PBML

The Prague Bulletin of Mathematical Linguistics
NUMBER 97 APRIL 2012 23–41

Mapping Semantic Information from FrameNet onto VALLEX

Václava Kettnerová, Markéta Lopatková, Eduard Bejček

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

Abstract

In this article, we introduce a project aimed at enhancing a valency lexicon of Czech verbs with semantic information. For this purpose, we make use of FrameNet, a semantically oriented lexical resource. At the present stage, semantic frames from FrameNet have been mapped to eight groups of verbs with various semantic and syntactic properties. The feasibility of this task has been verified by the achieved inter-annotator agreement measured on two semantically and syntactically different groups of verbs – verbs of communication and exchange (85.9% and 78.5%, respectively). Based on the upper level semantic frames from the relation of 'Inheritance' built in FrameNet, the verbs of these eight groups have been classified into more coherent semantic classes. Moreover, frame elements from these upper level semantic frames have been assigned to valency complementations of the verbs of the listed groups as semantic roles. As in case of semantic frames, the achieved interannotator agreement concerning assigning frame elements measured on verbs of communication and exchange has been promising (95.6% and 91.2%, respectively).

As a result, 1 270 lexical units pertaining to the verbs of communication, mental action, psych verbs, social interaction, verbs of exchange, motion, transport and location (2 129 Czech verbs in total if perfective and imperfective verbs being counted separately) have been classified into syntactically and semantically coherent classes and their valency complementations have been characterized by semantic roles adopted from the FrameNet lexical database.

1. Introduction

Information on syntactic and semantic properties of verbs, which are traditionally considered as the center of sentence, plays a key role in many rule-based NLP tasks

 Corresponding author: kettnerova@ufal.mff.cuni.cz
Cite as: Václava Kettnerová, Markéta Lopatková, Eduard Bejček. Mapping Semantic Information from FrameNet onto VALLEX. The Prague Bulletin of Mathematical Linguistics No. 97, 2012, pp. 23–41.
doi: 10.2478/v10108-012-0002-0.

as machine translation, information retrieval, text summarizing, question answering, etc. Lexical resources providing such information are designed within different theoretical frameworks and different theoretical assumptions are also reflected in their annotation schemes. As a result, there are great differences between individual lexical resources: each lexical resource captures different types of information. Consequently, interlinking information from several lexical resources represents an effective way of enriching a particular lexical resource.

However, differences in theoretical assumptions reflected in lexical resources bring several difficulties with mapping information: the different level of granularity in word sense disambiguation represents a typical example. Moreover, other requirements for harmonizing linguistic information are imposed on interlinking information from lexical resources of different languages: a fundamental prerequisite for successful mapping lies first of all in an accurate translation.

In this contribution, we introduce a project aimed at enhancing a valency lexicon of Czech verbs, VALLEX (Lopatková et al., 2008), with semantic information from FrameNet (Baker et al., 1998). This project can be seen as a pilot project focusing on mapping information from a lexical resource of a different language, namely from the English lexical resource (FrameNet) onto the Czech lexical resource (VALLEX). VALLEX and FrameNet are based on different theoretical assumptions: VALLEX takes primarily syntactic criteria in describing valency whereas FrameNet adopts more semantically oriented approach to valency. Moreover, in the project, we have to cope with the different levels of granularity in word sense disambiguation made in VALLEX and FrameNet.

The project consists of several steps. First, semantic frames from FrameNet were manually mapped onto valency frames of Czech verbs from the chosen groups of verbs, namely verbs of `communication`, `mental action`, `psych` verbs, verbs of `social interaction`, `exchange`, `motion`, `transport`, `location`. Second, frame elements from the assigned semantic frame were assigned to valency complementations of the given verbs. Then semantic frames from appropriate upper levels of abstraction based on the relation of 'Inheritance' built in FrameNet were used for classifying the verbs of the given groups into more coherent semantic classes. Moreover, frame elements from these upper level semantic frames were assigned to valency complementations of Czech verbs as semantic roles.

Manual annotation, despite being highly time consuming, seems to be indispensable at this stage of research as it brings necessary insight into the problem. Moreover, it allows us to reach the desired quality of the resulting annotation.

Two aspects are addressed in this project: (i) a practical aspect – providing data for NLP tasks, such as generation, information retrieval, or question answering, and (ii) a theoretical aspect – semantic classes allow us to observe the relation between semantic properties of verbs and their syntactic behavior; further, semantic roles enable us to make inference on lexical entailments that verbs impose on their valency complementations.

The present paper is structured as follows: in Section 2 we briefly describe two lexical resources VALLEX and FrameNet; moreover, we provide a motivation for introducing semantic information from FrameNet to VALLEX. Section 3 is focused on our experiment with mapping semantic frames and frame elements onto valency frames and valency complementations, respectively. Evaluation of both annotations is presented. In Section 4, use of the relation of 'Inheritance' built in FrameNet for classifying Czech verbs and assigning semantic roles to their valency complementations is discussed. Finally, the results of this experiment and future work are summarized.

2. Two Lexical Resources: VALLEX and FrameNet

In this section, we briefly characterize two lexical resources used in the project: VALLEX, which takes into account mainly syntactic criteria for the description of valency characteristics of verbs, and the semantically oriented FrameNet.

2.1. VALLEX – Valency Lexicon of Czech Verbs

VALLEX 2.5[1] provides information on the valency structure of verbs in their particular senses: on the number of valency complementations, on their type labeled by functors, and on their morphemic forms (Žabokrtský and Lopatková, 2007). VALLEX 2.5 describes 2 730 verb lexemes containing about 6 460 lexical units (henceforth LUs) typically corresponding to one verbal sense. At present, more than 44% of LUs are divided into heterogeneous 'supergroups', e.g., verbs of `communication`, `contact`, `emission`, `exchange`, `change`, `location`, `mental action`, `motion`, `perception`, `psych` verbs, verbs of `social interaction`, and `transport`, based primarily on similarities in morphosyntactic properties with regard to semantics. Key information on valency is stored in a valency frame.

