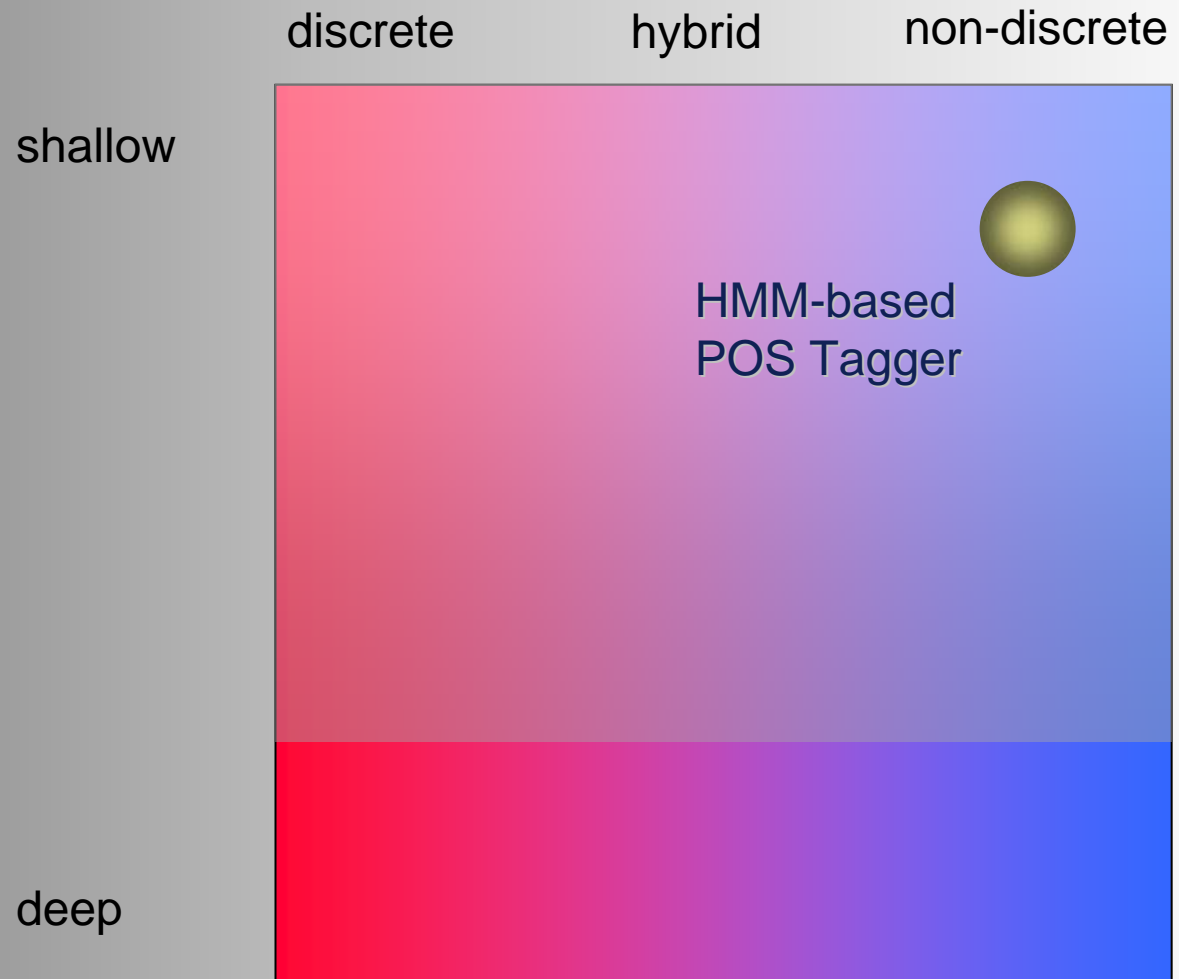
ASSUMPTIONS

- **Information extraction is not an alternative to deep processing but a continuum between classification and "full" semantic analysis**
- **Information Extraction via Text Enrichment**
- **We can detect topics, names, binary relations, complex relations, answers, etc.**
- **Question: At what point is deep processing needed?**

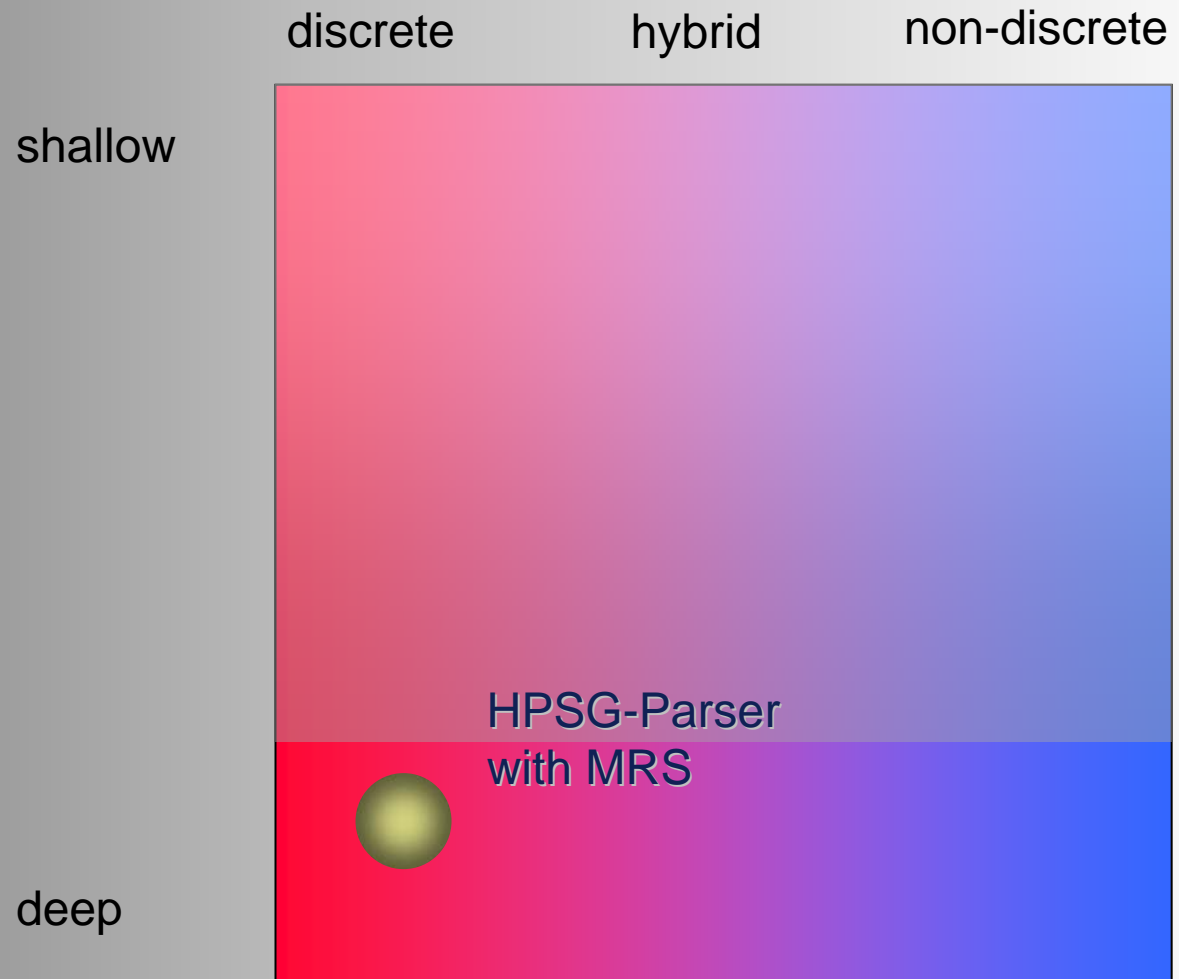
APPROACH

- **Lack of robustness and coverage remains a serious problem for deep processing.**
- **So we need to find applications, where deep processing can improve detection without spoiling the performance.**
- **Example: Relation extraction.**
- **Let deep processing assist shallow methods.**

