REAL GRAMMARS

- **LinGO - English Resource Grammar**
  - **8.000 types**
  - **100.000 lines of code**
  - **average feature structure > 300 nodes**
- **German Grammar of equal size**
- **Japanese and Norwegian grammars are getting close**

International Collaboration

Tokyo

- **Tsujii Lab at the University of Tokyo**
- Tsujii, Torisawa, Ninomiya, Taura, Yoshida, Mitsoishi,...

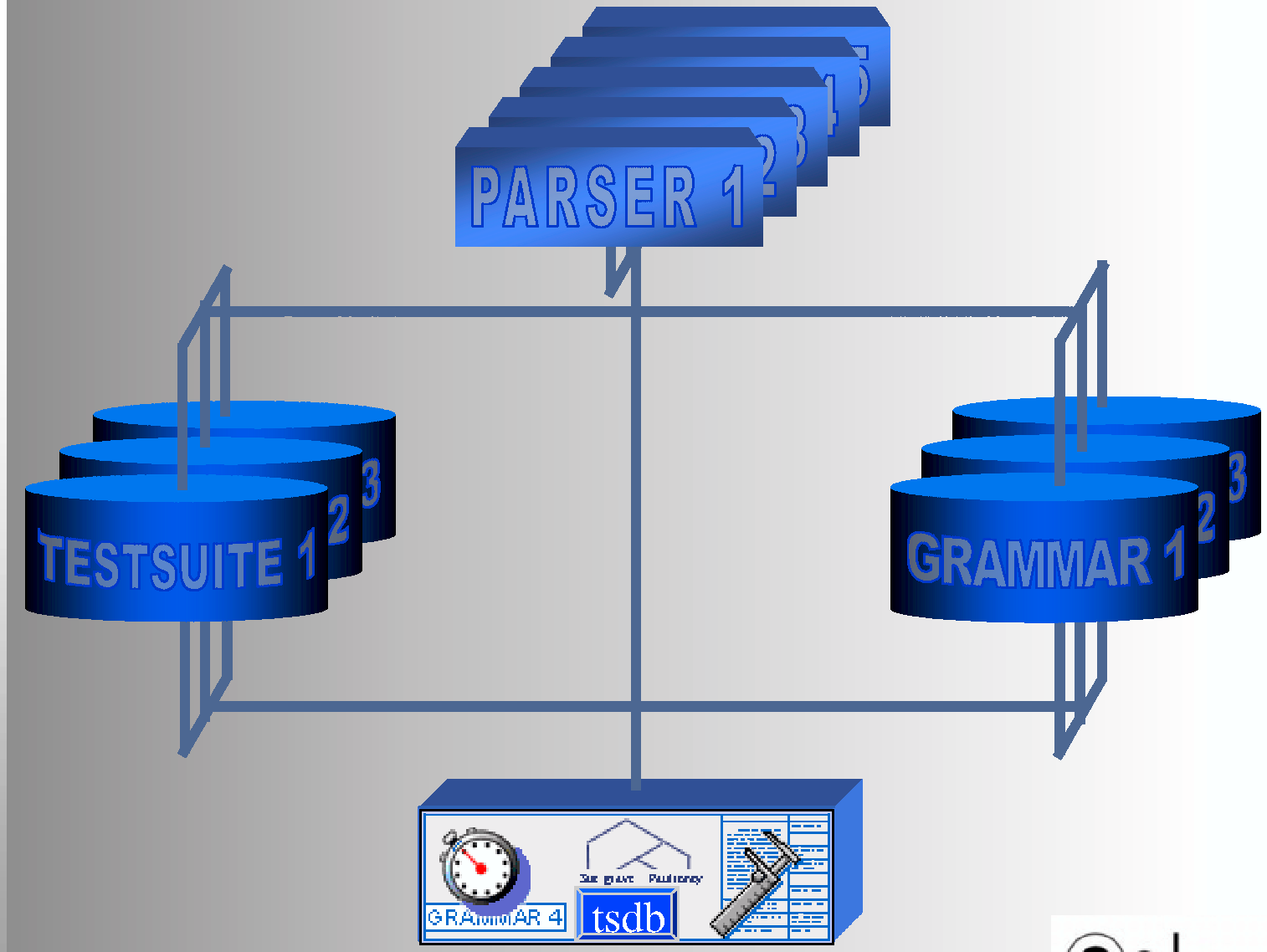
Stanford

- **HPSG Group at CSLI**
- Sag, Flickinger, Copestake, Malouf, Carroll (Brighton),...

Saarbrücken

- **LT Lab at DFKI and Dept. of CL**
- Oepen, Callmeier, Krieger, Kiefer, Ciortuz, Müller,...

THE EVALUATION SETUP



RESULTS

- All participating systems have benefitted from the systematic comparative evaluation
- Currently the fastest system is the runtime parser PET by Ulrich Callmeier (Saarbrücken)
- But the other parsers also improved drastically,e.g.:
  - LKB (Stanford, Cambridge)
  - LILFES (Tokyo)
  - PAGE (Saarbrücken)

RESULTS

- **HPSG Parsing is now 2000 times faster than before**
- **Normal-length sentences parse in 0.1 - 1.0 seconds**
- **Steady increase in hardware efficiency will also help**

REFERENCES

- D. Flickinger, S. Oepen, H. Uszkoreit, and J. Tsujii (eds.). 2000. *Journal of Natural Language Engineering* 6 (2000) 1. Special Issue on Efficient Processing with HPSG: Methods, Systems, Evaluation. Cambridge University Press. Cambridge.
- A. Copestake. 2002. Implementing Typed Feature Structure Grammars. CSLI Publications, Stanford. Building a Large Annotated Corpus of English:
- S. Oepen, D. Flickinger, J. Tsujii, and H. Uszkoreit. 2002. *Collaborative Language Engineering. A Case Study in Efficient Grammar-based Processing.* CSLI Publications, Stanford.

