

Using Valency Inheritance in Building a Valency Lexicon

Václava Kettnerová, Veronika Kolářová, Jiří Mírovský, Michal Olbrich

Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics
Malostranské náměstí 25, 118 00 Prague, Czech Republic
{kettnerova, kolarova, mirovsky, olbrich}@ufal.mff.cuni.cz

Abstract

Derived words often share certain characteristics with their base words, which leads to the idea that identical properties are inherited from the base words. These properties also cover valency. Valency inheritance has not been used to automatically build lexical resources providing information on valency, the manual annotation of which requires significant human effort. In this paper, we propose a procedure for generating valency frames of selected semantic categories of Czech nouns and adjectives exhibiting a significant level of valency inheritance, thus covering the productive and systemic core of the lexicon. Based on a semiautomatic comparison of the noun and adjectival valency frames from *NomVallex* and the verbal valency frames from *VALLEX*, rules describing valency changes in the valency frames of noun and adjectival derivatives are formulated. The conditions imposed by the rules on valency frames identify individual base lemmas in these lexicons for which direct noun and adjectival derivatives are searched in *DeriNet*. Based on the changes in valency determined in the rules, more than 23,000 valency frames assigned to more than 10,000 noun and adjectival derivatives were derived, achieving high accuracy. These valency frames were included in *DeriVallex*, a database providing a solid basis for extending current lexical resources.

1. Introduction

Valency (argument structure) can be defined as “the capacity a verb (or noun, etc.) has for combining with particular patterns of other sentence constituents” (Allerton, 2006, pp. 301). As a lexically determined property, relating to individual senses of words, it is often assumed not to be automatically predictable, and as such it has to be accounted for in a lexicon (e.g., Herbst et al., 2004). However, it has been observed that certain derivationally related words exhibit systemic correspondences in their valency. The observed correspondences support the idea that derived words inherit valency properties from their base words (see, esp., Booij, 2000; Bierwisch, 2015). Valency inheritance has become one of the key issues in studies concerning the interaction between morphology and syntax (see, e.g., Spencer, 2005, and a summary of this research introduced by Härtl, 2015). In these studies, the most prominent position is still occupied by deverbal nouns and partly adjectives, as correspondences between their valency and that of their base verbs are strikingly apparent (see, esp., Chomsky, 1970; Grimshaw, 1990; Alexiadou, 2007, for Czech, see Kolářová et al., 2019, and Kolářová et al., 2021). In contrast, other derivational categories, i.e., deadjectival and denominal derivatives, have remained of marginal interest (for deadjectival nouns, see, e.g., Roy, 2010; Borer, 2013; for Czech, see Kolářová, 2024). Moreover, valency inheritance has not been – to our best knowledge – exploited in building lexical resources providing information on valency, the manual annotation of which, though ensuring high-quality data, is time-consuming and demanding in terms of human re-

sources.¹

Focusing on Czech valent nouns and adjectives, our goal in this paper is thus two-fold: (i) to determine the extent of correspondence between base words and their direct noun and adjectival derivatives and (ii) to propose and implement a method enabling the generation of valency frames of noun and adjectival derivatives based on the valency frames of their base words. In this task, we adopt the following procedure. First, we semiautomatically compare the valency frames of noun and adjectival derivatives contained in *NomVallex*,² a valency lexicon of Czech nouns and adjectives (Kolářová et al., 2024), with the valency frames of their respective base words, contained either in *VALLEX*,³ a valency lexicon of Czech verbs (Lopatková et al., 2022), or in *NomVallex*. If changes in the valency frames of these derivatives are sufficiently systemic, rules underlying these changes are formulated and applied in an automatic generation of the valency frames of noun and adjectival derivatives found in *DeriNet*, a lexical network which models word-formation relations in the lexicon of Czech (Olbrich et al., 2025). The aim is to automatically create a valency lexicon of Czech derived nouns and adjectives that exhibit systemic valency correspondences with their base words, *DeriVallex*, thus laying the foundation for further extension of current lexical resources. Such

¹The first attempt for Czech, focused on action nouns (see Section 3.2.1), did not contribute to an automatic extension of a lexicon (see Lopatková et al., 2002).

²<https://ufal.mff.cuni.cz/nomvallex/2.5/>

³<https://ufal.mff.cuni.cz/vallex/4.5/>

a lexicon, by providing structured knowledge, can support linguistically informed NLP tasks, such as dependency parsing, semantic role labeling, and information extraction. As the proposed research is not limited to deverbal derivatives only but includes other derivational categories, i.e., deadjectival and denominal, it can – beyond its lexicographic potential – contribute to the theoretical question of how much systemic the valency correspondences between derivationally related words are.

The paper is structured as follows. First, the *NomVallex* lexicon is described (Section 2). Second, a semiautomatic comparison of valency frames identifying the systemic valency behavior of noun and adjectival derivatives is introduced (Section 3). Third, the rules describing systemic changes in their valency frames are proposed and applied in the generation of the valency frames of noun and adjectival derivatives obtained from *DeriNet* (Section 4).

2. NomVallex

Like verbs, some nouns and adjectives fulfil the role of predicates and thus are likely to be valent. Compared to the valency of verbs, the valency of nouns and adjectives has remained peripheral to linguistic inquiry and its description in valency lexicons is still rather limited; for English, see, e.g., (Herbst et al., 2004), for Polish (Alberski et al., 2018), and for Czech (Svozilová et al., 2005) and (Urešová et al., 2024). Unlike the aforementioned dictionaries, the Czech valency lexicon *NomVallex* not only specifies the valency of nouns and adjectives, but also captures derivational relations to their base words (Kolářová et al., 2024).

