# Enhancing Czech Valency Lexicon with Semantic Information from FrameNet: The Case of Communication Verbs

**Václava Benešová, Markéta Lopatková and Klára Hrstková**
Institute of Formal and Applied Linguistics
Faculty of Mathematics and Physics
Malostranské nám. 25, Prague
{benesova,lopatkova,hrstkova}@ufal.mff.cuni.cz

## Abstract

In this paper, we report on our attempt at assigning semantic information from FrameNet to lexical units in VALLEX, a valency lexicon of Czech verbs. We focus on the class of communication verbs. We experiment with assigning FrameNet semantic frames to lexical units for communication verbs. The second task consists in linking their valency complementations with FrameNet frame elements. The exact pairwise inter-annotator agreement reaches almost 69% on the semantic frames and 84.6% on the frame elements. We propose enhancing VALLEX with missing semantic information from FrameNet based on exploitation of the semantic relation 'Inheritance'.

## 1 Introduction

Syntactic and semantic behavior of verbs, which are considered to be the syntactic centers of sentences, is the key information for any rule-based tasks of NLP. Apparently, various theoretical approaches to valency are reflected in the annotation schemes of particular lexical resources (for a short survey and characteristics of the most important projects, see e.g. (Žabokrtský, 2005)). For linguists, their dissimilarity represents a challenging task of their comparison and mutual enriching.

In this article, we describe an experiment with enhancing VALLEX, a valency lexicon of Czech verbs (Žabokrtský et al., 2007), with FrameNet semantic information. In the experiment, we attempted at linking FrameNet (Baker et al., 1998) semantic features – semantic frames and frame elements – to the VALLEX data. First, lexical units in VALLEX were assigned semantic frames. Then their valency complementations were linked with frame elements.

VALLEX, using the Functional Generative Description as its theoretical background, see (Sgall et al., 1986), takes into account mainly the syntactic criteria for identifying arguments. However, the semantic information is crucial for handling such NLP tasks as generation, information retrieval, or question answering. Therefore, we have decided to obtain this missing information from the Berkeley FrameNet project. In spite of its being still in progress, it represents an elaborated semantically oriented lexical resource describing valency for English.

We exploit the tentative classification of verbs in VALLEX, sorting verbs into rough and heterogeneous 'supergroups'. We pursue three goals: First, we attempt at further structuring a particular group of verbs into more homogenous, subtler classes. Second, we expect to observe a hierarchy of semantic relations between Czech verbs analogous to the hierarchy of English verbs[1] as a result of assigning semantic frames to a semantically coherent group of Czech verbs. Last, we aim at choosing coarse-grained semantic information from FrameNet for enhancing VALLEX.

The paper is structured as follows: in Section 2, we present basic properties of VALLEX and FrameNet, Section 3 describes the method of our ex-

[1] Displayed by Frame Grapher Tool, see http://framenet.icsi.berkeley.edu/FrameGrapher/

periment with assigning semantic information from FrameNet to lexical units from VALLEX. Section 4 provides the analysis of the material obtained from the experiment. Lastly, the outline of further exploiting FrameNet semantic information for VALLEX follows in Section 5.

## 2 A Brief Characterization of Two Annotation Schemes: VALLEX and FrameNet

In this section, we briefly describe two approaches to the description of valency: VALLEX, which takes into account mainly syntactic criteria for identifying arguments, and semantically oriented FrameNet.

### 2.1 Valency Lexicon of Czech Verbs: VALLEX

The Valency Lexicon of Czech Verbs, VALLEX, version 2.0[2], provides information on the valency structure of verbs in their particular senses, lexical units (LUs).

LUs are considered, identically with (Cruse, 1986), as "form-meaning complexes with (relatively) stable and discrete semantic properties"; roughly speaking, 'the given word in the given sense'. Distinguishing senses is based on both syntactic and semantic properties (i.e., considerable shifting in meaning).