THE CORE MACHINERY

LKB
Development Platform

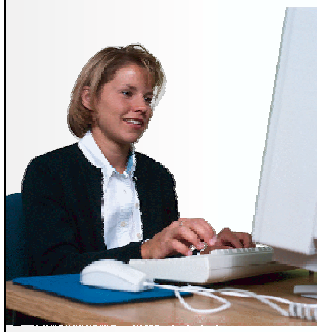
PET
Runtime Parser

English
Grammar

German
Grammar

Japanese
Grammar

Application

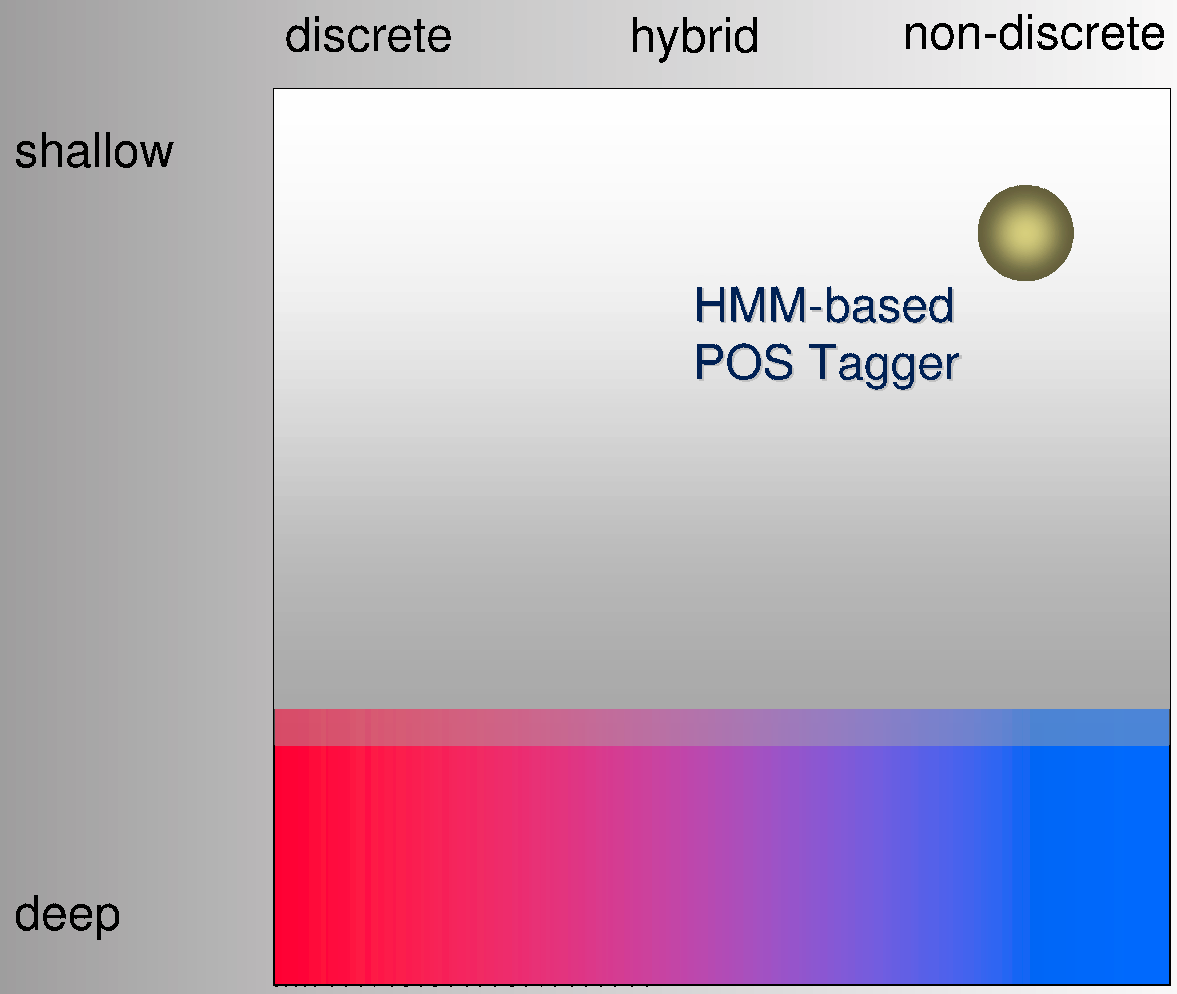


