

Detailed morphological description of word forms represents one of the most important conditions of a successful automatic processing of linguistic data. The system of categories and their values which are used for the description are the subject of the first part of the thesis.

The basic principle, so-called Golden rule of morphology, states that every word form has to be described by the system unambiguously. The existence of variants of word forms and whole paradigms, however, complicates the accomplishment of this rule. We introduce so called mutations as an extension of the variants to be able to include other sets of word forms with the same description (for instance multiple word forms of Czech personal pronouns). We divide mutations into two parts – global ones describing all word forms of a paradigm, and inflectional ones for the description on the word form level. This division enables us to express their various combinations. We do not use features of style for the mutation division, for they are subjective.

With a consistent use of the categories called Inflectional Mutation and Global Mutation, the Golden rule of morphology will always be valid.

The concept of multiple lemma is introduced in a chapter dealing with lemmatization. It describes lemma variants.

We give a detailed description of so-called compounds, which incorporate word forms of the type *za, pro-, koupilas, koliks*. The concept of multiple lemma is also used for their lemmatization. According to the word class of their components we divide the compounds into several types. We also deal with the problem of their searching in language corpora.

The second part of the thesis describes a system of patterns for word description. It is divided according to the part of speech. Each pattern has a special set of parameters that allow to grasp a large variability in word formation. We also deal with regular derivations of related words using suffixes.