Each LU covers both the perfective and the imperfective Czech verbs (if they exist) that create a single lexeme. The following Table 1 shows the basic statistics about VALLEX.

| | VALLEX 2.0 |
|---|---|
| Number of lexeme entries | 2731 |
| Number of verbs | 4250 |
| Number of LUs | 6462 |
| Number of LUs with a class | 2903 |
| Number of classes | 22 |
| Lexical units with a class (%) | 44.9% |

Table 1: Basic statistics about VALLEX 2.0

**Valency Frames.** In VALLEX, the key information on the valency structure of a given LU is encoded in the form of *valency frames*. Valency frames are formed as a sequence of slots; each slot stands for one valency complementation and consists of its type ('Actor', 'Addressee', etc.), morphemic realization and its obligatoriness (obligatory or optional), see below. An example of the lexeme for the verbs *doplňovat*$^{impf}$, *doplnit*$^{pf}$ 'to add' as captured in VALLEX can be seen in Figure 1.

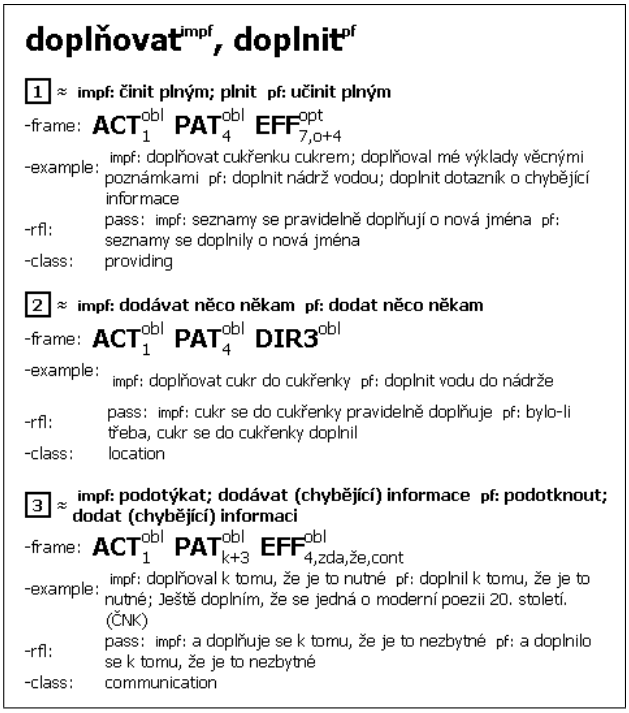


Figure 1: The lexeme for the verbs *doplňovat*$^{impf}$, *doplnit*$^{pf}$ 'to add' in VALLEX, consisting of three LUs.

**Valency Complementations.** In VALLEX, based on the *Functional Generative Description* (FGD, see esp. (Sgall et al., 1986), (Panevová, 1974)), valency complementations (VCs) are divided into arguments (inner participants) and free modifications (adjuncts). They both can be obligatory or optional.

*(Verbal) arguments* are distinguished rather on the basis of the syntactic behavior of verbs. Two criteria are applied (introduced in (Panevová, 1974)):

- each argument can modify only a more or less closed class of verbs (that can be listed),
- each argument can modify a particular verb only once (except for the case of coordination).

[2] http://ufal.mff.cuni.cz/vallex /2.0/

Moreover, possible morphemic realization(s) of these arguments is/are typically determined by the governing verb.

Five types of arguments have been determined – 'Actor' (label ACT), 'Patient' (PAT), 'Addressee' (ADDR), 'Origin' (ORIG), and 'Effect' (EFF).

According to FGD, see (Panevová, 1974), if a verb has only one argument, then this argument is 'Actor' and if a verb has two arguments, then these are 'Actor' and 'Patient'. PropBank (Palmer, 2005) has the similar approach at least to non-ergative verbs.[3] When determining the other three arguments, semantic criteria are also taken into account. As a consequence, the types of arguments do not always reflect the exact semantic relation between a verb and its arguments.

In contrast to the arguments, the *free modifications* are semantically distinctive, being identified on the basis of their syntactico-semantic functions.

**Verb Classes in VALLEX.** At the present time, a selected part of LUs in VALLEX are assigned semantic classes like 'motion', 'communication', 'perception', etc. These classes were built in a 'bottom-up' way mainly on the basis of the syntactic properties of LUs, with regard to their semantics. These rough heterogeneous 'supergroups' – although not based on a properly defined ontology – can represent an efficient starting point for building an applicable semantic classification of verbs.

**Communication verbs.** The reported project focuses on the 'communication verbs' (the so-called 'verba dicendi'). These verbs can generally be specified as verbs rendering situation when 'a speaker conveys information to a recipient'. They are characteristic by the entity of 'information' which can be expressed as a dependent content clause.

