

Meaning and Semantic Roles in CzEngClass Lexicon

Zdeňka Urešová, Eva Fučíková, Eva Hajičová, and Jan Hajič

Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics

Abstract. This paper focuses on Semantic Roles, an important component of studies in lexical semantics, as they are captured as part of a bilingual (Czech-English) synonym lexicon called CzEngClass. This lexicon builds upon the existing valency lexicons included within the framework of the annotation of the various Prague Dependency Treebanks. The present analysis of Semantic Roles is being approached from the Functional Generative Description point of view and supported by the textual evidence taken specifically from the Prague Czech-English Dependency Treebank.

1 Introduction

Since the Functional Generative Description (FGD) [10] has never systematically explored lexical semantics, it is not surprising that no description of lexical synonymy and semantic roles can be found in the pioneering works of this theory, neither there is a systematic theoretical description in the FGD follow-up works. However, some experiments with enhancing the valency lexicon of Czech verbs, starting with VALLEX 2.5 [2], [18], with semantic roles for the verbs of communication and the verbs of exchange [3], [4] building mainly on [5], where a lexicographic representation of lexical-semantic conversions is presented. The so far last version of VALLEX¹ [6] is divided into data and rule component, in order to present the representation of grammaticalized and lexicalized alternations [7]. Instead of semantic roles the term “situational participants” is used. We consider the approach of Kettnerová and Lopatková [5], [8] the fundamental starting point for our research and in various aspects we build upon it.

2 FGD approach to Meaning

The FGD’s understanding of meaning is consistent with the concept of meaning in European structural linguistics as firstly formulated in de Saussure’s works and his followers, specifically in the works of Prague scholars. As stated in [10], FGD considers the linguistic meaning distinct from (cognitive, ontological) content (or factual knowledge). FGD distinguishes two types of asymmetry: first “same content - different meaning” and second, “different content - same meaning” and makes it clear that for some correct interpretations (without ambiguity) of a sentence it is not enough to reach into the layer of linguistic meaning, but one needs to go into “the layer of cognitive content”.

Apart from the distinction of linguistic meaning and ontological content, FGD took over and substantially precised the concept of language layers. Units of the “lower” layer serve as a form, while the units on the “higher” layer aligned with these forms serve as their functions. This stratificational approach, while considering several

¹ <https://lindat.mff.cuni.cz/repository/xmlui/handle/11234/1-2307> and <http://ufal.mff.cuni.cz/vallex/3.5/>

layers, emphasizes the (deep syntactic) “tectogrammatical” (TG) layer as the main one, representing linguistically structured meaning.

As implied from the above, it is not unexpected that the deep („underlying”) syntactic constructions of which the TG layer consists of do not represent cognitive content. As noted in [16] (p. 326), “for many semanticists this (= TG) layer would not belong to the domain of meaning at all”.

3 FGD approach to Lexical Semantics

No special attention was paid to the issues related to lexical semantics within the FGD until the development of FGD-related valency lexicons started (PDT-VALLEX [17], VALLEX [2], EngVallex [27] and CzEngVallex [26]). The approach to valency in all these lexicons is based on the valency theory developed within the FGD [19], [20]. These lexicons mostly focus on verbs.

This FGD Valency Theory recognizes five “actants” of predicates (called also “inner participants,” or by other theories called „arguments”): ACT (Actor), PAT (Patient), ADDR (Addressee), EFF (Effect) and ORIG (Origin). In addition, FGD distinguishes free modifiers (by other theories called „adjuncts”) that capture circumstantial relations, such as manner-type, temporal, spatial, causal, etc. Valency characteristics of predicates are captured in their valency frames. Each valency frame consists of “(valency) slots” [9] corresponding to predicate-specific actants (obligatory or optional) and to obligatory free modifiers.²

Clearly, a given predicate verb may be ambiguous – it may have several different valency frames (verb senses), which may or may not differ in the number and type of slots. For example, the verb “to stay” may have an ACTor (who stays) and LOC (where he stays, as an obligatory free modifier, as in “John stays at home”), but also two actants – ACTor and PATient (who stays what, as in “the governor stayed the execution”). In the valency lexicons, these are considered two different entries and correspond to two different senses of the verb. The opposite is not necessarily true: if two potential valency frames have absolutely identical slots (including the morphosyntactic form associated with each slot), they are split into two entries only if their senses are clearly distinct. This is one case where the original valency lexicons “reach” for content to help distinguish the two or more senses (cf. 4.1, examples for the verb “*založit*”- “loan”, “shorten” or “bookmark”).

