

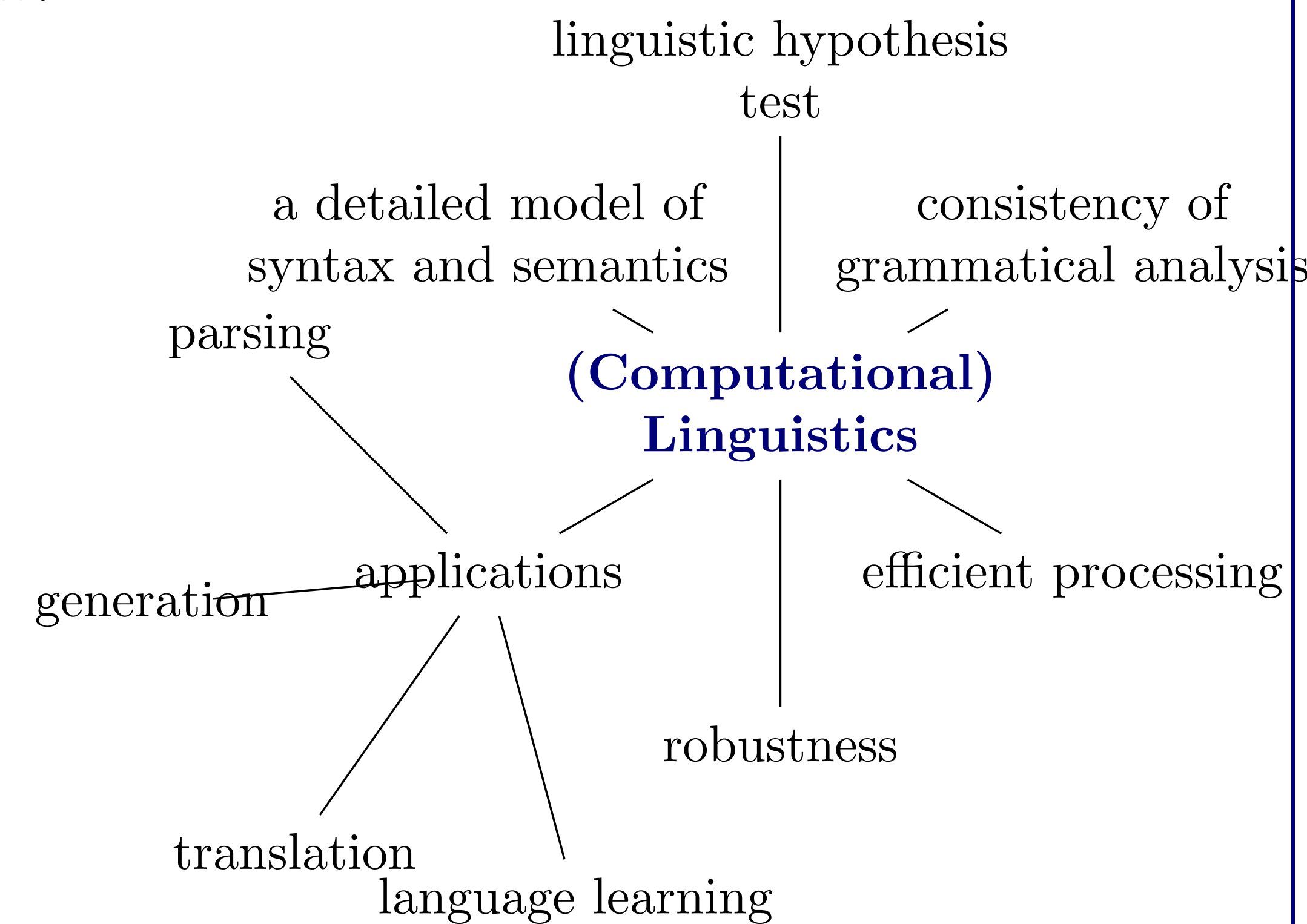


INDonesian Resource grAMmar (INDRA)