This class of verbs was convenient for our experiment for two reasons: first, they have specific syntactic behavior, and second, they represent a good-sized (large enough) class. Moreover, we assume that the results will be applicable also to other groups of verbs with a sentential complement.

[3]However, the status of ergative verbs in the Slavic languages are not clear.

### 2.2 FrameNet

The FrameNet lexical database[4] is an on-line lexical resource for English, see (Baker et al., 1998). Its aim is "to document the range of semantic and syntactic combinatory possibilities (valences) of each word in each of its senses, through a computer-assisted annotation of example sentences", see (Ruppenhofer et al., 1986).

As to the quantitative characteristics, FrameNet contains more than 10 thousand lexical units (a pair consisting of a word and its meaning)[5] in more than 825 semantic frames, exemplified by around 135 thousand annotated sentences. At present, the project focuses primarily on verbs, nouns and adjectives.

**Semantic Frames.** The descriptive framework of FrameNet is based on *frame semantics*. Each LU evokes a particular semantic frame underlying its meaning. Each *semantic frame* (SF) can be understood as a "conceptual structure describing a particular type of situation, object, or event", see (Ruppenhofer et al., 1986). Each SF contains the so-called frame elements, i.e., semantic participants which are seen as components of such situations.

For example, the SF 'Statement' is defined as follows: "This frame contains verbs and nouns that communicate the act of a Speaker to address a Message to some Addressee using language ...", see FrameNet webpage.

**Frame Elements.** Semantic frames consist of *frame elements* (FEs), semantic arguments of a predicating word evoking this frame.

Whereas, for instance, *Case Grammar*, see (Fillmore, 1968) assumes a fixed, relevant-across-the-board collection of underlying 'cases', FEs representing semantic information are understood in terms of roles in specific frames and not as a restricted set of universal semantic roles. It implies that the inventory of FEs is specific for each SF.

Three types of FEs are distinguished, *core FEs* (conceptually necessary FEs whose combination is characteristic for a particular SF), *peripheral FEs* (not unique for a given SF, they can occur in any

[4]http://framenet.icsi.berkeley.edu/

[5]We use the same abbreviation LU both for VALLEX and FrameNet because the same concepts are concerned in principle.

SFs) and *extra-thematic FEs* (that set a given event on the background of another event or state of the same type). E.g., the SF 'Statement' consists of core FEs 'Speaker', 'Topic', 'Message' and 'Medium' and peripheral FEs 'Addressee', 'Depictive', 'Degree', 'Epistemic_stance', 'Group', 'Internal_cause', 'Manner', 'Means', 'Occasion', 'Particular_iteration', 'Place' and 'Time'.

The following sentence illustrates the SF 'Statement' and its FEs:

> *President Kennedy*.Speaker *said to an astronaut*.Addressee: *("Man is still the most extraordinary computer of all.")*.Message.

**Hierarchy of Semantic Relations between SFs.** FrameNet builds a wide network of hierarchical relations between SFs and their FEs. The most important relations are the following, see (Ruppenhofer et al., 1986):

- '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 extra-thematic FEs) is related to a relevant FE in the child frame.

- 'Using' – the parent frame constitutes the background for its child frames. Not all FEs from the parent frame must be bound to the FEs from the child frame.

- 'Subframe' – the child frame instantiates a part of a complex event represented by the parent frame.

For the purpose of enhancing VALLEX with semantic information, we exploit the transitive relation 'Inheritance', as will be discussed in Section 5.

## 3 Assigning Semantic Information from FrameNet to Valency Frames in VALLEX

In this section, we report on assigning the semantic information from FrameNet to the VALLEX communication verbs. In the first step, we translated each LU from Czech to English.[6] The total number of translated Czech LUs was 341 (without idiomatic LUs). These LUs correspond to 531 Czech verbs, counting perfective and imperfective verbs separately.

### 3.1 Assigning Semantic Frames and Frame Elements

Two human annotators (referred $A_1$ and $A_2$ in the sequel) searched each translated English LU in FrameNet and indicated an appropriate SF (labeled as 'Unambiguous Annotations'). The annotators were allowed to assign more than one SF to a particular LU ('Ambiguous annotations') – if the English equivalents belonged to more than one SF.