Open Source
Public Domain

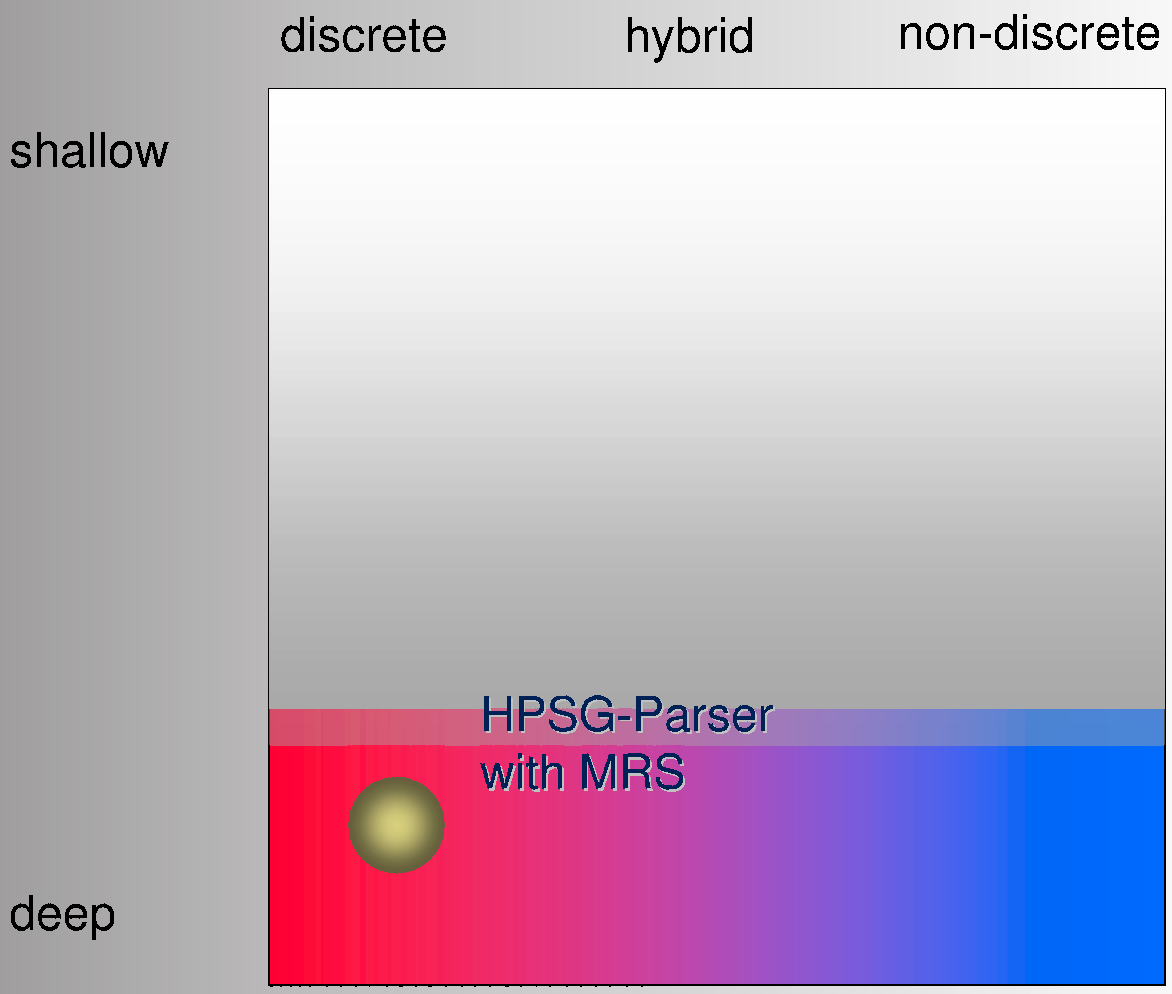
HOWEVER

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

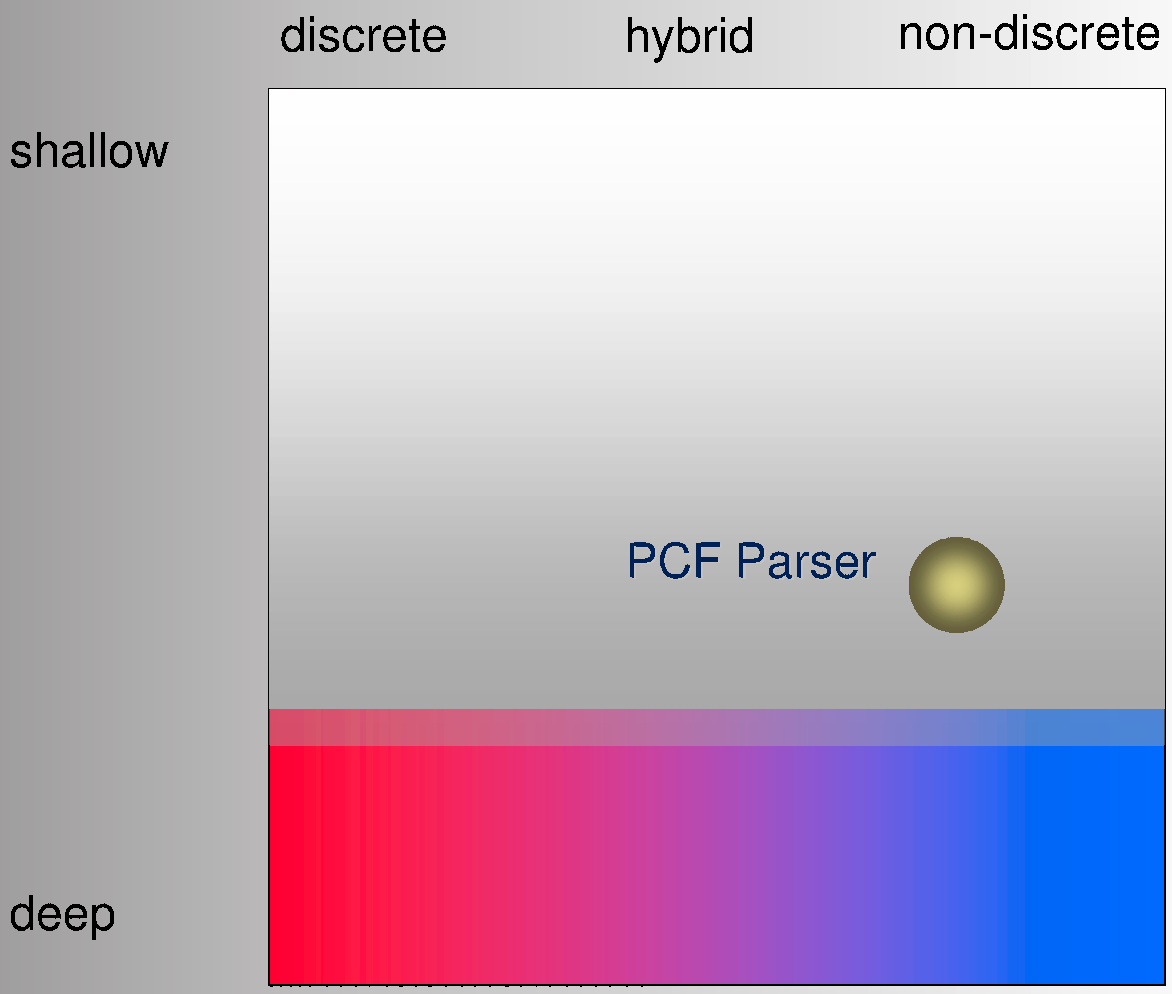
LT METHODS


discrete
hybrid
non-discrete
