Sentence Information Structure – Contextual Boundness of Subordinate Clauses in Czech Based on the Corpus Analysis

Kateřina Rysová

rysova@ufal.mff.cuni.cz

Charles University in Prague, Faculty of Mathematics and Physics,

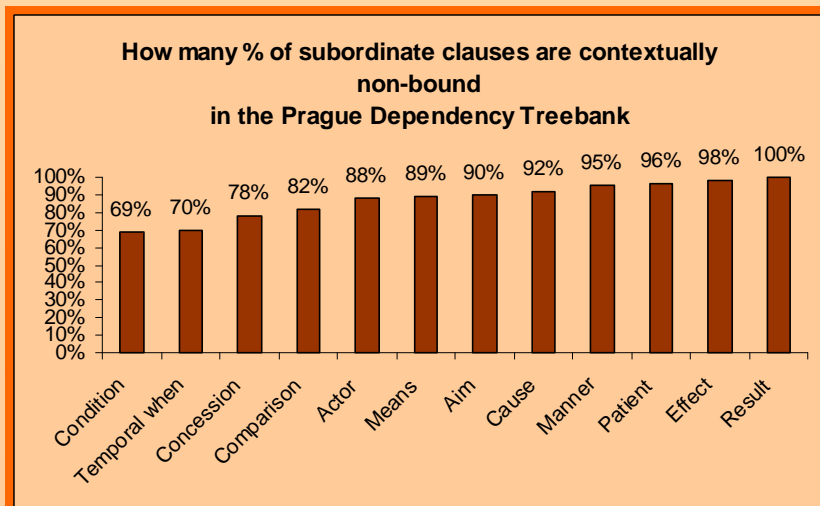Institute of Formal and Applied Linguistics

Aim of Research

To find out how often the various types of subordinate clauses are contextually bound and how often contextually non-bound in Czech newspaper texts.

Methods and Material

- The analysis was based on the data from the *Prague Dependency Treebank* (it contains almost 50,000 sentences).
- The theory of **Topic-Focus Articulation** of the **Prague Functional Generative Description** was used.
- **Topic** elements are mostly **contextually bound**.
- **Focus** elements are mostly **contextually non-bound**.

Example

Example of a contextually non-bound subordinate clause expressing condition in Czech:

- In Czech: *Ke zkouškám je nutné předložit: ... Vzorek výrobku v provozuschopném stavu. **Pokud je k provozu nutné i příslušenství**, předkládá se i toto příslušenství.*
- In English: *For testing, it is required to provide: ... Sample of the product in serviceable condition. **If also accessories are necessary for operation**, they shall be submitted too.*

Results of the Research

The "Scale of Contextual Boundness" for Czech

Type of subordinate clause	How many subordinate clauses are contextually bound in PDT	How many subordinate clauses are contextually non-bound in PDT	How many % of subordinate clauses are contextually non-bound in PDT
Condition	346	773	69 %
Temporal (when)	161	383	70 %
Concession	99	354	78 %
Comparison	33	154	82 %
Actor	132	977	88 %
Means	11	93	89 %
Aim	39	362	90 %
Cause	53	647	92 %
Manner	9	159	95 %
Patient	207	4463	96 %
Effect	70	3308	98 %
Result	0	129	100 %

How many % of subordinate clauses are contextually non-bound in the Prague Dependency Treebank

Conclusion

The results demonstrate that some types of subordinate clauses tend to appear more often as contextually bound (e.g. conditional clauses, temporal clauses) than others (e.g. object clauses, clause of manner, clause of cause etc.).

From this, we can deduce that some semantic categories expressed as subordinate clauses appear also more often as known for the addressee and some are supposed to be new or bringing the new relations among the facts in the majority of cases.

Data sources

BEJČEK, E. a kol. (2012): Prague Dependency Treebank 2.5 – a revisited version of PDT 2.0. In: *Proceedings of the 24th International Conference on Computational Linguistics* (Coling 2012). Mumbai, India.