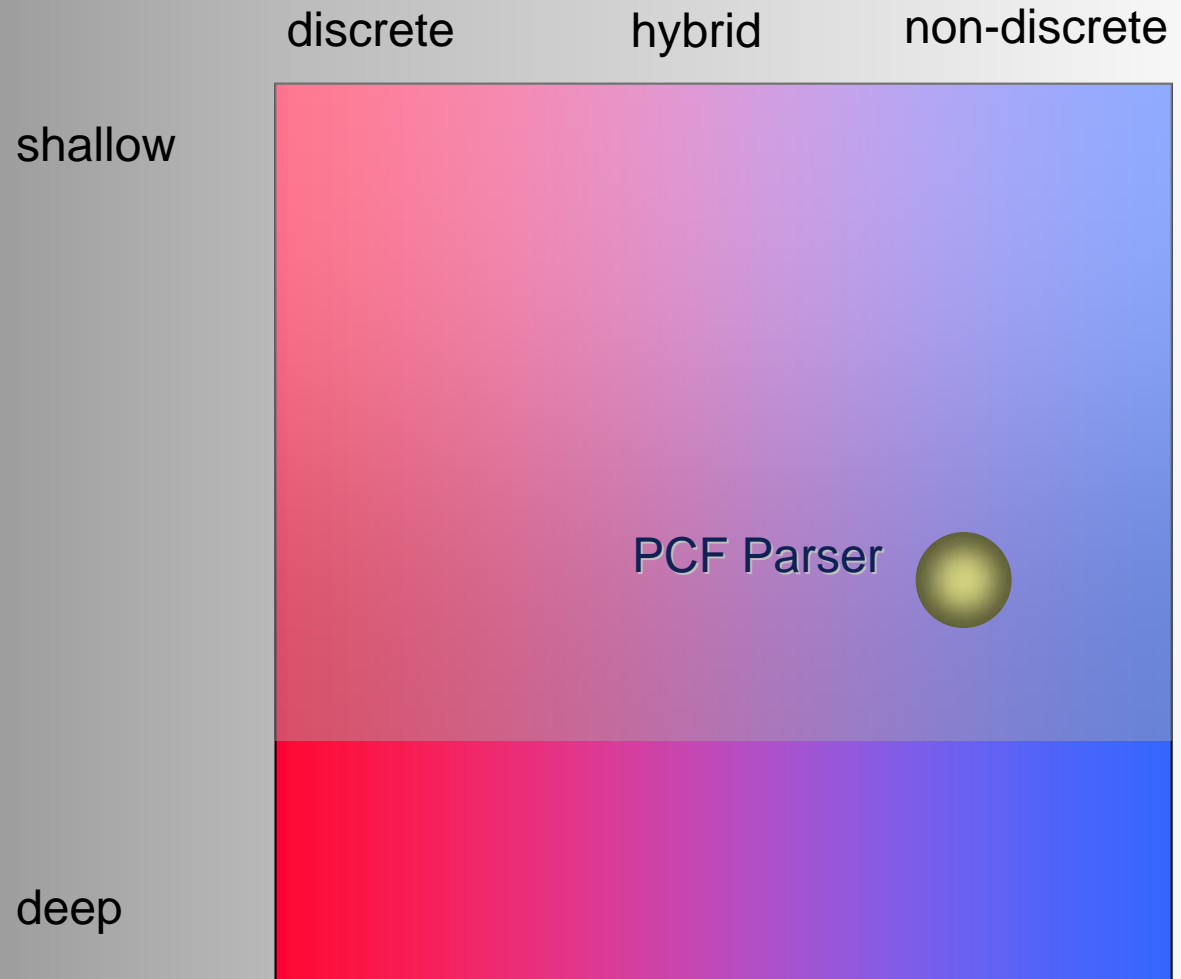
LT METHODS



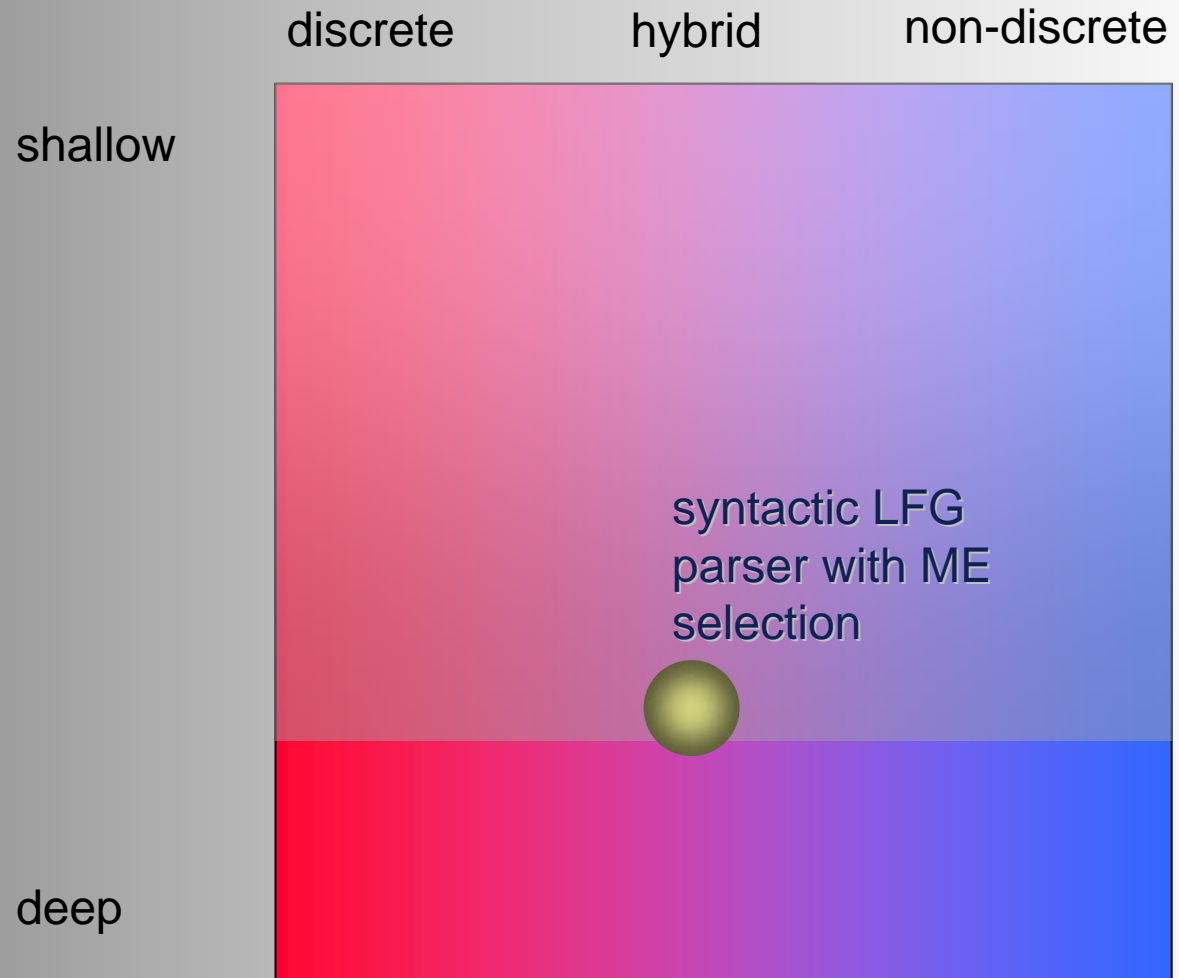
LT METHODS



LT METHODS



LT METHODS

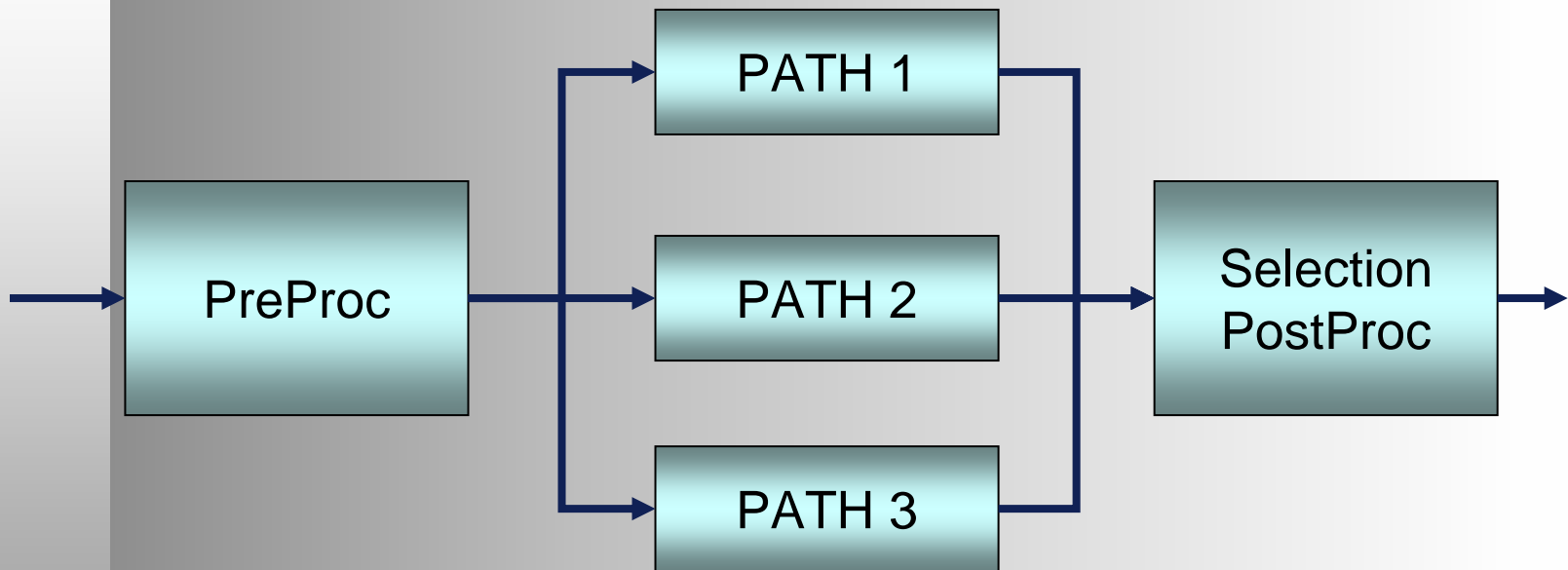


METHODS OF COMBINATION

- **Separate systems running in parallel with a selection function based on preference, confidence or assessment metrics**
- **All in one integrated system (such as a blackboard or manager architecture)**
- **Separate systems annotating the same input with different control schemes (whiteboard or pool architecture)**

Systems Running in Parallel

- **Separate systems running in parallel with a selection function based on preference, confidence or assessment metrics**