In two situations, the annotators could conclude that the given English LU is missing in FrameNet: (i) the corresponding lemma was missing in FrameNet at all, or (ii) the found English LU did not correspond to the meaning of the given Czech LU. The following Table 2 shows the basic statistics concerning assigning SFs.

| | $A_1$ | $A_2$ |
|---|---|---|
| Cz LUs | 341 | 341 |
| Eng equivalents | 653 | 653 |
| Annotations of SFs | 610 | 556 |
| Unambiguous annotations of SFs | 143 | 165 |
| Ambiguous annotations of SFs | 467 | 391 |
| Marked as missing Eng SFs | 11 | 19 |
| Marked as missing Eng LUs | 33 | 35 |

Table 2: Annotated data size and statistics of two annotations of SFs.

After having indicated the appropriate SF(s), the annotators had to assign the corresponding FE(s) of this/these SF(s) to each valency complementation (VC) of the given Czech LU. Similarly as in the case of assigning SFs, a valid answer indicated appropriate FE(s) ('Unambiguous/Ambiguous annotations'). In the cases when no suitable FE was found, the annotators used a special flag. Table 3 gives the numbers of VCs and FEs used in the experiment.

[6] The on-line dictionary at http://www.lingea.cz/ was used for manual translation.

| | $A_1$ | $A_2$ |
|---|---|---|
| Annotations of VCs from VALLEX | 1088 | 1088 |
| Annotations of FEs | 1322 | 1314 |
| Unambiguous annotations of FEs | 869 | 879 |
| Ambiguous annotations of FEs | 453 | 435 |
| Marked as missing Eng FEs | 47 | 34 |

Table 3: Annotated data size and statistics about two annotations of FEs.

### 3.2 Results: Inter-annotator Agreement

Table 4 summarizes the inter-annotator agreement (IAA) and Cohen's $\kappa$ statistics, see (Carletta, 1996). The exact match of answers relating to SFs reaches 68.8%. The $\kappa$ statistics compensates IAA for agreement by chance. The level relating to SFs that we achieved (0.47) represents a very moderate agreement, see (Krippendorff, 1980). However, the intersection match (if both annotators chose the same SFs regardless of other variants in the case of ambiguous annotations) gives a more satisfactory result (88.2%, $\kappa$ = 0.79). IAA relating to FEs is measured only in cases of an exact match of SFs (401 cases). IAA concerning FEs (84.6%, $\kappa$ = 0.83) is much better in comparison with SFs. The intersection match concerning FEs represents a significant result (93.3%, $\kappa$ = 0.92).

| | IAA [%] | $\kappa$ |
|---|---|---|
| Exact match of SFs | 68,8% | 0.47 |
| Intersection match of SFs | 88,2% | 0.79 |
| Exact match of FEs | 84,6% | 0.83 |
| Intersection match of FEs | 93,3% | 0.92 |

Table 4: Inter-annotator agreement and $\kappa$ statistics.

## 4 Analysis of Obtained Material

In this section, we describe the analysis of the material obtained from our experiment mainly from the linguistic point of view. Special attention is paid to ambiguous assignment of SFs to Czech LUs and FEs to valency complementations.

### 4.1 Analysis of Assigned Semantic Frames

The annotators assigned 100 SFs from FrameNet to 341 communication verbs from VALLEX. The following SFs belong to the most often assigned:

- 'Statement' (141 cases in 2 annotations), e.g., *dodat*$^{pf}$ 'to add', *oznámit*$^{pf}$ 'to announce', *poznamenat*$^{pf}$ 'to remark', *sdělit*$^{pf}$ 'to tell', ... ,
- 'Request' (76), e.g., *nakázat*$^{pf}$ 'to order', *naléhat*$^{impf}$ 'to urge', *žádat*$^{impf}$ 'to plead', ... ,
- 'Telling' (59), e.g., *povědět*$^{pf}$ 'to tell', *říci*$^{pf}$ 'to say', ... ,
- 'Communication_manner' (35), e.g., *křičet*$^{impf}$ 'to shout', *šeptat*$^{impf}$ 'to whisper', *zamumlat*$^{impf}$ 'to gabble', ... ,
- 'Reporting' (34), e.g., *nahlásit*$^{pf}$ 'to inform', *udat*$^{pf}$ 'to report', ... ,
- 'Attempt_suasion' (31), e.g., *povzbudit*$^{pf}$ 'to encourage', *vybízet*$^{impf}$ 'to urge', ... .

