

Representing Nominal Inflectional Morphology for Slavonic Languages in DATR

Velislava Stoykova

Institute of Bulgarian Language – BAS

Bulgarian Academy of Sciences

vili1@bas.bg

Introduction

- Presenting inflectional morphology is a key feature for a formal interpretation of any Slavonic language.
- Slavonic languages have had a long parallel historical development and as a result they share similar grammar features at the level of phonetics, morphology, and syntax. Thus, their formal interpretation can be presented using common logical frameworks.

Linguistic and computational approaches to inflectional morphology

- The traditional interpretation of inflectional morphology given at descriptive academic grammar works is a presentation of tables.
- Formal representations offer logical frameworks which allow computationally tractable encoding preceded by a related semantic analysis and suggest a subsequent architecture. Thus, representing inflectional morphology is, in fact, the representation a specific type of grammar knowledge.

Linguistic and computational approaches (cont.)

- The contemporary linguistic theories offer different approaches to formal presentation of a word segmentation and classification.
 - (i) The Word and Paradigm (WP) approach uses a paradigm as a central notion and a high-level constraint for a word segmentation.
 - (ii) The Item and Agreement (IA) uses subword units (morphotactics) and morphosyntactic units as a central notion and a high-level constraint for a word segmentation.

Linguistic and computational approaches (cont.)

pref 1	pref 2	base	suff 1	suff 2	ending
---------------	---------------	-------------	---------------	---------------	---------------

The standard computational approach to both derivational and inflectional morphology is to represent words as a rule-based concatenation of morphemes, and the main task is to construct relevant rules for their combinations. With respect to the number and types of morphemes, the different theories offer different approaches depending on variations of either stems or suffixes as follows:

- (i) Conjugational solution offers invariant stem and variant suffixes, and
- (ii) Variant stem solution offers variant stems and invariant suffix.

Linguistic and computational approaches (cont.)

	suff 1
stem	suff 2
	suff 3

stem 1	
stem 2	suff
stem 3	

Both these approaches are suitable for languages, which use inflection rarely to express syntactic structures, whereas for those using rich inflection some cases where phonological alternations appear both in stem and in concatenating morpheme a "mixed" approach is used to account for the complexity.

Also, some complicated cases where both prefixes and suffixes have to be processed require such approach.

Linguistic and computational approaches (cont.)

- We evaluate the "mixed" approach as a most appropriate for the task because it considers both stems and suffixes as variables and, also, can account for the specific phonetic alternations.
- The additional requirement is that during the process of the inflection all generated inflected rules (both using prefixes and suffixes) have to produce more than one type of inflected forms.
- We evaluate the DATR language for lexical knowledge presentation as a suitable formal framework for analyzing and presenting Slavonic nominal inflectional morphology.

The DATR Language

- The DATR language is a non-monotonic language for defining the inheritance networks through path/value equations.
- It has both an explicit declarative semantics and an explicit theory of inference allowing efficient implementation, and at the same time, it has the necessary expressive power to encode the lexical entries presupposed by the work in the unification grammar tradition.
- In DATR, information is organized as a network of nodes, where a node is a collection of related information.

The DATR Language

- Each node has associated with it a set of equations that define partial functions from paths to values where paths and values are both sequences of atoms.
- Atoms in paths are sometimes referred to as attributes.
- DATR is functional, it defines a mapping which assigns unique values to node attribute-path pair, and the recovery of this values is deterministic.
- DATR allows construction of various types of language models (language theories), however, our model is presented as a rule-based formal grammar and a lexical database, and the query to be evaluated is a related inflected word form.

Russian nominal inflectional morphology in DATR

- The ideas used for Russian nominal inflection interpretation offered by Corbett and Fraser underlay that of a paradigm and the encoding presents resolving of a tabular conceptualization encoding task.
- Network Morphology is a framework for describing inflection which offers a formally explicit account of lexical entries, declensional classes, word classes, and the relationships between them by giving a set of universal constraining principles of morphology.
- It is linguistically motivated. In particular, the underlying basic idea of the analysis is to reconsider the Russian declensional classes described in the Zaliznjak's dictionary, however, the approach adopted has implications well beyond the Russian.

Russian nominal inflection (cont)

- The interpretation uses declensional classes, i.e. the Word and the Paradigm framework and the features of case, number, and animacy as a starting point of the formal analysis, which is of theoretical value since it presents four declensional classes instead of three, presented traditionally.
- It consists of a formal grammar (inflectional rules) and a lexical database (nouns of all declensional classes) and the queries to be evaluated are all inflected word forms.