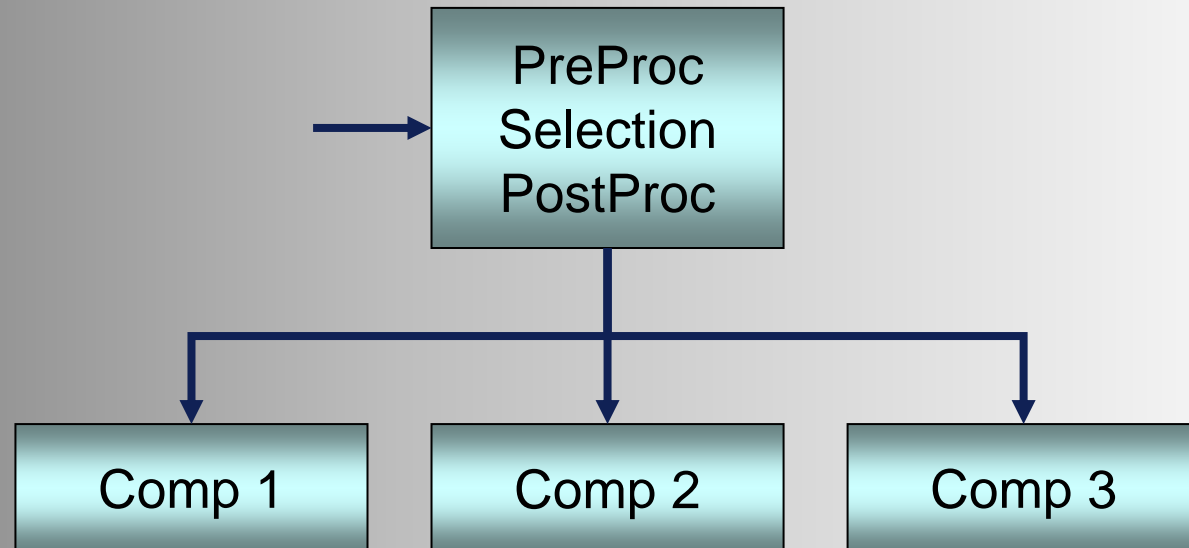


Selection Mechanism

- **Preference:**
most probable component wins (deep < shallow)
- **Confidence:**
most confident wins (each analysis gets a confidence measure)
- **Probability:**
most probable analysis wins (calibrated probabilities)
- **Mixed Approaches**

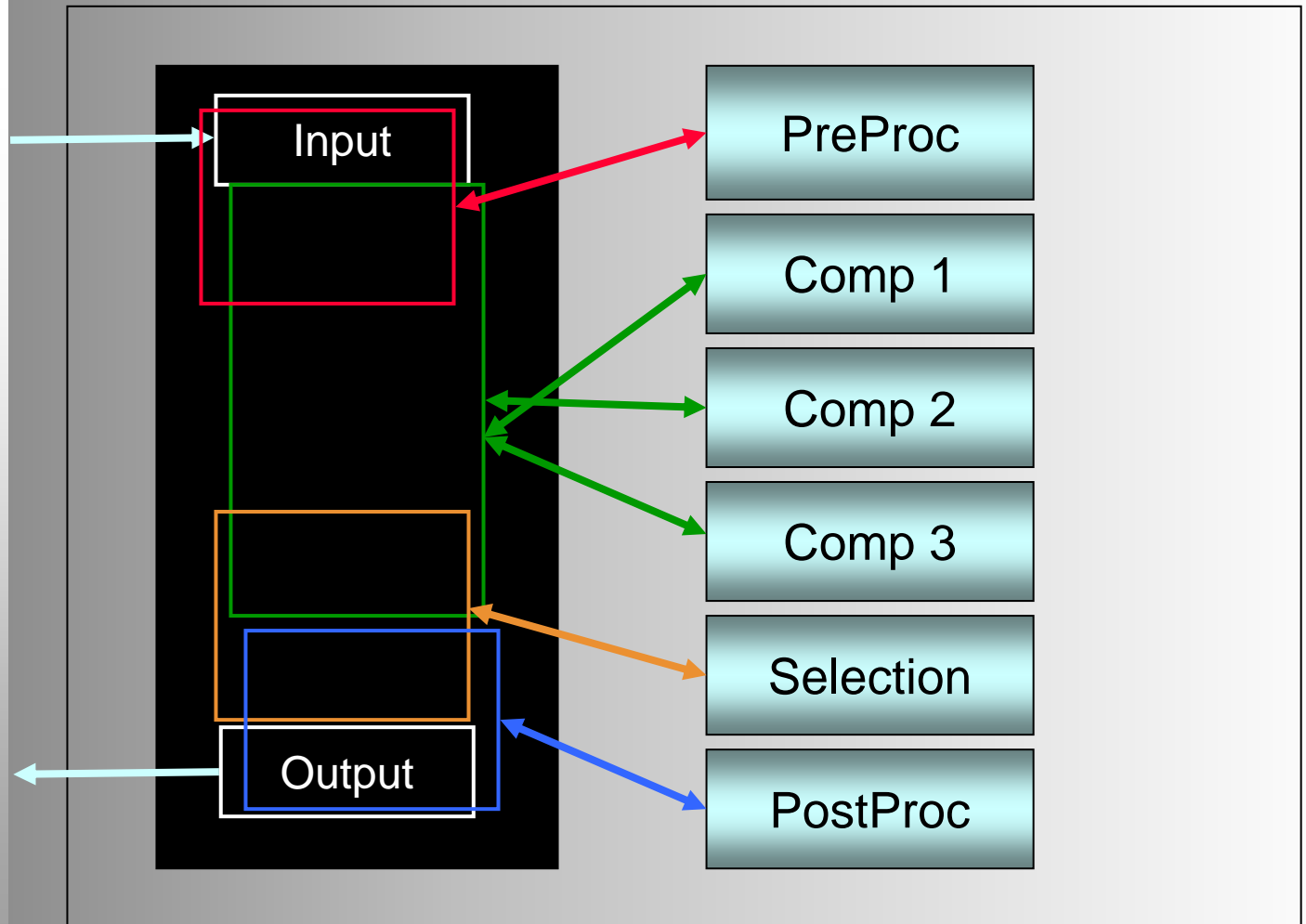
One Integrated System

- **Manager Architecture**



One Integrated System

■ Blackboard Architecture



- **Combination of shallow and deep processing by incremental text enrichment**

- **Shallow Methods**
 - **POS Tagging -- TnT**
 - **Morphology -- Morphix and MMorph**
 - **Word Nets -- GermaNet**
 - **Information Extraction -- SPPC and SPROUT**