An example LU entry in VALLEX is structured as follows:[2]

lemmas:	*impf:* ***vymýšlet***; *pf:* ***vymyslet/vymyslit*** 'to think up'
gloss:	*impf: myšlením vytvářet*; *pf: myšlením vytvořit* 'to invent or to imagine something'
frame:	ACT (obligatory) PAT (obligatory) AIM (typical) BEN (typical)
example:	*impf: vymýšlí novou metodu k léčení nádorových onemocnění*; *pf: vymyslel novou metodu k léčení nádorových onemocnění* 'he thinks up a new strategy to neoplasia treatment'
class:	mental action

VALLEX 2.5, which is closely related to the Prague Dependency Treebank 2.0 (Hajič et al., 2006), takes the Functional Generative Description (henceforth FGD) as its

[1] `http://ufal.mff.cuni.cz/vallex/2.5/`

[2] The example is simplified and translated.

theoretical background (Sgall et al., 1986). FGD applies more syntactically oriented approach to valency, see esp. (Panevová, 1994). Valency complementations are sorted out into inner participants (arguments) and free modifications (adjuncts). Both inner participants and free modifications may be obligatory or optional. Five verbal inner participants are determined rather on the basis of syntactic behavior of verbs: 'Actor' (labeled by functor ACT), 'Patient' (PAT), 'Effect' (EFF), 'Addressee' (ADDR) and 'Origin' (ORIG). In contrast to inner participants, free modifications are semantically distinctive, e.g., 'Location', 'Direction-where', 'Temporal-when', 'Cause' or 'Means', see (Mikulová et al., 2006).

2.2. FrameNet

FrameNet[3] is an on-line lexical database documenting semantic and syntactic combinatory possibilities (valences) of each word in each of its senses (Baker et al., 1998). FrameNet is based on frame semantics (Fillmore et al., 2003) and its annotation is supported by corpus evidence: each LU evokes a particular semantic frame (SF) underlying its meaning. Each SF is conceived as a "conceptual structure describing a particular type of situation, object, or event" (Ruppenhofer et al., 2006). Each SF contains the so-called frame elements (FEs), i.e., semantic participants which are understood as components of such situations. FrameNet contains more than 12 thousand LUs in 1 126 semantic frames, exemplified by more than 160 thousand lexicographic annotation sets.

FrameNet builds a wide network of hierarchical relations between SFs and their FEs. For the purpose of enhancing VALLEX with semantic information, we use the transitive relation of 'Inheritance', which is informally described as follows: "Inheritance – everything which is true about the semantics of the parent frame holds for the semantics of its child frame(s). Each FE from the parent frame (except for extrathematic FEs) is related to a relevant FE in the child frame" (Ruppenhofer et al., 2006).

2.3. Motivation for Introducing Semantic Information to VALLEX

In this section, we discuss the motivation for enhancing VALLEX with missing semantic information, namely semantic classes and semantic roles.

Semantic classes. Semantic classes provide information on relations between LUs. At present, VALLEX does not offer sufficient insight into the way a particular LU relates to another LU(s). For illustration, LUs sharing the same morphosyntactic characteristics may have the same valency frame. Thus they remain indistinct with respect to the valency structure, despite being semantically different, see the pairs of sentences (1)-(2) and (3)-(4).

[3] https://framenet.icsi.berkeley.edu/fndrupal/

(1) *Radní*.ACT *vymysleli nový plán*.PAT *rozvoje města.*
Eng. Councilmen.ACT thought a new plan.PAT for development of the city.

(2) *Turisté*.ACT *vyšli kopec*.PAT
Eng. The tourists.ACT climbed the hill.PAT

(3) *Matka*.ACT *vyprávěla dětem*.ADDR *pohádku*.PAT
Eng. The mother.ACT told the children.ADDR the fairy-tale.PAT

(4) *Jana*.ACT *přinesla otci*.ADDR *dárek*.PAT
Eng. Jane.ACT brought the father.ADDR a gift.PAT

Classifying LUs into semantic classes makes it possible to differentiate between semantically different verbs that exhibit a similar syntactic behavior. For instance, assigning SFs to the pairs of verbs in examples (1)-(2) and (3)-(4) allows us to differentiate between the given LUs: the verb *vymyslet* 'to think up' is classified as belonging to the SF 'Coming_up_with', example (1), whereas the SF 'Intentional_traversing' is assigned to *vyjít* 'to climb', example (2). Similarly, different SFs, the SF 'Statement' and the SF 'Bringing', correspond to the LUs from examples (3) and (4), respectively.

Further, semantic classes make it possible to generalize about syntactic behavior of LUs with similar semantic properties. We suppose that verbs that fall into the same class exhibit similar syntactic behavior, see also (Levin, 1993). For illustration, the verb *vystoupat* 'to ascend' and the verb *vyjít* 'to climb' appertaining to the SF 'Intentional_traversing' share the same valency frame. Similarly, other verbs, e.g. *navrhnout* 'to devise', *formulovat* 'to formulate', and *vynalézt* 'to invent' evoking by the SF 'Coming_up_with' are described by the same valency frame.

Semantic roles. Semantic roles represent one of the oldest linguistic constructs associated with a huge variety of sets of roles. These sets range from verb-specific roles, such as the 'Perpetrator' and 'Victim' for the verb 'to rape', or domain-specific roles, such as the 'Cook' and 'Produced_food' for the verbs 'to cook' or 'to bake', to general roles, such as the 'Agent', 'Theme', 'Beneficiary', or ''protoroles'', Proto-Agent and Proto-Patient, see (Dowty, 1991). FGD – using five functors for inner participants and more semantically specific functors for free modifications – lies in between these approaches, see Section 2.1. We suppose that identifying more specific semantic roles for valency complementations allows us to determine which role an individual complementation plays in a situation portrayed by a LU. Moreover, they enable us to draw inferences on lexical entailments imposed by LUs on their complementations.