shallow
HMM-based
POS Tagger
deep

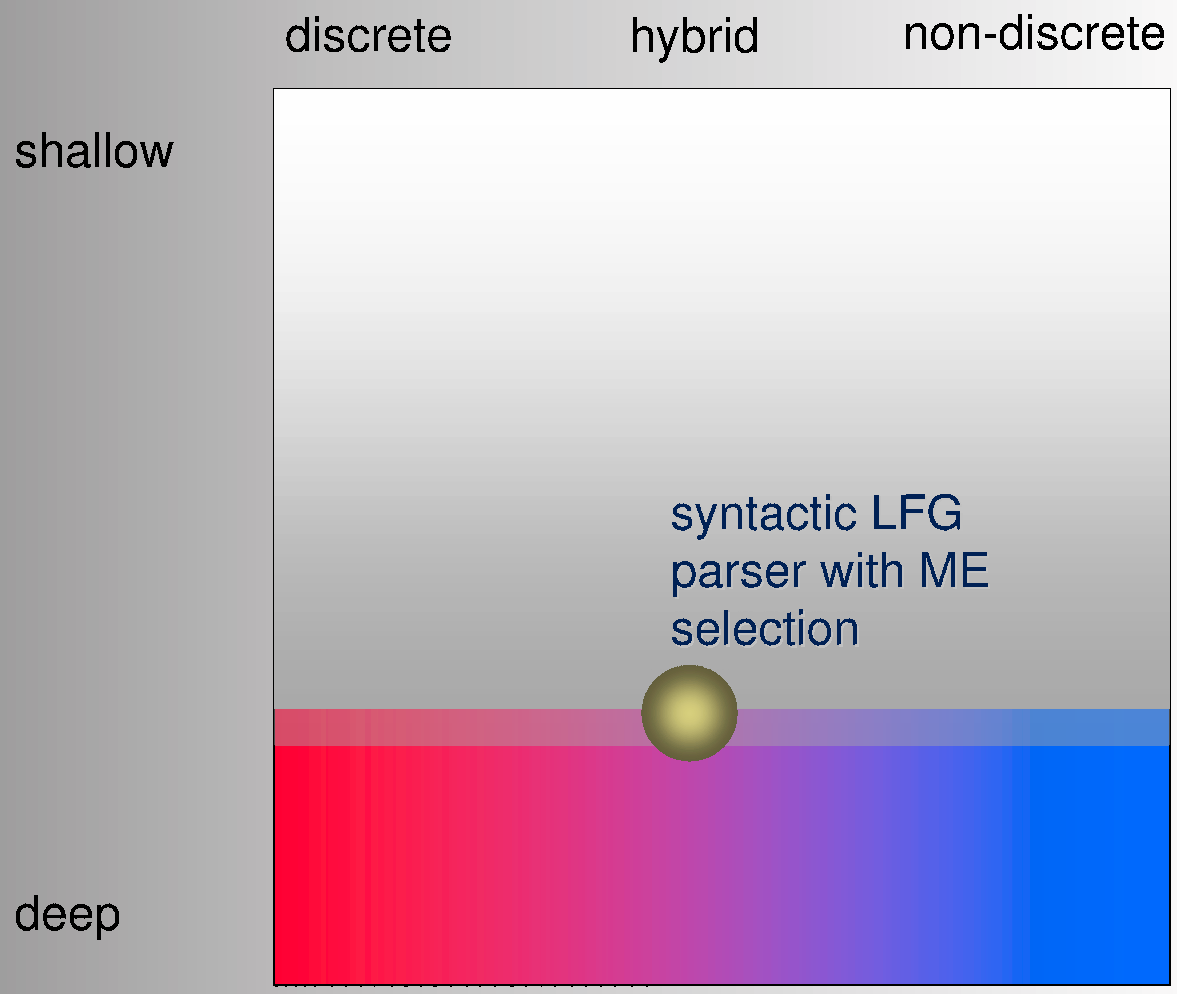
LT METHODS


discrete
hybrid
non-discrete
shallow
deep
HPSG-Parser
with MRS

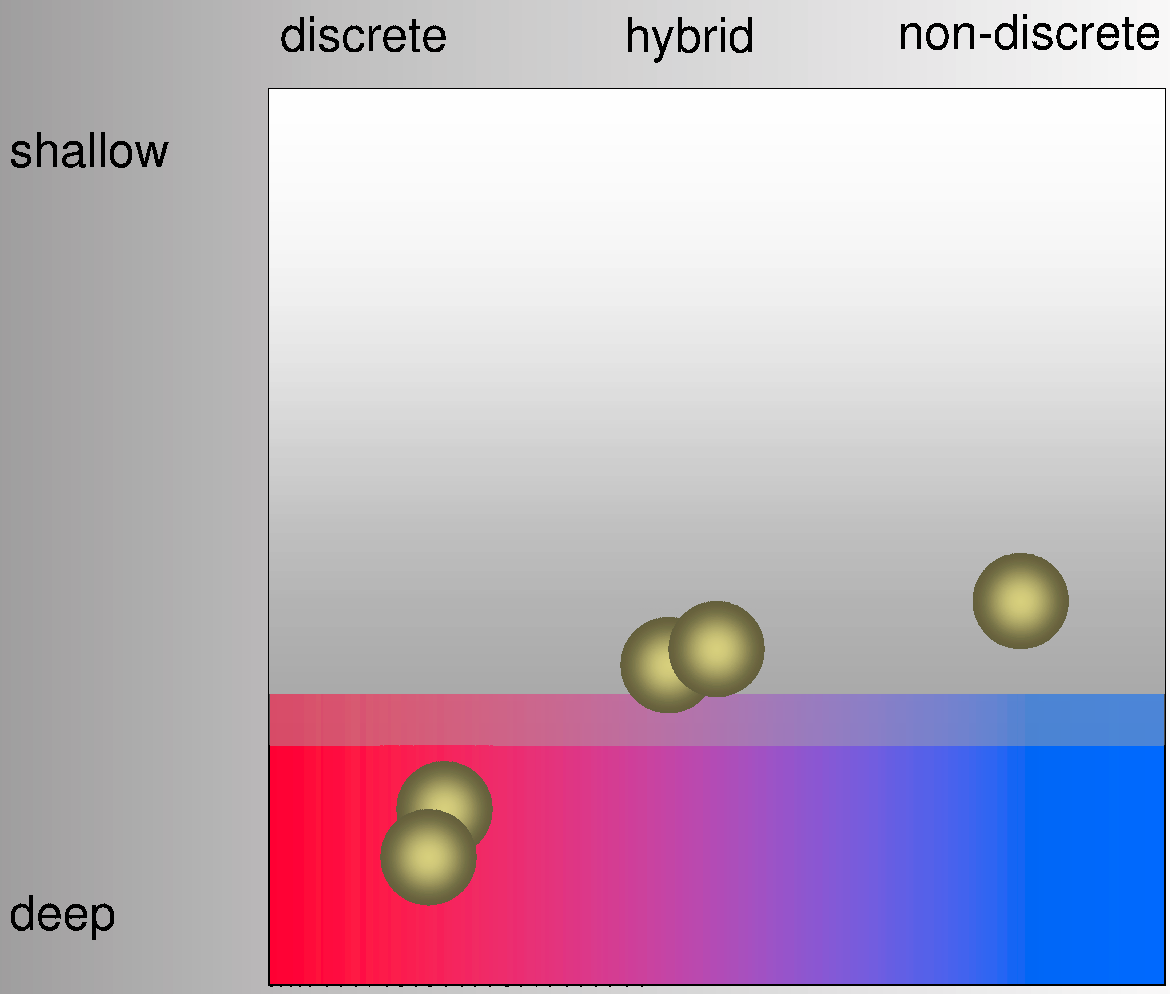
LT METHODS



LT METHODS


discrete
hybrid
non-discrete
shallow
deep
syntactic LFG
parser with ME
selection

COMBINATION OF 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)**

The Consortium

Saarland University (Coordinator)

