

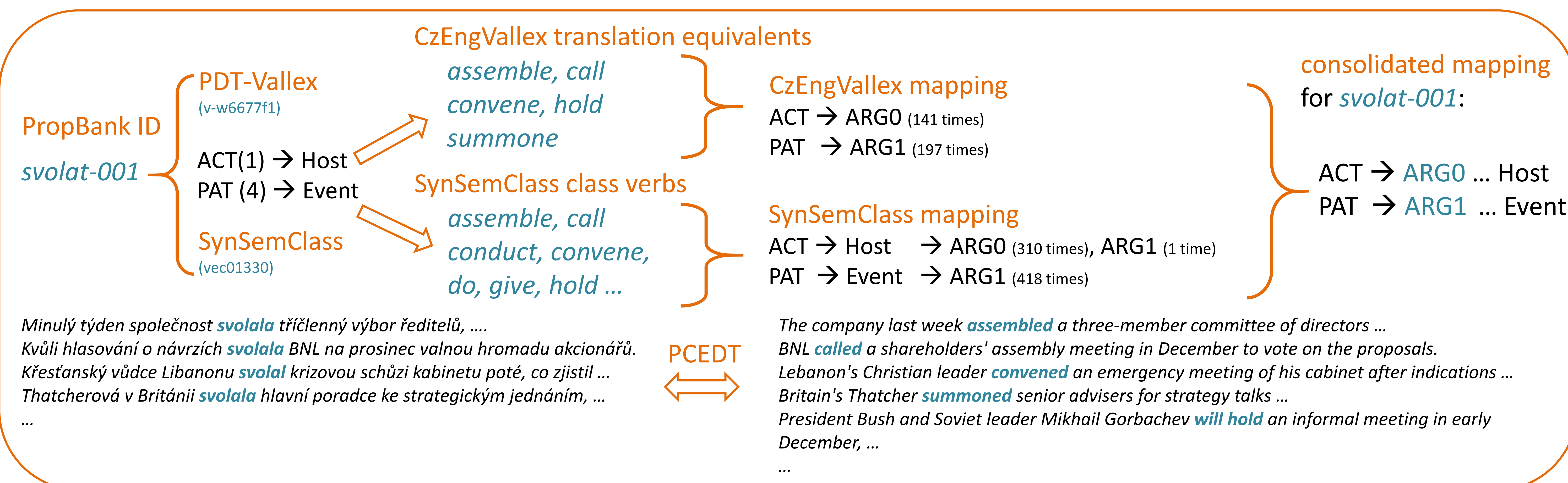
Goal and Research Questions

We present here a high-coverage pilot study of **mapping the labeling system used in PropBank (for English) to PDT-based annotation (for Czech)**, which has so far used mainly valency lexicons for annotation projects, under different levels of specification and different theoretical assumptions. The purpose of this study is both theoretical (**comparing the argument labeling schemes**) and practical (to be able **to annotate Czech under the standard UMR specifications**).

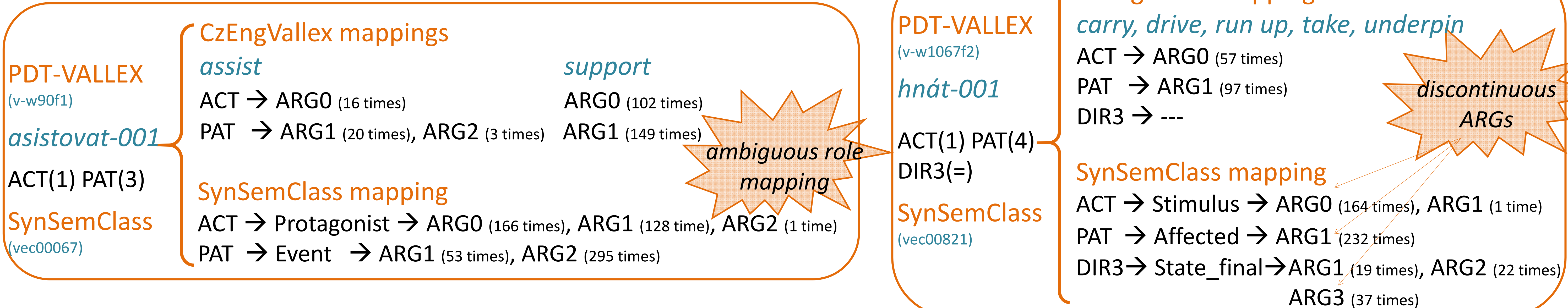
The Mapping Algorithm

1. For each Czech verb (sense), collect its valency frame from the valency lexicon (**PDT-Vallex**) & its semantic class and semantic roles (**SynSemClass**)
2. Collect its PCEDT translation equivalents (**CzEngVallex**) and all equivalents from the same class (**SynSemClass**)
3. Collect available argument mappings (**CzEngVallex**, **SynSemClass**, **EngVallex-PropBank** mapping) and **PCEDT** corpus frequencies
4. Consolidate the information and create the new, complete **PropBank-style roleset** for the given Czech verb sense

Example – Successful Mapping (verb *svolat* "assemble; call; convene")



Examples – Unsuccessful Mappings



Statistics: arguments (in fully mapped rolesets)

	unambiguous	preferred	unmapped	total
functors	9,465	8,579	24,072	42,116
percent	23	20	57	100

Statistics: rolesets

	auto-suggested	unassigned	total
rolesets	5,085	10,569	15,654
percent	32	68	100

Outputs

- **PropBank style rolesets for more than 5,000 Czech verbs** (= verb senses) obtained automatically as unambiguous or preferred mappings from different lexical resources.
- Collected data with the necessary valency / predicate-argument information and clickable links for annotators to finish the work manually in an efficient manner.
- The resulting Czech PropBank frame files for future Czech UMR annotation requiring PropBank-style argument labels.
- Possibility for more direct, large-scale comparison between the two approaches to predicate-argument labeling.