For illustration, the verb *vymyslet* 'to think' in (1) is classified as belonging to the SF 'Coming_up_with' and thus the valency complementations 'Actor' and 'Patient' are mapped onto the FEs 'Cognizer' and 'Idea', respectively; whereas in case of the verb *vyjít* 'to climb' in (2) appertaining to the SF 'Intentional_traversing', these complementations are interlinked with the FEs 'Self_mover' and 'Path', respectively. Similarly, the valency complementations 'Actor', 'Addressee' and 'Patient' are described by the FEs 'Speaker', 'Addressee' and 'Message' from the SF 'Statement' in case of the verb

vyprávět 'to tell' in (3) , and by the FEs 'Agent', 'Goal' and 'Theme' from the SF 'Bringing' in case of the verb *přinést* 'to bring' in (4), respectively.

3. Mapping Semantic Information from FrameNet onto VALLEX

In this section, we report on mapping semantic information from FrameNet onto VALLEX, namely interlinking Czech LUs in VALLEX with SFs from FrameNet (Section 3.1) and their valency complementations with FEs from these SFs (Section 3.2).

3.1. Mapping Semantic Frames onto Czech Lexical Units

As the first step, we translated each LU belonging to groups of verbs of communication (C), mental action (MA), psych verbs (P), verbs of social interaction (SI), exchange (E), motion (M), transport (T), and location (L) from Czech into English.[4] The total number of annotated Czech LUs was 1 881 (341 verbs of communication, 308 verbs of mental action, 83 psych verbs, 85 verbs of social interaction, 129 verbs of exchange, 347 verbs of motion, 189 verbs of transport, and 399 verbs of location).[5]

Then the annotators had to indicate an appropriate SF (unambiguous assignment of SF) or more than one SF (ambiguous assignment of SF) for these LUs in FrameNet. The annotators could also conclude that no SF corresponds to a given Czech LU. For the overall statistics see Table 1.

Group of verbs	C	MA	P	SI	E	M	T	L
Total Czech LUs for annotation	340/341	308	83	85	129/129	347	189	399
Czech LUs without SF	66/77	125	29	54	21/27	100	34	178
Czech LUs with SF	274/264	183	54	31	108/102	247	155	221
ambiguous assignment	100/57	74	14	6	27/35	157	87	90
unambiguous assignment	174/207	109	40	25	81/67	90	68	131
SFs evoked by English LUs	415/338	292	73	38	150/140	566	279	337
Unambiguous assignments of SF	174/207	109	40	25	81/67	90	68	131
Ambiguous assignments of SF	241/131	183	33	13	69/73	476	211	206

Table 1. Annotated data size and overall statistics on the annotations of SFs.

The most frequent SFs assigned to Czech LUs include the following ones:

- communication: 'Statement', 'Request', 'Telling', 'Communication_manner', Reporting, 'Attempt_suasion';

[4] The on-line dictionary available at http://www.lingea.cz/ was used. The annotators were instructed to use all translations of a given LU provided by the lexicon.

[5] Verbs of communication and exchange were annotated by two annotators in parallel.

- `mental action:` 'Cogitation', Coming_to_believe', 'Becoming_aware', 'Assessing', 'Scrutiny', 'Grasp', 'Awareness', 'Experiencer_subj', 'Categorization', 'Hear';
- `psych:` 'Experiencer_obj', 'Cause_to_experience', 'Experiencer_subj', 'Prevarication', 'Attempt_suasion', 'Subjective_influence', 'Suasion', 'Perception_ _body', 'Objective_influence', 'Influence_on_event_on_cognizer';
- `social interaction:` 'Congregating', 'Forming_relationships', 'Residence', 'Personal_relationship', 'Make_acquaintance', 'Getting', 'Contacting', 'Be_in _agreement_on_assessment', 'Visiting', 'Temporary_stay';
- `exchange:` 'Giving', 'Getting', 'Commerce_pay', 'Theft', 'Receiving', 'Exchange', 'Commerce_buy', 'Bringing', 'Supply', 'Transfer'';
- `motion:` 'Self_motion', 'Motion', 'Arriving', 'Traversing', 'Departing', 'Body _movement', 'Operate_vehicle', 'Motion_directional', 'Path_shape', 'Ride_vehicle';
- `transport:` 'Cause_motion', 'Bringing', 'Removing', 'Cotheme', 'Sending', 'Placing', 'Delivery', 'Smuggling', 'Import_export', 'Taking';
- `location:` 'Placing', 'Attaching', 'Removing', 'Cause_motion', 'Change_posture', 'Theft', 'Residence', 'Being_located', 'Temporary_stay', 'Posture'.

Inter-annotator agreement. The feasibility of the assignment of SFs to Czech LUs was confirmed by the achieved inter-annotator agreement (IAA) measured on the groups of verbs of `communication` and `exchange`; these groups of verbs were chosen with respect to their different syntactic and semantic properties (Kettnerová et al., 2008b,a). Table 2 summarizes the inter-annotator agreement (IAA) and Cohen's κ statistics, see (Carletta, 1996), on the total number of SFs assigned to verbs of `communication` and `exchange`.

Match of SFs	IAA	κ
`C (communication)`	85.9%	0.82
`E (exchange)`	78.5%	0.73

Table 2. Inter-annotator agreement and κ statistics (considering the annotations of individual SFs for a given Czech LU as independent tasks).

Ambiguous assignments of SFs. Ambiguous annotations (i.e., annotations where the annotator has indicated more than one SF to a particular LU) draw attention to the divergence in granularity of word sense disambiguation adopted by VALLEX and FrameNet, which represents a great setback in any project dealing with mapping lexical resources.

First, let us focus on the cases in which two (or more) SFs mapped to a single Czech LU are connected by the hierarchical relation of 'Inheritance' – in general, there are these cases that reveal the finer granularity of senses applied in FrameNet. For instance, the SFs 'Bringing' and 'Smuggling' are assigned to the single Czech LU *převézt* ^{pf}, *převážet* ^{impf} 'to transport' / 'to smuggle', as in *They transported grapes to the wine lodges* and *They smuggle cocain from Peru to Britain*, respectively. The SF 'Smuggling' inherits the characteristics from the SF 'Bringing', its ancestor in the relation of 'Inheritance'; i.e., although the LU 'to smuggle' from the SF 'Smuggling' is semantically more specified – the transport is typically illegal – it inherits semantic properties from the LU 'to transport' evoking the SF 'Bringing'. We will return to the problem of different level of granularity of word sense disambiguation in Section 4.1 where we propose a method of overcoming this difficulty. This method also settles the ambiguous annotation in which sibling SFs in the relation of 'Inheritance' (or SFs with a common ancestor on an appropriate level, see below) are assigned to a single Czech LU.

