Valency Lexicon for Czech: from Verbs to Nouns

Markéta Lopatková, Veronika Řezníčková, and Zdeněk Žabokrtský*

Center for Computational Linguistics, Charles University (MFF),
Malostranské nám. 25, CZ-11800 Prague, Czech Republic
{stranak,rez,zabokrtsky}@ckl.mff.cuni.cz
http://ckl.mff.cuni.cz

Abstract. Valency lexicon of Czech verbs has been intensively worked on for more than a year, and now we have at our disposal a detailed description of valency frames of several hundreds verbs. Presently, the challenge naturally arises, to use the existing lexicon for capturing valency of other word classes. In this paper, we focus on valency of nouns derived from verbs. We propose an algorithm for automatic prediction of valency frames of these nouns, and we test it on a sample of data.

1 Introduction

Valency Lexicon of Czech Verbs. (Only the basic features of our valency lexicon are mentioned here, see [5] for details). The lexicon contains a rich syntactic information based on Functional Generative Description (see [2] for references). Currently, the manual annotation of roughly 350 most frequent Czech verbs has been practically finished, the annotation of next 300 verbs is in progress.

The lexicon has the following structure: *(i)* There is a list of valency frames for each verb. Each verb has at least one valency frame (but usually more, which is typically caused by its polysemy). *(ii)* For each valency frame several attributes are specified: "pointer(s)" to EuroWordNet synset(s), semantic class (verba dicendi, verbs of motion etc.), synonymic expression(s), example sentence(s), lemma of its aspectual counterpart, and the sequence of frame slots are the most important ones. *(iii)* Each frame slot is associated with a functor (name of inner participant or free modifier), the type of relation (obligatory, optional, 'quasi-valency' or 'typical' modifier), and with its possible surface realization(s) (direct or prepositional case, infinitive, subordinating conjunction etc.).

Nominal Valency in Czech. In this paper, we discuss only nouns derived from verbs (non-deverbal nouns and their modifiers are studied e.g. in [7]). While describing valency frames of such nouns, J. Panevová differentiates (according to Kuryłowicz) two types of word-formative process: *syntactic derivation* (SD) and *lexical derivation* (LD). While in SD only syntactic function of derived word changes, in LD the lexical meaning of derived word is changed as well (see [4]). J. Panevová presumes the position of particular types of deverbal nouns on the

* The research reported on in this paper has been carried out under the projects GAČR 405/96/K214 and MŠMT LN00A063.

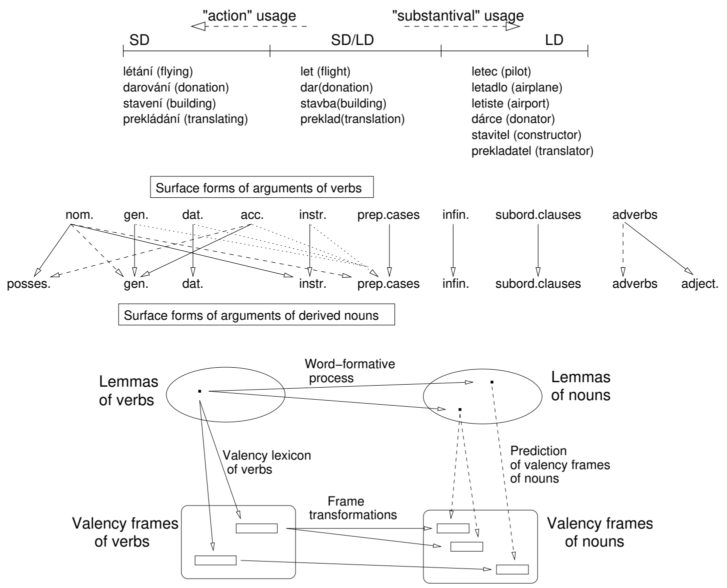
virtual axis: nouns derived from verbs by SD ↔ nouns derived from verbs by LD. While the most of nouns come at the either end of the axis, nouns that would be located "somewhere in the middle" exist as well (cf. [6]).

2 Verbal Frames versus Noun Frames

Morphological Means of Word-formative Process. The word-formative process is accompanied by using some typical suffixes with particular types of derivation: *(1)* Nouns derived by SD are produced especially by productive means, e.g. *-ní/-tí* (*létání* (flying)). *(2)* Nouns on the boundary between SD and LD are derived typically by non-productive means, e.g. with *-ba* (*stavba* (building)) or with zero suffix (*let* (flight)). *(3)* Nouns derived by LD belong for instance to: (i) actor names (derived from verbs by suffixes *-tel*,*-č* etc., e.g. *učitel* (teacher)), (ii) names denoting the place of action (derived by suffixes *-árna*, *-iště* etc., e.g. *umývárna* (washroom)) (iii) names denoting a tool (derived from verbs by suffixes *-dlo*, *-tko* etc., e.g. *ořezávátko* (a pencil sharpener).
Incorporated Role and its Deletion. Many deverbative nouns adopt all the functors of the "source" verb frame. We mark them PRED (they function similarly as the original "predicate"). On the other hand, there are nouns that take over a specific role from one of the participants or free modifiers of the source verb within the word-formative process (see [1]). The meaning of the noun itself occupies one valency slot of the original valency frame. Therefore the respective slot should be deleted when we try to transform the verbal frame into the frame of the noun. See Table 2 for the examples of incorporated participants.
Action vs. Substantival Usage. Not only the type of morphological derivation by specific suffixes is relevant for the belonging to the SD or LD, but especially the type of usage of a particular noun ("action usage" or "substantival usage", see Fig. 1 (a)). Thus the noun *psaní* (writing, a letter), derived from the verb *psát* (to write) can mean the action (*psaní dopisu Petrem trvalo asi 3 hodiny*/Writing a letter by Peter took him about three hours.), but also the result of this action - the letter (*Petrovo psaní leželo na stole*/Peter's letter lay on the table.). The noun *výhra* (a win) derived from the verb *vyhrát* (to win) can be used as an action noun (*Petrova výhra milionu šokovala celou rodinu*/Winning of one million crowns by Peter shocked the whole family.), but also again as a result of the action of winning—the amount of money (*Petrova výhra činila million korun*/Peter's win was one million crowns.). We will preferably study the nouns in the action usage, since they possess a richer valency behavior.
Change of Surface Realizations. The process of nominalization of a verb is accompanied with a change of surface realization of particular members of the valency frame (cf. [3]). The possible transitions are depicted in Fig. 1(b).

