Automatic mapping Lexical Resources: A Lexical Unit as a Keystone
Eduard Bejček, Václava Kettnerová and Markéta Lopatková

This research has been supported by the Czech Science Foundation, GA ČR, grant No. G20-16840S and by grant CDP70161317 of the Ministry of Education, Youth and Sports of the Czech Republic.

Task: Link two lexicons together. Both are valency lexicons. Both are for Czech language.

An example of VALLEX lexicon:
- Complex description of selected verbs.
- Verbs selected according to their frequency.
- Some supplement information (missing in PDT-Vallex).

An example of PDT-Vallex lexicon:
- Description of only selected lexical units.
- Verbs and lexical units selected from PDT corpora.
- Each LU is provided with a corpus evidence.

All possible mappings are shown. All of them will be assigned a score according to lemmas, valency frames and reciprocity.

Linking statistics
- VALLEX
  - Verbs in lexicon covered by both lexicons: 5,341
  - Average number of UEs per verb lemma: 8,836
- PDT-Vallex
  - Verbs in lexicon covered by both lexicons: 5,341
  - Average number of UEs per verb lemma: 8,836

Comparison with manual linking
- VALLEX
  - Verbs in lexicon covered by annotation: 1,154
  - Average number of UEs per verb lemma: 3.5
- PDT-Vallex
  - Verbs in lexicon covered by annotation: 769
  - Average number of UEs per verb lemma: 2.2

Evaluation against manual annotation
- VALLEX
  - Average weighted score per verb: 93
- PDT-Vallex
  - Average weighted score per verb: 92

Conclusion
- If there is:
  - a different granularity,
  - a border between lexical units set differently,
  - missing lexical unit or
  - not enough comparable information, it is very difficult to automatically link the verb.

(No matter which format we use for the data.)