THE STUDY OF VALENCY IS BIASED TOWARD MORE FREQUENT VERBS: A CORPUS STUDY OF THE VALENCY OF LESS FREQUENT VERBS IN CZECH

JAKUB SLÁMA

Czech Language Institute, Czech Academy of Sciences, Prague, Czech Republic

Abstract: Theories of valency and valency dictionaries are inevitably and understandably based on the valency behavior of frequent verbs. This paper scrutinizes 154 low-frequency Czech verbs and argues that they demonstrate that Czech verbs are more malleable in their valency behavior than suggested by the literature. It is argued that this fits better within a constructionist approach to valency rather than a lexicalist one. Furthermore, the paper illustrates two alternations, previously unrecognized for Czech as semantic diatheses, namely the causative-inchoative alternation and the Agent-Means alternation.

Keywords: valency, valency alternation, causativity, frequency

1   INTRODUCTION[1]

What we (think we) know about valency (in Czech) is somewhat biased toward more frequent verbs. Valency theories are based on examples featuring frequent predicates, and valency dictionaries understandably describe the valency behavior of the most frequent verbs (or words of other parts of speech, which are not the focus here, however). If we examine the behavior of less frequent verbs, we might encounter phenomena which might present difficulties for the traditional approaches to valency; consider, e.g., the following examples of the metaphorical sense of the infrequent verb *hypertrofovat*, roughly corresponding to 'grow':

(1) *Tento trend **hypertrofuje** zejména v posledních deseti letech.* (syn v8 [1])
     'This trend has been growing especially in the last ten years.'
(2) *Komplikují a **hypertrofují** legislativu.* (syn v8)
     'They are complicating the legislation and making it (grow) too complex.'
(3) *Ten **hypertrofoval** v podobu, kterou nelze finančně udržet.* (syn v8)
     'It grew into a form that is impossible to sustain financially.'

---

(4) *Tuhle svoji dětskou touhu jsem **hypertrofoval** do téhle chalupy.* (syn v8)
  'I transformed this childhood dream of mine into this cottage.'

The traditional approach is to say that since we are dealing with multiple (presumably four) different valency frames, we are dealing with multiple senses of the verb. However, this is a prime example of what has been described as the polysemy fallacy, that is, of viewing contextually-bound uses of a lexical item as instances of polysemy [2, p. 63]. Even more problematically, the reasoning is cyclical [3, p. 10], as it would have us believe that we are dealing with four different valency frames precisely because we have four different senses of the verb, while deducing the four senses of the verb on the basis of its use in four different valency frames. Furthermore, the idea that using the same verb in different valency frames (as, e.g., in the locative alternation) leads to polysemy has been challenged by psycholinguistic evidence [4].

2   THEORETICAL BACKGROUND

2.1 Two basic approaches to valency

There are two types of approaches to valency, or, argument structure [5, pp. 11–12]. The traditional, lexicalist, approach is characterized by the central belief that the argument structure of a predicate is determined by the predicate itself and by its semantics, and that if a verb occurs in various valency frames, these are associated with various senses of the verb. The traditional approaches to valency known in the Czech context – that of Daneš and Hlavsa [6], and the Functional Generative Description [7] – are lexicalist in nature.

An alternative approach originated within the framework of Construction Grammar (for a brief introduction, see [8]), based on the observation that argument structure cannot, in fact, be trivially reduced to knowledge tied to individual verbs (or other predicates), as illustrated by examples such as the following, cited after [9, p. 2]:

(5) *He **stared** her into immobility.*
(6) *Chess **coughed** smoke out of his lungs.*
(7) *Her nose was so bloodied that the ref **whistled** her off the floor.*
(8) *Navin **sneezed** blue pollen onto his shirt.*

