Hybrid NLP

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

- **The Grammar Matrix version 0.7 is available via CVS.**
- **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**

The Grammar Matrix

- **The Matrix was the direct basis for building up the Italian and the Norwegian grammars.**
- **It was used for the adaptation of the English, German and Japanese grammars to RMRS and SEM-I standards.**
- **Through the use of the matrix grammar, the needed effort in defining the Norwegian and the Italian grammar could be drastically reduced if compared to the development times of earlier grammars.**

Matrix-based multilingual grammar engineering

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

TEXT the man smiles
TOP h1

RELS {
- *prpstn_m_rel*: LBL *h1*; ARG0 *h5*
- *_the_q*: LBL *h8*; ARG0 *x9* pers=3 num=sg gender=n; RSTR *h10*; BODY *h12*
- *_man_n*: LBL *h13*; ARG0 *x9* pers=3 num=sg gender=n
- *_smile_v*: LBL *h15*; ARG0 *e2* tense=present; ARG1 *x9* pers=3 num=sg gender=n
}

HCONS {h5 qeq h15, h10 qeq h13}
ING {}

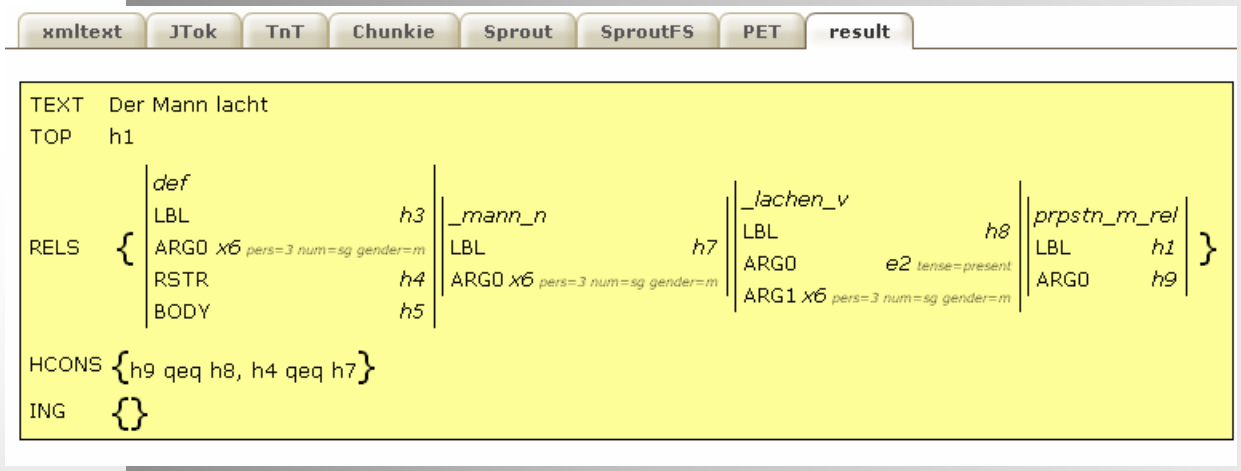
xmltext | PET | result

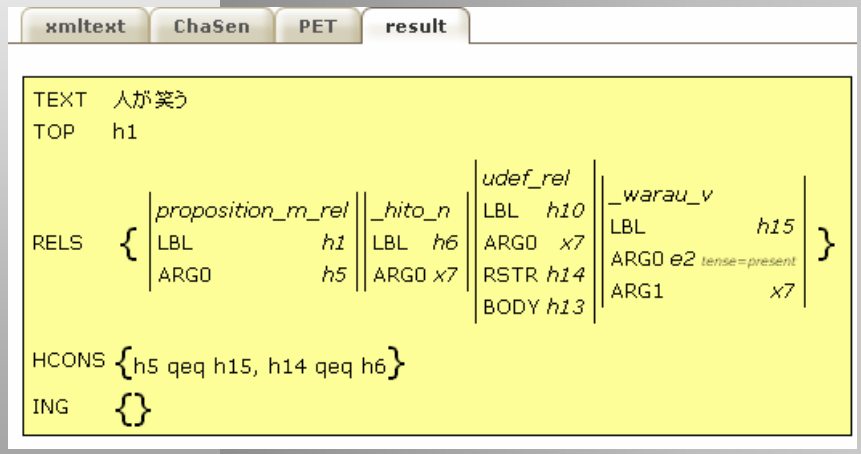
TEXT mannen smiler
TOP h1

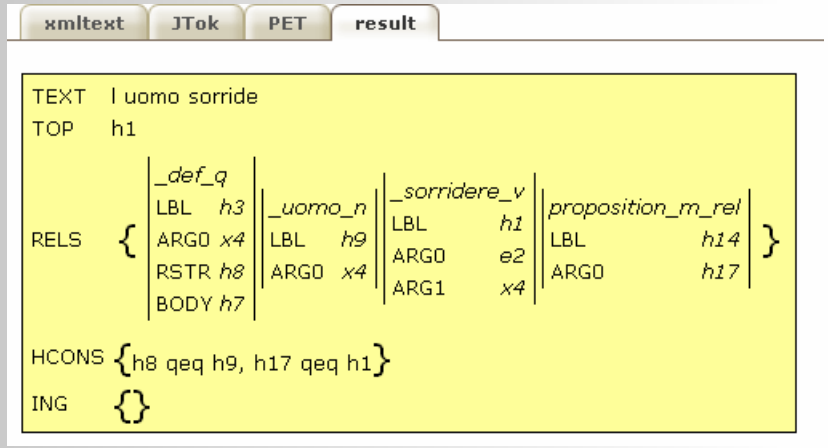
RELS {
- *mann-rel*: LBL *h3*; ARG0 *x4* bounded=+
- *def-q-rel*: LBL *h5*; ARG0 *x4* bounded=+; RSTR *h7*; BODY *h6*
- *card_rel*: LBL *h8*; ARG0 *x4* bounded=+; CARG *1-rel*
- *smile-rel*: LBL *h9*; ARG0 *e2* tense=present; ARG1 *x4* bounded=+
- *prpstn_rel*: LBL *h1*; ARG0 *h10*
}

HCONS {h7 qeq h3, h10 qeq h9}
ING {}

Matrix-based multilingual grammar engineering







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

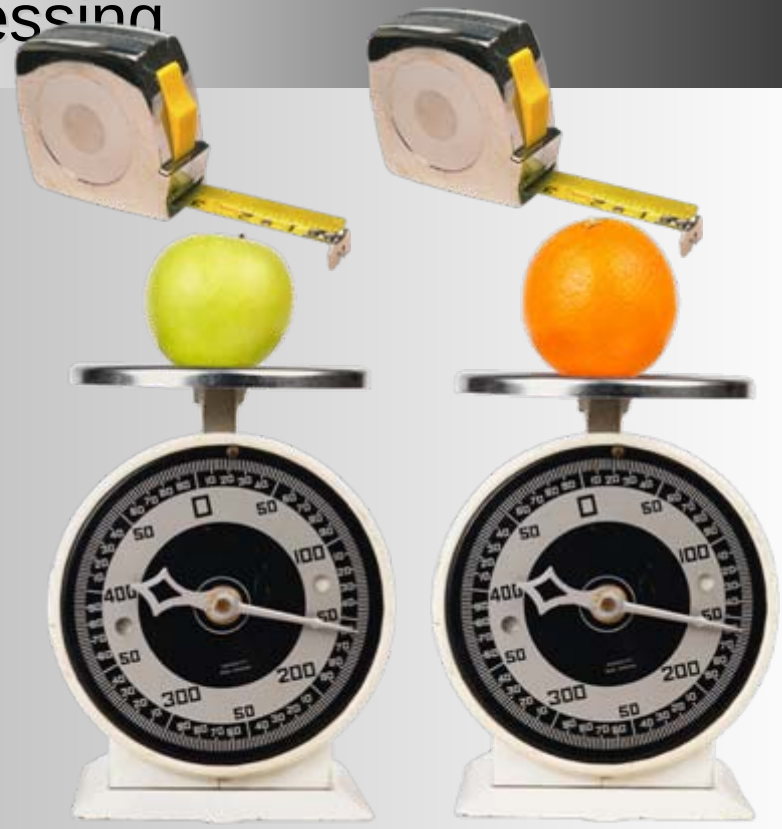
- **Including open-source resources:**
  - **LKB grammar development system (incl. generation)**
  - **PET grammar processing system**
  - **[incr tsdb()] grammar profiling system**
  - **ERG English HPSG**
  - **JACY Japanese HPSG**
  - **NorSource Norwegian HPSG**
  - **Modern Greek Resource Grammar**
  - **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.**