U. of Cambridge

U. of Sussex

U. of Trondheim

Stanford University

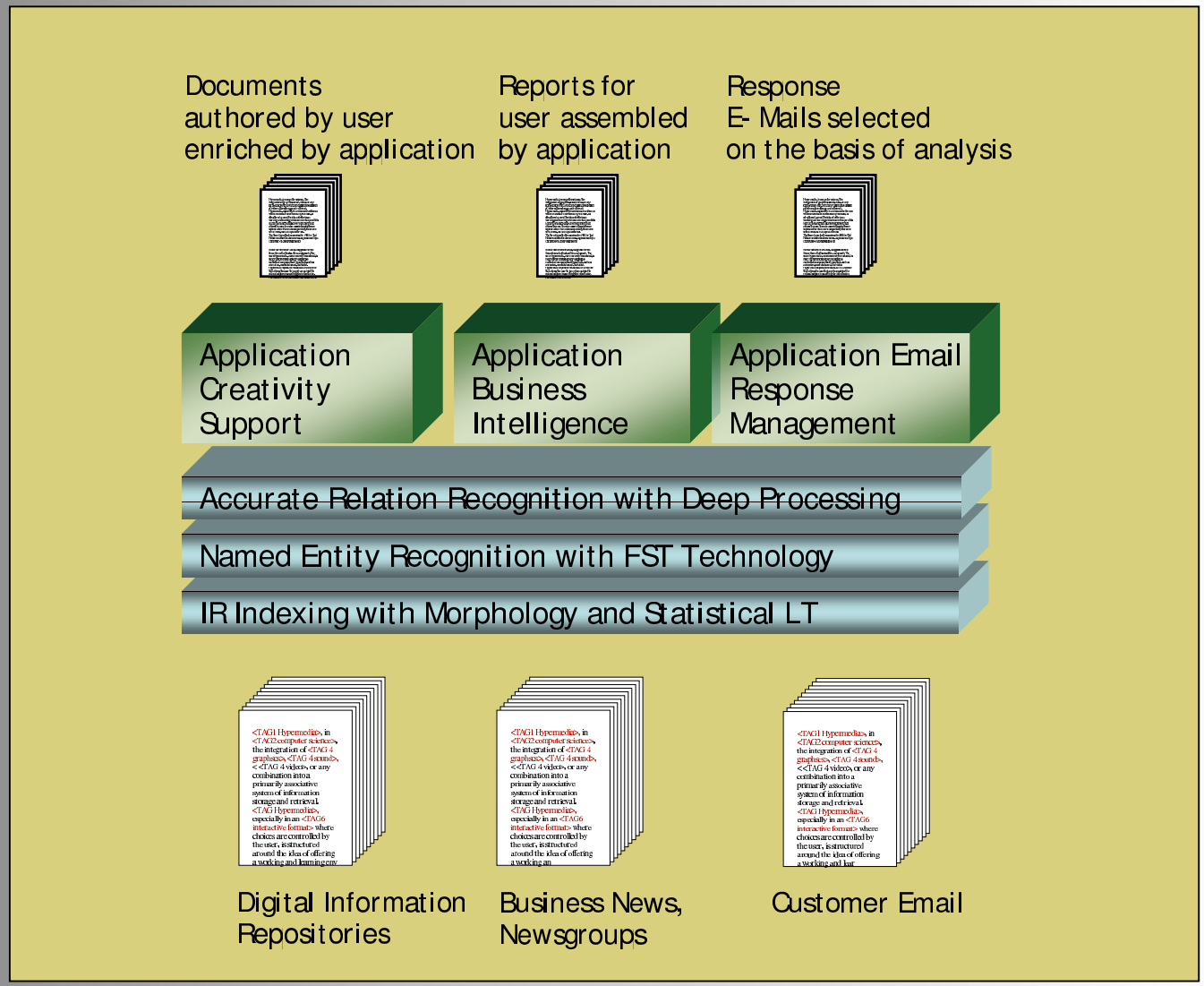
CELI Company of Torino

XtraMind GmbH of Saarbrücken

acrolinx gmbh, Berlin

DFKI Saarbrücken

Three Applications



DEEPTHOUGHT

Project Objectives

- 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

Main project results

- **RMRS provides a common representation for linguistic processing on all levels and for all languages. Shallow and deep processing produce compatible semantic representations.**
- **Different strategies for hybrid NLP have been exploited and implemented.**
- **Strategies for lexicon acquisition based on deep and statistical processing have been implemented.**
- **The theory and implementations are inherently multilingual.**
  - **The Grammar Matrix provides ways to implement new grammars for new languages based on insights from former implementations.**

DEEPTHOUGHT

NL Input

Every experienced vice president served on a committee.

Event-Based LF

every(*x*), experienced(*x*),VP(*x*),serve1(*e,x*), spast(*e*), on2(*e',e,y*), some(*y*), committee(*y*).

Robust Minimal Recursion Semantics (RMRS)

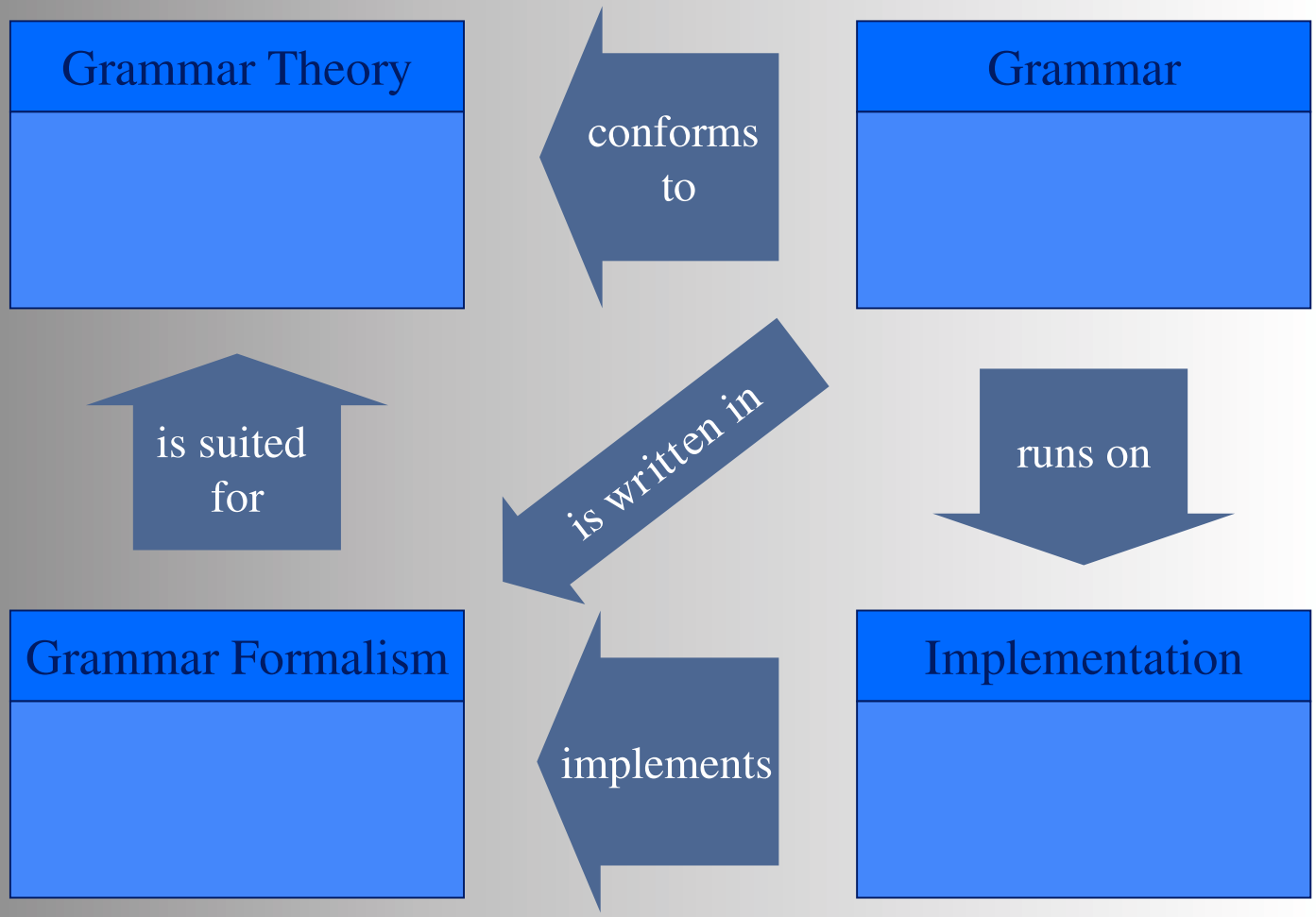
every(*x1*), experienced(*x2*),VP(*x3*),serve1(*e4*), spast(*e4*), on2(*e5*), some(*x6*), committee(*x7*)