These examples feature an unusual use of the verb in bold; in light of such examples, "the idea that argument structure is primarily knowledge about verbs loses some of its appeal" [9, p. 2]. The basic idea of the constructionist approach is that argument structure constructions (ASCs) – which are constructions in the sense of Construction Grammar, that is, they are Saussurean signs – exist independently of

verbs and have their own abstract meaning, and verbs might be combined with various ASCs. In English, a common verb like kick might be used in as many as some nine ASCs, without any apparent changes in its semantics [10, p. 394]. The constructionist approach has received a lot of experimental support (reviewed in [11]), and a fundamentally similar view of valency is espoused in various approaches other than Construction Grammar, which independently arrive at the conclusion that valency patterns (or, ASCs) exist as autonomous Saussurean signs (e.g., [12], [13], [14], [15]).

2.2   Bias toward more frequent verbs

Quite understandably, Czech valency dictionaries such as *Vallex* [16] include the most frequent verbs of Czech, with the latest version including 4,659 of them, which is roughly 22% of all verbs found in the representative corpus of written Czech syn2020 [17]. Nevertheless, it might be interesting to inspect less frequent verbs with respect to valency, since this can allow us to scrutinize the valency behavior of a verb in its entirety, without necessarily limiting our scope of attention to a sample of its uses and to its most typical uses, which is inevitable when compiling a (valency) dictionary. Furthermore, it has been known for a long time that more frequent words are more prone to polysemy [18, p. 109], and so inspecting the use of low-frequency verbs might allow us to study their valency behavior without the burden of polysemy. Finally, especially within the framework of usage-based (cognitive) linguistics, it has been abundantly demonstrated that frequency plays a crucial role in language, and more frequent units or expressions might behave quite differently from less frequent ones (cf. [19]). The intuition that frequency is relevant has also been present in valency research (e.g., [13, p. 59]), although rather marginally (but cf. e.g. [20]).

3   DATA

From the corpus syn2020 [17] I extracted the frequency list of all verbs, from which I selected the 118 verbs that occur in the corpus twenty times and the 36 verbs that occur seventy times (both of these numbers are arbitrary). All 4,880 occurrences of the 154 verbs were manually inspected, 36 of them were discarded (mostly as errors in lemmatization), and the 154 verbs were annotated for their valency behavior, especially for the number and type of valency frames in which they occurred in the data. Unless otherwise specified, all the examples cited in the paper are from syn2020. Occasionally, I use handier examples found in the bigger corpus syn v8 [1].

In describing valency frames, I mostly followed the Functional Generative Description. When a verb was found in multiple grammatical diatheses such as the passive (cf. [21]), these were naturally not taken as constituting different valency

frames of the verb. The same applies to definite (*nyní spolurozhodují obyvatelé Vranovic* 'now the residents of Vranovice are codeciding') and indefinite null objects (*právo spolurozhodovat* 'the right to codecide'); on both see, e.g., [22]. Similarly, I did not take examples of reflexive objects to constitute a new valency frame; that is, the same verb with a single valency frame is instantiated in *ověsí se šperky* 'she will decorate herself with jewels' and *ověsí svoji polovičku blýskavějšími diamanty* 'he will decorate his partner with more glittering diamonds'.

On the other hand, I took instances such as the following, in which the reflexive variant does not denote an action whose patient is expressed by the reflexive *se*, to represent two different valency frames of the same verb. In this respect, I diverge from most traditional accounts, which would see the verbs as two different lexical units, and I do so simply because I view examples such as these two as representing an identical meaning of one verb (which is further modulated by the syntactic context):

(9) *on se **poblil** na chodník*
    'he threw up on the sidewalk'
(10) *slibte mi, že **nepoblijete** doktora Reeda*
    'promise you will not puke all over Dr Reed'

Of course, if a verb always occurs with *se*, I do not diverge from traditional accounts. One crucial advantage of treating reflexive verbs in this partly unconventional way should become apparent when the causative-inchoative alternation is discussed in section 4.2.