Second, the ambiguous annotations of SFs that do not arise from the finer granularity (see above) may reveal mistakes in word sense disambiguation made in VALLEX. For instance, the SFs 'Grant_permission' and 'Permitting' are assigned to the Czech LU *dovolit* ^{pf}, *dovolovat* ^{impf} 'to allow', as in *Peter has allowed me to smoke here* and *This program allows data checking*, respectively. Although the verbal occurrences appear to be semantically close, the SFs evoked by them are not in the relation of 'Inheritance'. Thus this Czech LU represents a candidate for being split into two distinct senses. As a consequence, the FrameNet data can be used for checking word sense disambiguation in VALLEX.

3.2. Mapping Frame Elements onto Valency Complementations

If the human annotators indicated an appropriate SF for a Czech LU, they assigned the FE(s) from this SF to the valency complementation(s) (VCs in the following table) of the given Czech LU. Similarly as in case of mapping of SFs, more than one FE could be assigned to a single valency complementation ('Ambiguous annotation of FEs'). When no FE corresponded to a particular complementation, the annotators concluded that the given FE was missing. For the overall statistics see Table 3.

Inter-annotator agreement. As in case of SFs, the inter-annotator agreement (IAA) and κ statistics measured on the FEs assigned to the valency complementations of the verbs of `communication` and `exchange` gave satisfactory results, see Table 4.

Ambiguous assignments of FEs.
Type A. The first type of ambiguous assignments of FEs represents cases when an annotator concluded that more than one FE from a single SF corresponded to a single valency complementation due to a variety of lexical entailments imposed by a verb on such valency complementation. We can illustrate this case by the verb *zkontrolo-*

Group of verbs	C	MA	P	SI	E	M	T	L
Total VCs for annotation	1139/1142	861	259	215	522/522	1412	1024	1176
VCs without FEs (and without SF)	216/257	326	90	136	73/98	366	168	488
VCs without FEs (but with a SF)	30/22	32	8	2	38/37	32	26	26
VCs with FE(s)	893/863	503	161	77	411/387	1014	830	662
Unambiguous assignments of FE	427/534	242	116	59	276/211	275	271	289
Ambiguous assignments (type A)	212/194	98	8	2	50/75	268	218	189
Ambiguous assignments (type B)	351/195	219	46	16	112/150	680	497	279
Ambiguously assigned FEs (type A)	566/456	232	27	5	125/177	941	600	462
Ambiguously assigned FEs (type B)	952/526	537	111	32	277/309	2397	1341	735

Table 3. Annotated data size and overall statistics on the annotations of FEs.

Match of FEs	IAA	κ
C (communication)	95.6%	0.95
E (exchange)	91.2%	0.91

Table 4. Inter-annotator agreement and and κ statistics concerning assignment of FEs.

vat translated as 'to check', which belongs to the (only one) SF 'Inspecting' but has an ambiguous assignment of FEs, namely 'Patient' is labeled both with the FE 'Desired_state' and with the FE 'Ground', see (5)-(6):

(5) *Před odchodem zkontrolujte, (zda jsou zhasnutá světla)*.PAT-Desired_state
Eng. Before leaving check (that the lights are switched off).PAT-Desired_state

(6) *Zkontrolujte zámek*.PAT-Ground, *zda není porušen.*
Eng. Check the lock.PAT-Ground whether it is not damaged.

This case of the ambiguous assignment of FEs often results from the different approach to in/animateness which FrameNet and VALLEX take: VALLEX does not take into account in/animateness of the first and second inner participants, so 'Actor' and 'Patient' are often assigned ambiguously (in contrast to more semantically based valency complementations), see examples (7)-(8) in which the FEs 'Speaker' and 'Medium' are mapped onto 'Actor' of the verb *diktovat* 'to dictate':

(7) *Vzbouřenci*.ACT-Speaker *diktovali vládě*.ADDR-Addressee *své požadavky*.PAT-Message
Eng. The rebels.ACT-Speaker dictated their requirements.PAT-Message to the government.ADDR-Addresee

(8) *Mnichovská dohoda*.ACT-Medium *diktovala Československu*.ADDR-Addressee *(postoupit Německu pohraničí)*.PAT-Message

Eng. The Munich agreement.ACT-Medium ordered Czechoslovakia.ADDR-Addressee (to hand the border region over to Germany).PAT-Message

In case of verbs of `communication`, the ambiguous assignment of FEs to 'Patient' often follows from the fact that in Czech one abstract entity can express both 'theme' and 'what is said about the theme', see example (9):

(9) *Zprávy*.ACT-Medium *mluvily (o strašném zemětřesení, které zasáhlo v pátek ráno Turecko)*.PAT-Topic, Message
Eng. The news.ACT-Medium talked (about the horrible earthquake that struck Turkey on Friday morning).PAT-Topic, Message

Moreover, in Czech both 'Topic' and 'Message' can be expressed separately within a single structure (Daneš and Hlavsa, 1987), see example (10).

(10) *Cizinci*.ACT-Complainer *si stěžují starostovi*.ADDR-Addressee *na obchodníky*.PAT-Topic, *(že užívají dvojí ceny)*.EFF-Complaint
'foreigners – refl – complain – city mayor – about – sellers – that – use – double – prices'
Eng. The foreigners complain to the city mayor that the sellers use double prices.