3 Prediction of Noun Frames

Algorithm for Frame Generating. We propose an algorithm that automatically generates valency frames of deverbative nouns. The algorithm consists of



(a)

(b)

(c)

Fig. 1. (a) Distribution of nouns derived from verbs along the "derivational" axis. (b) Possible changes of surface realization (full arrows – the most typical transition, dashed arrows – lower frequency transitions, dotted arrows – really rare transitions). (c) The relation between verbs and nouns from the viewpoint of the valency predicting algorithm.

Role	#Occ.	Example noun and its frame
PRED	134	*hovor* ACT(2/pos) PAT(o+6) ADDR(s+7) (conversation with sb...)
ACT	61	*čtenář* PAT(2) (reader of st.)
PAT	9	*přínos* ACT(2/pos) BEN(pro+4) (sb's. contribution to st.)
EFF	6	*hra* ACT(2/pos) PAT(o+6) (sb's. play about st.)

Table 1. Number of occurrences of selected incorporated roles in the handcrafted sample of nouns derived from 240 verbs (nouns with ending *-ní, -tí* are not counted in this table).

the following steps: *(1)* find the verb from which the noun is derived, *(2)* select the appropriate frame of the verb, *(3)* determine the incorporated role of the noun with respect to the verb, and delete the respective slot from the verbal frame (if necessary), *(4)* transform the surface realization of the remaining slots. (The situation is illustrated in Fig. 1 (c): we emulate the valency lexicon of nouns (the dashed line) via traversing the picture in the anti-clockwise fashion.)

Several problems appeared during the implementation of the algorithm. First, it is not always clear, from which verb the noun was derived. Second, it is difficult to choose the appropriate frame. For instance, noun *odpověď* is related to one frame of verb *odpovídat* (to answer), while not to the second (to be responsible) nor to the third (to correspond). Third, the incorporated role can be only approximated using the primary meaning of suffixes (see paragraph Morphological means...), but exceptions exist. As for the change of the surface realizations, we use (slightly modified) transformations of frames from [6] as well as our own rules that capture some more regularities in surface realization change we are aware of (however, this set of rules is still being developed). The resulting frame transformation (verb→noun) may look like follows:

ACT(nom) ADDR(dat) PAT(acc) → ACT(instr,adjpos,od+gen) PAT(gen) ADDR(dat)

e.g., *dodat* (to deliver st to sb) → *dodání* ((sb's) delivery (from sb, by sb) of st to sb)

Experiment and Evaluation. We tested a part of the algorithm above on a small sample of data. We took 209 nouns with ending *-ní* or *-tí* (incorporated role PRED) and we generated their valency frames using the valency lexicon of verbs (in step (2) of the algorithm, we used all frames of the respective verb besides idiomatic frames). The algorithm was applicable on 197 nouns. Altogether, 567 frames have been generated. 170 frames obviously contained an error, the remaining ones (70 %) are acceptable.

4 Conclusion and Future Work

We have shown that it is possible to generate reasonable valency frames of deverbative nouns using valency lexicon of verbs. The error analysis of the results will hopefully lead us to improvements of the algorithm. Also some further linguistic observations have to be done in order to enhance the quality of the resulting frames. Simultaneously, we are also trying to automatically acquire valency frames of nouns from the Prague Dependency Treebank ([2]). In any case, in order to obtain a high-quality valency lexicon of nouns, certain amount of manual annotation seems to be unavoidable.

References

1. Dokulil, M.: Tvořeníslov v češtině I. Teorie odvozováníslov. (Word-formative process in Czech I. Theory of word-formation.) ČSAV, Praha (1962)
2. Hajičová E. et al: The Current Status of the Prague Dependency Treebank. Proceeings of the 4th Internation Conference Text, Speech and Dialogue, Masaryk University, Brno (2001)
3. Karlík, P., Nübler, N.: Poznámky k nominalizaci v češtině (Notes on nominalization in Czech). SaS, 59, 105-112. (1998)
4. Kuryłowicz, J.: Dérivation lexicale et dérivation syntaxique. Bulletin de la Société linguistique de Paris, 37, 79-92 (1936).
5. Lopatková M., Žabokrtský Z.: Valency Dictionary of Czech Verbs: Complex Tectogrammatical Annotation. Proceedings of Third International Conference on Language Resources and Evaluation, Las Palmas, Spain (2002)
6. Panevová, J: Poznámky k valenci podstatých jmen (Notes on the valency of nouns). In: Hladká, Z.,Karlík, P. (Eds.): Čeština - univerzália a specifika, Brno: MU, 173-180 (2000).
7. Piťha, P.: On the case frames of nouns. In: Prague Studies in Mathematical Linguistics, Vol. 7, Academia, Praha, 215-224 (1981).