EUROPEAN UNION
EUROPEAN REGIONAL DEVELOPMENT FUND

University of Latvia
Institute of Mathematics and Computer Science

Towards a Latvian Valency Lexicon

Gunta Nešpore
Baiba Saulīte
Normunds Grūzītis
Ginta Garkāje

5th International Conference on Human Language Technologies – the Baltic Perspective, Tartu, October 4-5, 2012

Introduction

- Valency: the behaviour (realization) of a word in a text
  - What are the typical valences of a verb?
- A crucial resource in NLP
  - Helps to reduce ambiguities in syntactic parsing
  - Vital for semantic parsing (semantic role labelling)
    - e.g. in information extraction
- There is no valency dictionary for Latvian
- We do not propose a new approach, but a new resource
  - For Latvian, we combine various approaches

Valency corpus

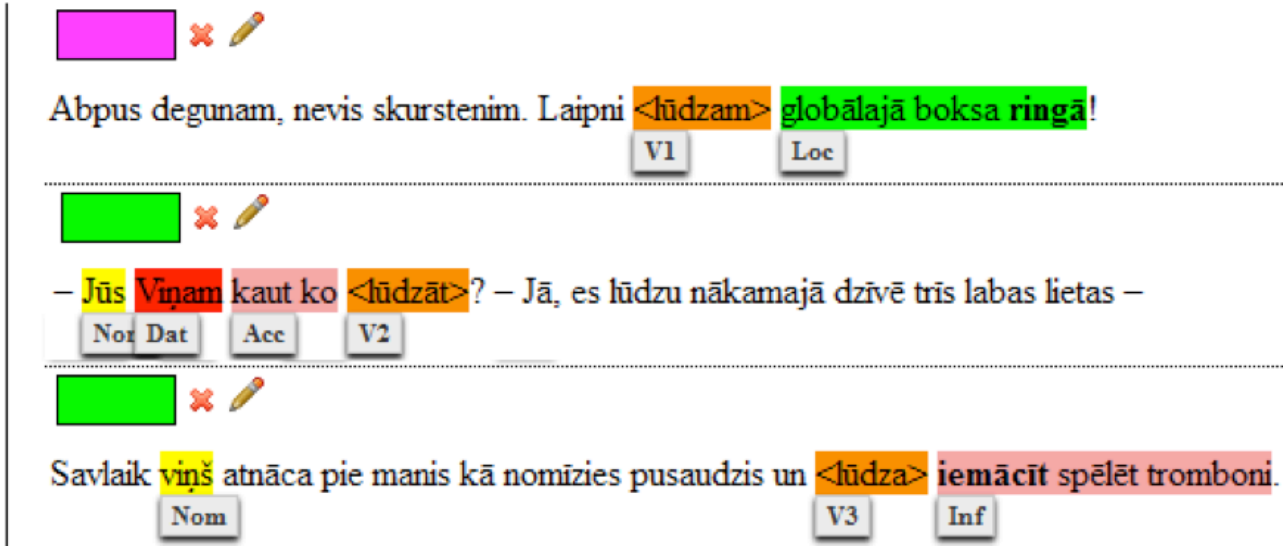
- To be annotated:
  - sets of concordances for ~250 most frequent verbs from the Balanced Corpus of Modern Latvian (www.korpuss.lv)
  - ~70 verbs expressing motion
- Progress: ~170 verbs/~22 000 corpus examples
- Finite forms (indicative mood, active voice) of verbs only

viņš(Nom) <lasa>(V3) žurnālus(Acc)
viņam(Dat) <jālasa>(Deb) žurnāli(Nom)

