

Linking VerbaLex with FrameNet

Case Study for the Indicate Verb Class

Jiří Materna

Centre for Natural Language Processing
Faculty of Informatics, Masaryk University
Botanická 68a, 602 00, Brno, Czech Republic
xmaterna@fi.muni.cz <http://nlp.fi.muni.cz>

Abstract. The aim of this work is to evaluate possibilities of linking FrameNet frames with Czech verb valency frames from VerbaLex on the class of Indicate verbs. This class is taken from VerbaLex. The motivation comes from the intention to build a FrameNet-like database of semantic frames in Czech.

Key words: VerbaLex, FrameNet, Indicate verb class

1 Introduction

This work is a part of much more complex task, which aims to build a large, domain independent lexicon of semantic frames in Czech, based on the Frame Semantic formalism, similar to the original Berkeley FrameNet [1].

A fundamental assumption of the methodology of building Czech FrameNet is that the most of Berkeley FrameNet frames can be reused for the semantic analysis of Czech. The assumption takes advantage of the nature of frames as coarse-grained semantic classes, which refer to prototypical situations. Nevertheless, the assumption that these situations are applicable across languages should be empirically verified. In general, a sense of a lemma can evoke a FrameNet frame if this sense is able to realize the conceptually necessary components of the frame (its core frame elements). Inversely, the FrameNet frames cannot be applicable to other languages if the sub-categorization properties of lemmas in this language differ significantly from their English translations.

In this work we try to reveal the most significant problems by carrying out manual linkage of all VerbaLex frames from the Indicate verb class with Berkeley FrameNet.

2 Frame Semantics and FrameNet

Frame semantics is an approach to the study of lexical meaning based on the work by Charles J. Fillmore and his collaborators [2]. The central idea of the frame semantics is that word meaning is described in a relation to *semantic frame*,

which consists of a target *lexical unit* (pairing of a word with a sense), *frame elements* (its semantic arguments) and relations between them.

FrameNet is a project in which the information about the linked semantic and syntactic properties of English words is extracted from a large electronic text corpora, using both manual and automatic procedures. The information about words and their properties is stored in an electronic lexical database. Possible syntactic realizations of the semantic roles associated with a frame are exemplified in the annotated FrameNet corpus.

2.1 Semantic Frames

A semantic frame is defined as a script-like conceptual structure that describes a particular type of situation, object or event along with its participants and properties [3].

Lexical unit is a pairing of a word with a meaning. Typically, each sense of a polysemous word belongs to a different semantic frame. For example, the *Commerce_sell* frame describes a situation in which a *seller* sells some goods to a *buyer*, and is evoked by lexical units such as *auction*, *retail*, *retailer*, *sale*, *sell*, etc. The semantic participants are called Frame Elements.

2.2 Frame Elements

The semantic valencies of a lexical unit are expressed in terms of the kinds of entities that can participate in frames of the type evoked by the lexical unit. The valencies are called *frame elements*. Frame elements (FEs) bear some resemblance to the argument variables used in first-order predicate logic, but have important differences came from the fact that frames are much more complex than logical predicates [4]. In the example above, the frame elements include *Seller*, *Goods*, *Buyer*, etc.

FrameNet distinguishes three types of frame elements – *core* FEs (the presence of such FEs is necessary to satisfy a semantic valence of a given frame), *peripheral* FEs (they are not unique for a given frame and can usually occur in any frame, typically expressions of time, place, manner, purpose, attitude, etc.) and *extra-thematic* FEs (these FEs have no direct relation to the situation identified with the frame, but add new information, often showing how the event represented by one frame is a part of an event involving another frame).

3 VerbaLex

VerbaLex is an electronic database of verb valency frames in Czech, which has been developed in the Centre for Natural Language Processing at the Faculty of Informatics MU recently. Entries in VerbaLex are formed by lemmata in synonymic relations followed by their sense numbers in standard Princeton WordNet notation. Verb valencies are realized on two levels – deep valency

which corresponds to the semantic role (or selectional restrictions) and surface layer reflecting information about syntactic and morphological valencies.