Type B. The second type of the ambiguous assignment of FEs arises from the ambiguous assignment of SFs. In case that more than one SF were assigned to one Czech LU, the valency complementations of such Czech LU got FEs from all these SFs. For illustration, the Czech verb *dodat*$^{\mathrm{pf}}$, *dodávat*$^{\mathrm{impf}}$ translated by English verbs 'to supply' and 'to deliver' falls into the SFs 'Supply' and 'Delivery'. Thus the valency complementations of this verb are linked both with the FEs 'Supplier, 'Theme' and 'Recipient' belonging to the SF 'Supply' and with the FEs 'Deliverer', 'Recipient', and 'Theme' coming from the SF 'Delivery', see example (11a)-(11b):

(11) a. *Farmáři*.ACT-Supplier *dodávali obchodníkům*.ADDR-Recipient *čerstvou zeleninu*.PAT-Theme ('Supply')
Eng. a. The farmers.ACT-Supplier supplied the retailers.ADDR-Recipient with fresh vegetable.PAT-Theme ('Supply')
b. *Farmáři*.ACT-Deliverer *dodávali obchodníkům*.ADDR-Recipient*čerstvou zeleninu*.PAT-Theme ('Delivery')
Eng b. The farmers.ACT-Deliverer delivered fresh vegetable.PAT-Theme to the retailers.ADDR-Recipient ('Delivery')

Similarly as for SFs, the affected FEs may be connected by the relation of 'Inheritance' (as a result of SFs being in this relation). These cases arise from finer-grained granularity of word sense disambiguation in FrameNet. We will focus on them in Section 4.

In cases when ambiguously assigned FEs do not come from SFs connected by the relation of 'Inheritance' they may point out to mistakes in word sense disambiguation in VALLEX. These cases are left aside here.

4. Enhancing VALLEX with Semantic Information

In this section, we propose a method of enriching VALLEX with semantic classes (Section 4.1) and with semantic roles (Section 4.2) based on upper level SFs and their FEs from the relation of 'Inheritance'.

4.1. Enhancing VALLEX with Semantic Classes

In classifying Czech LUs into semantic classes and assigning semantic roles to their valency complementations, the semantic relation of 'Inheritance' plays a key role. This relation links such SFs which share basic semantic properties: each child frame inherits semantics from its parent frame(s). As for semantic classes, SFs from the appropriate upper level of this relation are chosen (top level SFs – represented by non-lexical and abstract SFs or SFs indicating a very general event – were disregarded); i.e., each Czech LU was classified according to the selected ancestor of the assigned SF. This method allows us to overcome the problem with coarser level of granularity made in VALLEX.

Let us demonstrate the principles of this classification on the verb *vyhnout se*$^{\mathrm{pf}}$, *vyhýbat se*$^{\mathrm{impf}}$ 'to sidestep'. This verb belongs to the SF 'Dodging' whose upper level ancestor SF in the relation of 'Inheritance' is represented by the SF 'Avoiding'. Thus to the given Czech LU, the SF 'Avoiding' is assigned as a semantic class. The same class is assigned also to the verbs belonging to the other descendant SF of 'Avoiding', namely 'Evading' (e.g., *uhnout*$^{\mathrm{pf}}$, *uhýbat*$^{\mathrm{impf}}$ 'to dodge'). See Figure 1 displaying the relation of 'Inheritance' of the SFs 'Avoiding', 'Dodging' and 'Evading'.

However, in case a Czech LU exhibits different morphosyntactic properties than LUs assigned by the relevant ancestor SF, we use the SF from an appropriate lower level of the relation of 'Inheritance'. E.g., the verb *doprovodit*$^{\mathrm{pf}}$, *doprovázet*$^{\mathrm{impf}}$ 'to accompany' belongs to the SF 'Cotheme' with the ancestor SF 'Self_motion'. Since in Czech this verb has different valency frame (obligatory 'Patient') than verbs onto which the SF 'Self_motion' was mapped (e.g., *běhat* 'to run', *kráčet* 'to march', *létat* 'to fly'), the SF 'Cotheme' from the lower level of the relation of 'Inheritance' was used as semantic class.

We set 81 SFs in total as candidates for semantic classes for verbs from the above mentioned eight groups of verbs, the entire list can be found in Appendix A.

The coverage of selected groups of verbs with these semantic classes is summarized in Table 5 (the numbers indicate a percentage of the annotated verbs from individual 'supergroups' to which semantic classes based on the selected SFs were as-

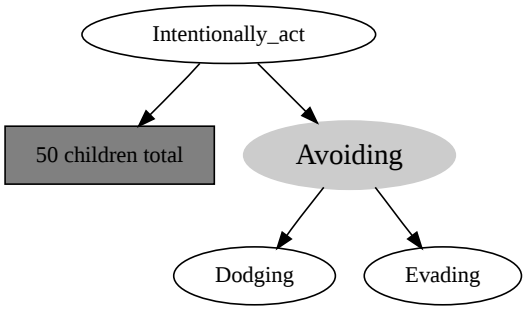

Figure 1. The relation of 'Inheritance' linking the SFs 'Avoiding', 'Dodging', and 'Evading'.

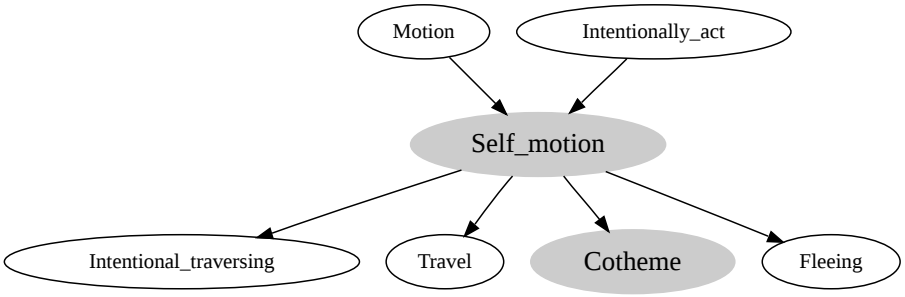

Figure 2. The relation of 'Inheritance' linking the SFs 'Cotheme' and 'Self_motion'.

signed). The differences in coverage are given primarily by the different coverage of the relation of 'Inheritance' in FrameNet.

The proposed method consisting in attributing the ancestor SFs of the assigned SFs as semantic classes allows us to overcome the problem with different granularity of verb senses in FrameNet and VALLEX. This method results in a usable set of syntactically and semantically homogeneous verb classes. Moreover, it represents also a solid basis for semantic classification of valency complementations, which is addressed in the following section.