Russian nominal inflection (cont)

- Further, we are going to analyze the fragment of encoding presenting the Russian nouns inflection for the features of case and number.
- NOMINAL:
- `<stem>` == "`<infl_root>`"
- `<phon stem hardness>` == hard
- `<mor stem hardness>` == "`<phon stem hardness>`"
- `<acc>` == "`<mor nom>`"
- `<acc pl animate>` == "`<mor gen pl>`"
- `<acc sg animate masc>` == "`<mor gen sg>`"
- `<mor acc $number>` == `<acc $number "<syn animacy>" "<syn gender>">`
- `<mor dat pl>` == "`<stem pl>`" "`<mor theme_vowel>`" `_m`
- `<mor inst pl>` == "`<stem pl>`" "`<mor theme_vowel>`" `_m'i`
- `<mor loc pl>` == "`<stem pl>`" "`<mor theme_vowel>`" `_x`.

Russian nominal inflection (cont)

- The node GENDER is introduced to differentiate between different types of gender assignment (including the semantic gender defined as 'formal').
- GENDER:
 - <male> == masc
 - <female> == fem
 - <undifferentiated> == "<formal gender>".

Russian nominal inflection (cont)

- The basic node which defines the general rules of nouns inflection is the node NOUN. It inherits the grammar rules of node NOMINAL but also defines new inflectional rules.
- NOUN:
- $\langle \rangle == \text{NOMINAL}$
- $\langle \text{mor loc sg} \rangle == \langle \text{stem sg} \rangle _e$
- $\langle \text{mor nom pl} \rangle == \langle \text{stem pl} \rangle _i$
- $\langle \text{mor gen pl} \rangle == \langle \langle \text{mor stem hardness} \rangle \text{ mor gen pl} \rangle$
- $\langle \text{soft mor gen pl} \rangle == \langle \text{stem pl} \rangle _ej$
- $\langle \text{mor theme_vowel} \rangle == _a$
- $\langle \text{syn cat} \rangle == n$
- $\langle \text{syn animacy} \rangle == \langle \text{sem animacy} \rangle$
- $\langle \text{syn gender} \rangle == \text{GENDER: } \langle \langle \text{sem sex} \rangle \rangle$
- $\langle \text{sem sex} \rangle == \text{undifferentiated.}$

Russian nominal inflection (cont)

- Node N O defines nouns which are assigned to declensional types I and IV and it inherits all grammar rules from node NOMINAL but introduces new inflectional rules.
- N_0:
- $\langle \rangle == \text{NOUN}$
- $\langle \text{mor gen sg} \rangle == \langle \text{stem sg} \rangle _a$
- $\langle \text{mor dat sg} \rangle == \langle \text{stem sg} \rangle _u$
- $\langle \text{mor inst sg} \rangle == \langle \text{stem sg} \rangle _om.$
- Node N I defines nouns which belong to I declension.
- N_I:
- $\langle \rangle == \text{N } 0$
- $\langle \text{formal gender} \rangle == \text{masc}$
- $\langle \text{mor nom sg} \rangle == \langle \text{stem sg} \rangle$
- $\langle \text{hard mor gen pl} \rangle == \langle \text{stem pl} \rangle _ov.$

Russian nominal inflection (cont)

- The example Russian word for law 'zakon' which uses the inflectional rules of node N I is defined as a separate node through the <infl root> and <sem animacy>.
- Zakon:
- <> == N_I
- <infl_root> == zakon
- <sem animacy> == inanimate.
- Zakon: <gloss> = law.
- Zakon: <mor nom sg> = zakon.
- Zakon: <mor acc sg> = zakon.
- Zakon: <mor gen sg> = zakon _a.
- Zakon: <mor dat sg> = zakon _u.
- Zakon: <mor inst sg> = zakon _om.
- Zakon: <mor loc sg> = zakon _e.
- Zakon: <mor nom pl> = zakon _i.
- Zakon: <mor acc pl> = zakon _i.
- Zakon: <mor gen pl> = zakon _ov.
- Zakon: <mor dat pl> = zakon _a _m.
- Zakon: <mor inst pl> = zakon _a _m'i.
- Zakon: <mor loc pl> = zakon _a _x.
- Zakon: <syn gender> = masc.
- Zakon: <syn animacy> = inanimate.

Russian nominal inflection (cont)

- The entire application of nominal inflectional morphology uses new insights into specific areas of Russian inflectional morphology like paradigm, gender assignment, case, number, and animacy.
- It presents the declensional classes as nodes of inheritance hierarchy and uses default inheritance hierarchy to model word structure by using a great deal of information sharing.
- It represents the inflectional morphology as a network of hierarchies and differentiates between the lexemic hierarchy and the inflectional hierarchy by using semantic principles.