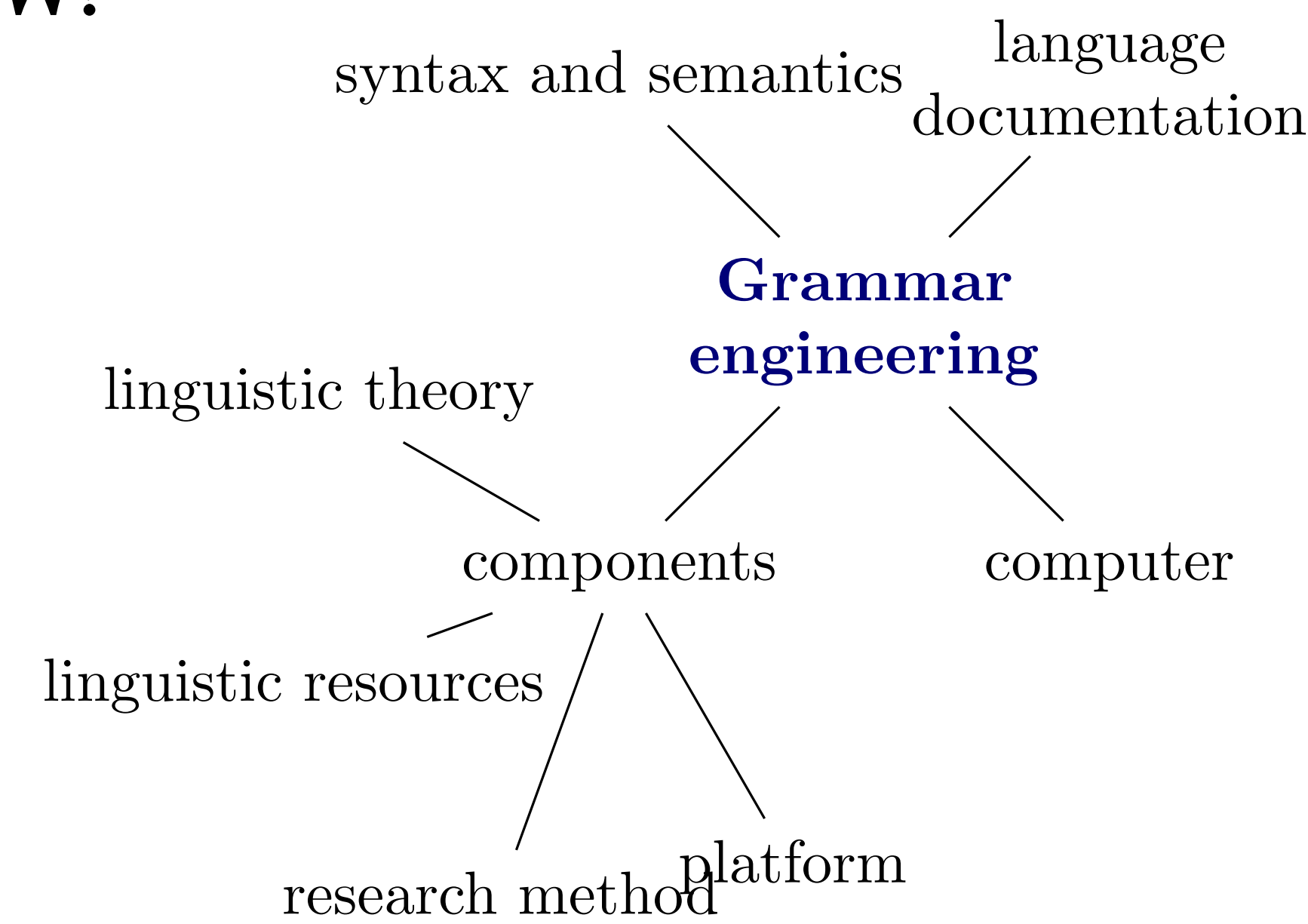
David Moeljadi — Faculty of Arts, Palacký University Olomouc

Indonesian computational grammar

a **grammar** (a set of syntactic and semantic production rules for strings in a formal language) written explicitly enough to be interpreted by a **computer**
WHY?



HOW?