There is another lexical-semantic issue, namely lexical-semantic conversions, which are also partly beyond the borders of the language system. There is “a content match conditioned by the sentence context (lexical occupancy), not coming out of the language-structured meaning” [15] but from the cognitive entities within extralinguistic reality. Lexical-semantic conversions are understood as relations linked to changes in valency frames of verbs resulting from the changes in the cross-referencing of the situational participants and valence modifiers. The reference to the cognitive entities within extralinguistic reality is the reason why this language phenomenon is considered as partly leaving the borders of the language system.

Lexical-semantic conversions were first addressed by Kováčová [21]. She attempts to extend the FGD by distinguishing two types of meaning: situational (cognitive,

² The valency frames were later enriched by quasi-valency and typical modifiers [1].

lexical) and structural (grammatical) and introduces the notion of cognitive role for a participant in a linguistically structured situation. Conversion is understood as a relationship based on the identity of the situational meaning of expressions, with a specific difference in their structural meaning.

Following (with some reservations) the work of Kováčová, Kettnerová [5] proposes a lexicographic representation of lexical-semantic conversions. The proposed representation, based on “lexical-conceptual structure,”³ captures the correspondence between situational participants and valency complementations. The author considers (similarly to Kováčová) the differentiation of two types of meaning (situational content and structural meaning) crucial for delimiting the lexical units [8], [2].

Despite these two endeavours outside of the “linguistically structured meaning” principle (i.e., distinguishing verb senses with identical valency frames and conversions), the area of lexical semantics has not yet been elaborated in a systematic way in the FGD, as Hajičová [22] (p. 142) herself points out: “we are aware that lexical semantics is a domain to be investigated.”

4 Semantic Roles

Although the notion of “semantic role” (SR) is generally accepted and largely used in linguistics, there is no consent about a unified definition of SRs nor real consensus about SRs’ inventory. SRs are however considered an appropriate basis for a lexical semantic representation [23]. According to the International Organization for Standardization (ISO),⁴ semantic role is defined as a mode of involvement of a participant (i.e., a conceptual semantic unit referred to by one or more lexical items in an utterance) in an eventuality (i.e., event, state, process, or action).

4.1 FGD and Semantic Roles

FGD mostly defines the tectogrammatic level of language description as the level of linguistically structured meaning. TG „functors,” assigned to every unit of the TG representation, are understood as functions of sentence members of the surface syntax layer. They are essentially defined on semantic basis, assuming regular (“standard”, “basic”) correspondence between the domain of cognitive (semantic) roles and the domain of functors at the level of linguistically structured meaning. The TG functors, however, are not the same as cognitive roles (it would of course contradict the FGD understanding of meaning because cognitive layer is considered already “beyond” the tectogrammatrics). For example, the TG actants are subject to shifting of cognitive roles (described in detail in [19] [20]) when the “standard” or “regular” mapping of actants to cognitive roles breaks. However, when no shifting is involved, then most TG functors (ADDR, ORIG and EFF and most free modifiers) can be well compared with cognitive, i.e., extralinguistic content. In other cases, however, the same free modifier is used for what is undoubtedly different from the cognitive point of view; e.g., LOC is used not only for place - “to be under the fence.LOC,” but also for

³ http://www.glottopedia.org/index.php/Lexical_conceptual_structure

⁴ <https://iss.isolutions.iso.org/obp/ui#iso:std:iso:24617:-4:ed-1:v1:en>

“State”, since “State” is often expressed by similar morphosyntactic forms as locations (“be in a state of ...”, e.g., “to be under pressure.LOC”). These examples show that the TG actants and in some cases, even free modifiers are defined not only on semantic but also on (morpho)syntactic basis. This is most strongly displayed for ACTor and even more often for PATient, which are simply defined as the first and second actant (i.e., syntactically) regardless of a possible application of the shifting principle and their mapping to the cognitive role.

For illustration, the verbs used in the following examples have the same linguistically structured meaning as expressed on the TG layer but different meaning from the cognitive perspective (i.e., different sense). If we consider valency to share the same basic principles with the FGD, then there should only be one “meaning” of the verb “založit” in the valency lexicon (since the two actants involved, ACT and PAT, are the same, including their morphosyntactic realization as nominative and accusative, respectively), but in fact the lexicon has three entries, since the sense distinctions are obvious (“loan”, “shorten” and “bookmark”).

Anna.ACT založila Marii. PAT – Ann loaned [money] to Mary.

Anna.ACT založila sukni.PAT – Ann shortened a skirt.

Anna.ACT založila stránku.PAT v knize. – Ann put a marker in the book.