**Ambiguous Assignment of Semantic Frames.** From the linguistic point of view, the cases when the annotators assigned two or more SFs to one Czech LU are the most interesting, as in the following scheme:

$$\text{Cz LU} \begin{cases} \nearrow \text{Eng LU}^{I} \rightarrow \text{SF}^{I} \\ \searrow \text{Eng LU}^{II} \rightarrow \text{SF}^{II} \end{cases}$$

The SFs which were systematically assigned ambiguously to Czech LUs refer to regular differences which FrameNet and VALLEX make in word sense disambiguation.

Describing valency frames, VALLEX leaves aside in/animateness of the entities occupying one valency position, in contrast to semantically based FrameNet. As a result, some Czech verbs represented by one LU in VALLEX belong to two (or even more) LUs in FrameNet.

For instance, the following instances of the verb *dokázat*$^{pf}$, *dokazovat*$^{impf}$ 'to prove' are described by one valency frame in VALLEX.

> *Peter has proved that the given solution was out-of-date.*
> *The facts have proved that the given solution was out-of-date.*

In FrameNet, the first instance corresponds to English LU$^{I}$, 'to prove' from SF$^{I}$ 'Reasoning', whereas the second instance belongs to English LU$^{II}$, 'to prove' from SF$^{II}$'Evidence'.

This fact concerns a number of other n-tuples of SFs, e.g., the SF 'Grant_permission' – 'Permitting', or 'Judgment_communication' – 'Judgment' – 'Notification_of_charges', etc.

In several cases, Czech verbs had more different translations belonging to the same SF, as in the following scheme.

$$\text{Cz LU} \nearrow \text{Eng LU}^{I} \searrow \text{SF}$$
$$\text{Cz LU} \searrow \text{Eng LU}^{II} \nearrow \text{SF}$$

In contrast to the above mentioned ambiguous assignment of SFs, we do not consider these cases interesting.

### 4.2 Analysis of Assigned Frame Elements

The annotators assigned 116 types of FEs to valency complementations. The most often assigned FEs are 'Speaker' (545 times in 2 annotations), 'Addressee' (485), 'Message' (393), 'Medium' (358), 'Topic' (330), 'Communicator' (92), 'Content' (81), etc.

Logically, the most often assigned FEs come from the most frequently assigned SFs. Moreover, some FEs are parts of more than one SF: 'Speaker' belongs to the SFs 'Statement', 'Telling', 'Request', 'Communication_manner', etc.

After having assigned FEs to the arguments of communication verbs, we can summarize which semantic properties are relevant for entities occupying their valency positions. The most significant are the following:

- 'Actor' matches esp. 'Speaker', 'Medium', 'Communicator', 'Voice', 'Author',
- 'Patient' matches esp. 'Topic', 'Message', 'Content',
- 'Addressee' matches esp. 'Authorities', 'Evaluee', 'Grantee'.

We can observe that both the animate (e.g., 'Speaker') and the inanimate (e.g., 'Voice') entities can occupy the positions of 'Actor' by communication verbs. Predominantly animate entities play the role of 'Addressee'. 'Patient' and 'Effect' are characterized by abstract semantic properties.

**Ambiguous Assignment of Frame Elements.** We observe two types of ambiguous assignment of FEs. The first type occurs if a particular Czech LU has only one English equivalent or it has more translations but all of them belong to one SF (see Section 4.1), as in the following scheme.

$$\begin{matrix} \text{VC}_1 \\ \text{VC}_2 \\ \text{VC}_3 \\ \cdots \end{matrix} \leftrightarrow \text{Cz LU} \begin{matrix} \nearrow \\ \\ \searrow \end{matrix} \begin{matrix} \text{Eng LU}^{I} \\ \\ \text{Eng LU}^{II} \end{matrix} \begin{matrix} \searrow \\ \\ \nearrow \end{matrix} \text{SF} \leftrightarrow \begin{matrix} \text{FE}_\alpha \\ \text{FE}_\beta \\ \text{FE}_\gamma \\ \cdots \end{matrix}$$