Valency layers

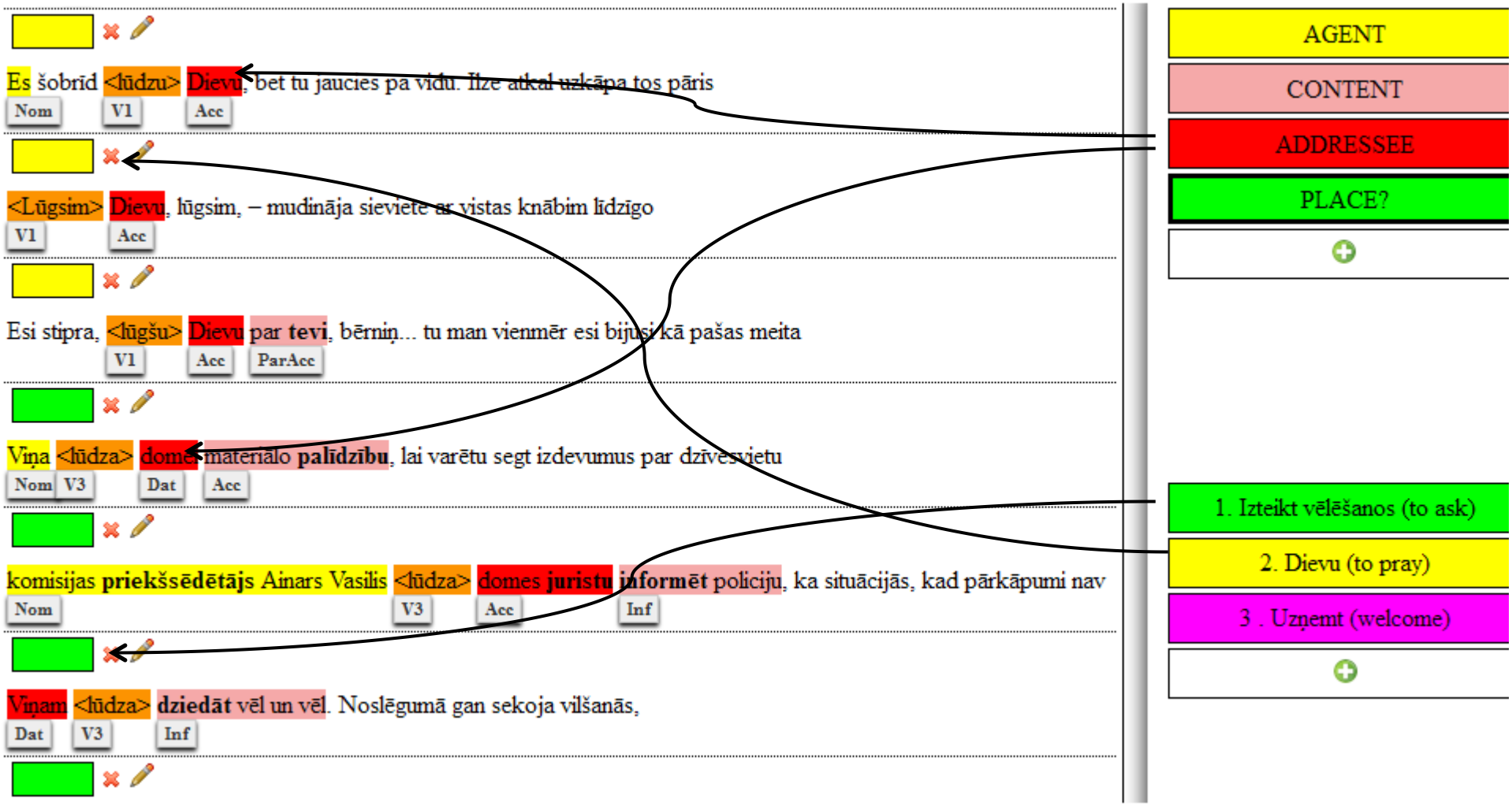
- **Morpho-syntactic valency** – the morphological description of the dependents of the target verbs
- **Semantic valency** – semantic roles
- Lexical valency (selectional restrictions) – future work

Morpho-syntactic valency



- The annotation of morphological forms that realize the semantic roles in the environment of the verb
- Indirectly reveals the syntactic function of the dependent:
  - a noun in nominative—the grammatical subject
  - a noun in accusative—the grammatical object
  - a noun in locative—an adverbial modifier

Semantic valency



Institute of Mathematics and Computer Science
University of Latvia

Mākslīgā intelekta laboratorija
LU MII

The set of semantic roles

- Granularity—fine-grained or coarse-grained semantic roles?
- **FrameNet**: a very fine-grained set of the semantic roles (*frame elements*) with specific titles (e.g. Arguer1, Arguer2, Arguers, Issue in the frame Quarrelling)
- **Vallex, VerbNet, Sližienė**: 20–35 general semantic roles (AGENT, PATIENT, EXPERIENCER etc.), assuming that the functions of the participants of the situation (semantic roles) may be similar in sentences with different verbs.
- **LV lexicon**: the general semantic roles:
  - the possibility to compare, group and analyze the verbal environment (valency patterns); to characterize each semantic role