NomVallex is manually annotated. It adopts and further modifies the annotation scheme of the valency lexicon of Czech verbs *VALLEX* (Lopatková et al., 2022). The lexicon entry contains a lexeme, an abstract unit associating lexical forms with their lexical units, i.e., word senses. Each lexeme is represented by one or more lemmas (e.g., in the case of aspectual counterparts, such as *přesvědčování_{impf}* – *přesvědčení_{pf}* ‘persuading’, and lemma or variants, e.g., *dotyk* / *dotek* ‘touch’). *NomVallex*, version 2.5, comprises 1,337 lexical units contained in 730 lexemes. The research presented here is based on the working version of *NomVallex* (*NomVallex* 2.6),⁴ comprising 1,497 lexical units in 812 lexemes.

As its theoretical background, *NomVallex* uses the valency theory of the Functional Generative Description (FGD) (see esp. Panevová, 1994). In FGD, actants (arguments) and free modifications

(adjuncts) are distinguished. In verbal constructions, actants are expressed on the surface as the subject or (in)direct objects, while free modifications correspond to adverbials. When determining actants and free modifications in noun and adjectival constructions, analogies with verbal constructions are taken into consideration. Five types of actants are determined: ACTor, PATient, ADDRessee, EFFect, and ORIGin (e.g., *Petrův_{ACT} smysl pro humor_{PAT}* ‘Peter’s_{ACT} sense of humor_{PAT}’, *výrobek prodejný mládeži_{ADDR}* ‘product marketable to young people_{ADDR}’, *opatření cigaret_{PAT} falešnými kolkyně_{EFF}* ‘stamping cigarettes_{PAT} with counterfeit tax stamps_{EFF}’, *vymahatelnost pokut_{PAT} od lidí_{ORIG}* ‘enforceability of fines_{PAT} from individuals_{ORIG}’). In contrast, free modifications are numerous and they are semantically grounded (e.g., *zaměstnancova přítomnost na pracovišti_{LOC}* ‘employee’s presence at the workplace_{LOC}’, *muž povoláný do armády_{DIR3}* ‘a man drafted into the army_{DIR3}’). Both actants and free modifications can be either optional or obligatory. Only actants (regardless of their obligatoriness) and obligatory free modifications belong to the valency frame of a lexical unit, thus representing its complements.

Valency frames are modeled as a sequence of valency slots, each standing for one valency complement. Each valency slot consists of a type of the valency complement and its obligatoriness (describing the deep valency), and it is supplemented with a list of morphosyntactic forms, i.e., complement morphosyntactic markers (indicating the surface valency).⁵

Further, each lexical unit of noun and adjectival derivatives is assigned the information on its derivational category and derivational type. A derivational category represents a set of derivatives that share the same part of speech and are derived from words belonging to one specific part of speech (for individual derivational categories see Table 1; for the sake of completeness, also unmotivated (primary) nouns and adjectives and compounds are introduced here). Within derivational categories, derivational types are formally distinguished in the attribute *type*, based on individual means (esp. affixes) used in derivation (in total, 18 types for nouns, e.g., deverbal root nouns such as *hroz-ba* ‘threat’, *žád-ost* ‘request’, *námít-ka* ‘objection’, *úsud-ek* ‘judgment’, and 13 types for adjectives, e.g., adjectives ending in *-telný* ‘able’). In addition, lexical units are semantically classified in the attribute *semcategory*, according to the meaning expressed (e.g., the noun *žádost* as an abstract result of an action in the meaning ‘request’ and as a material

⁴https://quest.ms.mff.cuni.cz/vallex/?version=data_release-NomVallex2.6

⁵Valency complements can be marked by prepositionless cases, prepositional cases, the possessive form of adjectives and pronouns, the infinitive, and complementizers introducing dependent clauses.

Derivational category	Lexical units/ Lexical units with links	Lexemes/ Lexemes with links
N-DV	699 / 624	334 / 297
N-DA	347 / 341	220 / 218
N-DN	53 / 46	30 / 26
N-P	11 / 0	3 / 0
A-DV	246 / 213	152 / 129
A-DA	12 / 12	10 / 10
A-DN	60 / 37	31 / 21
A-P	63 / 0	26 / 0
C	6 / 6	6 / 6
Total	1,497 / 1,279	812 / 707

Table 1: Statistics on lexical units of noun and adjectival derivatives, and lexemes containing these units in *NomVallex* grouped according to their derivational category, i.e., deverbal nouns (N-DV), deadjectival nouns (N-DA), denominal nouns (N-DN), primary nouns (N-P), deverbal adjectives (A-DV), deadjectival adjectives (A-DA), denominal adjectives (A-DN), primary adjectives (A-P), and compounds (C).

object in the meaning ‘application’).

Moreover, each lexical unit of a noun or adjectival derivative is linked with a lexical unit of its respective base word, if it is available either in *NomVallex* (in the case of base nouns and adjectives) or in *VALLEX* (in the case of base verbs); for the statistics see Table 1.⁶ The valency frame of each lexical unit with the link is automatically compared with the respective frame of its base word and the differences (if any) are specified in the *valdiff* attribute.

3. Semiautomatic Analysis of the *NomVallex* Data

In this section, we analyze derivational categories of Czech derived nouns and adjectives (Section 3.1), and within these categories their individual semantic categories, roughly corresponding to word-formation categories (see *Dokulil, 1962*), that typically exhibit systemic correspondences in valency compared to their base words, and their valency can thus be considered inherited (Section 3.2). In order to identify the systemic correspondences, the information provided in *NomVallex* by the *valdiff* at-

⁶There are several reasons why the link is missing for some lexical units, e.g., it is a primary noun or adjective (e.g., *síla* ‘strength’), the base verb is not frequent enough to be captured in the *VALLEX* lexicon where the verbs were selected based on their corpus frequency (e.g., *dychtit* ‘crave’ > *dychtivý* ‘avid’), or the base noun is a valent, so it is not treated in the *NomVallex* lexicon (e.g., *skála* ‘rock, cliff’ > *úskalí* ‘difficulty’).

tribute has been exploited (only noun and adjectival derivatives with a link to their base word were used, excluding compounds).