Finally, I distinguished clear lexical ambiguity: e.g., the verb *odsekávat*$_1$ is found in two valency frames (11–12), while the verb *odsekávat*$_2$ is found in one (13):

(11) ***odsekávala***$_1$ ***jsem** mu* – 'I kept snapping at him'
(12) *drze **odsekáváte**$_1$ repliky* – lit. '(you) rudely retort lines'
(13) ***odsekávat***$_2$ *maso od kostí* – 'chop the meat from the bones'

I might have distinguished valency patterns that could be lumped together by others, thus, e.g., considering the following examples as instantiations of three different valency frames:

(14) $ACT_1^{obl} PAT_4^{obl}$
    *plátky **napařujte** pod pokličkou*
    'steam the slices with the lid on'
(15) $ACT_1^{obl} PAT_4^{obl} BEN^{obl}$
    ***napařuju** si obličej*
    'I am steaming my face'

(16) ACT$_1^{obl}$ PAT$_4^{obl}$ DIR3^{obl}

   *na vrstvy polyesterového filmu se **napařují** různé kovy*
   'various metals are steamed onto the layers of polyester film'

Other people might conflate (14) and (15) by claiming that both represent the same frame with an optional, albeit typical BEN, which I did not do because while instances similar to (15) are always accompanied by a BEN, instances of (14) in the data never are.

4   SOME OBSERVATIONS

4.1   Czech verbs combine with various valency frames

Clear instances of lexical ambiguity – be it due to polysemy (*vybílit* 'whitewash' → 'clean (out), steal') or homonymy (*nadívat* 'stuff' vs. *nadívat se* 'get enough of looking') – are rather rare in the data, appearing only with 13 out of the 154 verbs. Despite that, each of the 154 plus 13 verbs appears on average with 2.006 valency frames, suggesting that Czech verbs might be more malleable in their valency behavior than presumed. For instance, while *Vallex* gives three valency frames for the relevant senses of the verb *foukat* 'blow,' with the much less frequent verb *profukovat* 'blow through,' we find six valency patterns (some of the corpus examples were shortened):

DIR2^{typ} DIR3^{typ} LOC^{typ}
(17a) *přece jen trochu profukuje* 'it's a bit windy after all'
(17b) *trošku tu profukuje* 'there is a breeze here'
(17c) *okny dovnitř profukuje* 'the wind comes in through the windows'

ACT$_1^{obl}$ DIR3^{obl}
(18)   *severák profukuje až do kostí* 'the north wind blows through the bones'

ACT$_1^{obl}$ DIR2^{obl} (BEN^{typ})
(19a) *vítr profukuje skulinami* 'the wind blows through the cracks'
(19b) *vítr jí profukoval košilí* 'the wind blew through her shirt'

ACT$_1^{obl}$ PAT$_4^{obl}$ (BEN^{typ})
(20a) *letadlo profukoval ledový vítr* 'an ice cold wind was blowing through the plane'
(20b) *studený vítr profukoval Reedovi bundu* 'a cold wind blew through Reed's jacket'

ACT$_1^{obl}$
(21)   *bunda profukovala* 'the wind was blowing through my jacket' (lit. 'the jacket blew through')

ACT$_1^{obl}$ PAT$_4^{obl}$ (MEANS^{typ})
(22) *profukuje trysku karburátoru* 'he blows air through the nozzle of a carburetor'

ACT$_1^{obl}$ PAT$_4^{obl}$ DIR2^{obl}
(23) *roztaveným železem se profukoval vzduch* 'air was blown through molten iron'

While the Functional Generative Description acknowledges, e.g., the systematic alternation between uses such as (19) and (20), it still analyzes these as separate senses of the verb. However, we can claim that all of the uses above are in fact instantiations of a single sense of the verb, and the fact that the verb is interpreted differently in different valency frames does not need to be ascribed to different senses of the verb but to the valency frames, by which we avoid falling prey to the polysemy fallacy mentioned above. Thus, if there is any systematic difference between, say, sentences such as (19) and (20), it can be ascribed to the valency pattern (or, ASC) while maintaining that the meaning of the verb remains the same. Along these lines, the seemingly different meaning of (22) (and of examples such as *okna profukovala* – lit. 'the windows blew through') can again be ascribed to the construction rather than the verb itself.