Selectional restrictions/lexical valency

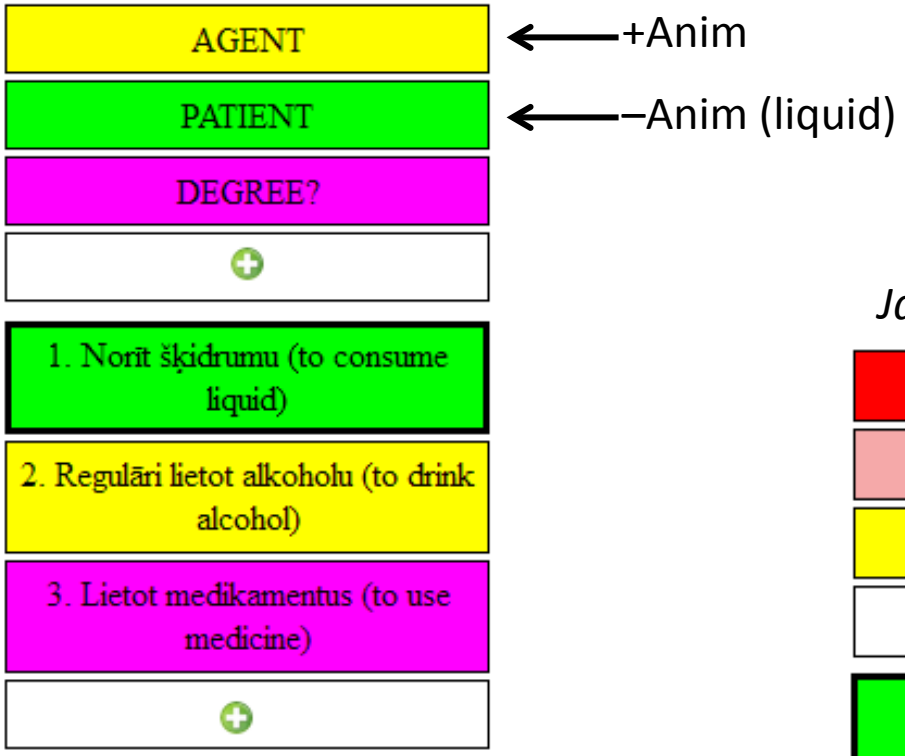
- To assign the general lexical meaning (semantic types) of the semantic roles—solely empirically from the corpus data:
  - the verb *dzert* 'to drink' requires an animate AGENT and an inanimate (liquid) object as PATIENT
  - AGENT and ADDRESSEE of the verb *jautāt* 'to ask' typically are humans
- How fine-grained or coarse-grained semantic types?
  - specific enough to facilitate the word sense disambiguation
  - general enough to avoid over-controlled usage of the verb
    - What is AGENT of *domāt* 'to think'?
      Human only? Or any animate object? Maybe even devices, e.g. computers in a figurative sense?

Selectional restrictions – LV

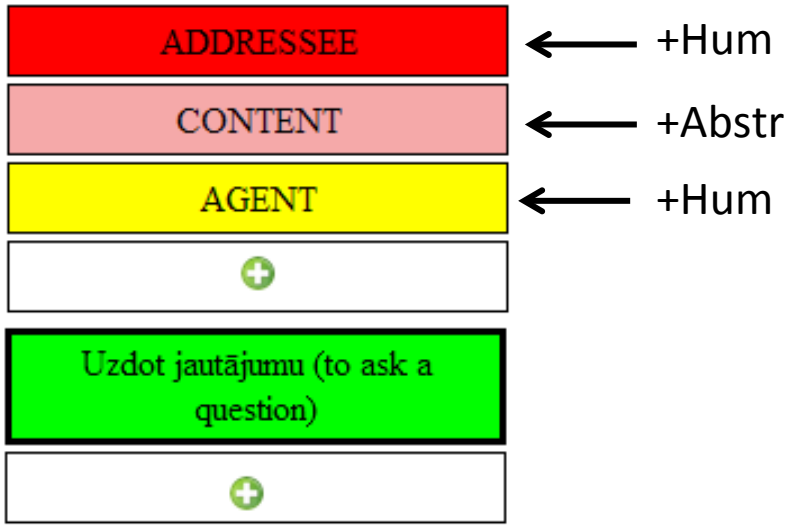
- A set of few general ontological categories: **animate object, inanimate object, abstract object, human**
- More specific information given when required by meaning of the verb:
  - e.g. the grammatical object of the verb *uzģērbt* 'to put on / to wear' is an inanimate object, specifically clothing
- Result: the set of categories obtained empirically by annotating the language data instead of choosing ontological categories from a predefined set