Polish nominal inflectional morphology in DATR

- Polish nominal inflectional morphology presented by Czuba uses the idea of paradigm to present the case morphology. It introduces almost the same architecture of the application as that for the Russian including the same name of the semantic network nodes as NOMINAL, GENDER, and NOUN but slightly different inflectional rules.
- It reflects the fact that Polish and Russian are closely related languages which use cases to represent syntactic relations.
- The principle difference in both encodings is the application of the idea for finite state transducers used in the Polish nominal inflection encoding to account for different inflectional sound alternations.

Bulgarian nominal inflectional morphology in DATR

- The standard Bulgarian language uses prepositions and a base word form instead of case declensions. It uses relatively free word order, so the subject can take every syntactic position in the sentence (including the last one). Another important grammar feature of Bulgarian is the feature of definite article which is an ending morpheme.
- The syntactic function of definiteness in Bulgarian is expressed by a formal morphological marker which is an ending morpheme. The following part-of-speech in Bulgarian take the definite article: nouns, adjectives, numerals (both cardinals and ordinals), possessive pronouns (the full forms), and reflexive-possessive pronoun (its full form). The definite morphemes are the same for all part-of-speech, however, in further description we are going to analyze only some general types of rules used for the interpretation of nominal inflectional morphology of definiteness in Bulgarian.

Bulgarian nominal inflection (cont)

- The analyzed application of nominal inflectional morphology of Bulgarian is linguistically motivated. In particular, the basic idea is that of a paradigm since morphemes are defined to be of semantic value and are considered as a realization of a specific morphosyntactic phenomenon.
- The words are encoded by introducing different roots to account for the related phonetic alternations, which are defined to be of semantic value as well.

The architecture of the application and the inflectional rules

- The architecture represents an inheritance network consisting of

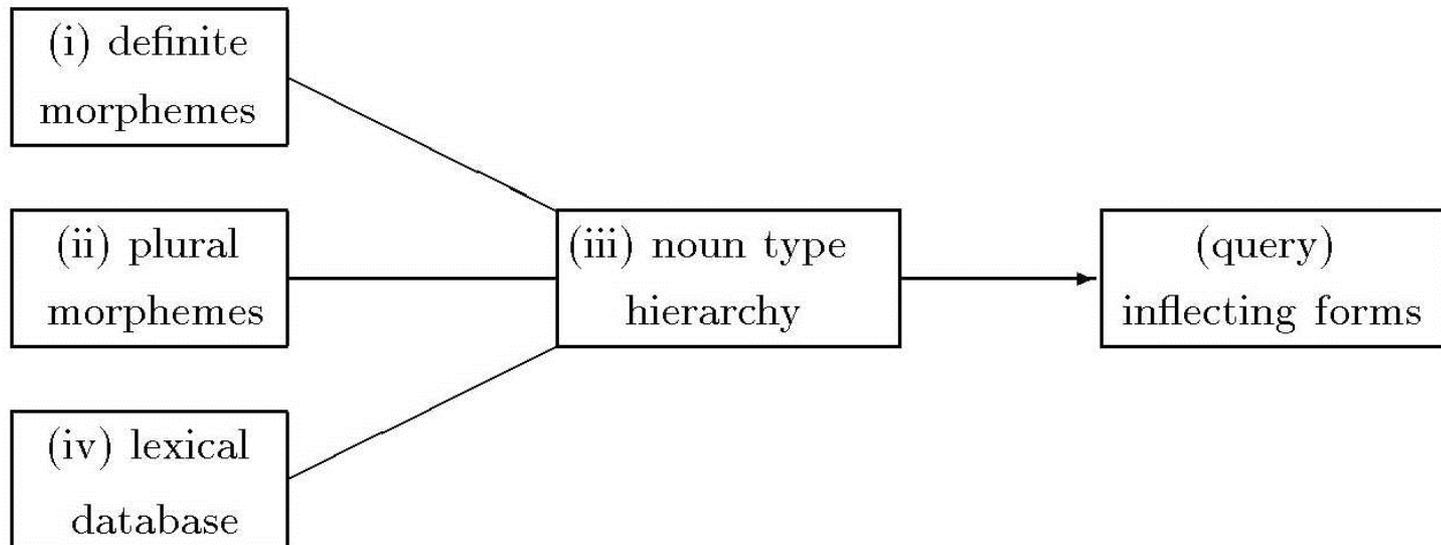


Fig. 1. The overall architecture of Bulgarian noun inflection.