Indonesian language

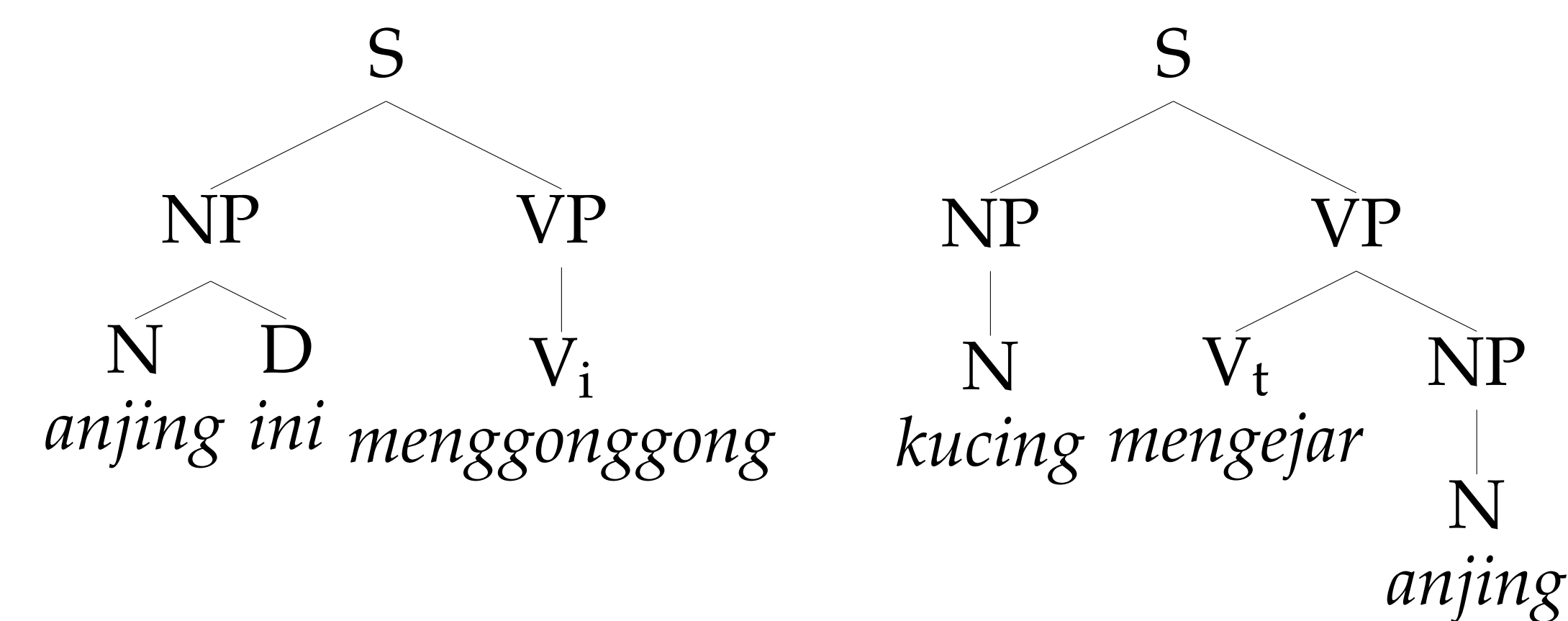
- Austronesian > Malayo-Polynesian > ... > Malay
- Speakers: 43 mio (L1), 156 mio (L2) (2010 census)
- Diglossic
- **Language resources:** KBBI Indonesian dictionary fifth edition, Indonesian Reference Grammar (Sneddon et al., 2010) etc.
- mildly agglutinative, SVO, N-Adj-Dem