- **Deep Methods**
 - **HPSG Parsing -- PET**

POOL ARCHITECTURE



POS Tagging
by TnT

Morphology
by Morphix
or M Morph

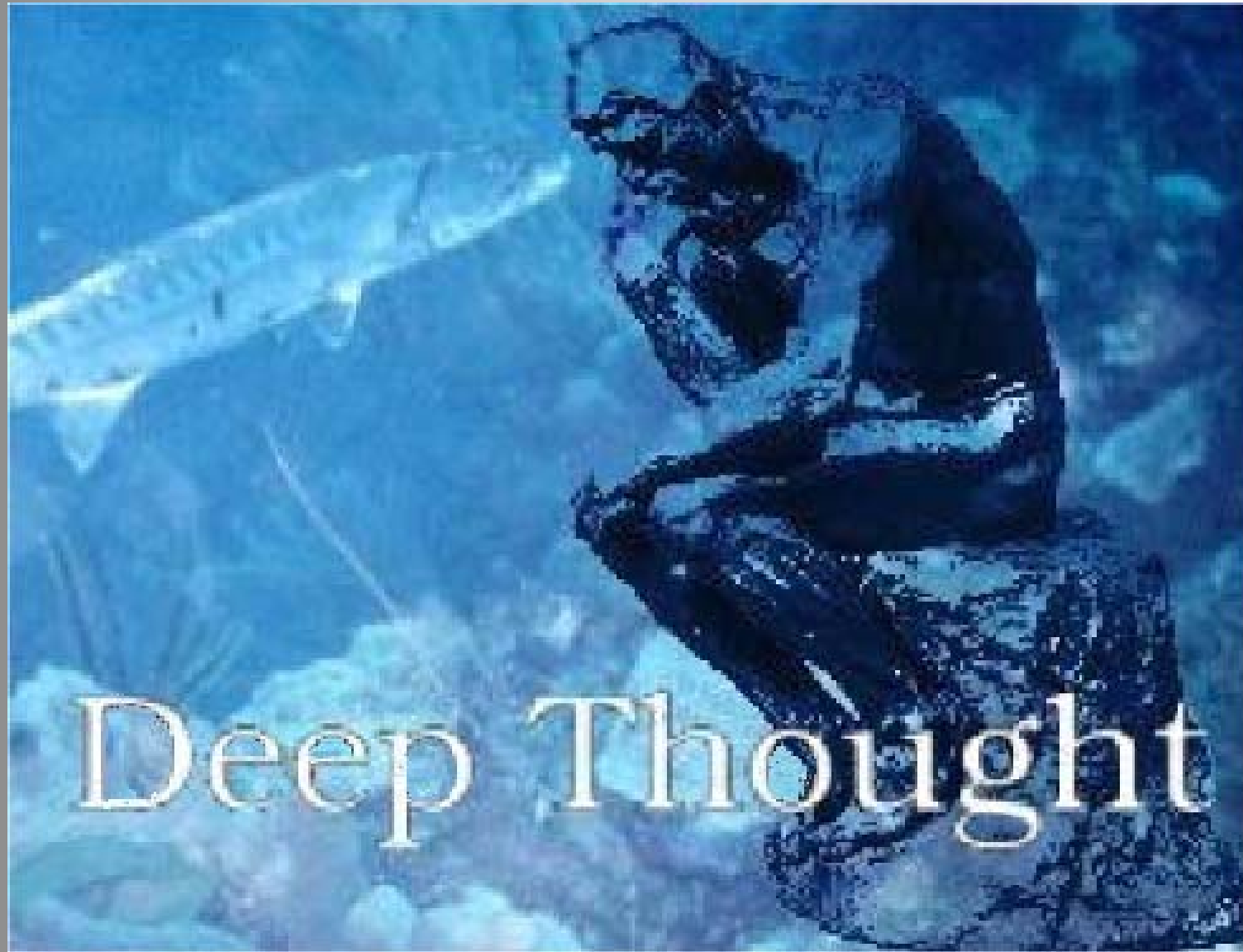
Word Senses
by GermaNet

Topological
Parsing by
TOPOPARSER

Information
Extraction by
SPPC - SPROUT

HPSG Parsing
by PET

B. Crysmann, A. Frank,
B. Kiefer, S. Müller,
G. Neumann, J. Piskorski,
U. Schäfer, M. Siegel,
H. Uszkoreit, F. Xu,
M. Becker, H.-U. Krieger





Saarland University (Coordinator)

U. of Cambridge

U. of Sussex

U. of Trondheim

Stanford University

CELI Company of Torino

XtraMind GmbH of Saarbrücken

acrolinx gmbh, Berlin

DFKI Saarbrücken

Three Applications

Documents authored by user enriched by application



Application Creativity Support

Reports for user assembled by application



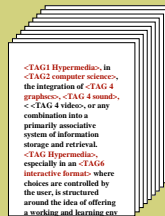
Application Business Intelligence

Response E- Mails selected on the basis of analysis

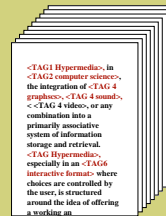


Application Email Response Management

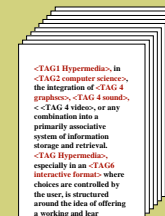
Accurate Relation Recognition with Deep Processing
 Named Entity Recognition with FST Technology
 IR Indexing with Morphology and Statistical LT



Digital Information Repositories



Business News, Newsgroups



Customer Email

- The lack of robustness and coverage is a serious problem for Information Extraction based on **deep processing** alone.

- Information Extraction with **shallow processing** methods is efficient and robust, but:
 - The lack of subcategorization information precludes extraction of conventional semantic representations.
 - There is insufficient information to recover scope information.

Hybrid approach to multilingual information extraction

NL Input

Every experienced vice president served on a committee.

Event-Based LF

$\text{every}(x), \text{experienced}(x), \text{VP}(x), \text{serve1}(e, x), \text{spast}(e), \text{on2}(e', e, y),$
 $\text{some}(y), \text{committee}(y).$

Robust Minimal Recursion Semantics (RMRS)

$\text{every}(x1), \text{experienced}(x2), \text{VP}(x3), \text{serve1}(e4), \text{spast}(e4), \text{on2}(e5),$
 $\text{some}(x6), \text{committee}(x7)$

$\text{ARG1}(e4, x1), \text{ARG1}(e5, e4), \text{ARG2}(e5, x6),$

$x1=x2, x2=x3, x6=x7.$

NL Input

Every experienced vice president served on a committee.

POS Tagger

every_DT(x1), experienced_AJ(x2), vice_president_N(x3), served_V(e4),
past(e4), on_PRP(e5), a_DT(x6), committee_N(x7)

Deep Processing

every(x1), experienced(x2), VP(x3), serve1(e4), spast(e4), on2(e5),
some(x6), committee(x7)

ARG1(e4,x1), ARG1(e5,e4), ARG2(e5,x6),

x1=x2, x2=x3, x6=x7.

The NP chunk parser outputs an (R)MRS representation in which the parts of NPs are connected by equations.

NL Input

Every experienced vice president served on a committee.

NP Chunker

every_DT(x1), experienced_AJ(x2), vice_president_N(x3), served_V(e4),
past(e4), on_PRP(e5), a_DT(x6), committee_N(x7),

$x1=x3$, $x2=x3$, $x6=x7$.

Certain grammatical relations or attachment may not be detected by the shallow parser. (Such as topicalization in the example below).

NL Input

That proposal every experienced committee member supported.

NP Chunker

that_DT(x1), proposal_N(x2), every_DT(x3), experienced_AJ(x4),
committee_member_N(x5), support_V(e6), past(e6),

ARG1(e6,x5), x1=x2, x3=x4, x4=x5

Main Idea

- **RMRS** provides a common representation for linguistic processing on all levels and for all languages. Shallow and deep processing produce compatible semantic representations.