The architecture of the application and the inflectional rules

- The DATR logical representation framework uses rule-based reasoning with non-monotonic inference and default inheritance to represent the inflectional rules in semantic network in which the grammar knowledge is encoded by the attachment of inflectional rules to the related nodes and allows to account for the grammar irregularities and to generate all possible inflected word forms within one application.
- The application uses a hierarchical structure in which the feature of gender is a trigger to change the values of the inflected forms.

The architecture of the application and the inflectional rules

- During the process of inflection, also, various phonetic alternations are taking place. The phonetic alternations at the morpheme boundary are interpreted either by defining new grammar rules or new nodes, and the phonetic alternations inside morphemes are interpreted by introducing different roots. It is possible, also, to use the technique of finite state transducers.
- The analyzed application interprets, also, more complicated cases of inflection, where both prefixes and suffixes can be processed by defining new nodes of the network.

Bulgarian nominal inflection (cont)

- The DATR analysis of nouns starts with node DET which defines all inflecting morphemes for the definite article.
- DET:
- <sing undef> ==
- <sing def_2 masc> == _ja
- <sing def_2 masc_1> == _a
- <sing def_1 masc> == _jat
- <sing def_1 masc_1> == _ut
- <sing def_1 femn> == _ta
- <sing def_1 neut> == _to
- <plur undef> ==
- <plur def_1> == _te.

Bulgarian nominal inflection (cont)

- Node Suff defines 12 inflecting morphemes for generating the plural.
- Suff:
- <suff_11> == _i
- <suff_111> == _ovci
- <suff_12> == _e
- <suff_121> == _ove
- <suff_122> == _eve
- <suff_123> == _ovce
- <suff_21> == _a
- <suff_22> == _ja
- <suff_211> == _ishta
- <suff_212> == _ta
- <suff_213> == _ena
- <suff_214> == _esa.

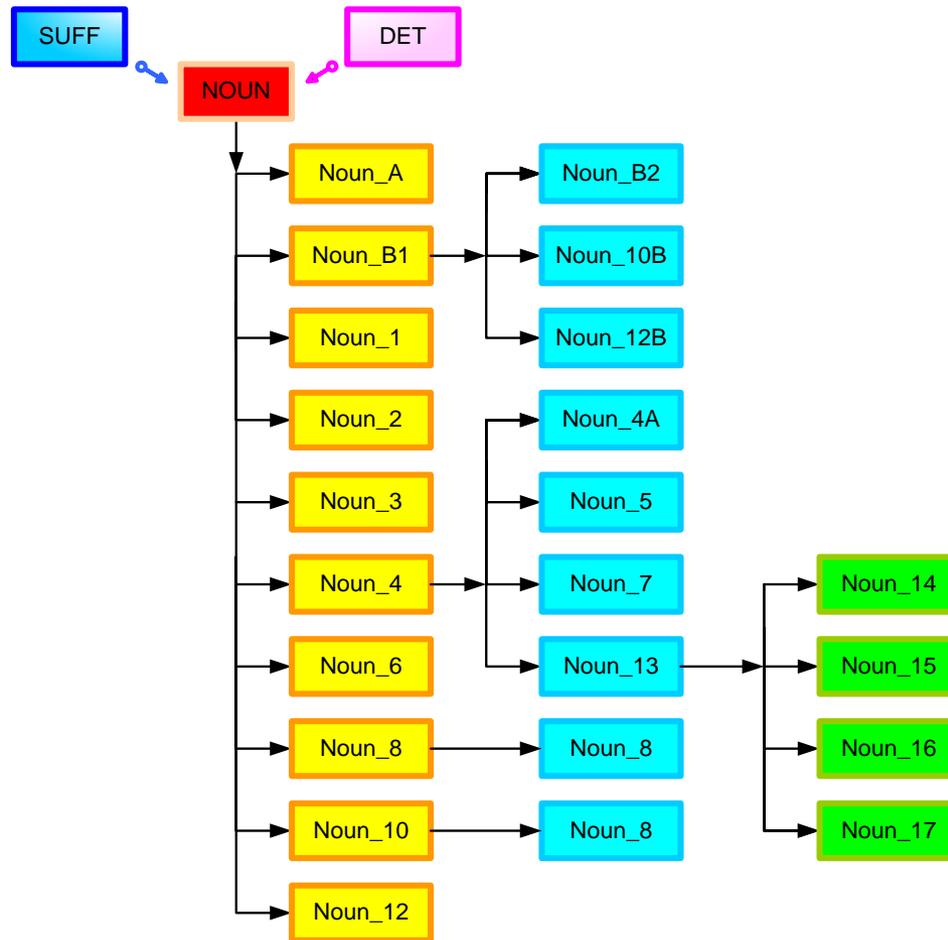
Bulgarian nominal inflection (cont)