Then there may be more FEs from this SF that correspond to a particular valency complementation. We can illustrate this case with the verb *děkovat*$^{impf}$ translated as 'to thank', which belongs to the (only one) SF 'Judgment_direct_address' but has an ambiguous assignment of FEs.

| VCs | corresponding FEs |
|---|---|
| Actor | Communicator, Medium |
| Addressee | Addressee |
| Patient | Reason, Topic |

This case of the ambiguous assignment of FEs often results from the different approach to in/animateness which FrameNet and VALLEX have. As VALLEX does not take into account in/animateness of the first and second arguments, 'Actor' and 'Patient' are often assigned ambiguously in contrast to 'Addressee' and 'Effect'.

The ambiguous assignment of FEs to 'Patient' often follows from the fact that one abstract entity can express both 'theme' and 'what is said about the theme' by Czech communication verbs, as in the following sentence:

> *The news talked (about the horrible earthquake that struck Turkey on Friday morning).*Topic, Message).

Moreover, both 'Topic' and 'Message' can be realized separately in one Czech sentence, as in the following example, see also (Daneš et al., 1987), (Panevová, 1974):

> *Cizinci si stěžují starostovi na obchodníky.*Topic, *(že užívají dvojí ceny).*Message.
> '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.*

The second type of the ambiguous assignment of FEs is closely related to the ambiguous assignment

of SFs (see Section 4.1). If one Czech LU was assigned more than one SF, then the valency complementations of such Czech LU were assigned FEs from all these SFs. Therefore, the ambiguous assignment of FEs automatically arises from the ambiguous assignment of SFs.

For instance, one Czech LU *dokázat*$^{pf}$, *dokazovat*$^{impf}$ translated as 'to prove' corresponds to two SF$^{I}$ 'Reasoning' and SF$^{II}$ 'Evidence' in FrameNet (see Section 4.1) – as a result, the valency complementations of this LU are assigned FEs from both SFs. In these cases, all valency complementations can be assigned more than one FE:

| VCs | FEs from SF$^{I}$ | FEs from SF$^{II}$ |
|---|---|---|
| Actor | Arguer | Support |
| Addressee | Addressee | Cognizer |
| Patient | Content | Proposition |

## 5 Exploiting Semantic Information from FrameNet for VALLEX

In this section, we outline a further exploitation of the semantic information from FrameNet for VALLEX. We propose enhancing VALLEX with coarse-grained semantic information based on the semantic relation 'Inheritance'.

### 5.1 Structuring Semantic Classes in VALLEX.

FrameNet distinguishes several types of semantic relations, on the basis of which the semantic information is provided on different levels of granularity (see Section 2.2).

We consider the semantic relation 'Inheritance' as the most important. The child frame, although more specific than its parent frame, inherits all semantic properties from it. This concerns FEs, their relations to each other and frame-to-frame relations.

We exploit the semantic information from the top levels of the 'Inheritance' relations. This method allows classifying 'supergroups' of communication verbs in VALLEX into well-defined, coarse-grained classes from FrameNet as 'Communication', 'Prohibiting', 'Judgment_communication', etc., see also Figure 2.

We have obtained 59 top SFs for Czech communication verbs. (However, more than a half of the total number of these SFs have not been integrated into the network of the relation 'Inheritance' yet. We suppose continuously complementing the top levels of SFs in the future. Therefore, the final number of coarse-grained semantic classes is assumed to be significantly lower.)

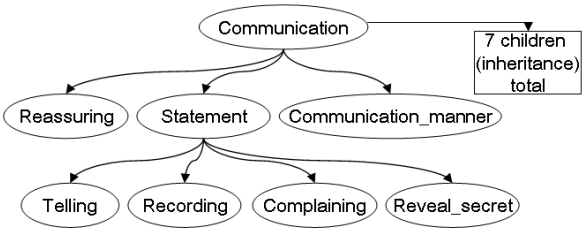


Figure 2: The relation of 'Inheritance' in FrameNet for the SF 'Communication'. (The SF 'Communication' represents the top SF in the relation 'Inheritance', e.g., for the SFs 'Reassuring', 'Communication_manner' and 'Statement', and transitively also for their children: 'Telling', 'Complaining', etc.)