Main results

- **New **architecture** for hybrid integrated multilingual natural language processing.**
 - **Combining efficient processing modules on different levels of preciseness and robustness and for multiple languages in a variable way.**

Main results: applications

- Information extraction for **business intelligence**.
 - Implementation where HPSG refines and verifies the output of shallow parsing.
- Information lookup for **document authoring and creative meetings**.
 - Server-based implementation with connection to external services (Google/Answerbus) and HoG.
- Information extraction in an **email autoresponse task**.
 - Implementation with connection to HoG using the deepest available analysis triggered by clues.
- Information extraction for a **web portal from hike descriptions**.
 - Implementation with semantically enriched grammar and connection via HoG .

Main results

- **Linguistically sound analyses** for a variety of phenomena and languages.
 - We have an integrated demonstration of large and wide-coverage grammars for five languages.
- **Lexical databases.**
- **Methods for unknown words and named entities.**
- **Efficient generation.**

Hybrid Natural Language Processing

- The analysis results of NLP tools at lower processing levels can be used by components at higher levels.
 - For example, the deep linguistic analysis module PET uses default lexicon entries for Named-Entities that Sprout delivers.

TEXT show me a picture of a Nokia 56 phone
 TOP h1

<table border="1"> <tr><td><i>imp_m_rel</i></td><td></td></tr> <tr><td>LBL</td><td>h1</td></tr> <tr><td>ARG0</td><td>h5</td></tr> </table>		<i>imp_m_rel</i>		LBL	h1	ARG0	h5	<table border="1"> <tr><td><i>pronoun_q_rel</i></td><td></td></tr> <tr><td>LBL</td><td>h8</td></tr> <tr><td>ARG0</td><td>x9 pers=2</td></tr> <tr><td>RSTR</td><td>h10</td></tr> <tr><td>BODY</td><td>h12</td></tr> </table>		<i>pronoun_q_rel</i>		LBL	h8	ARG0	x9 pers=2	RSTR	h10	BODY	h12	<table border="1"> <tr><td><i>pron_rel</i></td><td></td></tr> <tr><td>LBL</td><td>h13</td></tr> <tr><td>ARG0</td><td>x9 pers=2</td></tr> </table>		<i>pron_rel</i>		LBL	h13	ARG0	x9 pers=2	<table border="1"> <tr><td><i>_show_v</i></td><td></td></tr> <tr><td>LBL</td><td>h15</td></tr> <tr><td>ARG0</td><td>e2</td></tr> <tr><td>ARG1</td><td>x9 pers=2</td></tr> <tr><td>ARG2</td><td>x17 pers=3 num=sg</td></tr> <tr><td>ARG3</td><td>x16 pers=1 num=sg</td></tr> </table>		<i>_show_v</i>		LBL	h15	ARG0	e2	ARG1	x9 pers=2	ARG2	x17 pers=3 num=sg	ARG3	x16 pers=1 num=sg	<table border="1"> <tr><td><i>pron_rel</i></td><td></td></tr> <tr><td>LBL</td><td>h19</td></tr> <tr><td>ARG0</td><td>x16 pers=1 num=sg</td></tr> </table>		<i>pron_rel</i>		LBL	h19	ARG0	x16 pers=1 num=sg	<table border="1"> <tr><td><i>pronoun_q_rel</i></td><td></td></tr> <tr><td>LBL</td><td>h21</td></tr> <tr><td>ARG0</td><td>x16 pers=1 num=sg</td></tr> <tr><td>RSTR</td><td>h22</td></tr> <tr><td>BODY</td><td>h24</td></tr> </table>		<i>pronoun_q_rel</i>		LBL	h21	ARG0	x16 pers=1 num=sg	RSTR	h22	BODY	h24	<table border="1"> <tr><td><i>_a_q</i></td><td></td></tr> <tr><td>LBL</td><td>h25</td></tr> <tr><td>ARG0</td><td>x17 pers=3 num=sg</td></tr> <tr><td>RSTR</td><td>h26</td></tr> <tr><td>BODY</td><td>h28</td></tr> </table>		<i>_a_q</i>		LBL	h25	ARG0	x17 pers=3 num=sg	RSTR	h26	BODY	h28	<table border="1"> <tr><td><i>_picture_n</i></td><td></td></tr> <tr><td>LBL</td><td>h29</td></tr> <tr><td>ARG0</td><td>x17 pers=3 num=sg</td></tr> <tr><td>ARG1</td><td>x30 pers=3 num=sg</td></tr> </table>		<i>_picture_n</i>		LBL	h29	ARG0	x17 pers=3 num=sg	ARG1	x30 pers=3 num=sg	<table border="1"> <tr><td><i>_of_p</i></td><td></td></tr> <tr><td>LBL</td><td>h10001</td></tr> <tr><td>ARG0</td><td>e32</td></tr> <tr><td>ARG2</td><td>x30 pers=3 num=sg</td></tr> </table>		<i>_of_p</i>		LBL	h10001	ARG0	e32	ARG2	x30 pers=3 num=sg	<table border="1"> <tr><td><i>_a_q</i></td><td></td></tr> <tr><td>LBL</td><td>h35</td></tr> <tr><td>ARG0</td><td>x30 pers=3 num=sg</td></tr> <tr><td>RSTR</td><td>h36</td></tr> <tr><td>BODY</td><td>h37</td></tr> </table>		<i>_a_q</i>		LBL	h35	ARG0	x30 pers=3 num=sg	RSTR	h36	BODY	h37
<i>imp_m_rel</i>																																																																																																									
LBL	h1																																																																																																								
ARG0	h5																																																																																																								
<i>pronoun_q_rel</i>																																																																																																									
LBL	h8																																																																																																								
ARG0	x9 pers=2																																																																																																								
RSTR	h10																																																																																																								
BODY	h12																																																																																																								
<i>pron_rel</i>																																																																																																									
LBL	h13																																																																																																								
ARG0	x9 pers=2																																																																																																								
<i>_show_v</i>																																																																																																									
LBL	h15																																																																																																								
ARG0	e2																																																																																																								
ARG1	x9 pers=2																																																																																																								
ARG2	x17 pers=3 num=sg																																																																																																								
ARG3	x16 pers=1 num=sg																																																																																																								
<i>pron_rel</i>																																																																																																									
LBL	h19																																																																																																								
ARG0	x16 pers=1 num=sg																																																																																																								
<i>pronoun_q_rel</i>																																																																																																									
LBL	h21																																																																																																								
ARG0	x16 pers=1 num=sg																																																																																																								
RSTR	h22																																																																																																								
BODY	h24																																																																																																								
<i>_a_q</i>																																																																																																									
LBL	h25																																																																																																								
ARG0	x17 pers=3 num=sg																																																																																																								
RSTR	h26																																																																																																								
BODY	h28																																																																																																								
<i>_picture_n</i>																																																																																																									
LBL	h29																																																																																																								
ARG0	x17 pers=3 num=sg																																																																																																								
ARG1	x30 pers=3 num=sg																																																																																																								
<i>_of_p</i>																																																																																																									
LBL	h10001																																																																																																								
ARG0	e32																																																																																																								
ARG2	x30 pers=3 num=sg																																																																																																								
<i>_a_q</i>																																																																																																									
LBL	h35																																																																																																								
ARG0	x30 pers=3 num=sg																																																																																																								
RSTR	h36																																																																																																								
BODY	h37																																																																																																								
RELS {																																																																																																									
<table border="1"> <tr><td><i>compound_rel</i></td><td></td></tr> <tr><td>LBL</td><td>h39</td></tr> <tr><td>ARG0</td><td>e42</td></tr> <tr><td>ARG1</td><td>x30 pers=3 num=sg</td></tr> <tr><td>ARG2</td><td>x41 pers=3 num=sg</td></tr> </table>		<i>compound_rel</i>		LBL	h39	ARG0	e42	ARG1	x30 pers=3 num=sg	ARG2	x41 pers=3 num=sg	<table border="1"> <tr><td><i>named_abb_rel</i></td><td></td></tr> <tr><td>LBL</td><td>h43</td></tr> <tr><td>ARG0</td><td>x41 pers=3 num=sg</td></tr> <tr><td>CARG</td><td>Nokia_56</td></tr> </table>		<i>named_abb_rel</i>		LBL	h43	ARG0	x41 pers=3 num=sg	CARG	Nokia_56	<table border="1"> <tr><td><i>proper_q_rel</i></td><td></td></tr> <tr><td>LBL</td><td>h45</td></tr> <tr><td>ARG0</td><td>x41 pers=3 num=sg</td></tr> <tr><td>RSTR</td><td>h46</td></tr> <tr><td>BODY</td><td>h48</td></tr> </table>		<i>proper_q_rel</i>		LBL	h45	ARG0	x41 pers=3 num=sg	RSTR	h46	BODY	h48	<table border="1"> <tr><td><i>_phone_n</i></td><td></td></tr> <tr><td>LBL</td><td>h10002</td></tr> <tr><td>ARG0</td><td>x30 pers=3 num=sg</td></tr> </table>		<i>_phone_n</i>		LBL	h10002	ARG0	x30 pers=3 num=sg																																																																
<i>compound_rel</i>																																																																																																									
LBL	h39																																																																																																								
ARG0	e42																																																																																																								
ARG1	x30 pers=3 num=sg																																																																																																								
ARG2	x41 pers=3 num=sg																																																																																																								
<i>named_abb_rel</i>																																																																																																									
LBL	h43																																																																																																								
ARG0	x41 pers=3 num=sg																																																																																																								
CARG	Nokia_56																																																																																																								
<i>proper_q_rel</i>																																																																																																									
LBL	h45																																																																																																								
ARG0	x41 pers=3 num=sg																																																																																																								
RSTR	h46																																																																																																								
BODY	h48																																																																																																								
<i>_phone_n</i>																																																																																																									
LBL	h10002																																																																																																								
ARG0	x30 pers=3 num=sg																																																																																																								
HCONS {h5 qeq h15, h10 qeq h13, h22 qeq h19, h26 qeq h29, h36 qeq h39, h46 qeq h43}																																																																																																									
ING {h29 ing h10001, h39 ing h10002}																																																																																																									