- The basic node of noun inflectional types hierarchy is the node Noun and it defines the general inflectional rules for compilation of all possible inflected word forms.
- Noun:
- `<suff> == suff_11`
- `<gender> == masc_1`
- `<> == <stem> DET: <Idem "<gender>">`
- `<stem sing> == "<root sing>"`
- `<stem plur> == "<root plur>Suff:<"<suff>">.`

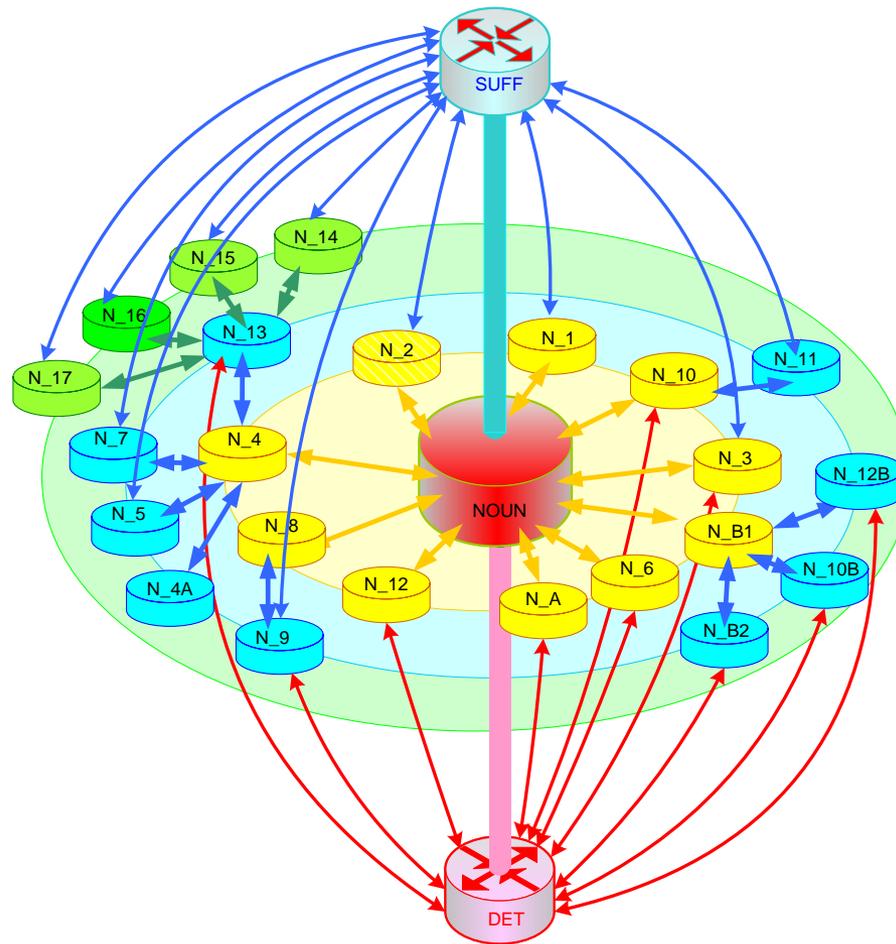
Bulgarian nominal inflection (cont)

- The other word inflectional types are defined either by changing the values of <gender> (the definite morphemes) and <suff> (the plural morphemes) or by introducing new inflectional rules.
- **Zakon:**
- <> = Noun.
- <root> = zakon.
- **Zakon:** <gender> == masc_1.
- **Zakon:** <sing undef> == zakon.
- **Zakon:** <plur undef> == zakon_i.
- **Zakon:** <sing def_1> == zakon_ut.
- **Zakon:** <sing def_2> == zakon_a.
- **Zakon:** <plur def_1> == zakon_i_te.

Bulgarian nominal inflection (cont)



Bulgarian nominal inflection (cont)



Conclusions

- The analyzed applications of Slavonic nominal inflectional morphology use the traditional grammar features of declension, gender, number, and animacy to encode the features of case and definiteness. The applications introduce the inheritance hierarchies for concise encoding and represent the declensional classes as nodes.
- The architecture of the interpretations differentiate between inflectional classes (types) hierarchy, lexemic hierarchy, and semantic hierarchy. It underlay the idea that related languages because of the fact that they share similar grammar features can be formally presented by using similar ideas and techniques for the encoding.

Thank you for your attention!