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

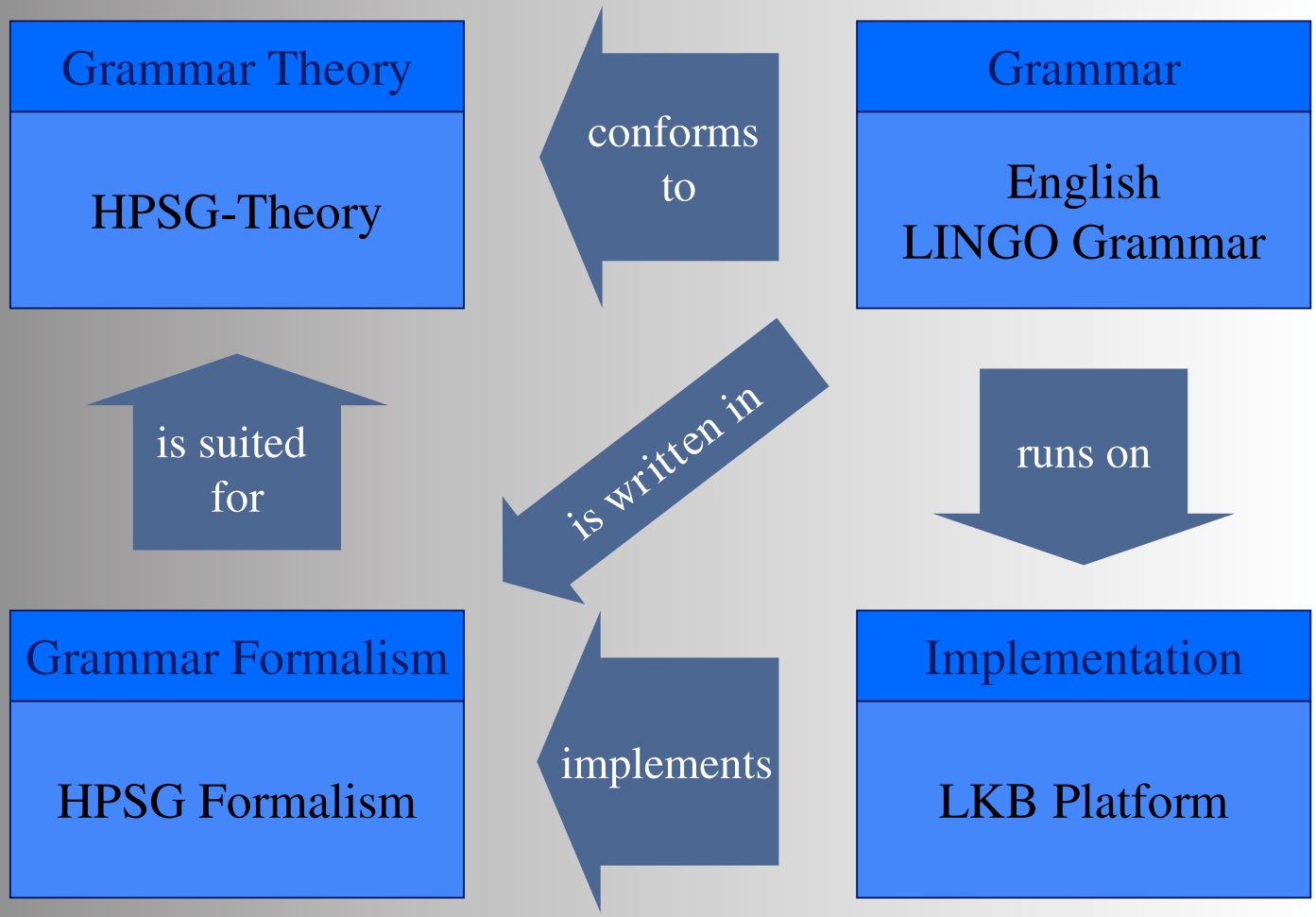
x1=x2, x2=x3, x6=x7.