As other examples show, variability in valency behavior is indeed linked to valency alternations that have been described by works such as [23]. However, this is often not the case: various verbs are used in various ASCs without any apparent change in the meaning of the verb even in cases which would not be described as alternations or diatheses, e.g.:

(24) *hlasitě **krkne*** – 'he burps loudly'
(25) ***krknula** mi do tváře dvě slova* – 'she burped two words in my face'

Here one could, indeed, posit one valency frame for the verb (featuring an optional PAT and perhaps an optional ADDR), but this would blur the fact that the (di)transitive use of the verb is marginal (which, however, should never be taken as a reason to discard it), and it does not seem possible to express either the PAT or the ADDR with most of the uses of the verb. The best analysis is in my view the constructionist one, which acknowledges that the verb has only one meaning and in the two examples, it is simply used in two different ASCs. One can posit tentatively the existence of a construction in which an ACT, an ADDR (often not expressed explicitly), and a PAT are required, which coerces the verb into the interpretation of a verbum dicendi, as witnessed by examples (12), (25), and many others, including both common expressions (*řekl mi to* 'he told me that') as well as creative uses of verbs, such as *zahalasí nějaký pozdrav* 'he shouts a greeting,' in which the verb, usually not used with a direct object, takes one.

4.2 Alternations previously unobserved in Czech

Several of the verbs display what seem to be semantic alternations previously unrecognized in Czech (cf. [23]). In the sample of 154 verbs, two of them, discussed in the following sections, recur across at least a dozen verbs, which suggests that these might be relatively common in Czech. Notably both of them feature especially (albeit perhaps not exclusively) change-of-state verbs. There were other interesting examples of what seem to be previously unrecognized alternations, but these occurred only with one or two verbs, and so for reasons of space, they will not be discussed here.

4.2.1 Causative-inchoative alternation

Multiple verbs in the data allow what is known in the literature as the causative-inchoative alternation [24, pp. 27–30]:

(26a)   *pampy se **zazelenaly***
        'the pampas turned green'
(26b)   *vykrojil z vody souš a **zazelenal** ji bylinami*
        'he cut out a patch of land out of the water and turned it green with herbs'
(27a)   *...by měla hmota **zesvětlovat***
        'the matter should get lighter'
(27b)   *Slunce je **zesvětluje**...*
        'the sun makes them lighter'
(28a)   *směs by měla **napěnit***
        'the mixture should foam'
(28b)   *my jsme ještě **nenapěnili** mýdlo*
        lit. 'we haven't foamed the soap yet'
(29a)   *prostě se **přežrala** k smrti*
        'she has just eaten herself to death'
(29b)   *buď chci nakrmit armádu králíků, nebo se těch pár pokouším **přežrat** k smrti*
        'either I want to feed an army of rabbits, or I am trying to make the few eat themselves to death [lit. trying to eat the few to death]'

While in sentences (a) the verb, often (but not necessarily), accompanied by the reflexive morpheme *se*, has an inchoative meaning, in (b) the meaning of the verb is causative, and the PAT of the causative construction[2] corresponds semantically to the ACT of the inchoative construction.[3] Although traditionally these examples would

---

[2] This is not to be confused with what has been described as the causative diathesis in the Functional Generative Description, illustrated e.g. by *dala/nechala dětem spravit boty* 'she had her children's shoes repaired' [25, p. 157].

[3] Examples of this alternation in Czech, although not treated as examples of an alternation, are however mentioned for instance in [26, pp. 223–225] and [27, p. 15].

be treated as featuring two different lexical units (verbs), I believe that both sentences in each pair in fact feature the same sense of the verb, which is, however, further modulated by the ASC in which it is used [11].