On the other hand, the verb „pocházet” [*originate*] used in the following examples has different linguistically structured meaning as well as different cognitive content. It is thus natural and within the FGD’s [linguistic-meaning-only] principle that these verb senses have separate valency frames, with different functors at the individual slots.

Anna.ACT pochází z Prahy. DIR1 - Ann comes from Prague.

Dům. ACT pochází z roku 1950. TFRWH - The house dates from 1950.

As already discussed in Sect. 3, both Kováčová [21] as well as Kettnerová a Lopatková [5], [8], need to relate TG functors between two (or more) valency frames when studying lexical-semantic conversions. Kováčová’s definition of cognitive role does not take into account the mapping between deep syntax (i.e., valency in the FGD framework) and ontological meaning. By contrast, Kettnerová and Lopatková [5], [8] work with the term “situational participants” appearing in “situations” called “situational content.” Situational content of a verb is supposed to be an abstraction (generalization) of the event situation expressed by this verb.

To sum up, the view adopted by Kováčová, who works with the term “cognitive role” referring to content, goes beyond the TG layer, i.e., beyond the FGD framework, whereas the approach of Kettnerová and Lopatková, who refer to the “situational participant” and “situational content” and their abstraction and do use it for the description of lexical conversions, is not claimed to be part of or a direct extension of FGD. Such a reference is only used as a “guidance,” and thus it is balancing on the boundary of the FGD framework, but their approach still remains largely “within the language system.”

4.2 CzEngClass approach to Semantic Roles

Whereas the FGD in detail elaborated the representation of linguistically structured meaning of verbs, the representation of cognitive content of verbs within the FGD was (for principled reasons, as discussed in Sect. 2, 3 and 4) missing. Nevertheless, CzEngClass approach to semantic roles (SRs) is based on the FGD framework and is inspired mainly by the formal representation of lexical-semantic conversions as elaborated in [5], [8] and incorporated into the newer versions of VALLEX.⁵

CzEngClass [25] strives to extend the concept of SRs to cover the whole lexicon. However, the use of SRs in CzEngClass is not the starting point: the goal is to build a bilingual Czech and English lexicon of synonym classes of verbs. SRs are an important, but not the only part of the description of the lexicon entries. The project aims primarily at delimitation of classes of synonymous verb senses by studying their semantic ‘equivalence’ in Czech-English translational context. Finding the appropriate set of SRs that characterizes each synonym class is considered to be an important tool for the specification of synonymous verb senses. The set of SRs is shared by every class member, both English and Czech. Class members are not verbs as “words” (or lemmas), but verb senses as represented by their distinct valency frames in the valency lexicons. Every SR from the given common set of SRs (Roleset) in a particular synonym class is mapped to a valency slot (represented by a TG functor) captured in the valency frame.⁶

Class: *soupeřit* – *compete*

Roleset (semantic roles)	Participant_1 – Participant_2 – Prize
<i>soupeřit</i> (PDT-Vallex-ID-v-w6280hsa_1181)	ACT – ADDR – PAT
<i>compete</i> (EngVallex-ID-ev-w616f1)	ACT – ADDR – PAT
<i>vie</i> (EngVallex-ID-ev-w3553f1)	ACT – ADDR – PAT
<i>...fight, scrap, wrangle, wrestle, ...</i>	
<i>soutěžít</i> (PDT-Vallex-ID-v-w6295f1)	ACT – ADDR – PAT
<i>bojovat</i> (PDT-Vallex-ID-v-w178f1)	ACT – ADDR – PAT
<i>...utkat se, závodit, ...</i>	

This setup, especially the introduction of SRs as the unifying element for each synonym class, is a necessary step, since otherwise it would be difficult to relate the valency (as represented in their valency frames) of the synonymous verb senses to each other, which in turn is necessary as a guidance to determine if two verb senses are synonymous or not. We believe that this is similar reason that led to the introduction of “situational participants” in the representation of cognitive content of verbs for the purposes of describing lexical-semantic conversions. Just as Kettnerová [5] refers to the layer of “situational participants” (see Fig.1),⁷ CzEngClass also links the layer of SRs to the layer of TG functors by an explicit mapping provided for the individual members of the synonymous class. This allows to relate possibly distinct valency slots (or even other complementations of the verb, i.e., free modifiers) among

⁵ <http://ufal.mff.cuni.cz/vallex/3.5/>

⁶ Or outside of it, in cases when the valency frame does not list the counterpart of the SR.

⁷ <http://ufal.mff.cuni.cz/vallex/3.0/theory.html#sec-sect-valence-alternace>

the class members, providing not only a (semi)formalized criterion for determining which verb sense should be part of the synonym class, but also to use this information in various language processing tasks.

