

RRGparbank in a nutshell

- Parallel corpus of RRG (Role and Reference Grammar) annotated sentences
- **Source:** George Orwell's '1984' (≈ 6700 sentences) and its translations
- **Languages:** English, Farsi, French, German, Russian (+ Hungarian, Turkish)
- **Largest RRG-based treebank available**
- **Genuine annotation**, i.e. all sentences are manually validated
- **Three classes of annotation:**
 - ◊ gold: validated by at least two annotators
 - ◊ silver: validated by one annotator
 - ◊ bronze: automatically generated parses
- **11 annotators** from 2019
- Overall **average annotator agreement** of 95.3% EVALB f-score
- **Multilingual data to browse, query, and download**

Motivation

- **Corpus-based investigations of RRG structures** across languages thanks to query and download features
- **Data-driven syntactic parsing** for several languages
- Exploration of **linguistic phenomena not yet covered** by the RRG theory

First Release Statistics

- English, German, Farsi, French, and Russian seed data
- High degree of parallelism and a broad coverage of linguistic constructions

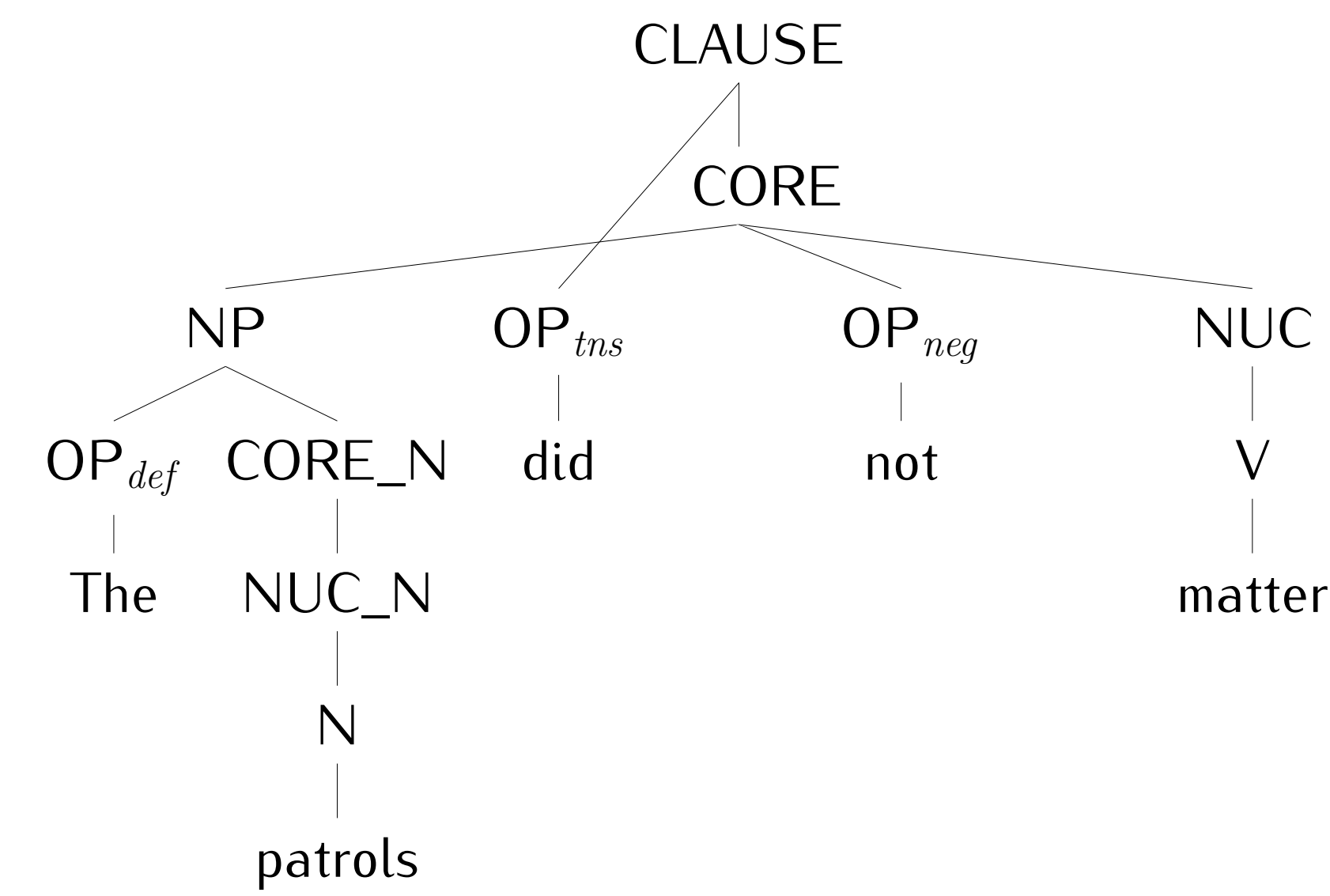
	EN	EN-SEED	DE-SEED	FR-SEED	RU-SEED	FA-SEED
Number of sentences	6 737	1 450	1 454	1 555	1 416	1 476
Number of tokens	122 843	23 750	23 444	24 670	17 697	22 456
Average sentence length	18.2	16.4	16.1	15.9	12.5	15.2
Not yet annotated	0	0	0	0	0	1 010
Silver	348	0	889	1 309	1 019	589
Gold with 1 annotation	2 691	575	286	112	183	0
Gold with ≥ 2 annotations	3 698	875	279	134	214	0