### 5.2 Exploitation of Top Frame Elements as Semantic Roles.

Each FE from a child frame represents a subtype of the corresponding FE in its parent SF. Thus it allows us to assign FEs from the top SF in the relation 'Inheritance' to valency complementations of Czech communication verbs which were assigned SFs from the lower levels in this relation. Figure 3 shows the relevant relations between FEs of the SFs 'Communication', 'Statement' and 'Telling'.

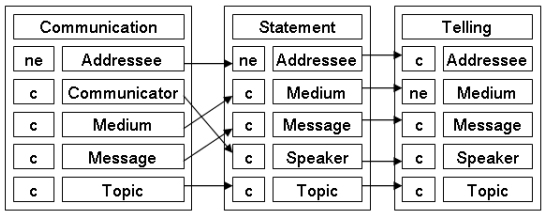


Figure 3: The FEs-to-FEs relations between the SFs 'Communication', 'Statement' and 'Telling' relevant for the valency complementations of Czech communication verbs assigned with these SFs.

Exploiting FEs as semantic roles from the top levels of 'Inheritance', we enrich the lexicon with more general, coarse-grained but extensive FEs. They provide VALLEX lexicon with sufficient informa-

tion on the selectional preferences of the individual arguments. The following example shows assignment of FEs to valency complementations of the verb *informovat*$^{biasp}$ 'to inform' from the SF 'Telling', a subtype of the SF 'Communication':

> *Učitel.*ACT-Communicator *informoval rodiče.*ADDR-Addressee, *(že jejich syn má špatné známky).*PAT-Message.
> Eng. *The teacher has informed the parents that their son has bad marks.*
>
> *Noviny*.ACT-Medium *informovaly čtenáře*.ADDR-Addressee, *(že ve věznici panují otřesné podmínky).*PAT-Message.
> Eng. *The newspapers have informed readers that outrageous conditions reign in the prison.*

Therefore, the enhanced valency frame for the corresponding LU for the verb *informovat*$^{biasp}$ 'to inform' has the following form:

| VCs | corresponding FEs |
|---|---|
| ACT | Communicator \| Medium |
| ADDR | Addressee |
| PAT | Topic |
| EFF | Message |

## 6 Conclusions and Future Work

We have presented an experiment in which VALLEX data were assigned semantic frames and frame elements from FrameNet. We attained a satisfactory inter-annotator agreement, especially concerning the FEs. We have proposed a method of enhancing VALLEX with the semantic information from FrameNet based on the relation 'Inheritance'. Focusing on communication verbs, we obtained applicable top level hierarchies.

For future work, we plan to assign the semantic information from FrameNet to other verbs with a sentential complement, namely to the verbs included in the VALLEX classes 'mental action', 'psych verb' and 'perception'. The significant inter-annotator agreement on assigning FEs promises good results for (semi)automatic assigning FEs as semantic roles.

**Acknowledgement.** The research reported in this paper is carried under the project of the Ministry of Education, Youth and Sports of Czech Republic No. MSM0021620838, under the grants LC536 and GA UK 7982/2007.

## References

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

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

D. Alan Cruse. 1986. *Lexical Semantics*. Cambridge University Press, Cambridge, UK.

František Daneš and Zdeněk Hlavsa et al. 1987. *Větné vzorce v češtině*. (2$^{nd}$ ed.) Academia, Praha.

Charles J. Fillmore. 1968. *The Case for Case*. In Bach, E., Harms R.T., eds. *Universals in Linguistic Theory*. Holt, Rinehart and Winston, 1–88.

Klaus Krippendorff. 1980. *Content Analysis: An Introduction to its Methodology*. Beverly Hills CA: Sage.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31(1):71–106.

Jarmila Panevová. 1974. On Verbal Frames in Functional Generative Description. *The Prague Bulletin of Mathematical Linguistics*, 22:3–40.

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

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

Zdeněk Žabokrtský. 2005. *Valency Lexicon of Czech Verbs*, Ph.D. thesis. Faculty of Mathematics and Physics, Charles University, Prague.

Zdeněk Žabokrtský and Markéta Lopatková. 2007. Valency Information in VALLEX 2.0. Logical Structure of the Lexicon. *The Prague Bulletin of Mathematical Linguistics*, 87:41–60.