This study explores the mutual relationship between linguistic theories, data and applications. We focus on one particular theory, Functional Generative Description (FGD), one particular type of linguistic data, namely valency dictionaries and one particular application: machine translation (MT) from English to Czech.

First, we examine methods for automatic extraction of verb valency dictionaries based on corpus data. We propose an automatic metric for estimating how much lexicographers’ labour was saved and evaluate various frame extraction techniques using this metric.

Second, we design and implement an MT system with transfer at various layers of language description, as defined in the framework of FGD. We primarily focus on the lexicographical (deep syntactic) layer.

Third, we leave the framework of FGD and experiment with a rather direct, “phrase-based” MT system. Comparing various setups of the system and specifically treating target-side morphological coherence, we are able to significantly improve MT quality and out-perform a commercial MT system within a pre-defined text domain.

The concluding chapter provides a broader perspective on the utility of lexicons in various applications, highlighting the successful features.

Ondřej Bojar

ISBN 978-80-904175-8-8