Demo: RRGparbank First Release



<https://rrgparbank.phil.hhu.de>

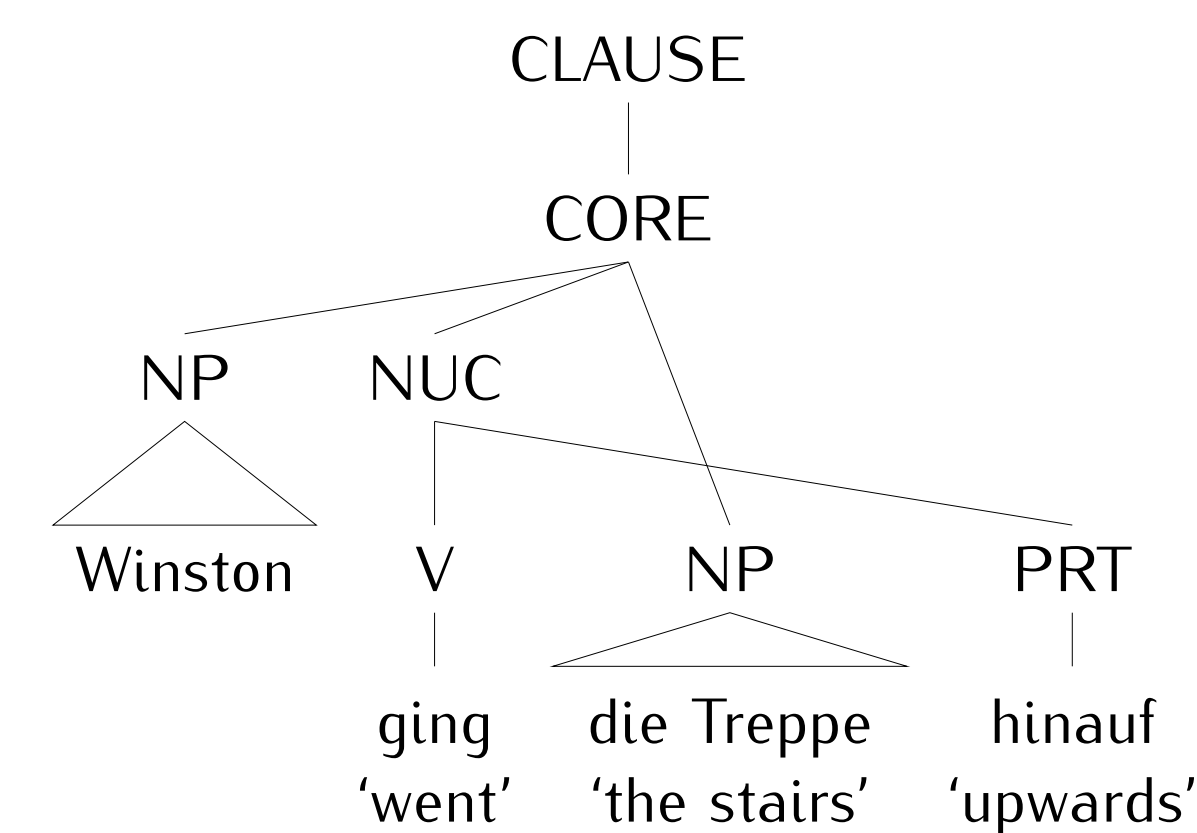
Role and Reference Grammar (RRG) [1,2,3]