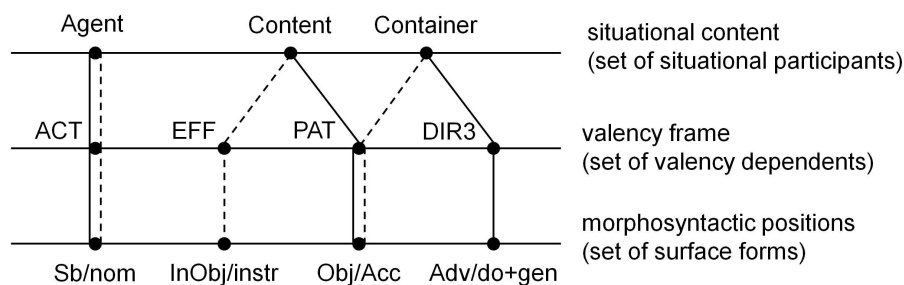


Fig. 1. Realization of locative conversion for “naplnit (fill)” (from [5])

Thanks to the provided mapping of SRs and TG functors there is enough information to also explicitly relate SRs to the surface layer (which might be useful, e.g., in natural language generation). CzEngClass SRs reflect the cognitive (extralinguistic) characteristics of the verbal complementations as activated in the “standard” contexts that we imagine to generalize across (or “abstract from”) many possible situations which are described by the utterances that use or might use the verbs from a single class. Therefore, CzEngClass’ SRs are *context-dependent semantic relations*. In this respect, our concept is close to that of the Frame Semantics⁸, where the study of meaning is considered to be the study of cognitive scenes that are created or activated by utterances [24] (p.192).

Since the CzEngClass project is a work in progress, there are still unanswered questions, such as whether the term semantic role is appropriate, or should TG functors be defined as forms for cognitive functions (by introducing a separate semantic (cognitive) layer “above” the TG layer), what will be the exact relation of this layer to the FGD, and more.

5 Conclusions

In our paper, we have presented some considerations regarding the term “semantic roles” in relation to the Functional Generative Description theory. The introduction of “semantic roles” is, in our opinion, well motivated by the need to define lexical synonymy (as approached in the CzEngClass project), especially when different valency frames are to be related for different verbs. This has been already discussed, albeit in more or less different contexts, by Kováčová and Kettnerová and Lopatková for similar reasons while studying lexical conversions. We have concluded that such a notion is indeed important and necessary for the aforementioned goals.

In the future, we will primarily focus on the relation of Semantic Roles across the synonym classes. The question here is whether they can be shared across them, and under which circumstances. In doing so, we will continue to relate them to the FGD

⁸ <http://lingo.stanford.edu/sag/papers/Fillmore-Baker-2011.pdf>

notions related to valency as well as to the way they are captured in the FGD-based valency lexicons.

6 Acknowledgments

This work has been supported by the grant No. GA17-07313S of the Grant Agency of the Czech Republic. It uses resources hosted by the LINDAT/CLARIN (LINDAT/CLARIAH-CZ) Research Infrastructure, projects No. LM2015071 and LM2018101, supported by the Ministry of Education and Youth of the Czech Republic.

References

- [1] Lopatková M. (2003): Valency in the Prague Dependency Treebank: Building the Valency Lexicon. In: *The Prague Bulletin of Mathematical Linguistics*, No. 79-80, MFF UK, pp. 37-60
- [2] Lopatková, M., Kettnerová, V., Bejček, E., Vernerová, A., Žabokrtský, Z., Barančíková, P. (2018): VALLEX 3.5 - Valenční slovník českých sloves. Charles University, Prague, <http://ufal.mff.cuni.cz/vallex/3.5/>
- [3] Kettnerová, V., Lopatková, M., Hrstková, K. (2008): Semantic Classes in Czech Valency Lexicon: Verbs of Communication and Verbs of Exchange. In: LNCS Vol. 5246, *Proceedings of the 11th International Conference TSD 2008*, Springer, Berlin/Heidelberg, pp. 109-116
- [4] Kettnerová, V., Lopatková, M., Bejček, E. (2012): Mapping Semantic Information from FrameNet onto VALLEX. In: *The Prague Bulletin of Mathematical Linguistics*, Vol. 97, Univerzita Karlova v Praze.
- [5] Kettnerová, V. (2012): *Lexikálně-sémantické konverze ve valenčním slovníku*. Ph.D. thesis, Charles University, Prague, Czech Republic, 220 pp.
- [6] Lopatková, M., Kettnerová, V., Bejček, E., Vernerová, A., Žabokrtský, Z. (2016): *Valenční slovník českých sloves VALLEX*. Nakladatelství Karolinum, Praha, Czechia, 698 pp.
- [7] Kettnerová, V., Barančíková, P., Lopatková, M. (2016): Lexicographic Description of Complex Predicates in Czech: Between Lexicon and Grammar. In: *Proceedings of the XVII EURALEX International Congress: Lexicography and Linguistic Diversity*, Tbilisi University Press, Tbilisi, Georgia, pp. 893-904.
- [8] Kettnerová, V., Lopatková, M., Bejček, E. (2012): The Syntax-Semantics Interface of Czech Verbs in the Valency Lexicon. In: *Proceedings of the 15th EURALEX International Congress*, Department of Linguistics and Scandinavian Studies, University of Oslo, Oslo, Norway, pp. 434-443.
- [9] Hajičová, E., Panevová, J. (1985): Valency (case) frames of verbs. In: *Contributions to Functional Syntax, Semantics and Language Comprehension. Linguistic and Literary Studies in Eastern Europe* 16. pp. 147-181.
- [10] Sgall, P. – Hajičová, E. – Panevová, J. (1986): *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Dordrecht, Reidel, and Prague, Academia.
- [11] Dokulil, M. and Daneš, F. (1958). K tzv. významové a mluvnické stavbě věty [On the so-called semantic and grammatical structure of the sentence], *O vědeckém poznání soudobých jazyků* 231—46 (Prague).