3.1. Systemic Correspondences in Valency across Derivational Categories

When comparing the valency frames of noun and adjectival derivatives with the valency frames of their base words, systemic correspondences can be observed in the deep valency as well as in the surface valency. In the deep valency, a derived word preserves the number and type of valency complements of its base word. E.g., the noun *darování* ‘giving’ is characterized by three obligatory valency complements, ACTor, ADDRessee, and PATient, as its base verb *darovat* ‘to give’ (cf., *Hospodinovo_{ACT} darování tóry_{PAT} Mojžíšovi_{ADDR}* ‘God’s_{ACT} giving of the Torah_{PAT} to Moses_{ADDR}’ with *Hospodin_{ACT} daroval tóru_{PAT} Mojžíšovi_{ADDR}* ‘God_{ACT} gave the Torah_{PAT} to Moses_{ADDR}’).

The surface valency is indicated by morphosyntactic markers of valency complements (Section 2). Two types of systemic correspondences can be detected in these markers: they either remain the same or undergo such changes that are conditioned by the change of part-of-speech membership of derived words. The following ones are attested in the *NomVallex* data in individual derivational categories:⁷

(i) Deverbal nouns

- Nom > Gen, Ins, Poss, *od* ‘from’+Gen
- Acc and Gen > Gen, Poss

e.g., *policisté_{Nom} kontrolují auta_{Acc}* ‘policemen are checking vehicles’ > *kontrolování aut_{Gen} policisty_{Ins} / od policistů_{od+Gen}* ‘checking of vehicles by policemen’, *jejich_{Poss} kontrolování policisty_{Ins} / od policistů_{od+Gen}* ‘their checking by policemen’

(ii) Deverbal adjectives

- Systemic surface ellipsis of one of the complements of the base verb (the corresponding adjectival complement is present in the deep valency of the adjective, but cannot be expressed in the surface structure as an adjectival modifier; the antecedent of the elided complement is expressed on the surface either as the governor of the adjective or as the subject of the copular construction)⁸

⁷The following abbreviations and numbers stand for particular cases: the nominative – Nom/1, the genitive – Gen/2, the dative – Dat/3, the accusative – Acc/4, the instrumental – Ins/7, the adjectival or pronominal possessive modifier – Poss, the infinitive – Inf.

⁸The antecedent to which the elided complement refers is marked in the examples by an upward arrow.

e.g., *jedinec*_{Nom} *se odlišuje od ostatních*_{od+Gen} ‘an individual differs from others’ > *jedinec*_↑ *odlišný od ostatních*_{od+Gen} ‘an individual different from others’

- Nom (if not elided) > Ins, *od* ‘from’+Gen⁹

e.g., *hacker*_{Nom} *vydírá podnikatele*_{Acc} ‘a hacker is blackmailing an entrepreneur’ > *podnikatel*_↑ *vydíráný hackerem*_{Ins} / *od hackera*_{od+Gen} ‘an entrepreneur blackmailed by a hacker / from a hacker’

(iii) Deadjectival nouns

- Reactivation of the systemic surface ellipsis (the systematically elided complement of the base adjective is expressed as an adnominal modifier)

e.g., *odlišný jedinec*_↑ ‘a distinct individual’ > *odlišnost jedince*_{Gen} ‘the distinctiveness of an individual’

(iv) Denominal adjectives

- Systemic surface ellipsis of one of the complements of the base noun

e.g., *odvaha klientů*_{Gen} *investovat*_{Inf} ‘courage of clients to invest’ > *klienti*_↑ *odvážní investovat*_{Inf} ‘clients courageous to invest’

(v) Denominal nouns (systemic correspondences in valency attested only in modifying types, e.g., in diminutives)

- the same part of speech and thus no change

e.g., *Petrova*_{Poss} *slabost pro opuštěná zvířata*_{pro+Acc} ‘Peter’s weakness for abandoned animals’ > *Petrova*_{Poss} *slabůstka pro opuštěná zvířata*_{pro+Acc} ‘Peter’s (little) weakness for abandoned animals’

(vi) Deadjectival adjectives (systemic correspondences in valency attested only in modifying types, e.g., in direct negation)

- the same part of speech and thus no change

e.g., *ochotný pomoci*_{Inf} ‘willing to help’ > *neochotný pomoci*_{Inf} ‘not willing to help’

Table 2 introduces the statistics on systemic correspondences (i.e., the number and type of complements, including their morphosyntactic markers, are preserved, or change regularly) and on non-systemic differences between valency frames of noun and adjectival derivatives in *NomVallex* and valency frames of their base words contained either in *VALLEX* or in *NomVallex*. It follows from the statistics that derivational categories are permeated with non-systemic changes. The highest degree of systemic correspondences is exhibited by

For the systemic surface ellipsis, see further (Kolářová, 2024).

⁹The status of Acc in deverbal adjectives requires further research.

Type	Systemic	Non-systemic		Total
		Deep valency	Surface valency	
N-DV	356 / 57%	102 / 16%	166 / 27%	624
N-DA	211 / 62%	73 / 21%	57 / 17%	341
N-DN	16 / 35%	16 / 35%	14 / 30%	46
A-DV	107 / 50%	47 / 22%	59 / 28%	213
A-DA	0 / 0%	11 / 92%	1 / 8%	12
A-DN	4 / 11%	19 / 51%	14 / 38%	37
Total	694 / 55%	268 / 21%	311 / 24%	1,273

Table 2: Statistics on the number of the valency frames of noun and adjectival derivational categories in *NomVallex* displaying systemic correspondences (col. Systemic) and non-systemic differences in the number and type of complements (col. Non-systemic, Deep valency), and in their morphosyntactic marking (col. Non-systemic, Surface valency).

deadjectival nouns (62%), deverbal nouns (57%), and deverbal adjectives (50%). As the valency behavior of a word is to a great extent determined by its semantic properties (Levin, 1993), this finding gives evidence that derivation brings about semantic shifts that are not predictable based on the word’s derivational category membership.