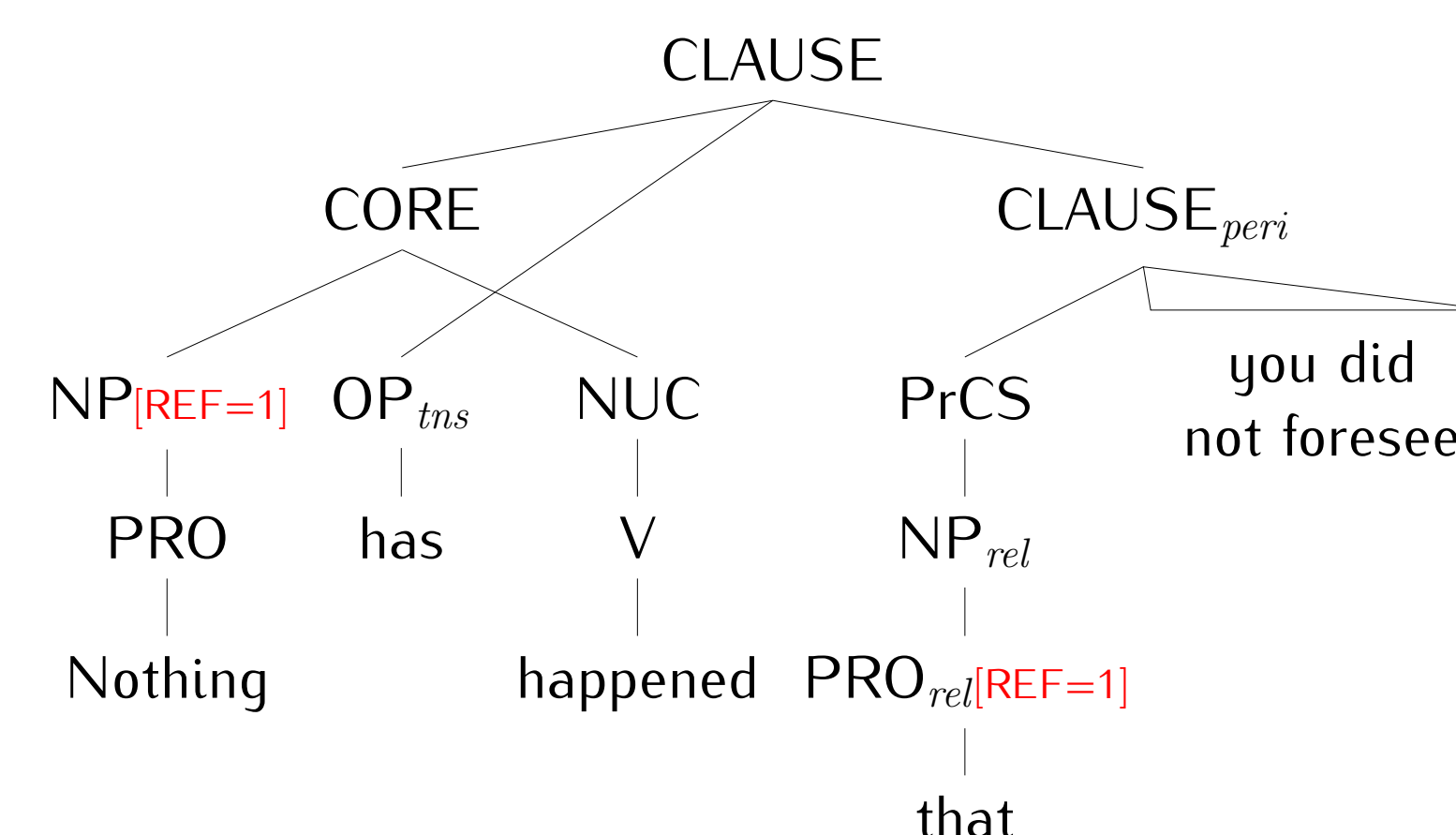
- Framework inspired by typological concerns
- Key assumption: layered structure of the clause
- RRG categories:
 - nucleus:** (verbal) predicate
 - core:** nucleus + arguments
 - clause:** core + extracted arguments
 - peripheries:** modifiers
 - operators:** elements encoding tense, definiteness, aspect or negation; attach to layer over which they take semantic scope