Linguistic theory

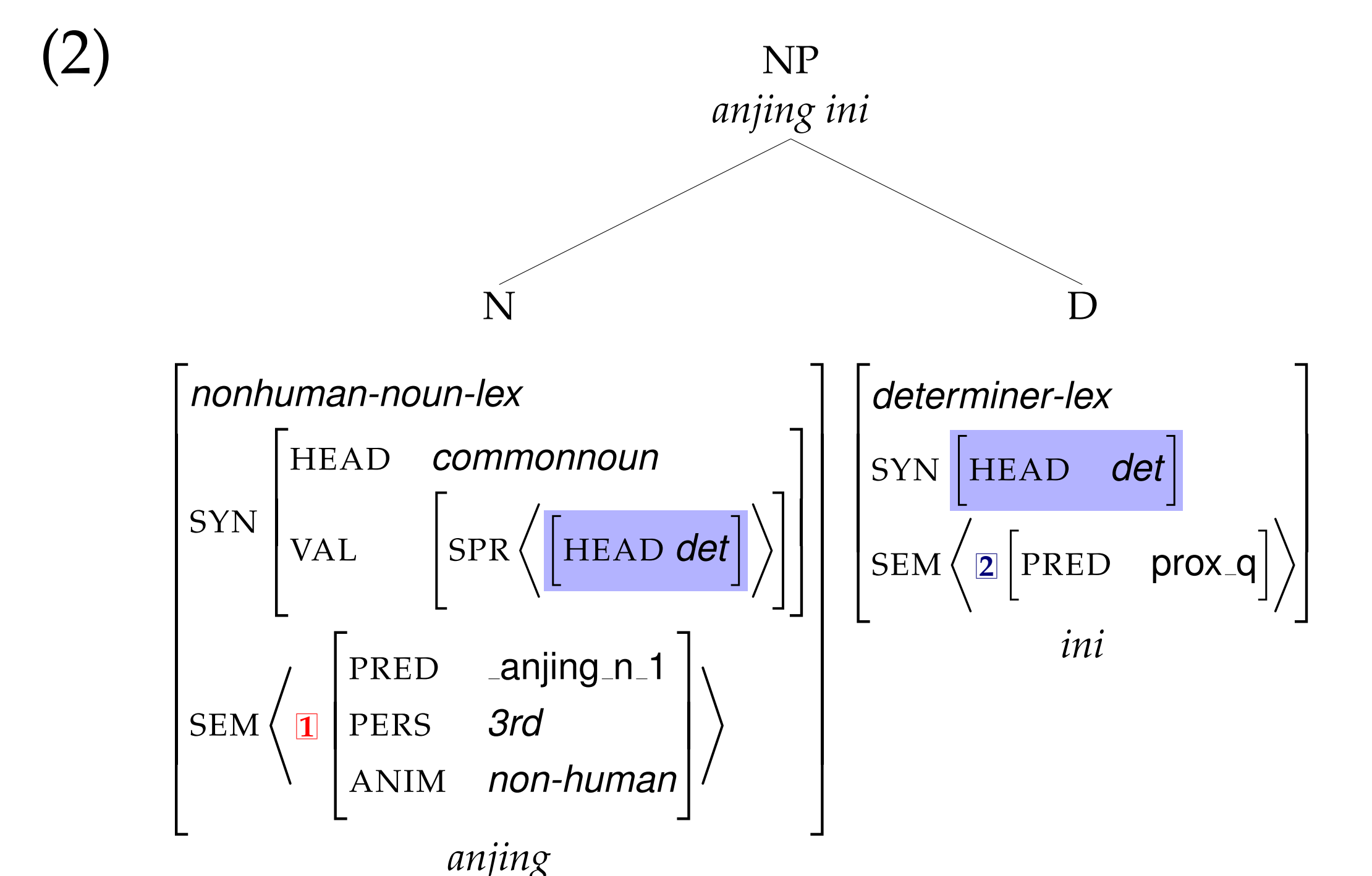
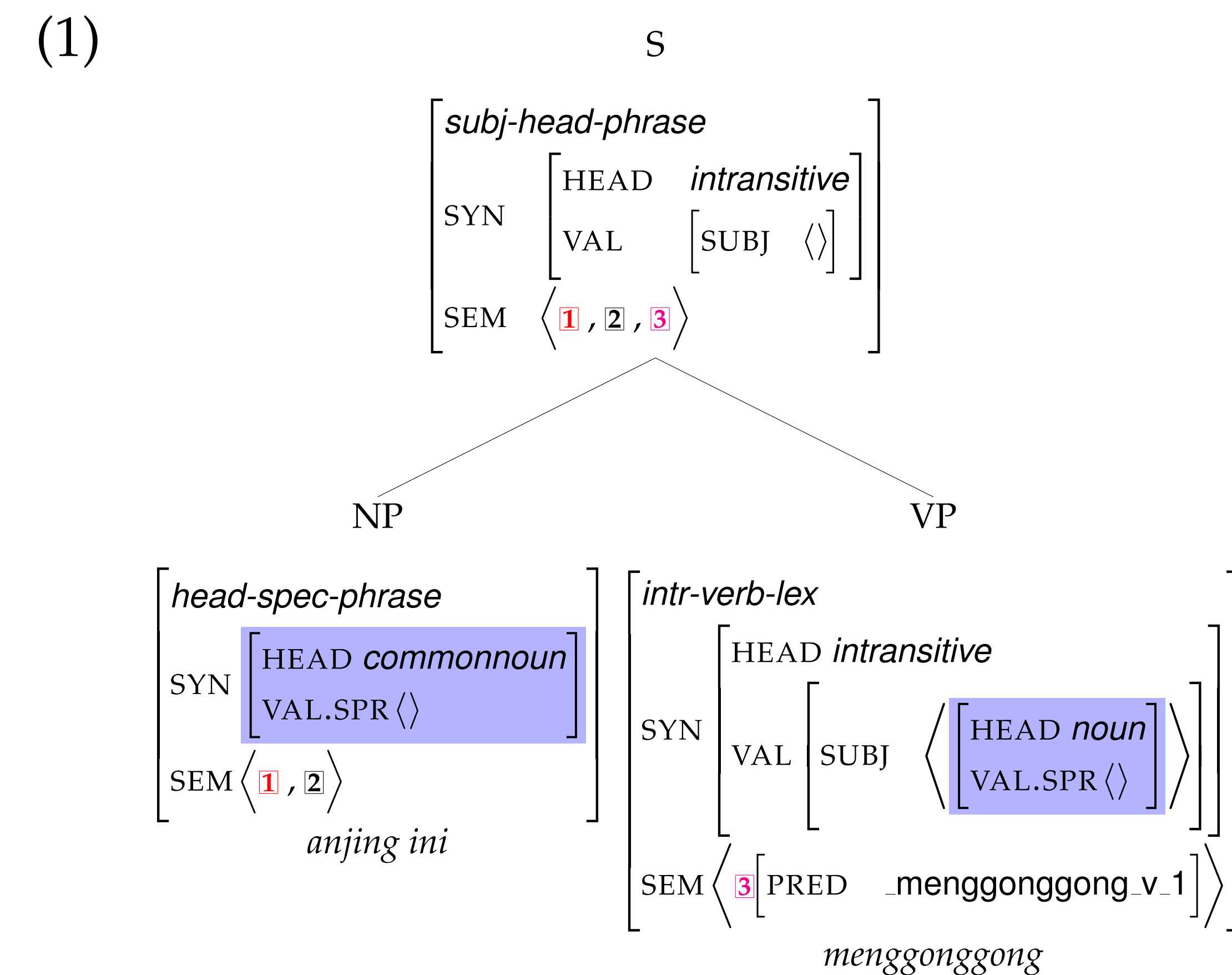
- **Head-Driven Phrase Structure Grammar (HPSG);** Pollard and Sag (1994); Sag et al. (2003)
- **Minimal Recursion Semantics (MRS);** Copestake et al. (2005)

