Valency as a Mediator between Grammar and Lexicon: Integrated Annotation of Verb Valencies in Bulgarian through Knowledge Transfer from English Resources (WG1 and WG2)

Petya Osenova and Kiril Simov

Institute of Information and Communication Technologies, Bulgarian Academy of Sciences

1. Overview

- Our ultimate goal is to integrate the available to us language resources for Bulgarian
    - on monolingual level between different types of resources
    - on multilingual level between resources from the same type
- A key prerequisite for providing more complex language technologies and developing more complex linguistic research on multilevel and multilanguage directions
- The main benefit from such an integration effort is the verification of distinct resources, simultaneous usage of complementary knowledge, and transfer of knowledge
- Here we report on our first attempts at integrating:
    - a treebank (providing syntactic structure)
    - a valency dictionary (providing syntagmatic potential)
    - BTB-Wordnet (providing lexical senses)

with a valency dictionary of English (VerbNet) as well as with other English resources related to it

2. Probes towards integration

- In this work we would like to discuss an integrated annotation of verb valencies in Bulgarian:
    - with the usage of Pustejovsky's ideas on argument structure
    - WordNet lexicographic classes
    - Verb senses from BTB-Wordnet and
    - Valency frames from VerbNet
- The frequencies of the roles in one file with 688 extracted roles (the number of all valency frames is around 4000. The lexicographic classes with most frequent Agent role are verb communication (62)
- and verb.social (37)

Role	Freq	Role	Freq
Agent	243	Source	10
Theme	211	Co-Agent	8
Destination	34	Pivot	8
Patient	29	Recipient	8
Attribute	20	Stimulus	8
Topic	20	Result	7
Beneficiary	17	Co-Patient	6
Experiencer	16	Asset	2
Location	16	Cause	1
Co-Theme	11	Trajectory	1
		Product	1

3. The approach

- Based on previous research we got evidence that Bulgarian prefers default arguments in comparison to true arguments and arguments in shadow.
- Thus, our representation of the Valency frames equals the HPSG representation of the argument structure (ARG-ST) which covers all the possible argument realizations in texts
- The lexicographic classes were transferred to the verbs in the respective meanings through BTB-Wordnet
- The adaptation went into several directions such as: the number and names of the roles, the verb grammatical behavior, the treatment of metaphorical usages, etc

4. Example for the verb CELEBRATE
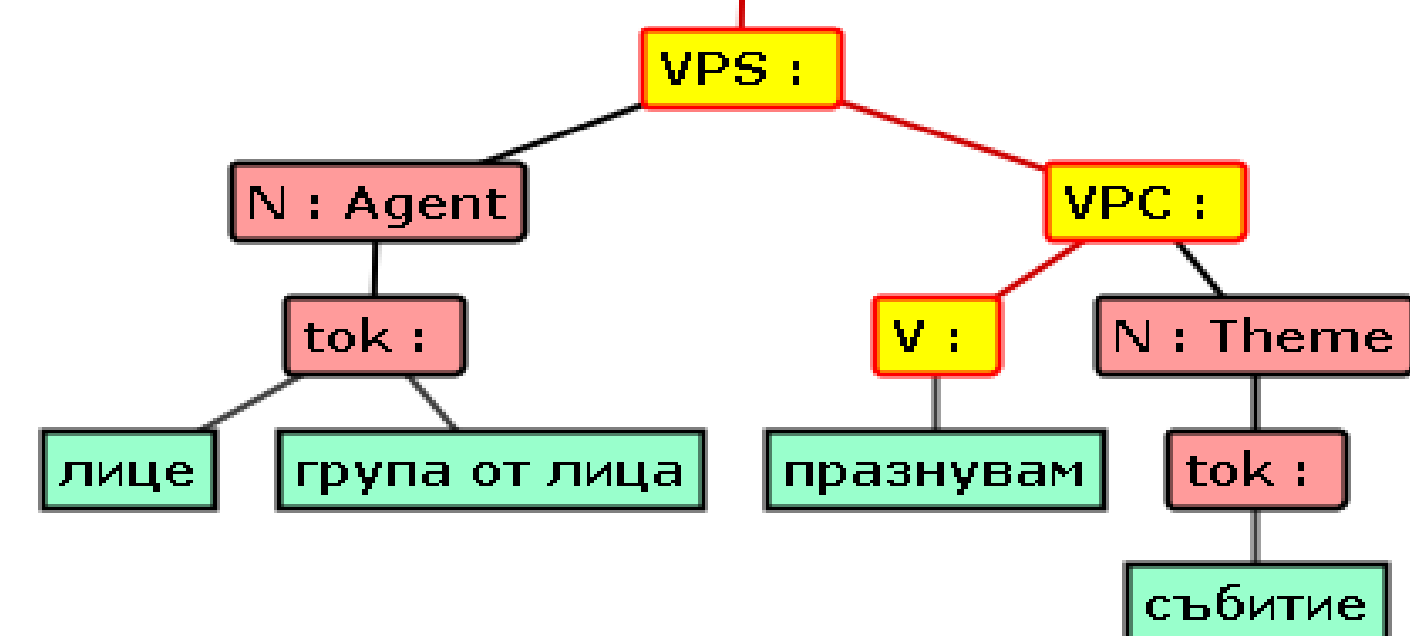

```
Litse praznuvam sabitie
Person celebrate-1SG event
A person celebrates an event
```

5. Modeling MWEs

1. Igraya na kotka i mishka

Play-1SG on cat and mouse

I play cat and mouse

Player = *Agent* and both pseudocomplements = coordinated *Theme*.

The lexicographic category of the construction is *verb.creation*

2. Podlagam nyakogo na stres

Make-1SG someone on stress

I am stressing someone out

Stressor = Stimulus

Stressed = Experiencer

Stress = Theme

The lexicographic category of the construction is *verb.emotion*

6. Conclusions

The main challenges in mapping resources are:
- diversity in valencies within synsets (due to usage of a different preposition or no preposition)
- the idiosyncracy of MWEs in wordnets and valency dictionaries
- reflexive verbs
- various aspectual nuances of verbs
- missing meaning and/or frame
- differing behaviour of the verb in the two languages
- blurred boundaries among some semantic roles
- ellipses in the examples, etc.