Semantic Roles in Valency Lexicon of Czech Verbs: Verbs of Communication and Exchange*

Václava Kettnerová, Markéta Lopatková and Klára Hrstková

Charles University in Prague, Institute of Formal and Applied Linguistics
{kettnerova,lopatkova,hrstkova}@ufal.mff.cuni.cz

Abstract. We introduce a project to enhance a valency lexicon of Czech verbs with semantic roles. For this purpose, we make use of FrameNet. At the present stage, frame elements from FrameNet have been mapped to valency complementations of verbs of communication and verbs of exchange. The feasibility of this task has been proven by the achieved inter-annotator agreement – 95.6% for the verbs of communication and 91.2% for the verbs of exchange. As a result, we have obtained 37 semantic roles for the verbs of communication and 34 for the verbs of exchange, based on frame elements of upper level semantic frames from FrameNet.

1 Introduction

Semantic roles play a key role in NLP tasks in which semantic interpretation is necessary, as information extraction, question answering, or summarization [1]. In this paper, we report on labeling VALLEX valency complementations with more verb-specific semantic roles. For this purpose, we exploit frame elements from FrameNet.

As a first step, we experimented with two groups of verbs with divergent semantic and morphosyntactic properties, verbs of communication and verbs of exchange. First semantic frames from FrameNet were manually assigned to these verbs.[1] Then their valency complementations were linked with frame elements. Manual annotation is highly time consuming, however, it allows us to reach the desired quality.

2 Two Lexical Resources: VALLEX and FrameNet

In this section, we briefly characterize two lexical resources: VALLEX which takes into account mainly syntactic criteria and semantically oriented FrameNet.

* The research reported in this paper is carried under the project of the Ministry of Education, Youth and Sports No. MSM0021620838 (Objects of Research), under the grants LC536 (Center for Computational Linguistics II) and GA UK 7982/2007.

[1] This part of the experiment is described in [2].

Valency Lexicon of Czech verbs VALLEX. The Valency Lexicon of Czech Verbs, Version 2.5 (VALLEX 2.5)[2] provides information on the valency structure of Czech verbs in their particular senses: primarily, the number of valency complementations, their type (labeled with functors), and their possible morphological forms. VALLEX 2.5 describes 2,730 lexeme entries containing around 6,460 lexical units, 'senses' (LUs), see [3].

VALLEX 2.5 has a rather syntactic approach to valency, see [4]. Five functors are determined for verb arguments: 'Actor', 'Patient', 'Addressee', 'Effect', and 'Origin'. However, not having verb-specific meaning, this tight set does not reflect similarities and differences in verb meaning. E.g., the following verbs remain indistinct, despite being semantically different:

(1) *Petr.*ACT *prodal Pavlovi.*ADDR *motorku.*PAT
Eng. *Peter.*ACT *has sold Paul.*ADDR *the motorbike.*PAT
(2) *Učitel.*ACT *vysvětlil dětem.*ADDR *pravidla.*PAT *hry*
Eng. *The teacher.*ACT *has explained the rules.*PAT *of the game to the children.*ADDR

Thus introducing verb-specific semantic roles to VALLEX allows us to capture relations between semantically similar verbs. Moreover, it enables us to make inferences on lexical entailments of verbs.

FrameNet. FrameNet[3] is an on-line lexical resource for English. It documents semantic and syntactic properties of each word in each of its senses, see [5]. FrameNet covers more than 10,000 LUs, i.e., pairs consisting of a word and its meaning.

The descriptive framework of FrameNet is based on *frame semantics*. The *semantic frame* (SF) represents a schematic representation of a particular situation involving various participants, *frame elements* (FEs). These are defined for each SF separately.

FrameNet records frame-to-frame relation (including FEs-to-FEs relation) in the form of a hierarchical network. The relation of 'Inheritance', i.e., the hyperonymy / hyponymy relation, represents the most important one – the semantics of the parent frame corresponds equally or more specifically to the semantics of its child frames.