4.2. Assigning Semantic Roles to Valency Complementations

Based on SFs mapping, we enhanced the valency lexicon with semantic roles. For this purpose, we use FEs from the ancestor SFs of the relation of 'Inheritance' that were

Groups of verbs		Coverage
C	(communication)	57/51%
MA	(mental action)	69%
P	(psych verbs)	13%
SI	(social interaction)	42%
E	(exchange)	58/56%
M	(motion)	88%
T	(transport)	84%
L	(location)	60%
Overall		76%

Table 5. Coverage of semantic classes.

chosen as semantic classes. For illustration, the valency complementations of the verb *vyhnout se*$^{\mathrm{pf}}$, *vyhýbat se*$^{\mathrm{impf}}$ 'to sidestep', included in the semantic class 'Avoiding' (representing the selected ancestor for the assigned SF 'Dodging') were labeled with FEs belonging to the SF 'Avoiding', namely 'Agent', 'Undesirable_situation', and the others, see Figure 3.

We obtained 327 FEs in total as candidates for semantic roles for the mentioned 8 'supergroups' of Czech verbs (only core FEs[6] as the most important ones are counted). The entire list can be found in Appendix B.

Similarly as in the case of semantic classes, there are differences in coverage of semantic roles, which are mainly given by the different coverage of the relation of 'Inheritance' in FrameNet.

5. Conclusion

We introduced the project aimed at enhancing the valency lexicon with missing semantic information – semantic classes and semantic roles. For this purpose, we made use of FrameNet data. We proposed a method of overcoming the problem with finer granularity of word sense disambiguation made in FrameNet. This method is based on the relation of 'Inheritance' built in FrameNet. As a result, 8 'supergroups' of Czech verbs, verbs of `communication`, `mental action`, `psych` verbs, verbs of `social interaction`, `exchange`, `motion`, `transport`, and `location` (specifically, 1 270 lexical units covering 2 129 Czech verbs in total if perfective and imperfective verbs being counted separately) were classified into syntactically and semantically coherent classes and

[6] According to (Ruppenhofer et al., 2006), core FEs are those FEs which are conceptually necessary and whose combination is characteristic of a particular SF.

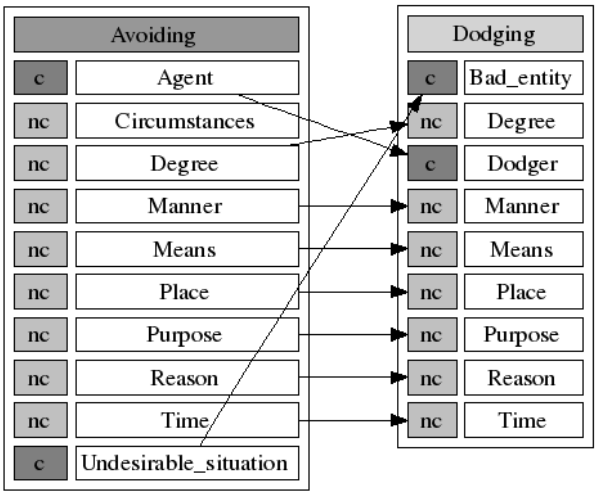

Figure 3. The relation of 'Inheritance' of the FEs belonging to the SF 'Avoiding' and 'Dodging'.

their valency complementations have been characterized by semantic roles adopted from the FrameNet lexical database.

As for future work, we intend to experiment with other groups of verbs and to increase the coverage of semantic information following the progress made in FrameNet.

5.1. Acknowledgement

This work has been using language resources developed and/or stored and/or distributed by the LINDAT-Clarin project of the Ministry of Education of the Czech Republic (project LM2010013). The paper reports on the research supported by the grant of GAČR No. P406/12/0557 and partially by the grant of GAČR P406/10/0875.

Bibliography

Baker, Collin F., Charles J. Fillmore, and John B. Lowe. The Berkeley FrameNet Project. In *Proceedings of the COLING-ACL*, pages 86–90, Montreal, Canada, 1998.

Carletta, Jean. Assessing agreement on classification tasks: The Kappa statistic. *Computational Linguistics*, 22(2):249–254, 1996.

Daneš, František and Zdeněk Hlavsa. *Větné vzorce v češtině.* Academia, Praha, 1987.

Dowty, David. Thematic Proto-Roles and Argument Selection. *Languauge*, 67:547–619, 1991.

Fillmore, Charles J., Christopher Johnson, and Miriam R. L. Petruck. Background to framenet. *International Journal of Lexicography*, 16(3):235–250, 2003.

Hajič, Jan, Eva Hajičová, Jarmila Panevová, Petr Sgall, Petr Pajas, Jan Štěpánek, Jiří Havelka, and Marie Mikulová. *Prague Dependency Treebank 2.0*. Linguistic Data Consortium, Philadelphia, PA, USA, 2006.

Kettnerová, Václava, Markéta Lopatková, and Klára Hrstková. Semantic roles in valency lexicon of czech verbs: Verbs of communication and exchange. In Ranta, Aarne and Bengt Nordström, editors, *Advances in Natural Language Processing (6th International Conference on NLP, GoTAL 2008)*, volume 5221 of *LNAI*, pages 217–221, Berlin / Heidelberg, 2008a. Springer.

Kettnerová, Václava, Markéta Lopatková, and Klára Hrstková. Semantic Classes in Czech Valency Lexicon: Verbs of Communication and Verbs of Exchange. In Sojka, Petr, Aleš Horák, Ivan Kopeček, and Karel Pala, editors, *Proceedings of Text, Speech and Dialog International Conference, TSD 2008*, volume 5246 of *LNAI*, pages 109–116, Berlin / Heidelberg, 2008b. Springer.

Levin, Beth C. *English Verb Classes and Alternations: A Preliminary Investigation*. The University of Chicago Press, Chicago and London, 1993.

Lopatková, Markéta, Zdeněk Žabokrtský, and Václava Kettnerová. *Valenční slovník českých sloves*. Nakladatelství Karolinum, Praha, 2008.