Hybrid Natural Language Processing

- The analysis results of NLP tools at lower processing levels can be used by components at higher levels.
 - For example, the deep linguistic analysis module PET uses default lexicon entries for Part-of-Speech tags that the POS tagger TnT delivers.

```

TEXT  what are polyphone ringtones ?
TOP   h1

RELS  {
  int_m_rel      | prpstn_m_rel | thing_rel      | which_q_rel    | _be_v          | idiom_q_rel    | generic_adj_rel | generic_nom_rel
  LBL            | LBL            | LBL            | LBL            | LBL            | LBL            | LBL            | LBL
  ARG0          | ARG0          | ARG0 x3 pers=3 num=sg | ARG0 x3 pers=3 num=sg | ARG0 e2 tense=present | ARG0 x20 pers=3 num=pl | ARG0 e27 tense=present | ARG0 x20 pers=3 num=pl
  TPC x3 pers=3 num=sg | ARG0 h10 | ARG0 x3 pers=3 num=sg | RSTR h17 | ARG1 x20 pers=3 num=pl | RSTR h24 | ARG1 x20 pers=3 num=pl | CARG ringtones
  BODY          | h18 | BODY            | h18 | ARG2 x3 pers=3 num=sg | BODY h25 | ARG1 x20 pers=3 num=pl | ARG1 x20 pers=3 num=pl
}

HCONS {h10 qeq h19, h17 qeq h13, h24 qeq h26}
ING   {h26 ing h10001}
    
```

Hybrid Natural Language Processing

Deliver the deepest result.

The Email Autoresponse application follows this approach

xmltext JTok TnT Chunkie Sprout SproutFS RASP PET **result**

TEXT what are polyphone ringtones ?

TOP h1

RELS {	$\left\{ \begin{array}{l} \text{int_m_rel} \\ \text{LBL} \quad h1 \\ \text{ARGO} \quad h6 \\ \text{TPC} \quad x3 \text{ pers}=3 \text{ num}=sg \end{array} \right\}$	$\left\ \begin{array}{l} \text{prpstn_m_rel} \\ \text{LBL} \quad h6 \\ \text{ARGO} \quad h10 \end{array} \right\ $	$\left\ \begin{array}{l} \text{thing_rel} \\ \text{LBL} \quad h13 \\ \text{ARGO} \quad x3 \text{ pers}=3 \text{ num}=sg \end{array} \right\ $	$\left\ \begin{array}{l} \text{which_q_rel} \\ \text{LBL} \quad h15 \\ \text{ARGO} \quad x3 \text{ pers}=3 \text{ num}=sg \\ \text{RSTR} \quad h17 \\ \text{BODY} \quad h18 \end{array} \right\ $	$\left\ \begin{array}{l} _be_v \\ \text{LBL} \quad h19 \\ \text{ARGO} \quad e2 \text{ tense}=present \\ \text{ARG1} \quad x20 \text{ pers}=3 \text{ num}=pl \\ \text{ARG2} \quad x3 \text{ pers}=3 \text{ num}=sg \end{array} \right\ $	$\left\ \begin{array}{l} \text{idiom_q_rel} \\ \text{LBL} \quad h22 \\ \text{RSTR} \quad h24 \\ \text{BODY} \quad h25 \end{array} \right\ $	$\left\ \begin{array}{l} \text{generic_adj_rel} \\ \text{LBL} \quad h26 \\ \text{ARGO} \quad e27 \text{ tense}=u \\ \text{ARG1} \quad x20 \text{ pers}=3 \text{ num}=pl \end{array} \right\ $	$\left\ \begin{array}{l} \text{generic_nom_rel} \\ \text{LBL} \quad h10001 \\ \text{ARGO} \quad x20 \text{ pers}=3 \text{ num}=pl \\ \text{CARG} \quad \text{ringtones} \end{array} \right\}$
--------	--	---	---	--	--	--	--	--

HCONS {h10 qeq h19, h17 qeq h13, h24 qeq h26}

ING {h26 ing h10001}

Hybrid Natural Language Processing

Deliver partial results, whenever a complete analysis is not available. Partial results are taken from the deepest module that delivers results.

xmltext JTok TnT Chunkie Sprout SproutFS RASP PET result

TEXT I want a new phone do you ?
TOP h1

RELS {	{	<i>prpstn_m_rel</i>	<i>pron_rel</i>	<i>pronoun_q_rel</i>	<i>_want_v</i>	<i>_a_q</i>	<i>_new_a</i>	<i>_phone_n</i>	}
		LBL h1	LBL h8	LBL h11	LBL h15	LBL h18	LBL h22	LBL h10001	
		ARG0 h5	ARG0 x9 pers=1 num=sg	ARG0 x9 pers=1 num=sg	ARG0 e2 tense=present	ARG0 x16 pers=3 num=sg	ARG0 e24 tense=u	ARG0 x16 pers=3 num=sg	
			RSTR h12	RSTR h12	ARG1 x9 pers=1 num=sg	RSTR h19	ARG1 x16 pers=3 num=sg	ARG0 x16 pers=3 num=sg	

HCONS {h5 qeq h15, h12 qeq h8, h19 qeq h22}

ING {h22 ing h10001}

TEXT I want a new phone do you ?
TOP h1

RELS {	{	<i>_do_v</i>	<i>pron_rel</i>	<i>pronoun_q_rel</i>	}
		LBL h1	LBL h7	LBL h9	
		ARG0 e2 tense=u	ARG0 x3 pers=2	ARG0 x3 pers=2	
		ARG1 u4	RSTR h10	BODY h12	







HCONS {h10 qeq h7}

ING {}

Hybrid Natural Language Processing

Combine modules and grammars for different languages. Each language has its own configuration of valid modules and grammars.

Heart of Gold Settings:

Depth: 10 Languages: English 
 20 German 
 30 Italian 
 50 Norwegian 
 80 Japanese 
 100 Greek 

Input text:

Please note that JavaScript™ is required.

Hybrid Natural Language Processing

The different modules use a compatible output formalism, RMRS.