Selected Linguistic Phenomena

Discontinuous NUC for German particle verbs

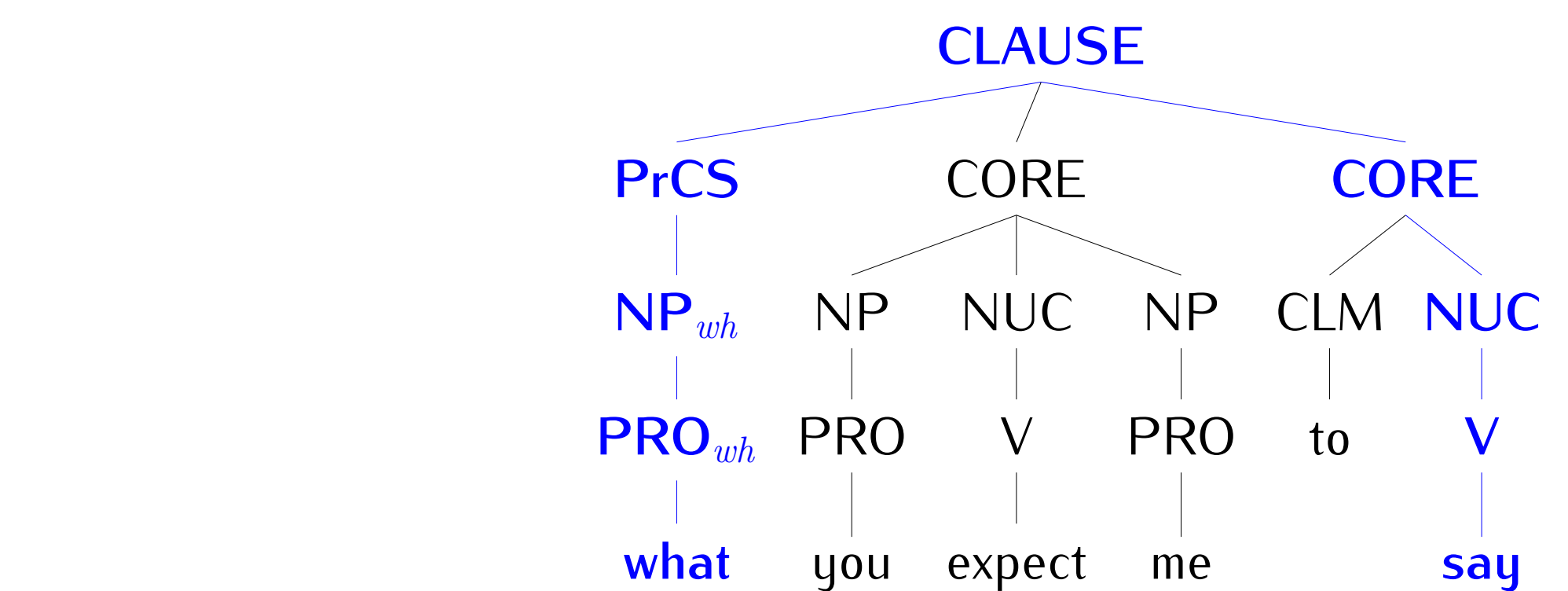
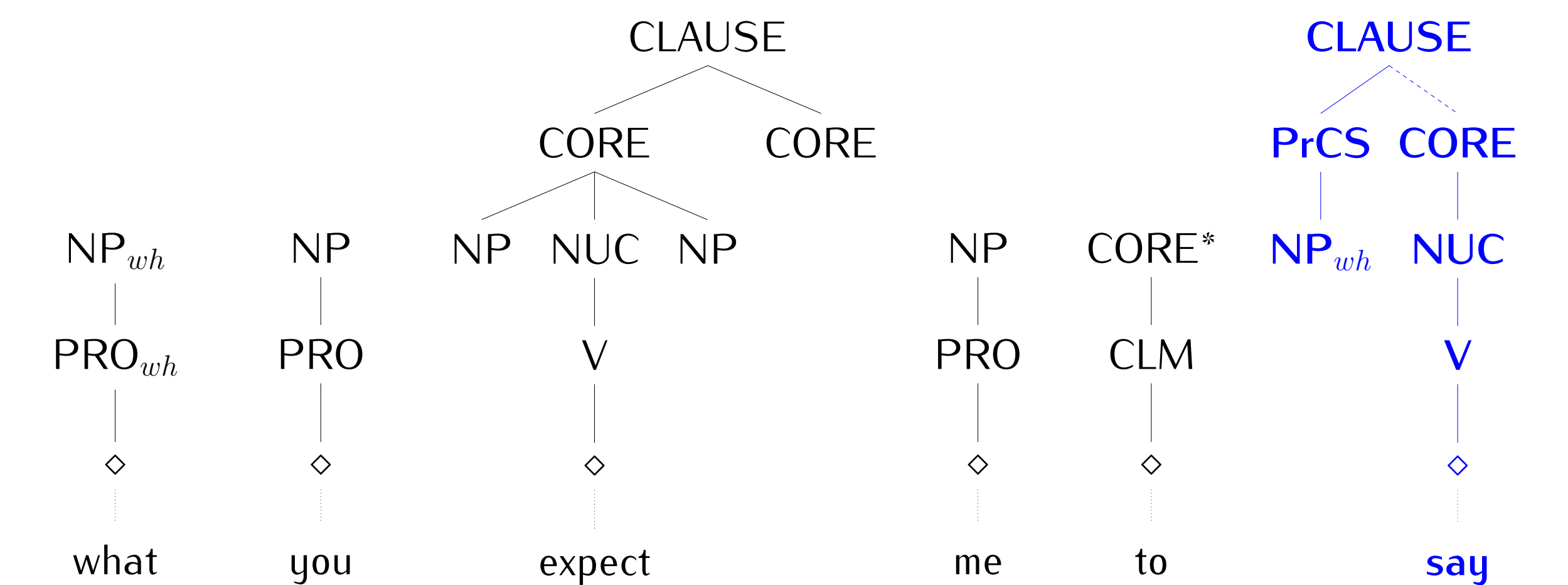