McConville, Mark and Myroslava O. Dzikovska. Using inheritance and coreness sets to improve a verb lexicon harvested from framenet. In *Proceedings of the Second Linguistic Annotation Workshop (LAW II)*, Marrakech, 2008. LREC.

Mikulová, Marie, Alevtina Bémová, Jan Hajič, Eva Hajičová, Jiří Havelka, Veronika Kolářová, Lucie Kučová, Markéta Lopatková, Petr Pajas, Jarmila Panevová, Magda Razímová, Petr Sgall, Jan Štěpánek, Zdeňka Urešová, Kateřina Veselá, and Zdeněk Žabokrtský. Annotation on the tectogrammatical level in the prague dependency treebank. annotation manual. Technical Report TR-2006-30, ÚFAL MFF UK, Prague, 2006.

Panevová, Jarmila. Valency Frames and the Meaning of the Sentence. In Luelsdorff, Philip A., editor, *The Prague School of Structural and Functional Linguistics*, pages 223–243. John Benjamins Publishing Company, Amsterdam, Philadelphia, 1994.

Ruppenhofer, Josef, Michael Ellsworth, Miriam R. L. Petruck, Christopher R. Johnson, and Jan Scheffczyk. *FrameNet II: Extended Theory and Practice*. University of California, Berkeley, 2006. http://framenet.icsi.berkeley.edu/book/book.html.

Sgall, Petr, Eva Hajičová, and Jarmila Panevová. *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Reidel, Dordrecht, 1986.

Žabokrtský, Zdeněk and Markéta Lopatková. Valency information in VALLEX 2.0: Logical structure of the lexicon. *The Prague Bulletin of Mathematical Linguistics*, (87):41–60, 2007.

Appendix A: List of Semantic Frames Assigned to Lexical Units as Semantic Classes

- **communication:** 'Communication', 'Statement', 'Communication_response', 'Judgment_communication', 'Chatting, 'Prohibiting', 'Request', 'Reporting', and 'Commitment';
- **mental action:** 'Cogitation', 'Assessing', 'Memorization', 'Coming_to_believe', 'Becoming_aware', 'Awareness', 'Categorization', 'Scrutiny', 'Desiring', 'Differentiation', 'Opinion', 'Forgiveness', 'Certainty', 'Purpose', 'Memory', 'Judgment', 'Resolve_problem', 'Attention', and 'Deciding';
- **psych:** 'Cause_to_experience', and 'Eventive_cognizer_affecting';
- **social interaction:** 'Forming_relationships', 'Make_acquaintance', 'Be_in_agreeement_on_assessment', 'Visiting', 'Rewards_and_punishments', 'Hostile_encounter', and 'Finish_competition';
- **exchange:** 'Giving', 'Getting', 'Replacing', 'Exchange', 'Robbery', 'Hiring', 'Transfer', 'Frugality', 'Taking', and 'Supply';
- **motion:** 'Departing', 'Self_motion', 'Motion', 'Traversing', 'Motion_directional', 'Change
_posture', 'Cause_to_move_in_place', 'Avoiding', 'Surpassing',
'Cause_impact', 'Arriving', and 'Touring';
- **transport:** 'Cause_motion', 'Bringing', 'Cotheme', 'Filling', 'Firing', and 'Releasing';
- **location:** 'Placing', 'Attaching', 'Removing', 'Residence', 'Being_located', 'Inhibit_movement', 'Gathering_up', 'Aiming', 'Hiding_objects', 'Appointing', 'Cause_to_amalgamate', 'Being_attached', 'Arranging', 'Preserving', 'Emptying', and 'Amalgamation'.

Appendix B: List of Frame Elements Assigned to Valency Complementations as Semantic Roles

- **`communication`:**
  1. 'Communication': 'Communicator', 'Medium', 'Message', and 'Topic';
  2. 'Statement': 'Medium', 'Message', 'Speaker', and 'Topic';
  3. 'Communication_response': 'Addressee', 'Message', 'Speaker', 'Topic', and 'Trigger';
  4. 'Judgment_communication': 'Communicator', 'Evaluee', 'Expressor', 'Medium', 'Reason', and 'Topic;
  5. 'Chatting': 'Interlocutor_1', and 'Interlocutor_2';
  6. 'Prohibiting': 'Principle', and 'State_of_affairs';
  7. 'Request': 'Addressee', 'Medium', 'Message', 'Speaker', and 'Topic';
  8. 'Reporting': 'Authorities', 'Behavior', 'Informer', and 'Wrongdoer';
  9. 'Commitment': 'Addressee', 'Medium', 'Message', 'Speaker', and 'Topic'.
- **`mental action`:**
  1. 'Cogitation':' Cognizer', and 'Topic';
  2. 'Assessing': 'Assessor', 'Feature', 'Medium', 'Method', and 'Phenomenon';
  3. 'Memorization': 'Cognizer', and 'Pattern';
  4. 'Coming_to_believe': 'Cognizer', 'Content', 'Evidence', 'Medium', 'Means', and 'Topic';
  5. 'Becoming_aware': 'Cognizer', 'Instrument', 'Means', 'Phenomenon', and 'Topic';
  6. 'Awareness': 'Cognizer', 'Content', 'Topic', and Expressor;
  7. 'Categorization': 'Cognizer', 'Criteria', 'Item', and 'Category';
  8. 'Scrutiny': 'Cognizer', 'Ground', 'Instrument', and 'Medium';
  9. 'Desiring': 'Event', 'Experiencer', 'Focal_participant', and 'Location_of_event';
  10. 'Differentiation': 'Cognizer', 'Phenomena', 'Phenomenon_1', 'Phenomenon_2', and 'Quality';
  11. 'Opinion': 'Cognizer', and 'Opinion';
  12. 'Forgiveness': 'Judge', 'Evaluee', and 'Offense';
  13. 'Certainty': 'Cognizer', 'Content', 'Expressor', and 'Topic';
  14. 'Purpose': 'Agent', 'Attribute', 'Goal', 'Means', and 'Value';
  15. 'Memory': 'Cognizer', 'Content', and 'Topic';
  16. 'Judgment': 'Cognizer', 'Evaluee', 'Reason', and 'Expressor';
  17. 'Resolve_problem': 'Agent', 'Cause', and 'Problem';
  18. 'Attention': 'Expressor', 'Figure', and 'Perceiver';
  19. 'Deciding': 'Cognizer', and 'Decision'.
- **`psych`:**
  1. 'Cause_to_experience': 'Agent', and 'Experiencer';
  2. 'Eventive_cognizer_affecting': 'Cognizer', 'Content', and 'Event'.
- **`social interaction`:**
  1. 'Forming_relationships': 'Partner_1', 'Partner_2', and 'Partners';
  2. 'Make_acquaintance': 'Individuals', 'Individual_1', and 'Individual_2';
  3. 'Be_in_agreeement_on_assessment': 'Cognizer_1', 'Cognizer_2', 'Cognizers', 'Opinion', 'Question', and 'Topic';
  4. 'Visiting': 'Agent', and 'Entity';