3.2. Systemic Correspondences in Valency across Semantic Categories

It thus seems more viable to identify within individual derivational categories those semantic categories of derivatives whose members exhibit a semantic change that is stable across the category to the degree that it manifests itself in systemic valency changes. For this purpose, we make use of the information provided in the attribute *semcategory* (Section 2).

3.2.1. Noun Derivatives

Noun derivatives in *NomVallex* are classified into action (e.g., *přesvědčování* ‘persuading’), abstract result of an action (e.g., *beznaděj* ‘despair’), quality (e.g., *nevinost* ‘innocence’), quantity (e.g., *počet* ‘number’), and material object (e.g., *pozornost* ‘gift’).

The highest degree of systemic correspondences (87%) is displayed by the noun derivatives denoting action (act in Table 3), falling within deverbal nouns, which supports the findings which are cross-linguistically attested.

Mismatches in the valency of action nouns follow from the fact that *NomVallex* captures only those valency complements and their morphosyntactic markers that are corpus-attested; unattested cases that, according to the annotator’s linguistic intuition,

Type	Systemic	Non-systemic		Total
		Deep valency	Surface valency	
act	289 / 87%	12 / 3%	33 / 10%	334
res	259 / 45%	122 / 21%	201 / 34%	582
qual	30 / 65%	16 / 35%	0 / 0%	46
quant	1 / 20%	4 / 80%	0 / 0%	5
obj	4 / 9%	37 / 84%	3 / 7%	44
Total	583 / 58%	191 / 19%	237 / 23%	1,011

Table 3: Statistics on systemic correspondences and non-systemic differences in the valency frames of noun derivatives, distinguished based on their semantics as captured in *NomVallex* in the attribute *semcategory*.

are unlikely to be used are left aside. For example, the valency complement ORIGIN, captured in the valency frame of the verb *dočkat se* (e.g., *dočkat se od dcery_{ORIG} nevděku* ‘to receive ingratitude from one’s daughter_{ORIG}’), is not part of the valency frame of the noun *dočkání se* ‘receiving’ (e.g., *dočkání se ?(od dcery) nevděku* ‘receiving ingratitude from one’s daughter’) as it is not attested in the corpus and the construction sounds odd. Other discrepancies are due to differences in the annotation approach to predicates expressing psychological and mental processes adopted in *NomVallex* and *VALLEX*.

The second highest agreement is attested by quality nouns (qual, 65%), pertaining to all derivational categories. If we limit their derivational category only to deadjectival nouns, the extent of systemic correspondences in their valency frames increases to 77%.

In contrast, the valency frames of noun derivatives denoting abstract results of action (res), quantity (quant) and material objects (obj) show high disagreement with the valency frames of their base words. However, the quantity type is insufficiently represented in the data to draw any conclusions about their systemic behavior.¹⁰

3.2.2. Adjectival Derivatives

Adjectival derivatives are categorized into simultaneous action (e.g., *rozhodující* ‘deciding’), anterior action (e.g., *dotknuvší se* ‘having touched’), passive action (e.g., *obdivovaný* ‘admired’), potentiality (e.g., *ovladatelný* ‘controllable’), event property

¹⁰The valency frames of agent nouns (e.g., *zaměstnavatel* ‘employer’) and patient nouns (e.g., *zaměstnanec* ‘employee’), both representing deverbal nouns, may be hypothesized to exhibit systemic correspondences as well. However, as *NomVallex* primarily focuses on nouns with abstract meanings, these semantic categories of nouns do not yet occur in the data.

(e.g., *podezřívavý* ‘suspicious’), purpose (e.g., *zasunovací* ‘sliding’), qualitative property (e.g., *bez-mocný* ‘powerless’), relational property (e.g., *přednostní* ‘priority’), and material object (e.g., *přeživší* ‘survivor’). Adjectives denoting simultaneous action, anterior action, passive action, potentiality, and purpose represent deverbal adjectives. In contrast, adjectives expressing qualitative property are denominal or deadjectival, and adjectives expressing relational property only represent denominal adjectives. Finally, adjectives denoting material objects may fall within any of the deverbal, deadjectival or denominal categories.

Table 4 summarizes the degree of correspondences between the valency frames of adjectival derivatives, grouped into the individual semantic categories, and the valency frames of their base words. A complete correspondence is attested by adjectives denoting simultaneous action (sim). In the case of passive action (pass), the correspondence exceeds three quarters, amounting to 81%. A lesser degree of correspondences is attested by adjectives expressing anterior action (anter, 67%) and potentiality (pot, 64%).

Manual checking revealed that a substantial portion of the differences in the valency frames of these adjectives can be attributed to different approaches to individual cases adopted in the annotation in *NomVallex* and *VALLEX*. Moreover, it showed that some discrepancies are motivated by semantic class membership of a predicate, e.g., adjectives of the semantic category potentiality falling within psychological or mental predicates tend to have their ACTor expressed by the prepositional group *pro* ‘for’+Acc, which was assessed as non-systemic behavior (see Section 3.1).

In contrast, the agreement in the valency frames of adjectives denoting event property (eprop) and qualitative property (qprop) compared to the valency frames of their base words is very low (20% and 10%, respectively). Finally, the valency frames of adjectives expressing relational property (rprop) and those denoting material objects (obj) display total disagreement.