What is deep processing

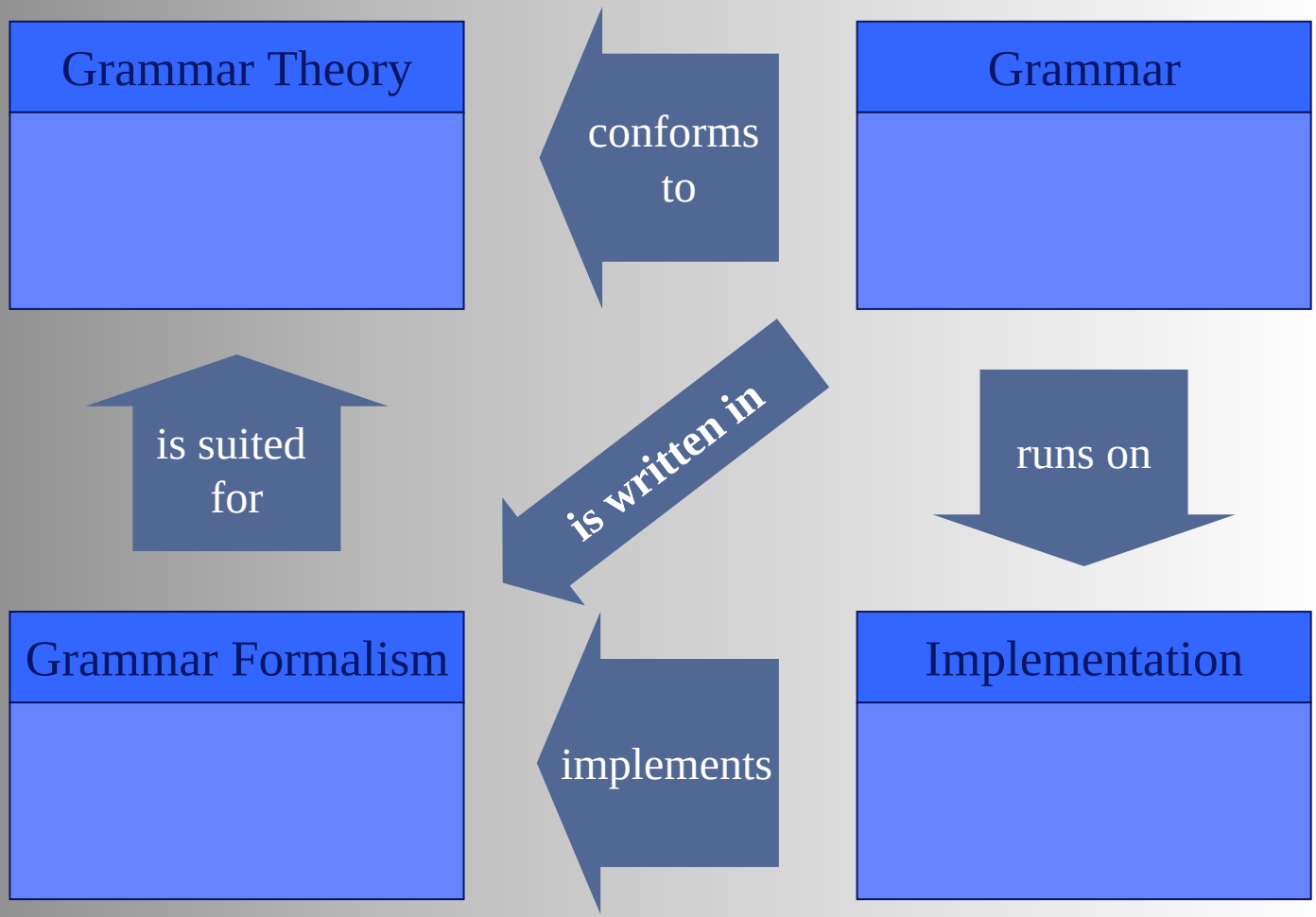


An example

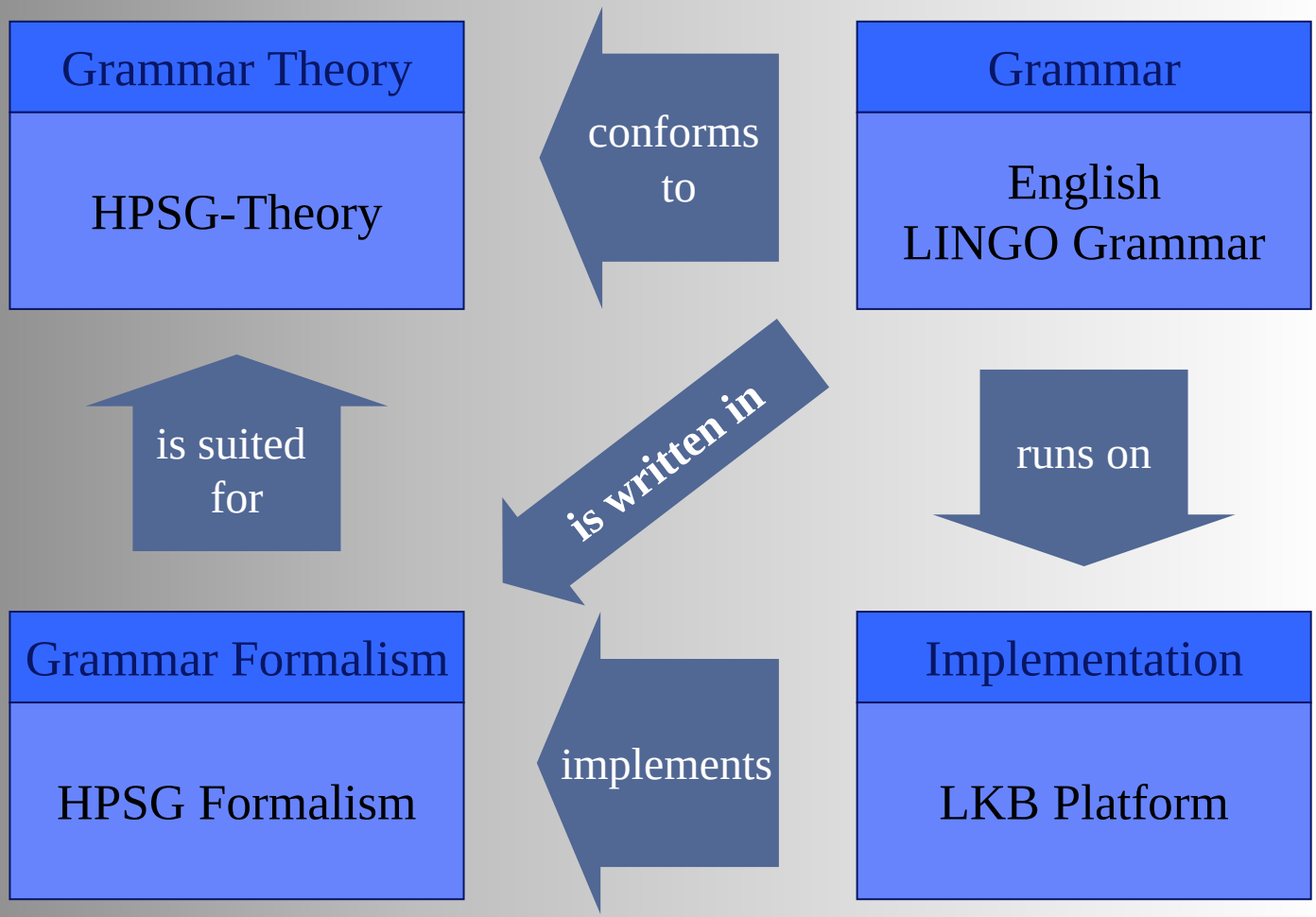
- Whom was this stock
- his stock was easy to forget to sell#

- Peter bekommt das Auto verrosted.
- Peter bekommt das Auto repariert.