The current version of VerbaLex contains more than 6,000 synsets, more than 21,000 verb senses and about 10,500 verb lemmata in 19,500 valency frames.

3.1 Complex Valency Frames

The valency frame represents verb valencies on both syntactic and semantic level. In the centre of the complex frame, there is a symbol representing the verb position, surrounded by the left-hand and right-hand arguments in the canonical word-order. The type of valency relation for each constituent element is marked as obligatory or optional. Semantic information about the verbal complement is represented by two-level semantic roles.

The first level contains semantic roles mainly based on the EuroWordNet [5] first-order and second-order top ontology entities, arranged in a hierarchical structure. The list of first level semantic entities is closed and currently contains 33 concepts. On the second level, selected lexical units from the set of EuroWordNet base concepts with relevant sense numbers are used. The list of second level semantic roles is open and currently contains about 1,200 literals.

The complex valency frame comprises basic valency frame and other additional information about verbs. The additional information includes:

- definition of verb meaning
- verb ability to create passive form
- number of meanings for homonymous verbs
- semantic class a verb belongs to
- aspect (perfective or imperfective)
- example of verb use
- types of reflexivity for reflexive verbs

4 Case study for the *Indicate* verb class

In order to discover main typological differences between Berkeley FrameNet frames and Czech valency frames in VerbaLex we have selected VerbaLex frames belonging to the *Indicate* class and carried out their complete linkage to FrameNet frames. The *Indicate* class is one of 111 semantic classes defined in VerbaLex and consists of 136 verb senses in 27 CZWN synsets evoking 119 valency frames.

4.1 Annotation Process

The annotator proceeds one VerbaLex frame at a time and is asked to assign at most one FrameNet frame to it. If the annotator is not able to find appropriate FrameNet frame, the VerbaLex frame will not be annotated and a new Czech FrameNet frame should be defined in future work. There are at least two reasons of the necessity to define a completely new frame [6]:

1. Inadequacy of frame definitions in the corresponding semantic domain or area.
2. Insufficient coverage of the domain in Berkeley FrameNet (i.e. English lexical units and corresponding frames have not been defined yet).

If it is possible to find an appropriate frame from FrameNet, the process of annotation continues with linking semantic roles from VerbaLex with frame elements from FrameNet. At most one frame element can be chosen for a semantic role and no more than one semantic role can be linked to a FrameNet frame element.

If the appropriate FrameNet frame element for a semantic role from VerbaLex frame does not exist, the semantic role should be connected to a newly-defined frame element in future. Nevertheless, the VerbaLex specific addition of frame elements to a FrameNet frame results in a different and more restricted frame. This more specific non-English frame could be related to the English one by a cross-lingual Inheritance relation, whereby it would become a cross-lingual Child frame of the English frame [6].

4.2 Statistics and Typological Divergences

In our experience, the most of VerbaLex frames can be directly linked to a semantic frame from FrameNet, nevertheless, some of the VerbaLex frames require an adaptation or creating a completely new FrameNet frames. In Table 1 there is a list of the most frequently assigned FrameNet frames with numbers of corresponding VerbaLex frames.

Table 1. Assigned FrameNet frames.

FrameNet frame name	VerbaLex frames
Telling	30
Reasoning	22
Reveal_secret	14
Gesture	11
Expressing_publicly	7
Forgiveness	5
Communication	4
Sign	3
Others	9
None	9

During the annotation phase we have identified 15 FrameNet frames belonging to 119 VerbaLex frames. It means that approximately 8 VerbaLex frames are linked to one FrameNet frame. Among all 119 VerbaLex frames there are 9 frames that cannot be linked to any known frame from FrameNet. For these frames a new FrameNet frame will need to be defined. The coverage of the Indicate class by FrameNet frames is more than 99%.