Adjectives expressing purpose (purp) cannot be evaluated, as they are represented by only a single adjective in the data.

4. Generation of Valency Frames

Based on the statistics provided in Section 3, we selected the following semantic categories of nouns as those the valency frames of which will be generated by rules: action deverbal nouns and quality deadjectival nouns. From adjectival derivatives, we chose those categories that denote simultaneous action, anterior action, passive action, and poten-

Type	Systemic	Non-systemic		Total
		Deep valency	Surface valency	
sim	8 / 100%	0 / 0%	0 / 0%	8
anter	14 / 67%	3 / 14%	4 / 19%	21
pass	38 / 81%	1 / 2%	8 / 17%	47
pot	30 / 64%	2 / 4%	15 / 32%	47
eprop	17 / 20%	34 / 41%	32 / 39%	83
purp	0 / 0%	1 / 100%	0 / 0%	1
qprop	4 / 10%	24 / 57%	14 / 33%	42
rprop	0 / 0%	4 / 80%	1 / 20%	5
obj	0 / 0%	8 / 100%	0 / 0%	8
Total	111 / 42%	77 / 30%	74 / 28%	262

Table 4: Statistics on systemic correspondences and non-systemic differences in the valency frames of adjectival derivatives, distinguished based on their semantics as captured in *NomVallex* in the attribute *semcategory*.

tiality, all representing deverbal adjectives.¹¹

First, based on the semiautomatic analysis described in Section 3, systemic correspondences in valency were determined for the individual selected types of noun and adjectival derivatives. Based on these correspondences, rules describing changes in their valency frames were formulated (Section 4.1). Second, the rules include the conditions under which they apply to the valency frames of base words. Based on these conditions, the valency frames of base words were identified either in *VALLEX* (in the case of base verbs) or in *NomVallex* (in the case of base adjectives) (Section 4.2); valency frames of idioms and complex predicates with light verbs were excluded from the data. For each of the lemmas that represented the valency frame meeting the conditions, a direct noun or adjectival derivative(s) was searched in *DeriNet*. Third, if found there, its valency frame(s) was generated based on the changes determined in the rules (Section 4.3).

4.1. Rules Describing Valency Changes

For each selected semantic category of noun and adjectival derivatives, the rule(s) determining changes in their valency frames was formulated. Each rule has two parts. The condition part specifies the conditions under which the rule is applied to the valency frame of the base word. The action part changes the valency frame of the base word into the valency frame of the respective derivative.

¹¹As a pilot study, adjectives expressing potentiality have been addressed in the experiment described by Kettnerová et al. (2025). For comparison with other noun and adjectival semantic categories, we also report the results for these adjectives here.

Passive action adjectives		Pass . a . 1
condition	id-v	$X_1[ACT] Y_{4,2}[PAT ADDR EFF]$
X	*	$\rightarrow 7, ?od+2$
Y	*	$\rightarrow \uparrow$
Z	<i>jako</i> +4	$\rightarrow jako+1$

Table 5: The rule describing changes in the valency frames of passive action adjectives whose base verbs have prototypical valency frames. Arabic numerals stand for the respective cases.

Let us illustrate the principles applied in the rules with one of two rules governing changes in valency frames of passive action adjectives (e.g., *dosazený* 'appointed'). The rule *Pass . a . 1*, presented in Table 5, is applied to the valency frames of verbs that are prototypical, i.e., in these frames, ACTor has the form of the nominative (and is thus expressed as the subject).

The conditions prescribed in the rule state that the rule is applied to each lexical unit of a verb (indicated by the id-v) that has ACTor (represented by the variable X) in the nominative (regardless of its other possible forms) and PATient, ADDResssee, or EFFect (the variable Y) expressed either in the prepositionless accusative or in the prepositionless genitive (regardless of its other forms).

The action part of the rule changes the valency frame of the lexical unit of a verb satisfying the conditions as follows. First, all morphosyntactic markers of the valency complement represented by the variable X (the sign *) are substituted by the instrumental or by the prepositional group *od* 'from'+2 (the sign ? indicates that this form serves as a stylistic variant for the instrumental).¹² Second, the rule determines that the valency complement expressed by the accusative or the genitive (the variable Y) is subject to the systemic surface ellipsis with adjectives (the sign \uparrow replaces all markers of the complement, which is indicated by the sign *).¹³ Third, if the valency frame of the lexical unit contains a predicative complement expressed by a phrase *jako* 'as' (the variable Z), its form changes in the nominative. Other valency complements, including their morphosyntactic markers and obligatoriness, remain preserved.

See the example of the application of rule *Pass . a . 1* to the verb *dosadit* 'to appoint':

- *dosadit* 'to appoint'
 $ACT_1^{obl} PAT_4^{obl} EFF_{jako+4, za+4}^{obl}$
*soud*_{ACT-Nom} *dosadil* *advokátku*_{PAT-Acc} *jako* *vy-*
rovnávací *správkyni*_{EFF-Acc}

¹²This variant is largely applicable to passive action adjectives, but not without exceptions.

¹³For the surface ellipsis, see Section 3.1.

‘the court_{ACT} appointed the lawyer_{PAT} as an insolvency trustee_{EFF}’

- *dosazený* ‘appointed’
ACT_{7,?od+2}^{obl} PAT_↑^{obl} EFF_{jako+1,za+4}^{obl}
*advokátka*_↑ *dosazená soudem*_{ACT-Ins} *jako vy-*
*rovnávací správkyně*_{EFF-Nom}
‘a lawyer_↑ appointed by the court_{ACT} as an insolvency trustee_{EFF}’

In total, 11 rules were formulated for the selected types of derivatives. The largest number of rules was required by action nouns (3 in total), as those action nouns that are derived from verbs denoting mental and psychological processes display specific valency changes that require special rules.