In case of shallower modules, this robust semantic structure allows for underspecification of, e.g., argument structure.

xmltext JTok TnT Chunkie Sprout SproutFS RASP PET result

TEXT Kim intended to buy the phone
TOP h31

RELS {		prpstn_m_rel		proper_q_rel		named_rel		_intend_v		prpstn_m_rel		_buy_v		_the_q		_phone_n	}											
				LBL		h1		LBL				h3		ARGO				e7 tense=past	LBL	h10	ARGO	x13	LBL	h14				
				ARGO		h33		RSTR				h4		ARGO				x2	ARG1	x2	ARGO	h26	ARGO	e11	RSTR	h19	ARGO	x13
				BODY		h5		CARG				kim		ARG2				h24	ARG2	x13	BODY	h16						

HCONS {h33 qeq h6, h4 qeq h3, h26 qeq h10, h19 qeq h14}

ING {}

xmltext JTok TnT Chunkie Sprout SproutFS RASP PET result

TEXT Kim intended to buy the phone
TOP h1

RELS {		prpstn_m_rel		proper_q_rel		named_rel		_intend_v		prpstn_m_rel		_buy_v		_the_q		_phone_n	}											
				LBL		h8		LBL				h13		ARGO				e2 tense=past	LBL	h15	ARGO	e25 tense=u	LBL	h27	LBL	h31		
				ARGO		h1		ARGO				x10 num=sg		ARGO				x10 num=sg	ARG1	x10 num=sg	ARGO	h17	ARGO	x10 num=sg	ARGO	x24 num=sg	LBL	h31
				ARGO		h5		RSTR				h9		CARG				kim	ARG2	h17	ARGO	h20	ARG1	x10 num=sg	RSTR	h28	ARGO	x24 num=sg

HCONS {h5 qeq h15, h9 qeq h13, h20 qeq h23, h28 qeq h31}

ING {}

Multilingual HPSG Grammar Engineering

- **Available HPSG grammars :**
 - **German (50.000 lexical entries)**
 - **English (12.300 lexical entries)**
 - **Japanese (35.000 lexical entries)**
 - **Norwegian (84.240 lexical entries)**
 - **Italian (4.850 lexical entries)**

- **We have a Grammar Matrix that allows an efficient implementation of new grammars with compatible and correct output.**

MULTILINGUAL GRAMMAR DEVELOPMENT

- Existing Grammars in English, German, Japanese
- Sizeable Grammar of Norwegian built in the project Deep Thought by Lars Hellan and others at Trondheim U.
- Italian Grammar by company CELI built in Deep Thought
- Greek grammar being set up by Valia Kordoni and Julia Neu at Saarland University
- Korean grammar being build by Jong-Bok Kim
- New Portuguese Grammar project at University of Lisbon headed by Antonio Branco
- Spanish Grammar converted from ALEP format at U. Barcelona
- New: Beginning of a Chinese Grammar at Saarland U.

The Grammar Matrix

- The Matrix for grammars of multiple languages:
 - A system of types that is *directly included* into new and existing grammars.
 - Reduced start-up costs.
 - Common feature descriptions.
 - *Shared insights* on analyses of phenomena.
 - Support for *multilingual* applications.
 - Robust treatment of *real corpora*.

The Grammar Matrix

- **It contains 19 files and documentation:**
 - **Basic types and features for multilingual HPSG development.**
 - **Basic types and features for multilingual semantic construction.**
 - **Settings for working with LKB, [incr tsdb()] and PET.**
 - **Basic lexical types**
 - **Basic rule types**

Matrix-based multilingual grammar engineering

xmltext JTok TnT Chunkie Sprout SproutFS RASP PET result

TEXT the man smiles
TOP h1

RELS {

{	prpstn_m_rel		_the_q		h8		_man_n		h13		_smile_v		h15	}			
			LBL				ARGO x9 pers=3 num=sg gender=n				LBL				ARGO x9 pers=3 num=sg gender=n	LBL	ARGO e2 tense=present
			h1				RSTR				h10				ARGO x9 pers=3 num=sg gender=n	e2	
			ARGO				h5				h12				ARGO x9 pers=3 num=sg gender=n	ARGO	ARGO x9 pers=3 num=sg gender=n

BODY h12
 ARG1 x9 pers=3 num=sg gender=n

HCONS {h5 qeq h15, h10 qeq h13}

ING {}

xmltext PET result

TEXT mannen smiler
TOP h1

RELS {

{	mann-rel		def-q-rel		h5		card_rel		h8		smile-rel		h9		prpstn_rel		h1	}					
			LBL				ARGO x4 bounded=+				LBL				ARGO x4 bounded=+				LBL	ARGO e2 tense=present	LBL	ARGO h10	
			h3				RSTR				h7				CARG				1-rel	ARG1	x4 bounded=+	ARGO	
			ARGO x4 bounded=+				h6																

BODY h6
 ARG1 x4 bounded=+

HCONS {h7 qeq h3, h10 qeq h9}

ING {}

Matrix-based multilingual grammar engineering

xmltext JTok TnT Chunkie Sprout SproutFS PET result

TEXT Der Mann lacht
TOP h1

RELS {

<i>def</i>								
LBL		<i>h3</i>	<i>_mann_n</i>			<i>_lachen_v</i>		<i>prpstn_m_rel</i>
ARG0	<i>x6</i>	<i>pers=3 num=sg gender=m</i>	LBL		<i>h7</i>	ARG0	<i>e2 tense=present</i>	LBL <i>h1</i>
RSTR		<i>h4</i>	ARG0	<i>x6</i>	<i>pers=3 num=sg gender=m</i>	ARG1	<i>x6</i>	ARG0 <i>h9</i>
BODY		<i>h5</i>						

HCONS {*h9* qeq *h8*, *h4* qeq *h7*}

ING {}

xmltext ChaSen PET result

TEXT 人が笑う
TOP h1

RELS {

<i>proposition_m_rel</i>		<i>_hito_n</i>	<i>udef_rel</i>		<i>_warau_v</i>		
LBL	<i>h1</i>	LBL <i>h6</i>	LBL <i>h10</i>		LBL <i>h15</i>		
ARG0	<i>h5</i>	ARG0 <i>x7</i>	ARG0 <i>x7</i>		ARG0 <i>e2 tense=present</i>		
			RSTR <i>h14</i>		ARG1 <i>x7</i>		
			BODY <i>h13</i>				

HCONS {*h5* qeq *h15*, *h14* qeq *h6*}

ING {}

xmltext JTok PET result

TEXT I uomo sorride
TOP h1

RELS {

<i>_def_q</i>		<i>_uomo_n</i>	<i>_sorridere_v</i>		<i>proposition_m_rel</i>
LBL	<i>h3</i>	LBL <i>h9</i>	LBL <i>h1</i>		LBL <i>h14</i>
ARG0	<i>x4</i>	ARG0 <i>x4</i>	ARG0 <i>e2</i>		ARG0 <i>h17</i>
RSTR	<i>h8</i>		ARG1 <i>x4</i>		
BODY	<i>h7</i>				

HCONS {*h8* qeq *h9*, *h17* qeq *h1*}

ING {}



HOME	html_page
INTRODUCTION	html_page
RESEARCH PROGRAM	pdf_doc
SYSTEMS & GRAMMARS	html_page
PUBLICATIONS	html_page
NEWS & EVENTS	html_page
LINKS	html_page
CONTACT US	mail_to

delphinsite

DEEP LINGUISTIC PROCESSING WITH HPSG

Computational linguists of several research sites have joined forces in a collaborative effort aimed at deep linguistic processing of human language. The goal is the combination of linguistic and statistical processing methods for getting at the meaning of texts and utterances. The partners have adopted Head-Driven Phrase Structure Grammar (HPSG) and Minimal Recursion Semantics (MRS), two highly advanced and influential models of formal linguistic analysis. They have also committed themselves to a shared format for grammatical representation and to a rigid scheme of evaluation.

The collaboration involves researchers from the following institutions:

- ◆ [Cambridge University \(UK\), Computer Laboratory](#)
- ◆ [DFKI Saarbrücken GmbH \(Germany\), Language Technology Lab \(co-founder\)](#)
- ◆ [Norwegian University of Science and Technology \(Norway\), Lingvistisk Institutt](#)
- ◆ [Saarland University \(Germany\), Department for Computational Linguistics](#)
- ◆ [Stanford University \(US\), LinGO Laboratory at CSLI \(co-founder\)](#)
- ◆ [Tokyo University \(Japan\), Tsujii Laboratory](#)
- ◆ [University of Oslo \(Norway\), MT Research Group](#)
- ◆ [University of Sussex \(UK\), School of Cognitive and Computing Sciences](#)
- ◆ [University of Washington \(US\), Computational Linguistics Laboratory](#)

The DELPH-IN collaboration is open to additional partners who share our ambitious goals and commitments and who can dedicate the necessary resources to the common task.