Other verbs allowing this alternation include *mutovat* 'mutate'; *vymanévrovat* 'maneuver from'; *rozcinkat* (*se*) '(start to) tinkle'; *vyháknout* (*se*) 'unhook'; *přisouvat* (*se*) 'move, push closer'; *rozesadit* (*se*) 'seat, take seats'; *přetrhat* (*se*) 'break, tear, sever'; *vyplést* (*se*) 'untangle'; *popíchat* (*se*) 'prick' (note that the translations are inevitably somewhat imprecise, especially with respect to Aktionsart).

4.2.2 Agent-Means alternation

Another salient type of alternation previously unrecognized in Czech[4] is one in which a verb takes either an ACT and a PAT, or an ACT, a PAT (corresponding to the PAT in the first configuration), and a MEANS (corresponding to the ACT in the first configuration); cf. [24, p. 80]:

(30a) *lichořešnice rychle **ozelení** plot*
    'the nasturtium quickly covers [lit. makes green] the fence'
(30b) *plot můžeme **ozelenit** některou z popínavek*
    'we can cover the fence with some vines'
(31a) *krev nepřátel se vsakuje do půdy a **zúrodňuje** ji*
    'the blood of the enemies soaks (into) the soil and fertilizes it'
(31b) *bez býložravců, kteří by svými výkaly **zúrodňovali** půdu…*
    'without herbivores which would with their dung fertilize the soil'
(32a) *sprej **zohyzdil** kašnu* [syn v8]
    'the spray damaged the fountain'
(32b) *sprejem **zohyzdil** fasády* [syn v8]
    'he damaged the facades with spray'

Other verbs allowing this alternation include *posilňovat* (*se*) 'strengthen, snack on'; *popíchat* 'prick'; *zahalasit* 'resound'; *zesvětlovat* 'lighten'; *vykurýrovat* 'cure'; *znejisťovat* 'make insecure'; *vystínovat* 'shade'; *ovonět* 'perfume'; *napěnit* 'foam'; *nastřihnout* 'incise'; *rozčleňovat* 'subdivide'; *odbouchnout* 'blow up, shoot' (again, the translations are inevitably somewhat imprecise, especially with respect to Aktionsart).

5 CONCLUSION

Scrutinizing a relatively small sample of verbs of relatively low frequency has shown that the repertoire of valency alternations available in Czech might be richer

---

[4] Note that in the Functional Generative Description, this alternation is not recognized because of the principle of shifting (i.e., the first participant is always an ACT, irrespective of its semantics).

than previously thought, and that verbs might be somewhat more malleable than is acknowledged by Czech valency theory in that they often seem to combine with multiple valency frames (or, argument structure constructions) without necessarily changing their meaning, much in the spirit of what Construction Grammar has demonstrated for English.

ACKNOWLEDGEMENTS

References

[1]   syn v8: Křen, M. et al. (2019). Korpus SYN [version 8]. Prague: Institute of the Czech National Corpus. Accessible at: www.korpus.cz.

[2]   Haugen, T. A. (2013). Adjectival valency as valency constructions: Evidence from Norwegian. Constructions and Frames, 5(1), pages 35–68.

[3]   Goldberg, A. E. (1995). Constructions. A Construction Grammar Approach to Argument Structure. Chicago: The University of Chicago Press.

[4]   Carlson, G. N., and Tanenhaus, M. K. (1988). Thematic Roles and Language Comprehension. In W. Wilkins (ed.), Thematic Relations, pages 263–288. New York: Academic Press.

[5]   de Almeida, R. G., and Manouilidou, C. (2015). The Study of Verbs in Cognitive Science. In R. G. de Almeida and C. Manouilidou (eds.), Cognitive Science Perspectives on Verb Representation and Processing, pages 3–39. Cham: Springer.