Selectional restrictions: LV

Dzert ‘to drink’



Jautāt ‘to ask’



Valency patterns

- Automatically generated valency patterns:
  - **combinations of semantic roles**
  - the mapping between the morpho-syntactic categories and semantic roles
  - **frequency information**
  - in future: selectional restrictions

Lūgt 'to ask/to pray'

1. Izteikt vēlēšanos (to ask) 325

AGENT ADDRESSEE verbs 5

(AGENT) verbs ADDRESSEE CONTENT 112

(AGENT) verbs CONTENT 199

AGENT verbs 9

2. Dievu (to pray) 20

(AGENT) verbs ADDRESSEE CONTENT 10

(AGENT) verbs ADDRESSEE 6

verbs CONTENT AGENT 3

verbs 1

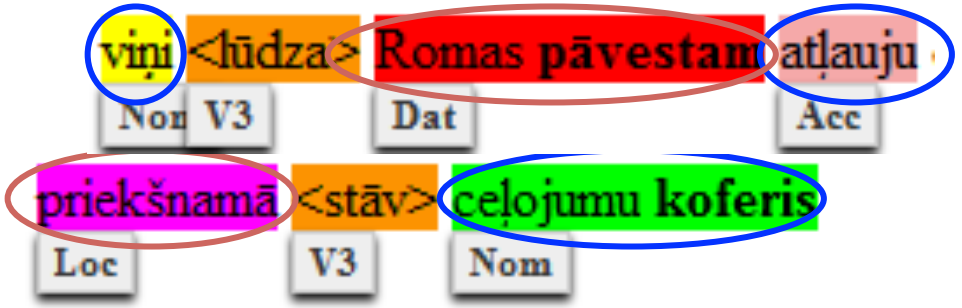
Valency patterns

- Automatically generated valency patterns:
  - combinations of semantic roles
  - **the mapping between the morpho-syntactic categories and semantic roles**
  - **frequency information**
  - in future: selectional restrictions

Kategorija	Skaits	Grupas	
AGENT	248	**Grupa**	**Skaits**
		Nom	248
CONTENT	324	**Grupa**	**Skaits**
		Adv	1
		Inf	187
		S	52
		Acc	73
		ParAcc	8
		Loc	3
ADDRESSEE	133	**Grupa**	**Skaits**
		Dat	54
		Acc	79
PLACE?	1	**Grupa**	**Skaits**
		Loc	1
Verbs	346	**Grupa**	**Skaits**
		V3	226
		V1	106
		V2	14

Contents of valency patterns

- Valency patterns are built from the **core semantic roles** [FrameNet]
- **Core semantic roles** roughly correspond to the semantically obligatory valent elements:
  - arguments of predicate: obligatory (subject, direct object), optional (e.g. indirect object)
  - obligatory free modifiers



- **non-core semantic roles**
  - can appear in any semantically appropriate situation, e.g. TIME, MANNER (free modifiers)
  - they are not included in valency patterns; specifically marked (if annotated), displayed only in the annotated corpus

Conclusion

- As a result, the following data will be available:
  - valency patterns of the annotated verbs—with semantic roles as central elements
  - mappings between semantic roles and their morpho-syntacic patterns as well as selectional restrictions for each semantic role (for each verb/sense)
  - a corpus of manually annotated sentences
- Draft results: http://tezaurs.lv/valence/preview

Future work

- FrameNet:
  - A robust (frame-)semantic interlingua
  - (Being) implemented in several languages (English, German, Swedish, Italian, Spanish, Bulgarian, etc.)
- Mapping Latvian valency patterns to FrameNet frames and their core elements
  - Reuse of semantics defined by FrameNet
  - (Implicit) valency mapping among languages

Thank you!

Institute of Mathematics and Computer Science
University of Latvia

Mākslīgā intelekta laboratorija
LU MII