3 Mapping Frame Elements from FrameNet to Valency Complementations in VALLEX

As a first step, we translated each LU belonging to the verbs of communication and to the verbs of exchange from Czech to English.[4] The total number of translated Czech LUs was 341 for the verbs of communication and 129 for the verbs of exchange.

[2] `http://ufal.mff.cuni.cz/vallex/2.5/`
[3] `http://framenet.icsi.berkeley.edu/`
[4] The on-line dictionary at `http://www.lingea.cz/` was used.

Two human annotators were asked to indicate an appropriate SF for each given Czech LU. Then they assigned FE(s) of this SF to argument(s) of the given Czech LU. More than one FE could be assigned to a single argument ('Ambiguous annotation of FEs'). When no FE corresponded to a particular argument, the annotators concluded that the given FE was missing. For the overall statistics, see Table 1.

Table 1. Overall statistics on the annotations of FEs.

	annotator 1		annotator 2	
	Com	Exch	Com	Exch
Annotations of arguments from VALLEX	1088	479	1088	479
Unambiguous annotations of FEs	869	435	879	416
Ambiguous annotations of FEs	453	88	435	142
Marked as missing FEs	47	47	34	50

Inter-Annotator Agreement. Table 2 summarizes the inter-annotator agreement (IAA) and Cohen's κ statistics [6] on the total number of assigned FEs. IAA was measured only in the cases of match of SFs. Both the exact and intersection match of FEs (when both the annotators chose the same FEs regardless of other variants of ambiguous annotations) gave satisfactory results for both the verbs of communication (84.6% and 95.6%) and the verbs of exchange (85.4% and 91.2%). The κ statistics represents an evaluation metric that reflects average pairwise agreement corrected for chance agreement. The achieved levels represent significant results even in case of the exact match of FEs.

Table 2. Inter-annotator agreement and κ statistics.

	IAA [%]		κ	
	Com	Exch	Com	Exch
Exact match of FEs	84.6	85.4	0.83	0.84
Intersection match of FEs	95.6	91.2	0.95	0.91

4 Exploiting Frame Elements as Semantic Roles

1185 FEs (in which the annotators concurred) were mapped to 1088 arguments of the verbs of communication and 433 FEs were assigned to 479 arguments of the verbs of exchange.

As for **ambiguous assignment of FEs**, the annotators mapped more than one FE to a single argument especially due to a variety of lexical entailments

imposed by a verb on such an argument. E.g., two valency slots of the verb *zmínit se*^{pf} 'to mention' were assigned ambiguously – (i) the FEs 'Speaker' and 'Medium' were mapped to 'Actor' and (ii) 'Message' and 'Topic' to 'Patient':

(1) *Peter*^{Speaker} *did not mention (that he had moved away from her wife.)*^{Message}
(2) *This resolution*^{Medium} *mentions the problem*^{Topic} *of the refugee camp.*

Frame Elements as Semantic Roles. We enhanced VALLEX with semantic roles based on the FEs from the SFs from upper levels of the relation of 'Inheritance' – we made use of the ancestor FEs belonging to the SFs from the appropriate level of the relation of 'Inheritance', see [2].

Figure 1 illustrates the relation of 'Inheritance' between core FEs from the SFs 'Giving', 'Commerce_sell', and 'Renting_out'. We mapped the FEs 'Donor', 'Recipient', and 'Theme' from the ancestor SF 'Giving' to the appropriate arguments of the Czech LUs to which the descendant SFs 'Renting_out' and 'Commerce_sell' were assigned.

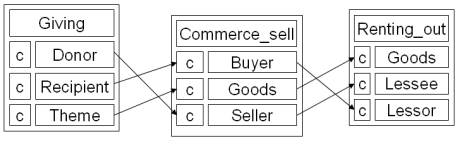

Fig. 1. The FEs-to-FEs relation of 'Inheritance' of the SFs 'Giving', 'Commerce_sell', and 'Renting_out'.