Phrase Structure Grammar

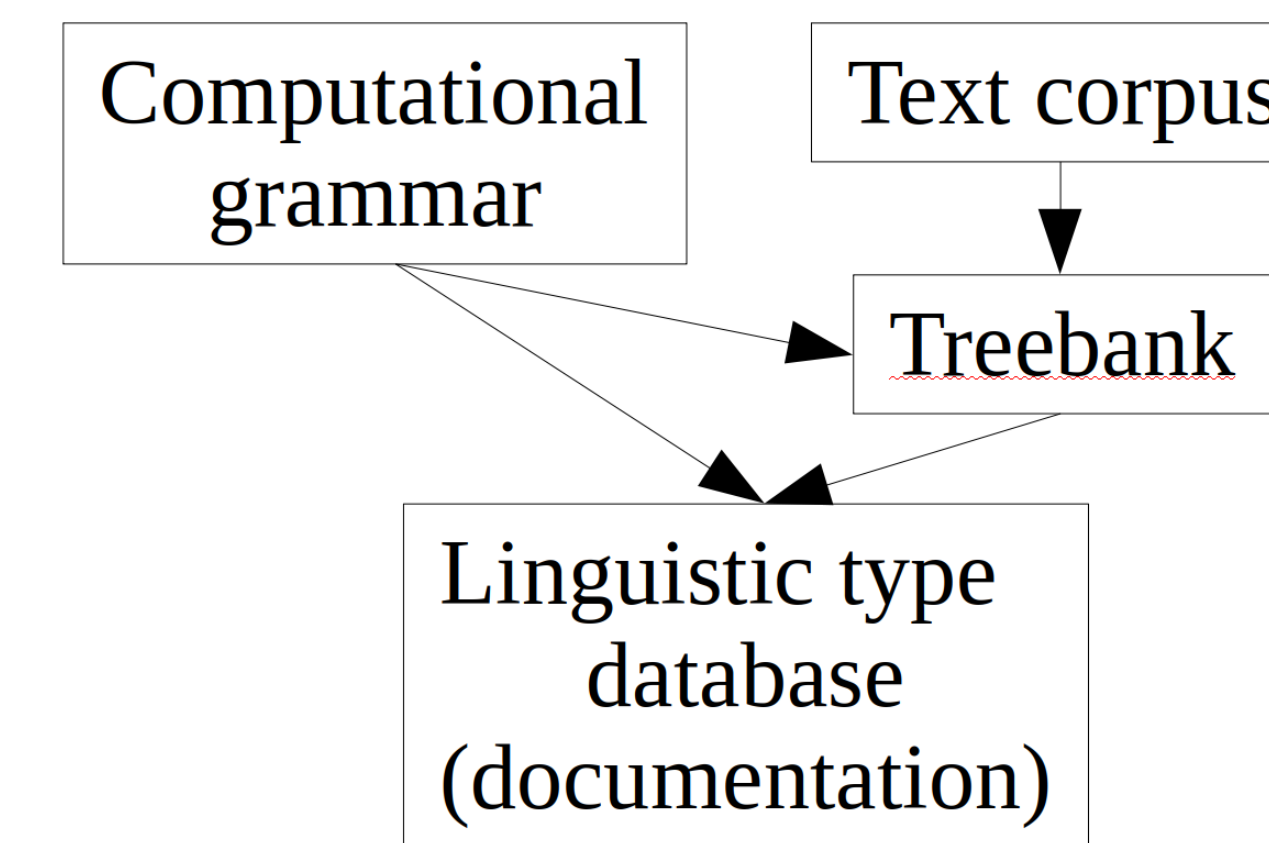
1. a set of **rules**
 $S \rightarrow NP VP$
 $NP \rightarrow N, NP \rightarrow N D$
 $VP \rightarrow V_i, VP \rightarrow V_t NP$
2. a **lexicon** of parts-of-speech (POS) and words
