

Sofia Update

Petya Osenova
IICT-BAS

DELPH-IN meeting, 14-18.07.2014, Tomar,
Portugal

Main activities

- Lexicon Enrichment
- Mapping of current verb types in Bulgarian Grammar to MultiVal model

Concerning enrichment

The *Good News*:

- Verbs are in the lexicon for real! 😊 But - need curation.
- Adjectives are in the lexicon for real!
- Nouns are prepared to enter

The somewhat *Bad News*:

- Yes, I keep the ‘ugly’ way of doing the morphology in order to reuse as much as possible from our existing resources, i.e. I put the load on the inflectional rules instead of the introduction of types.

Lexicon Verb Enrichment: Data

- 18 000 verbs (22 000 frames – 6000 from valency lexicon)
- 18 existing frame types and 28 new ones (still not added, but identified)
- Current grammar types cover all verb paradigms with the exception of 225 verbs (like *take*, *enter*, *live*, *wish*, *think*, *choose*, etc.)
- The mapping was done in two ways:
 - Integration from valency dictionary (6000 frames)
 - Addition from Morphological Dictionary (the remaining verbs)

Lexicon Enrichment: the ‘kitchen’ things

- Some frames from the valency dictionary need further refinement (for example, the semantic typology of complement PPs – dative, locative, etc.).
- Some frames, generated from the Morphological dictionary, also need checks and refinements (for example, the intransitive and transitive verbs are generated straightforwardly, but ditransitives and PPs should be checked)

The verb: Abdicate (v_-_p_le) is OK as it is when added!

```
abdikiram_v1 := v_-_p_le &¶
[ STEM < "абдикирам" >,¶
  SYNSEM [LKEYS.KEYREL.PRED "абдикирам_v_1_rel",¶
    LOCAL.CAT.HEAD.MCLASS [FIN-PRESENT finite-present-017,¶
      FIN-AORIST finite-aorist-072,¶
      PART-AORIST participle-aorist-059,¶
      FIN-IMPERF finite-imperf-020,¶
      PART-IMPERF participle-imperf-021,¶
      PART-PRESENT participle-present-010,¶
      GERUND gerund-m-jki,¶
      FIN-IMPERATIVE imperative-017]¶
    ]¶
  ].
```

The verb Wish (v_np-pp_i_le) needs specification for the semantic type of the preposition in PP

```
zhelaya_v3 := v_np-pp_i_le &
[ STEM < "желая" >,
  SYNSEM [LKEYS.KEYREL.PRED "желая_v_3_rel",
    LOCAL.CAT.HEAD.MCLASS [FIN-PRESENT finite-present-002,
      FIN-AORIST finite-aorist-012,
      PART-AORIST participle-aorist-020,
      FIN-IMPERF finite-imperf-006,
      PART-IMPERF participle-imperf-006,
      PART-PRESENT participle-present-005,
      GERUND gerund-ya-ejki,
      PART-PASSIVE participle-passive-011,
      FIN-IMPERATIVE imperative-003]
    ]
  ]
```

Adjectives

- 14 524 entries for adjectives
- 333 inflection rules

What to curate?

- Valency (now all are intransitives) to be added
- The participles within adjective groups to be deleted

(Nouns): to be added

- 12 000 feminine nouns
- 11 380 masculine nouns
- 13 480 neuter nouns
- 266 pluralia tantum nouns

- Operated by 425 inflectional rules.

Mapping the existing verb types to MultiVal schema: this is step 1 towards the full integration

v_np_i_le¶

SAS: "NP"¶

FCT: transitive¶

SIT: imperfective¶

Example: ¶

знаю2 (to know), искаю1 (to want), читаю (to read)¶

¶

v_dali_i_le¶

SAS: "Squest[dali]"¶

FCT: transitive¶

SIT: imperfective¶

Example: ¶

знаю3 (to know)¶

¶

Other related activities

- Ongoing work on projecting (R)MRSEs from Dependencies
- Thus, improving the dependency parsers for Bulgarian (a paper accepted at SPMRL-SANCL Workshop at COLING 2014 - <http://www.spmrl.org/spmrl-sancl2014.html>)