As a result, 37 core FEs from nine SFs – belonging to the upper levels of the relation of 'Inheritance', [2] – were applied as semantic roles to the arguments of the **verbs of communication**. (We introduce only core FEs as the most important ones.):

1. 'Communication': 'Communicator', 'Medium', 'Message', 'Topic'
2. 'Statement': 'Medium', 'Message', 'Speaker', 'Topic'
3. 'Communication_response': 'Addressee', 'Message', 'Speaker', 'Topic', 'Trigger'
4. 'Judgment_communication': 'Communicator', 'Evaluee', 'Expressor', 'Medium', 'Reason', 'Topic'
5. 'Chatting': 'Interlocutor_1', 'Interlocutor_2'
6. 'Prohibiting': 'Principle', 'State_of_affairs'
7. 'Request': 'Addressee', 'Medium', 'Message', 'Speaker', 'Topic'
8. 'Reporting': 'Authorities', 'Behavior', 'Informer', 'Wrongdoer'
9. 'Commitment': 'Addressee', 'Medium', 'Message', 'Speaker', 'Topic'

(Note that the FEs with the same name cannot be confused across different SFs.) The arguments of the **verbs of exchange** were labeled with 34 core FEs as semantic roles, arisen from ten SFs from the upper levels of the relation of 'Inheritance':

1. 'Giving': 'Donor', 'Recipient', 'Theme'
2. 'Getting': 'Recipient', 'Theme'
3. 'Replacing': 'Agent', 'New', 'Old'
4. 'Exchange': 'Exchanger_1', 'Exchanger_2', 'Theme_1', 'Theme_2'
5. 'Robbery': 'Perpetrator', 'Source', 'Victim'
6. 'Hiring': 'Employee', 'Employer', 'Field', 'Position', 'Task'
7. 'Transfer': 'Donor', 'Recipient', 'Theme', 'Transferors'
8. 'Frugality': 'Behavior', 'Resource', 'Resource_controller'
9. 'Taking': 'Agent', 'Source', 'Theme'
10. 'Supply': 'Purpose_of_recipient', 'Recipient', 'Supplier', 'Theme'

As a result, the FEs from the upper level SFs cover 95.4% of arguments of the verbs of exchange and almost 53% of arguments of the verbs of communication. The considerable difference is due to the low coverage of the relation of 'Inheritance' for the verbs of communication (only 68% of assigned SFs are connected by this relation for the time being). In the future, we plan to continuously increase the coverage following the progress made in FrameNet.

5 Conclusion

We have presented an experiment with enhancing the valency lexicon of Czech verbs, VALLEX 2.5, with semantic roles derived from FrameNet. As a first step, we mapped frame elements to arguments of the verbs of communication and the verbs of exchange. The attained inter-annotator agreement has proved the feasibility of this task. Then we labeled their arguments with semantic roles based on the frame elements from the upper level semantic frames of the relation of 'Inheritance' – 37 and 34 semantic roles are determined for the verbs of communication and the verbs of exchange, respectively. In the future, we plan to expand this experiment to other groups of verbs and we intend to exploit the obtained data for summarization of Czech texts.

References

1. Gildea, D., Jurafsky, D.: Automatic Labeling of Semantic Roles. Computational Linguistics **28** (2002) 245–288
2. Kettnerová, V., Lopatková, M., Hrstková, K.: Semantic Classes in Czech Valency Lexicon: Verbs of Communication and Verbs of Exchange. In: Proceedings of the 11th International Conference on Text, Speech and Dialogue. (in print)
3. Žabokrtský, Z., Lopatková, M.: Valency Information in VALLEX 2.0: Logical Structure of the Lexicon. The Prague Bulletin of Mathematical Linguistics 87 (2007) 41–60
4. Panevová, J.: Valency Frames and the Meaning of the Sentence. In Luelsdorff, P.L., ed.: The Prague School of Structural and Functional Linguistics. John Benjamins, Amsterdam-Philadelphia (1994) 223–243
5. Ruppenhofer, J., Ellsworth, M., Petruck, M.R.L., Johnson, C.R., Scheffczyk, J.: FrameNet II: Extended Theory and Practice. (2006) http://framenet.icsi.berkeley.edu/book/book.html.
6. Carletta, J.: Assessing agreement on classification tasks: The Kappa statistic. Computational Linguistics **22** (1996) 249–254