- [12] Sgall, P. and Panevová, J. (1976): Obsah, význam a gramatika se sémantickou bází *Slovo a slovesnost*, 37, pp. 14-25.
- [13] Hajičová, E. and Sgall, P. (1980): Linguistic meaning and knowledge representation automatic understanding of natural language, *PBML* 34: 5-21
- [14] Panevová, J. (2010): Ke vztahu kognitivního obsahu a jazykového významu. In: *Korpus – gramatika – axiologie*, Vol. 1, No. 1, Gaudeamus, Hradec Králové, Czech Republic, pp. 30-40
- [15] Panevová, J., Hajičová, E., Sgall P. (2002): *Úvod do teoretické a počítačové lingvistiky I. -- Teoretická lingvistika*. Karolinum.
- [16] Sgall, Petr (2006). Language in its multifarious aspects. Univerzita Karlova.
- [17] Hajič, J., Panevová J., Urešová Z., Bémová, A., Kolářová, V., Pajas, P. (2003): PDT-VALLEX: Creating a Large-coverage Valency Lexicon for Treebank Annotation. In: *Proceedings of The Second Workshop on Treebanks and Linguistic Theories*, Vaxjo University Press, Sweden, pp. 57–68.
- [18] Lopatková, M., Žabokrtský, Z., Kettnerová, V. (2008): *Valenční slovník českých sloves*. Nakladatelství Karolinum, Praha, 382 pp.
- [19] Panevová, J. (1974): On verbal frames in Functional Generative Description. *The Prague Bulletin of Mathematical Linguistics*, 22:3–40.
- [20] Panevová, J. (1975): On verbal frames in Functional Generative Description. *The Prague Bulletin of Mathematical Linguistics*, 23, s. 17-52.
- [21] Kováčová, Kateřina. (2005): Konverzivnost jako systémový vztah (disertační práce). Praha: Univerzita Karlova, Filozofická fakulta.
- [22] Hajičová Eva: *Syntax-Semantics Interface*. Karolinum, Praha, Czechia, 300 pp., Dec 2017
- [23] Levin, B., Rappaport H. M. (2005). *Argument Realization*. Cambridge: Cambridge University Press.
- [24] Fillmore, C. J. (1977): The Case for Case Reopened. In: P. Cole – J. M. Sadock (eds.), *Syntax and Semantics 8: Grammatical relations*. New York, San Francisco, Londyn: Academic Press, pp. 59-81.
- [25] Urešová Z., Fučíková E., Hajičová E. (2018): *CzEngClass – Towards a Lexicon of verb Synonyms with Valency linked to Semantic Roles*. In: *Jazykovedný časopis / Journal of Linguistics*, Vol. 68, No. 2, pp. 364-371, 201.
- [26] Urešová, Z., Fučíková, E., Šindlerová, J. (2016): CzEngVallex: a bilingual Czech-English valency lexicon. *The Prague Bulletin of Mathematical Linguistics*, 105:17-50.
- [27] Cinková, S. (2006): From PropBank to EngValLex: Adapting the PropBank-Lexicon to the Valency Theory of the Functional Generative Description. In: *Proceedings LREC 2006*, Genova, pp. 2170-2175.