5. 'Rewards_and_punishments': 'Agent', 'Evaluee', and 'Reason';
6. 'Hostile_encounter': 'Side_1', 'Side_2', 'Sides', 'Purpose', and 'Issue';
7. 'Finish_competition': 'Competition', 'Competitor', 'Opponent', and 'Competitors'.

- **exchange:**
  1. 'Giving': 'Donor', 'Recipient', and 'Theme';
  2. 'Getting': 'Recipient', and 'Theme';
  3. 'Replacing': 'Agent', 'New', and 'Old';
  4. 'Exchange': 'Exchanger_1', 'Exchanger_2', 'Theme_1', and 'Theme_2';
  5. 'Robbery': 'Perpetrator', 'Source', and 'Victim';
  6. 'Hiring': 'Employee', 'Employer', 'Field', 'Position', and 'Task';
  7. 'Transfer': 'Donor', 'Recipient', 'Theme', and 'Transferors';
  8. 'Frugality': 'Behavior', 'Resource', and 'Resource_controller';
  9. 'Taking': 'Agent', 'Source', and 'Theme';
  10. 'Supply': 'Purpose_of_recipient', 'Recipient', 'Supplier', and 'Theme'.
- **motion:**
  1. 'Departing': 'Source', and 'Theme';
  2. 'Self_motion': 'Area', 'Direction', 'Goal', 'Path', 'Self_mover', and 'Source';
  3. 'Motion': 'Area', 'Direction', 'Distance', 'Goal', 'Path', 'Source', and 'Theme';
  4. 'Traversing': 'Area', 'Direction', 'Distance', 'Goal', 'Path', 'Path_shape', 'Source', and 'Theme';
  5. 'Motion_directional': 'Area', 'Direction', 'Goal', 'Path', 'Source', and Theme';
  6. 'Change_posture': 'Protagonist';
  7. 'Cause_to_move_in_place': 'Agent', 'Body_part_of_agent', 'Cause', and 'Theme';
  8. 'Avoiding': 'Agent', and 'Undiserable_situation';
  9. 'Surpassing': 'Attribute', 'Profiled_attribute', 'Profiled_item', 'Standard_attribute', and 'Standard_item';
  10. 'Cause_impact': 'Agent', 'Cause', 'Impactee', 'Impactor', and 'Impactors';
  11. 'Arriving': 'Goal', and 'Theme';
  12. 'Touring': 'Attraction', and 'Tourist'.
- **transport:**
  1. 'Cause_motion': 'Agent', 'Area', 'Cause', 'Goal', 'Initial_state', 'Path', 'Result', 'Source', and 'Theme';
  2. 'Bringing': 'Agent', 'Area', 'Carrier', 'Goal', 'Path', 'Source', and 'Theme';
  3. 'Cotheme': 'Area', 'Cotheme', 'Direction', 'Goal', 'Path', 'Road', 'Source', and 'Theme';
  4. 'Filling': 'Agent', 'Cause', 'Goal', and 'Theme';
  5. 'Firing': 'Employee', 'Employer', 'Position', and 'Task';
  6. 'Releasing': 'Agent', 'Location_of_confinement', and 'Theme'.
- **location:**
  1. 'Placing': 'Agent', 'Cause', 'Goal', and 'Theme';
  2. 'Attaching': 'Agent', 'Connector', 'Item', 'Items', and 'Goal';
  3. 'Removing': 'Agent', 'Cause', 'Source', and 'Theme';
  4. 'Residence': 'Resident', 'Co_resident', and 'Location';
  5. 'Being_located': 'Theme', and 'Location';
  6. 'Inhibit_movement': 'Agent', 'Cause', 'Theme', and 'Holding_location';
  7. 'Gathering_up': 'Agent', 'Aggregate', and 'Individuals';

8. ‘Aiming’: ‘Agent’, ‘Instrument’, ‘Targeted’, and ‘Target_location’;
9. ‘Hiding_objects’: ‘Agent’, ‘Hidden_object’, and ‘Hiding_place’;
10. ‘Appointing’: ‘Selector’, ‘Role’, ‘Official’, ‘Function’, and ‘Body’;
11. ‘Cause_to_amalgamate’: ‘Agent’, ‘Part_1’, ‘Part_2’, ‘Parts’, and ‘Whole’;
12. ‘Being_attached’: ‘Item’, ‘Items’, ‘Goal’, and ‘Connector’;
13. ‘Arranging’: ‘Agent’, ‘Configuration’, and ‘Theme’;
14. ‘Preserving’: ‘Agent’, ‘Medium’, and ‘Undergoer’;
15. ‘Emptying’: ‘Agent’, ‘Cause’, ‘Source’, and ‘Theme’;
16. ‘Amalgamation’: ‘Parts’, ‘Part_1’, ‘Part_2’, and ‘Whole’.

Address for correspondence:
Václava Kettnerová
kettnerova@ufal.mff.cuni.cz
Institute of Formal and Applied Linguistics
Charles University in Prague
Malostranské náměstí 25
118 00 Praha 1, Czech Republic