[6]   Daneš, F., and Hlavsa, Z. et al. (1987). Větné vzorce v češtině. Prague: Academia.

[7]   Sgall, P., Hajičová. E., and Panevová, J. (1986). The Meaning of the Sentence in Its Semantic and Pragmatic Aspects. Dordrecht: Reidel.

[8]   Goldberg, A. E. (2003). Constructions: a new theoretical approach to language. Trends in Cognitive Sciences, 7(5), pages 219–224.

[9]   Perek, F. (2015). Argument Structure in Usage-Based Construction Grammar. John Benjamins Publishing Company.

[10] Kemmerer, D. (2015). Cognitive Neuroscience of Language. New York: Psychology Press.

[11] Goldberg, A. E. (2019). Explain Me This. Creativity, Competition, and the Partial Productivity of Constructions. Princeton University Press.

[12] Åfarli, T. A. (2007). Do verbs have argument structure? In E. Reuland, T. Bhattacharya and G. Spathas (eds.), Argument Structure, pages 1–16. Amsterdam: John Benjamins.

[13] Schøsler, L. (2007). The status of valency patterns. In T. Herbst and K. Götz-Votteler (eds.), Valency: Theoretical, Descriptive and Cognitive Issues, pages 51–66. Mouton de Gruyter.

[14] Ickler, I. (2007). Sentence patterns and perspective in English and German. In T. Herbst and K. Götz-Votteler (eds.), Valency: Theoretical, Descriptive and Cognitive Issues, pages 253–269. Mouton de Gruyter.

[15] Borer, H. (2013). Structuring Sense Volume III: Taking Form. Oxford: Oxford University Press.

[16] Lopatková, M. et al. (2020). VALLEX 4.0. Prague: Faculty of Mathematics and Physics, Charles University. Accessible at: https://ufal.mff.cuni.cz/vallex/4.0.

[17] syn2020: Křen, M. et al. (2020). SYN2020: reprezentativní korpus psané češtiny. Prague: Institute of the Czech National Corpus. Accessible at: https://www.korpus.cz.

[18] Zipf, G. K. (1949): Human Behavior and the Principle of Least Effort. Cambridge, Mass.: Addison-Wesley Press.

[19] Divjak, D. (2019). Frequency in Language: Memory, Attention and Learning. Cambridge: Cambridge University Press.

[20] Yi, E., Koenig, J., and Roland, D. (2019). Semantic similarity to high-frequency verbs affects syntactic frame selection. Cognitive Linguistics, 30(3), pages 601–628.

[21] Kettnerová, V., and Lopatková, M. (2010). The Representation of Diatheses in the Valency Lexicon of Czech Verbs. In H. Loftsson, E. E. Rögnvaldsson and S. Helgadóttir (eds.), Proceedings of the 7th International Conference on Advances in Natural Language Processing, IceTAL 2010, pages 185–196. Heidelberg: Springer.

[22] Gillon, B. S. (2015). Optional Complements of English Verbs and Adjectives. In R. G. de Almeida and C. Manouilidou (eds.), Cognitive Science Perspectives on Verb Representation and Processing, pages 67–75. Cham: Springer.

[23] Kettnerová, V. (2014). Lexikálně-sémantické konverze ve valenčním slovníku. Prague: Karolinum.

[24] Levin, B. (1993). English Verb Classes and Alternations. A Preliminary Investigation. The University of Chicago Press.

[25] Urešová, Z. (2011). Valence sloves v Pražském závislostním korpusu. Prague: Institute of Formal and Applied Linguistics.

[26] Kettnerová, V. (2020). Derived lexical reciprocal verbs in Czech. Prace filologiczne, 75(1), pages 215–240.

[27] Veselý, L. (2020). Český vid v kostce. In L. Veselý et al. (eds.), Kapitoly o slovesném vidu nejen v češtině, pages 11–104. Prague: Czech Language Institute of the Czech Academy of Sciences.