4.2. Identification of Derived Lemmas

Based on the conditions imposed by each of the rules (Section 4.1), the valency frames of verbs and the valency frames of adjectives satisfying the conditions were extracted from *VALLEX* and *Nom-Vallex*, respectively. For each of the extracted valency frames, the lemma(s) representing the lexical unit to which the frame is assigned was retrieved as well. For each of these lemmas, its direct derivative(s) was identified in *DeriNet*, if available, based on the derivational means (most often a suffix) characteristic of the given semantic category.

This task, however, poses a challenge as there is not one-to-one correspondence between semantic categories and derivational means. One and the same semantic category can be realized by more than one derivational means (see, e.g., the adjectives denoting potentiality *přemísti-telný* ‘movable’, *rozpust-ný* ‘soluble’ and *jed-lý* ‘eatable’), and conversely, a single derivational means can contribute to more than one semantic category, i.e., it can be ambiguous (see, e.g., *chráně-ný* ‘protected’ as a passive action adjective, *rozpust-ný* ‘soluble’ as a potentiality adjective, *krm-ný* ‘fodder (beet)’ as a purpose adjective, and *lad-ný* ‘graceful’ as an event property adjective).

Nevertheless, each semantic category is typically characterized by a certain derivational means which is central to that type (supplemented with other, usually peripheral means). In order to eliminate the effect of ambiguity of derivational means as much as possible, we thus relied solely on the central derivational means when extracting the lemmas from *DeriNet*. The following derivational means were identified for the individual semantic categories: *-ní* and *-tí* for action nouns, *-ost* for quality nouns, *-ící* and *-oucí* for simultaneous action adjectives, *-lý* and *-(v)ší* for anterior action adjectives, *-ný* and *-tý* for passive action adjectives, and *-telný* for adjectives expressing potentiality. Table 6 provides the statistics on the number of the lemmas of

noun and adjectival derivatives extracted from *DeriNet* and the number of lexical units represented by these lemmas. The obtained lemmas were manually checked and non-relevant lemmas and those that were not attested in Czech corpora¹⁴ were excluded from the dataset.¹⁵

The highest number of the retrieved noun lemmas falls within action nouns. For adjectives, the most amply represented adjectival lemmas rank among passive action adjectives. The lowest number of derivatives was found for quality nouns due to the limited number of adjectives covered in *Nom-Vallex*.

The number of lexical units is higher than that of the lemmas, as a single lemma may represent more than one lexical unit, corresponding to individual senses of the derived noun or adjective represented by the lemma (see Section 2). On average, a single derived lemma represented more than 2 lexical units (exactly 2.3 lexical units, when only relevant cases are taken into account).¹⁶

The low proportion of non-relevant lemmas suggests that the identification of the individual semantic categories based on their central derivational means was well-founded. One of the sources of non-relevant lemmas was the above-mentioned ambiguity of derivational means. E.g., the purpose adjective *zápalný* ‘flammable’ was identified as a passive action adjective, and the noun *schopnost* ‘ability’, expressing an abstract result of an action, was recognized as a quality noun. The other source originated from incorrect lemmas contained in *DeriNet*, especially, those with a low number of corpus occurrences often resulted from incorrect lemmatization of misspelled words.

4.3. Generated Valency Frames

As valency frames are assigned to individual lexical units (Section 4.2.), the number of the generated valency frames corresponds to the number of the identified lexical units, see the column ‘Relevant Lexical units’ in Table 6. As a result, the highest number of valency frames was generated for action

¹⁴The information on the frequency of lemmas in the Czech National Corpus is available directly in *DeriNet*.

¹⁵For those action nouns, simultaneous action adjectives, and anterior action adjectives ending in *-(v)ší* that are derived from reflexive verbs, the reflexive was added to the derived lemma as it is appropriate in these cases but not covered in *DeriNet*.

¹⁶*VALLEX* 4.5 contains 4,664 verb lemmas (without iterative verbs), representing 11,083 lexical units. For nearly three quarters of lexical units in *VALLEX*, an action noun has been found, and for nearly one half of lexical units, a passive action adjective was obtained. *Nom-Vallex* covers 244 adjectival lemmas corresponding to 425 lexical units. More than one third of adjectival lexical units have been assigned a quality noun.

Type	Relevant		Non-relevant		Unattested		Total	
	Lemmas	Lexical units	Lemmas	Lexical units	Lemmas	Lexical units	Lemmas	Lexical units
act	3,020	8,010	45	193	62	100	3,127	8,303
qual	114	147	10	45	7	13	131	205
sim	2,305	4,564	18	80	131	281	2,454	4,925
anter	1,144	2,280	41	116	1,074	2,014	2,259	4,410
pass	2,439	5,152	268	676	0	0	2,707	5,828
pot	1,198	2,910	2	10	1,024	2,326	2,224	5,246
Total	10,220	23,063	384	1,120	2,298	4,734	12,902	28,917

Table 6: Statistics on the number of the lemmas of the noun and adjectival derivatives and the lexical units represented by these lemmas, obtained from *DeriNet*. The lemmas and the lexical units are classified by their semantic categories. The number of lemmas/lexical units is divided into relevant, non-relevant, and unattested; the latter includes those derived lemmas that are covered in *DeriNet* but are indicated there as unattested in the used corpus data.

nouns, passive action adjectives, and simultaneous action adjectives, whereas the lowest number was derived for quality nouns.