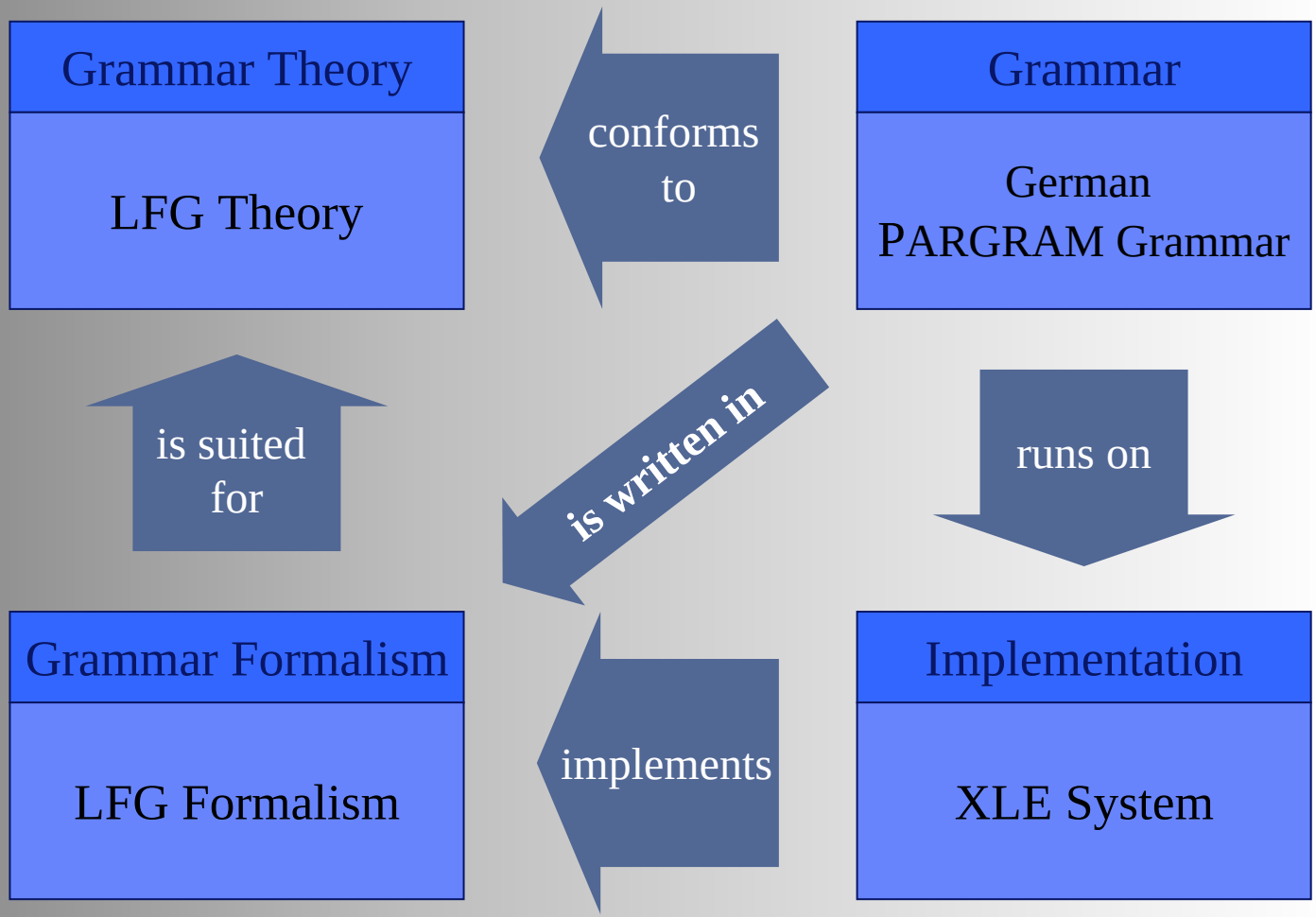
GRAMMAR[4]


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

GRAMMAR⁴


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