Blue Delph-In: Open-Source Deep Processing - Netscape
File Edit View Go Bookmarks Tools Window Help
http://www.delph-in.net/
Go
Search
Mail Home Netscape Bookmarks Melanie Siegel :siegel Google LEO English/G... Japanese <->... WhoWhere? E... Jim Breen's W... Mainichi-Kanji
Blue Delph-In: Open-Source Deep Processing
DELPHIN
HOME 0 html_page
INTRODUCTION 1 html_page
RESEARCH PROGRAM 2 pdf_doc
SYSTEMS & GRAMMARS 3 html_page
PUBLICATIONS 4 html_page
NEWS & EVENTS 5 html_page
LINKS 6 html_page
CONTACT US 7 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.
(Created by hansu, last modified 2004-09-19 | comments@delph-in.net)

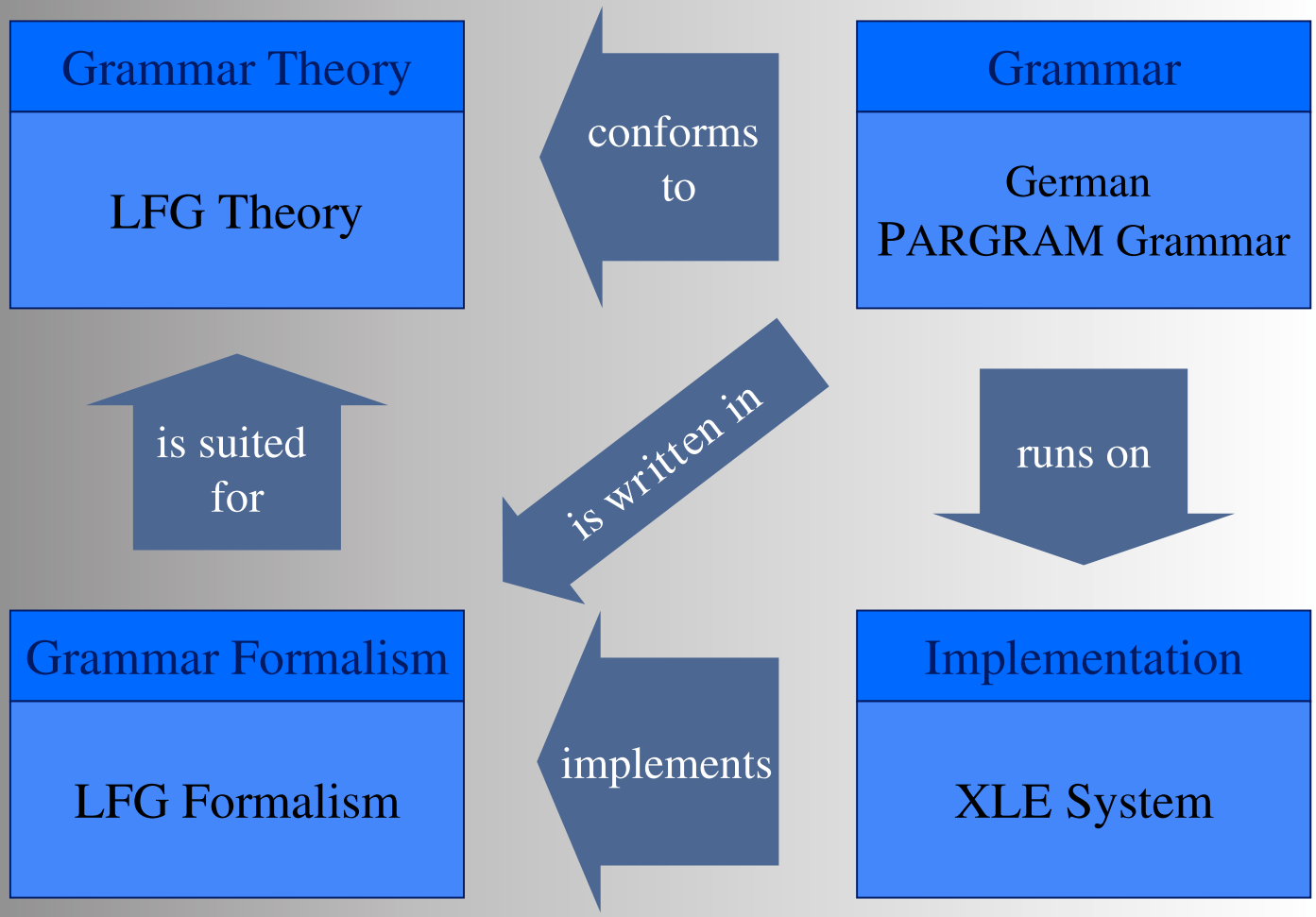
GRAMMAR[4]


Grammar Theory
conforms to
Grammar
is suited for
is written in
runs on
Grammar Formalism
implements
Implementation

GRAMMAR[4]


Grammar Theory
HPSG-Theory
conforms to
Grammar
English LINGO Grammar
is suited for
is written in
runs on
Grammar Formalism
HPSG Formalism
implements
Implementation
LKB Platform

GRAMMAR[4]


Grammar Theory
LFG Theory
conforms to
Grammar
German PARGRAM Grammar
is suited for
is written in
runs on
Grammar Formalism
LFG Formalism
implements
Implementation
XLE System