For each semantic category, 300 randomly selected valency frames were evaluated by an expert, deciding whether or not each of the selected valency frames contains errors in the type and number of valency complements and/or in their morphosyntactic markers. In the case of quality nouns, the evaluation was limited to 100 generated frames due to their small number. As a result, 1,600 valency frames were evaluated. The evaluation shows that the formulated rules (Section 4.1) correctly determine changes in the valency frames of the selected semantic categories to a large extent, see Table 7. The generated valency frames describing simultaneous action adjectives, anterior action adjectives, passive action adjectives, and action nouns achieve high accuracy. Errors in these frames were limited to morphosyntactic markers of valency complements (with the exception of simultaneous action adjectives where errors in valency complements occurred).

The lower accuracy was exhibited by the valency frames of quality nouns and adjectives denoting potentiality. As for quality nouns, their generated valency frames contained more errors in valency complements than in their morphosyntactic markers, as in some cases, the valency frames comprise an extra valency complement. For example, PATient of the adjective *mocný* ‘capable’ (e.g., *slovo mocné zranit_{PAT} druhého* ‘a word capable of hurting_{PAT} another person’) is not realized in the derived noun *mocnost* ‘ability’ (e.g., *mocnost slova *zranit_{PAT} druhého* ‘the ability of a word to hurt another person’).

In contrast, in the generated valency frames of adjectives denoting potentiality, errors in morphosyntactic markers of valency complements oc-

curred, as their ACTors can be expressed by the dative usually alternating with the prepositional group *pro* ‘for’+Acc (e.g., *příměr pochopitelný laikovi_{Dat} / pro laika_{pro+Acc}* ‘an analogy understandable to a layman’). These forms are hardly predictable from the valency frame of their base verbs; rather, they are determined by their semantic class membership. Specifically, these verbs denote mental and psychological processes or states.

As a result of this experiment, a core of *DeriVallex* has been created, containing more than 8,000 lexical units of noun derivatives and nearly 15,000 lexical units of adjectival derivatives of the selected semantic categories, which are assigned automatically derived valency frames, exhibiting a high degree of accuracy. *DeriVallex*, version 1.0, is released through the LINDAT/CLARIAH-CZ repository under the Creative Commons license (Ketterová et al., 2026).¹⁷

5. Conclusion

In contrast to the valency of verbs, the valency of nouns and adjectives is not sufficiently covered in lexical resources. We have identified several semantic categories of the noun and adjectival derivatives in Czech the valency frames of which exhibit changes that are regular enough to be described on the rule basis. For these types of derivatives, the rules have been formulated, allowing for the generation of their valency frames based on the valency frames of their base verbs or adjectives. The generated valency frames form the basis of *DeriVallex 1.0*, providing the automatically generated information on the valency of derived nouns and adjectives. The advantage of this approach

¹⁷<https://ufal.mff.cuni.cz/derivalex/1.0/>

Type	Correct frames	Incorrect frames	
		Valency complements	Morphosyntactic markers
act	0.94	0.00	0.06
qual	0.70	0.20	0.10
sim	0.99	0.01	0.00
anter	0.99	0.00	0.01
pass	0.99	0.00	0.01
pot	0.88	0.00	0.12
Average	0.92	0.03	0.05

Table 7: Proportion of correctly and incorrectly generated valency frames of the selected semantic categories of derivatives (col. Correct frames and Incorrect frames, respectively). Incorrectly generated valency frames are distinguished according to errors either in valency complements or in morphosyntactic markers of complements.

is relatively high number of units with assigned valency frames, and its high accuracy. The disadvantage is that it requires a large amount of precise linguistic information. Further, its limit lies in the fact that if we narrow down our focus to the relation between the base word and its direct derivative, regular valency changes are identified in a limited set of semantic categories of derivatives. For future work, the focus should be thus shifted to the valency patterns within whole derivational families, including also indirect relations, which may bring to light regularities that would otherwise remain unnoticed. Finally, it is worth noting that manual annotation of valency is still necessary for nouns and adjectives displaying irregular (non-systemic) valency behavior.

6. Acknowledgements

The research reported in the paper was supported by the Czech Science Foundation under the project No. 25-16716S. The work described herein has been using data and tools provided by the LINDAT/CLARIAH-CZ Research Infrastructure (<https://lindat.cz>), supported by the Ministry of Education, Youth and Sports of the Czech Republic (Project No. LM2023062), and the creation of the *DeriVallex* database has been partially supported by this project too.

7. Bibliographical References

Artemis Alexiadou. 2007. Argument structure in nominals. In A. Alexiadou, L. Haegeman, and M. Stavrou, editors, *Noun Phrases in the Generative Perspective*, page 477–546. Walter de Gruyter, Berlin – New York.

David John Allerton. 2006. Valency grammar. In *Encyclopedia of Language & Linguistics*, pages 301–314. Elsevier.

Manfred Bierwisch. 2015. Word-formation and argument structure. In P. O. Müller, I. Ohnheiser, O. Susan, and F. Rainer, editors, *Word-Formation: An International Handbook of the Languages of Europe*, volume 2, page 1057–1099. Walter de Gruyter, Berlin — Boston.

Geert Booij. 2000. Inheritance. In G. Booij, Ch. Lehmann, and J. Mugdan, editors, *An International Handbook on Inflection and Word-Formation*, volume 1, pages 857–865. Walter de Gruyter, Berlin – New York.

Hagit Borer. 2013. *Structuring Sense: Volume III: Taking form*. Oxford University Press, Oxford.

Noam Chomsky. 1970. Remarks on nominalization. In I. Jacobs and P. Rosenbaum, editors, *Readings in English Transformational Grammar*, volume 2, page 184–221. Ginn & Co, Waltham.

Miloš Dokulil. 1962. *Tvoření slov v češtině 1. Teorie odvozování slov*. Nakladatelství Československé akademie věd, Praha.

Joan Grimshaw. 1990. *Argument Structure*. The MIT Press, Cambridge.