 N : *anjing* "dog", *kucing* "cat", D : *ini* "this"
 V_i : *menggonggong* "bark", V_t : *mengejar* "chase"



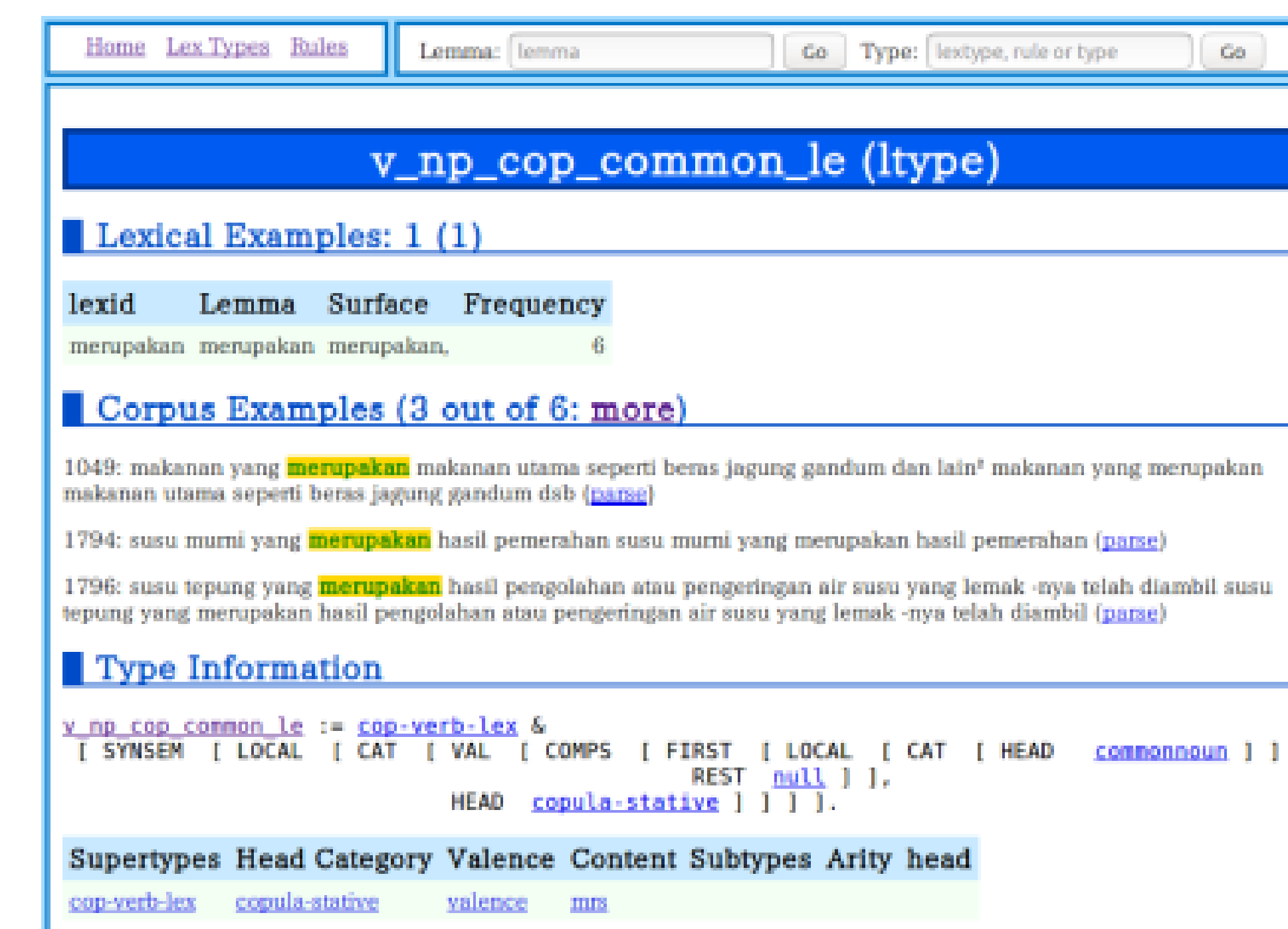
HPSG - Feature structures and unification



