# System of Pronominal Words in Czech with Respect to German and English*

Magda Razímová, razimova@ufal.mff.cuni.cz
Institute of Formal and Applied Linguistics, Charles University
Malostranské nám. 25, 118 00 Prague 1, Czech Republic

***Abstract***

Pronouns and other pronominal words are unproductive, 'closed' classes with (at least to a certain extent) transparent derivational relations not only in Czech, but also in other languages. In the first part of our contribution, we introduce the representation of Czech pronominal words at the underlying syntactic layer (so called tectogrammatical layer, t-layer) of the Prague Dependency Treebank version 2.0,[1] the annotation scenario of which was built on the theoretical basis of the Praguian Functional Generative Description (Sgall et al., 1986). In the second part, we try to apply this representation to English and German pronominals in order to illustrate some of the common (universal) properties of the pronominal systems.

**(1)** At the t-layer, pronominal words are represented by a reduced set of underlying representations (lemmas). The surface forms are generated from the lemma and the values of several attributes. **(a)** All personal pronouns are represented by a single, 'artificial' lemma #PersPron in combination with attributes Person, Number, Gender and Politeness (the last one for discerning between common and polite usage of the 2nd person pronouns). For example, the pronoun *vy* in '*vy* jste přišel' ('*you* came' said politely to a single person) is generated from the representation '#PersPron+2nd(person)+singular+masc.anim.+polite'. **(b)** Also indefinite, negative, interrogative, and relative pronouns can be viewed as surface forms derived from the same underlying representation. Since an artificial lemma like the lemma #PersPron in (a) has not yet been introduced for this type, all these pronouns are represented by the lemma corresponding to the relative pronoun at the t-layer; i.e. for example, the indefinite pronoun *někdo* (somebody) as well as the negative pronoun *nikdo* (nobody) are represented by the t-lemma *kdo* (who). The semantic feature completing the reduced lemma is stored in the attribute indeftype (its value indef corresponds to indefinite pronouns, the value negat to negative pronouns etc.). **(c)** Since pronominal adverbs (e.g. *nějak* (somehow) or *nikde* (nowhere)) express the same semantic features like pronouns in (b) in Czech, they are represented in the same way at the t-layer. Another derivational relation is seen between pronominal adverbs with directional meaning and those of location. E.g., the adverb *odněkud* (from somewhere) is represented by the lemma *kde* (where), the value indef (of the attribute indeftype) and the functor DIR1 (capturing the directional meaning).

**(2)** The presented reduced representation of pronominal words can be applied also to English or German pronominal systems. Such unified representation of pronominal expressions in several languages could be useful for instance for machine translation.

With regard to the Czech system, similar relations in the set of English and German pronouns (e.g. *who–somebody–nobody*, and *wer–jemand–niemand*, respectively)[2] will be shown. The suggested representation also enables to represent equivalent expressions in different languages by the same means. For example, both the English expression *from somewhere* and its Czech one-word counterpart *odněkud* will be represented by the analogous underlying representations (lemma *where/kde* + indef + functor DIR1). In spite of these resemblances, the application of the representation developed for Czech to other languages will not be straightforward and various subtle differences have to be taken into account.

***References:***

Helbig, H. (2001), *Die semantische Struktur natürlicher Sprache*, Berlin/Heidelberg/New York, Springer.
Sgall, P., et al. (1986), *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*, Prague, Academia.

---

* The research reported in this paper was supported by the projects 1ET101120503, GA-UK 352/2005 and GD201/05/H014.

[1] See http://ufal.mff.cuni.cz/pdt2.0/

[2] By means of the attribute indeftype, the German pronoun *nichts* will be represented 'lemma *was*+negat'. A similar treatment of indefinite and negative pronouns as of two subtypes of the same entity was introduced also in the MultiNet knowledge representation system (Helbig, 2001).