Thomas Herbst, David Heath, Ian F. Roe, and Dieter Götz. 2004. *A Valency Dictionary of English. A Corpus-Based Analysis of the Complementization Patterns of English Verbs, Nouns and Adjectives*. Mouton de Gruyter, Berlin – New York.

Holden Härtl. 2015. Argument-structural restrictions on word-formation patterns. In P. O. Müller, I. Ohnheiser, and O. Susan and F. Rainer, editors, *Word-Formation: An International Handbook of the Languages of Europe*, volume 2, pages 876–894. Walter de Gruyter, Berlin — Boston.

Václava Kettnerová, Jiří Mírovský, and Michal Olbrich. 2025. Modelling valency frames using inheritance: The case of Czech adjectives -telný ‘able’. *Jazykovedný časopis / Journal of Linguistics*, 76(1):312–321.

Veronika Kolářová. 2024. Active and passive syntax of Czech deverbal and deadjectival nouns. *Lingua*, 307:1–27.

Veronika Kolářová, Anna Vernerová, and Jana Klímová. 2021. Systemic and non-systemic valency behavior of Czech deverbal adjectives. *Jazykovedný časopis*, 72(2):371–382.

- Veronika Kolářová, Anna Vernerová, and Jonathan Verner. 2019. Non-systemic valency behavior of Czech deverbal nouns based on the NomVallex lexicon. *Jazykovedný časopis*, 70(2):424–433.
- Beth C. Levin. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. The University of Chicago Press, Chicago and London.
- Markéta Lopatková, Veronika Řezníčková, and Zdeněk Žabokrtský. 2002. Valency lexicon for Czech: From verbs to nouns. In *Text, Speech and Dialogue. TSD 2002, Lecture Notes in Computer Science*, vol. 2448, pages 147–150, Dordrecht. Springer.
- Jarmila Panevová. 1994. Valency Frames and the Meaning of the Sentence. In P. A. Luelsdorff, editor, *The Prague School of Structural and Functional Linguistics*, pages 223–243. John Benjamins, Amsterdam/Philadelphia.
- Isabelle Roy. 2010. Deadjectival nominalizations and the structure of the adjective. In A. Alexiadou and M. Rathert, editors, *The Syntax of Nominalizations across Languages and Frameworks*, page 129–158. Mouton de Gruyter, Berlin.
- Andrew Spencer. 2005. Word-formation and syntax. In P. Štekauer and R. Lieber, editors, *Handbook of Word-Formation*, page 73–97. Springer, Dordrecht.
- Nada Svozilová, Hana Prouzová, and Anna Jirsová. 2005. *Slovník slovesných, substantivních a adjektivních vazeb a spojení*. Academia, Praha.
- Zawistawska, Magdalena and Zgondek, Maciej and Żochowska, Natalia and Żurowski, Sebastian. 2018. *Walenty (2018-06-29)*. PID <http://hdl.handle.net/11321/586>. CLARIN-PL digital repository.
- Kettnerová, Václava and Mírovský, Jiří and Kolářová, Veronika and Olbrich, Michal. 2026. *DeriVallex 1.0*. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL). PID <http://hdl.handle.net/11234/1-6109>.
- Kolářová, Veronika and Kettnerová, Václava and Klímová, Jana and Mírovský, Jiří and Vernerová, Anna. 2024. *NomVallex 2.5*. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL). PID <http://hdl.handle.net/11234/1-5826>.
- Lopatková, Markéta and Kettnerová, Václava and Mírovský, Jiří and Vernerová, Anna and Bejček, Eduard and Žabokrtský, Zdeněk. 2022. *VALLEX 4.5*. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL). PID <http://hdl.handle.net/11234/1-4756>.
- Olbrich, Michal and Brezinová, Viktória and Dohnalová, Šárka and John, Vojtěch and Kyjánek, Lukáš and Papáček, Aleš and Svoboda, Emil and Ševčíková, Magda and Vidra, Jonáš and Žabokrtský, Zdeněk. 2025. *DeriNet 2.3*. PID <http://hdl.handle.net/11234/1-5846>. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL).
- Urešová, Zdeňka and Bémová, Alevtina and Fučíková, Eva and Hajič, Jan and Kolářová, Veronika and Mikulová, Marie and Pajas, Petr and Panevová, Jarmila and Štěpánek, Jan. 2024. *PDT-Vallex: Czech Valency lexicon linked to treebanks 4.5 (PDT-Vallex 4.5)*. PID <http://hdl.handle.net/11234/1-5814>. LINDAT/CLARIAH-CZ digital library at Institute of Formal and Applied Linguistics, Charles University, Prague.

8. Language Resource References

- Alberski, Bartłomiej and Andrejewicz, Jędrzej and Andrejewicz, Urszula and Andrzejczuk, Anna and Batko, Piotr and Brodzińska, Magdalena and Bukowiecka, Halina and Drabik, Lidia and Filipczak, Joanna and Grzeszak, Anna and Hajnicz, Elżbieta and Itoya, Bożena and Kaczmarska, Elżbieta and Kaluźna-Gołąb, Marta and Kocyba, Natalia and Kozłowska, Matylda and Linsztet, Barbara and Łodzińska, Agnieszka and Maciejewska, Małgorzata and Norwa, Agnieszka and Opacki, Marcin and Patejuk, Agnieszka and Przepiórkowski, Adam and Rabiega-Wiśniowska, Joanna and Rosalska, Paulina and Skubida, Natalia and Skwarski, Filip and Stankiewicz, Anna and Sulich, Adrian and Szczyszek, Michał and Szymczak, Jakub and Świdziński, Marek and Wiśniakowska, Lidia and Woliński, Marcin and Wójcicka, Alicja and Zagajewska, Anna and