Applications: Treebank & LTDB



- a **linguistically annotated corpus** that includes some grammatical analysis beyond the POS level
- to find an example of a certain linguistic construction or a counter-example to a claim about syntactic structure, as a source of information about frequencies and co-occurrences, to build a statistical model, **grammar development**
- semi-automatic (parse and select by hand)
- **JATI:** 1,253 items of KBBI Fifth Edition dictionary definition sentences related to food and drinks
- **Cendana:** 552 items of Traveloka (Indonesian online travel agent) customer service chat data
- **Linguistic type database** (Hashimoto et al., 2007)

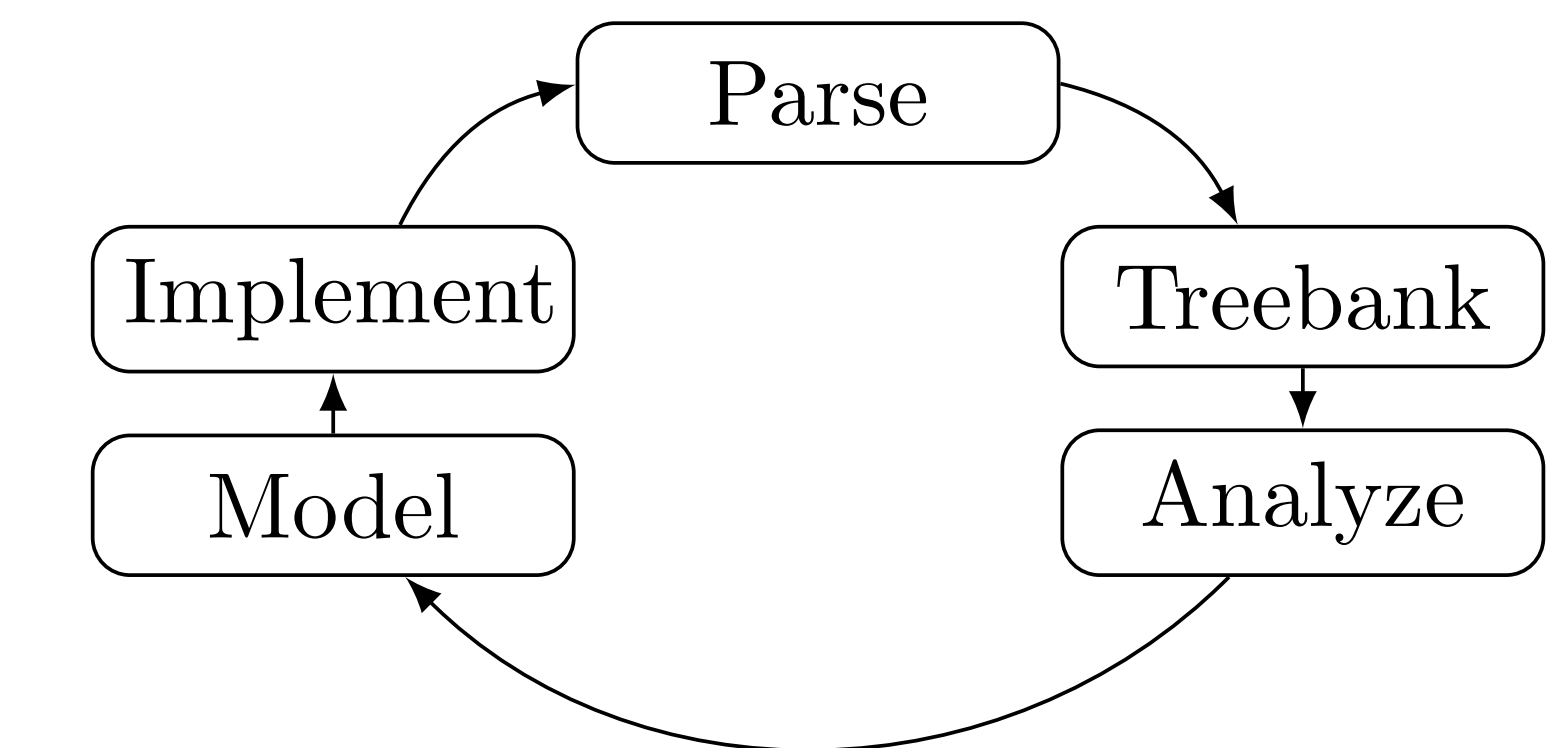


Grammar development

Tools

- DEep Linguistic Processing with HPSG - INitiative (DELPH-IN) (www.delph-in.net)
- **Open-source grammars:**
 - ERG English grammar (Flickinger, 2000)
 - Jacy Japanese grammar (Siegel et al., 2016) etc.
- **Tools:**
 - The Linguistic Knowledge Builder (LKB) grammar development system (Copestake, 2002) etc.

Grammar development cycle



Summary

- INDRA is an **open-source** HPSG Indonesian computational grammar (MIT license)
- phenomena covered: voice, copulas, full noun reduplication, noun & verb subcategorization etc.
- homepage: <http://moin.delph-in.net/IndraTop>, grammar sources: <https://github.com/davidmoeljadi/INDRA>
- demo pages: (1) <http://delph-in.github.io/delphin-viz/demo>, (2) <http://chimpanzee.ling.washington.edu/demophin/indra/>
- applications: sentence parsing and generation, treebanking, linguistic type database etc.

Acknowledgements

I gratefully acknowledge the support of the European Regional Development Fund-Project "Sino-phone Borderlands - Interaction at the Edges" CZ.02.1.01/0.0/16_019/0000791

References

Copestake, A. (2002). *Implementing Typed Feature Structure Grammars*. CSLI Publications, Stanford.

Copestake, A., Flickinger, D., Pollard, C., and Sag, I. A. (2005). Minimal Recursion Semantics: An Introduction. *Research on Language and Computation*, 3(4):281–332.

Flickinger, D. (2000). On Building a More Efficient Grammar by Exploiting Types. 6(1):15–28.

Hashimoto, C., Bond, F., and Flickinger, D. (2007). The Lextype DB: a web-based framework for supporting collaborative multilingual grammar and treebank development. In *International Workshop on Intercultural Collaboration*, pages 76–90. Springer.

Moeljadi, D., Bond, F., and Song, S. (2015). Building an HPSG-based Indonesian Resource Grammar (INDRA). In *Proceedings of the GEAF Workshop, ACL 2015*, pages 9–16.

Pollard, C. and Sag, I. A. (1994). *Head-driven phrase structure grammar*. University of Chicago Press.

Sag, I. A., Wasow, T., and Bender, E. M. (2003). *Syntactic Theory: A Formal Introduction*. CSLI Publications, Stanford, 2 edition.

Siegel, M., Bender, E. M., and Bond, F. (2016). *Jacy: An implemented grammar of Japanese*. CSLI Publications.

Sneddon, J. N., Adelaar, A., Djenar, D. N., and Ewing, M. C. (2010). *Indonesian Reference Grammar*. Allen & Unwin, New South Wales, 2 edition.