When evaluating linkage between semantic roles from VerbaLex and frame elements from FrameNet, we have found 19 frames where a new frame element has to be added, two or more frame elements have to be put together, or some frame elements have to be restructuralized. It has at least three reasons:

1. The corresponding frame element is missing.
2. The frame element is too general and has to be divided into more specific ones.
3. The frame element is too specific and has to be replaced by a more general one.

An example of the missing frame element case can be a sentence

Jeho pohled nám naznačil, že nemluví pravdu.
(His look signed us that he is not telling the truth.),

which corresponds to VerbaLex frame

AG(co1;<quality:1>) VERB PAT(komu3;<person:1>) INFO(const;<info:1>)

and FrameNet Frame Sign. This FrameNet frame allows frame elements Indicated (INFO), Indicator (AG) and Degree but does not allow any FE, which could be connected to the patient (PAT).

An example of case 2, where a frame element has to be divided can be a sentence

Ta zpěvačka demonstrovala svou lásku ke zvířatům.
(The singer demonstrated her love of animals.),

which corresponds to VerbaLex frame

AG(kdo1;<person:1>) VERB ACT(co4;<act:2>) PAT(komu3;<animal:1>)

and FrameNet frame Expressing_publicly. This frame allows core frame elements Communicator (AG), Content and Medium. In order to connect this FrameNet frame to the VerbaLex frame, we would have to split Content FE into two parts or join ACT and PAT arguments of the VerbaLex frame together.

The last example illustrates case 3, where a frame element is too specific and should be replace by a more general one

Ten pes cenil zuby na kočku za stromem.
(The dog showed the teeth to the cat behind the tree.),

which corresponds to VerbaLex frame

AG<animal:1> VERB DPHR<zuby> PAT <person:1|animal:1>

and FrameNet frame *Gesture*. This frame allows core frame elements *Addressee* (PAT?), *Body_part* (DPHR), *Communicator* (AG), *Indicated_entity* and *Message*. The problem is in the definition of frame element *Addressee*, which says 'This is the person to whom a non-verbal Message is communicated', therefore, an animal is not allowed.

5 Conclusions

The presented work describes an evaluation of linking possibilities between Czech verb valency lexicon *VerbaLex* and FrameNet on the domain of *Indicate* verb class. The class consists of 136 verb senses in 27 CZWN synsets evoking 119 valency frames. The results showed that the coverage of the *Indicate* class by the FrameNet frames is more than 99% and more than 82% of linkable FrameNet frames can remain without any modifications of their frame elements.

For the future work, the goal is to build a core of Czech FrameNet based on a complete linkage of *VerbaLex* to FrameNet. Such FrameNet based lexicon can be used for information retrieval and searching semantic relations in texts. Also other challenging tasks come into consideration, namely in the area of the Semantic Web.

Acknowledgements

This work has been partly supported by the Ministry of Education of CR within the Center of basic research LC536 and in the National Research Programme II project 2C06009.

References

1. Ruppenhofer, J., Ellsworth, M., Petruck, M.R.L., Johnson, C.R., Scheffczyk, J.: *FrameNet II: Extended Theory and Practice* (2006) <http://www.icsi.berkeley.edu/framenet>.
2. Fillmore, C.J.: *Frame Semantics and the Nature of Language*. In: *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech*. Volume 280. (1976) 20–32.
3. Fillmore, C.J.: *Frame Semantics*. In: *Linguistics in the Morning Calm*, Seoul, Hanshin Publishing Co. (1982) 111–137.
4. Fillmore, C.J., Johnson, C.R., Petruck, M.R.L.: *Background to FrameNet*. In: *International Journal of Lexicography*, Oxford University Press (2003) 235–250.
5. Vossen, P., Hirst, G.: *EuroWordNet: A Multilingual Database with Lexical Semantic*. Kluwer Academic Publisher (1998).
6. Lönneker-Rodman, B.: *Multilinguality and FrameNet*. Technical report, International Computer Science Institute, Berkeley (2007).