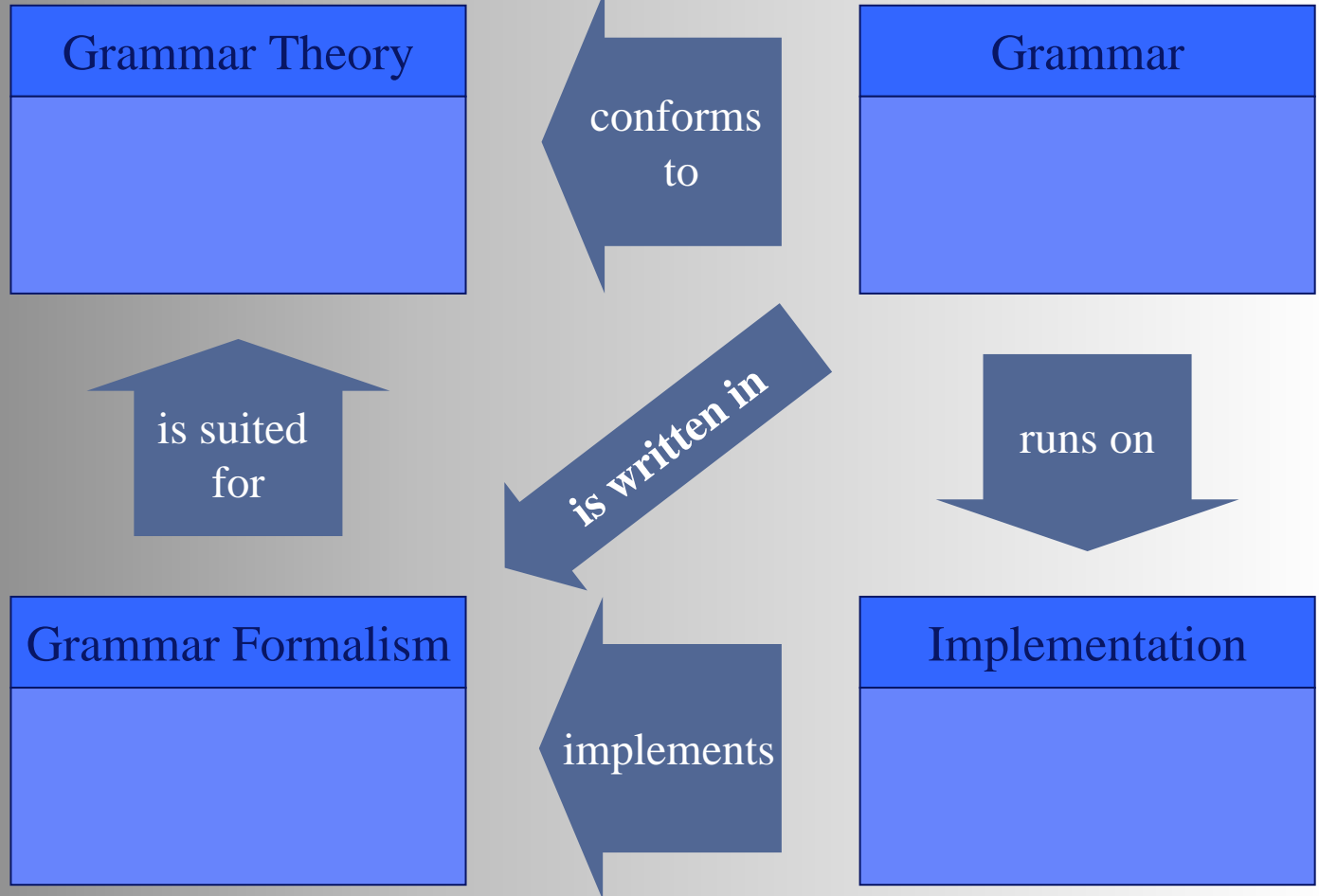
Scientific Impact: DELPH-IN

- **Including open-source resources:**
 - **LKB grammar development system (incl. generation)**
 - **PET grammar processing system**
 - **[incr tsdb()] grammar profiling system**
 - **ERG English HPSG**
 - **JACY Japanese HPSG**
 - **NorSource Norwegian HPSG**
 - **Modern Greek Resource Grammar**
 - **Lingo Grammar Matrix**
 - **Redwoods treebank**
- (DeepThought Heart of Gold will be part of DELPH-IN)**

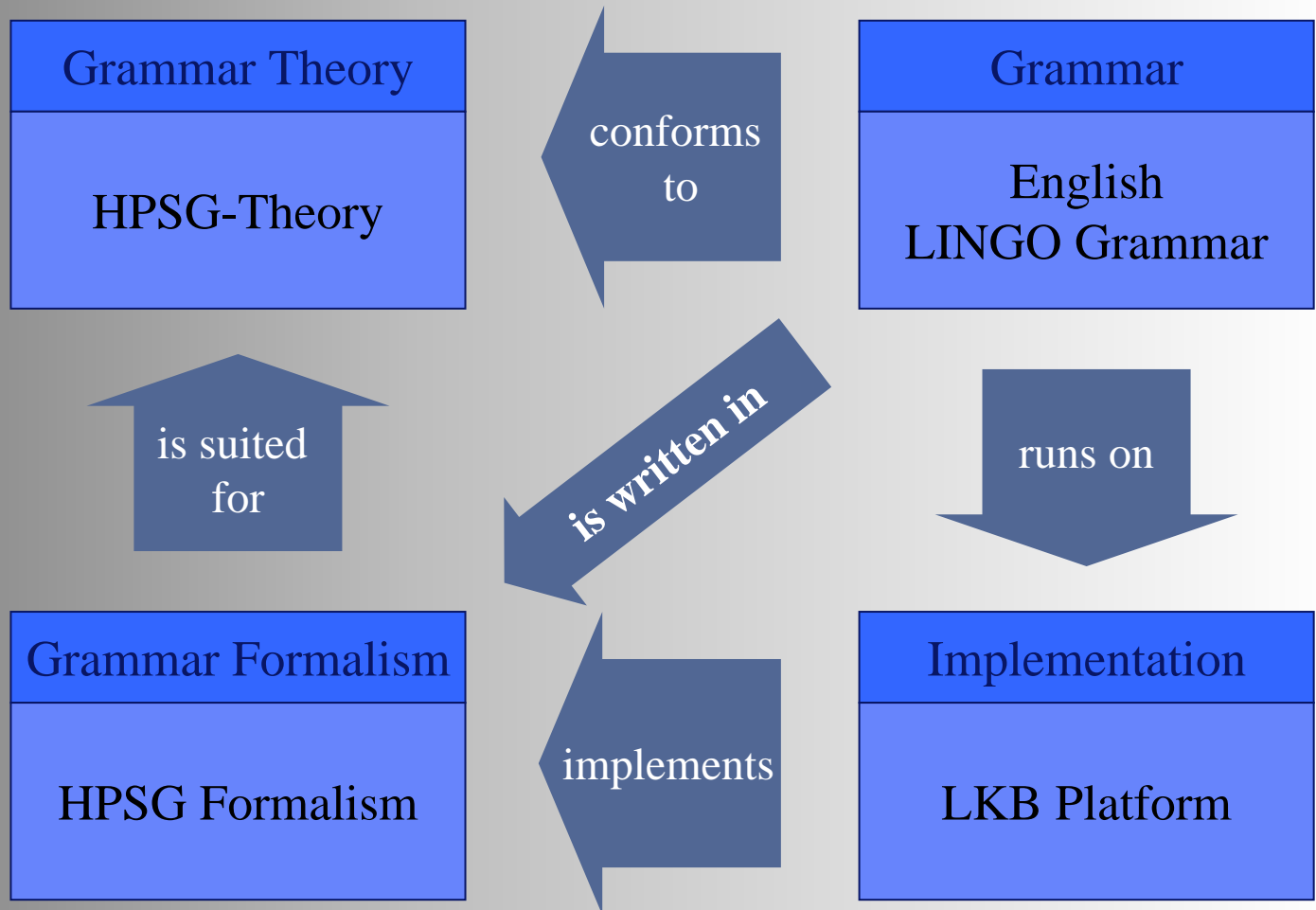
Conclusion and Outlook

- **There has been considerable progress in the area of deep linguistic processing.**
- **However, deep processing methods have to be combined with discrete and non-discrete shallow methods for sufficient performance.**
- **Flexible and scalable platform for the composition of hybrid systems.**
- **Test of the platform in real world applications.**
- **A better integration of statistical and deep linguistic methods is still badly needed.**

GRAMMAR⁴



GRAMMAR⁴



GRAMMAR⁴