Extrapolated relative clause



Applications

Formalization of RRG: Tree Wrapping Grammar (TWG) [4,5,6]



Statistical parsing for several languages [7]

- Statistical TWG parser based on a fine-tuned multilingual BERT model and single-language BERT models
- Parser available at <https://rrgparser.phil.hhu.de>

	multilingual model	single-language models	exact match (mult. model)	exact match (sing. model)	# sents	∅ len.
en	86.27	86.56	122	155	566	15.43
de	85.19	84.15	95	80	561	13.86
fr	85.68	85.21	66	71	289	11.66
ru	86.16	84.74	115	108	486	9.68
fa	80.80	74.37	37	17	127	8.66

Future Work

- Semantic annotations and frame-semantic parsing
- Finish annotation of non-English data

References [1] Van Valin, Jr., R. D. and LaPolla, R. J. (1997). Syntax: Structure, meaning, and function. Cambridge University Press. [2] Van Valin, Jr., R. D. (2005). Exploring the syntax-semantics interface. Cambridge University Press. [3] Van Valin, Jr., R. D. (2010). Role and Reference Grammar as a framework for linguistic analysis. In Bernd Heine et al., editors, The Oxford Handbook of Linguistic Analysis, pages 703–738. Oxford University Press, Oxford. [4] Kallmeyer, L., Osswald, R., and Van Valin, Jr., R. D. (2013). Tree Wrapping for Role and Reference Grammar. In G. Morrill et al., editors, Formal Grammar 2012/2013, volume 8036 of LNCS, pages 175–190. Springer. [5] Kallmeyer, L. and Osswald, R. (2017). Combining predicate-argument structure and operator projection: Clause structure in role and reference grammar. In Proceedings of the 13th International Workshop on Tree Adjoining Grammars and Related Formalisms, pages 61–70. [6] Osswald, R. and Kallmeyer, L. (2018). Towards a formalization of Role and Reference Grammar. In Rolf Kailuweit, et al., editors, Applying and Expanding Role and Reference Grammar, pages 355–378. Albert-Ludwigs-Universität, Universitätsbibliothek. [NIHIN studies], Freiburg. [7] Bladier, T., Waszczuk, J., and Kallmeyer, L. (2020). Statistical parsing of tree wrapping grammars. In Proceedings of the 28th International Conference on Computational Linguistics, pages 6759–6766.