

Interpretation of Possessive and Reflexive-Possessive Pronouns of Bulgarian Language in DATR

Velislava Stoykova

► **To cite this version:**

Velislava Stoykova. Interpretation of Possessive and Reflexive-Possessive Pronouns of Bulgarian Language in DATR. 10th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI), Sep 2014, Rhodes, Greece. pp.528-536, 10.1007/978-3-662-44654-6_52 . hal-01391355

HAL Id: hal-01391355

<https://hal.inria.fr/hal-01391355>

Submitted on 3 Nov 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Interpretation of Possessive and Reflexive-possessive Pronouns of Bulgarian Language in DATR

Velislava Stoykova

Institute for Bulgarian Language, Bulgarian Academy of Sciences,
52, Shipchensky proh. str., bl. 17, 1113 Sofia, Bulgaria
email: vstoykova@yahoo.com

Abstract. The paper presents a formal interpretation of possessive and reflexive-possessive pronouns in Bulgarian language using non-monotonic approach to formal logical representation. It offers a semantic network to model inflectional morphology of definite article for related pronouns using DATR language for lexical knowledge representation. The proposed encoding combines non-monotonic orthogonal semantic networks hierarchical representation and relational database approach. Finally, some conclusions about presenting inflectional morphology using non-monotonic semantic networks approach are presented.

Keywords: Non-monotonic Reasoning, Knowledge Acquisition and Representation, Natural Language Processing.

1 Introduction

Non-monotonic approach has been widely used for formal representation in various Natural Language Processing tasks. Its main advantage is that it allows formal account for both regular and irregular cases and is suitable for modeling any type of grammar feature. Also, some knowledge representation languages which allow non-monotonic interpretation have been successfully used to model various types of language features (like inflection, syntax, etc.) for related and non-related natural languages. Further, we are going to present a formal interpretation of inflectional morphology of possessive and reflexive-possessive pronouns in Bulgarian language by using DATR language for lexical knowledge representation as a suitable formal framework.

2 The DATR language

The DATR language for lexical knowledge representation is a non-monotonic language for defining the inheritance networks through path/value equations [4]. It has both an explicit declarative semantics and an explicit theory of inference allowing efficient implementation, and at the same time, it has necessary expressive power to encode lexical entries presupposed by the work in the unification grammar tradition [2, 3].

In DATR information is organized as a network of nodes, where a node is a collection of related information. 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 mapping which assigns unique values to node attribute-path pair, and the recovery of these values is deterministic. With respect to its universality, DATR's formal properties and techniques underlay both rule-based inference and non-monotonic inference by default, and allow to account for language phenomena such as regularity, irregularity, and subregularity by using deterministic parsing.

The semantics of DATR uses non-monotonic inference and default inheritance, and allows generalization-capturing representation of inflectional morphology. DATR has expressive power which is capable to encode and process both syntactic and morphological rules and it allows representation of grammar knowledge by using semantic networks.

The DATR language for lexical knowledge representation offers syntagmatic operators (which can be used to define concatenation procedures) and paradigmatic operators (which can be used to define specific structure of the inflecting rules, and for further development of part-of-speech interpretations).

The DATR language has a lot of implementations, however, our application was made by using QDATR 2.0 (consult URL <http://www.cogs.susx.ac.uk/lab/nlp/datr/datrnode49.html> for a related file `bul_det.dtr`). This PROLOG encoding uses Sussex DATR notation [13].

DATR allows construction of various types of language models (language theories), however, the analysed interpretations underlay architecture of a rule-based formal grammar and a lexical database. The particular queries to be evaluated are all related inflected word forms, and the implementation allows to process words in Cyrillic alphabet.

The DATR language has been applied for developing inflectional morphology for a lot of languages including Russian [1] and Polish [5]. The ideas used for Russian nominal inflection interpretation underlay that of a paradigm and the encoding presents a resolving of a tabular conceptualization encoding task.

3 Bulgarian possessive pronouns in DATR

The semantics of possessive pronouns in Bulgarian language includes various types of relation like: possession (depending whether it is an object or a subject of possession), part-of-whole, relational, etc. The main semantic relationship of possession varies depending whether it is referred to the possessor or to the thing being possessed [7]. Only the full forms of possessive pronouns have inflection. They are: 'moj' (my), 'tvoj' (your), 'negov' (his), 'nein' (her), 'nash' (our), 'vash' (your), 'tehen' (their).

They have grammar features of person, number, gender and definiteness. The grammar feature of person is not inflectional and expresses information both at the level of syntax and at the hypertext level through agreement. The full forms imply information both about the possessor and the object being possessed using agreement in number and gender.

The grammar feature of definiteness implies information about the possession at the syntactic level using agreement and is expressed by a formal morphological marker which is an ending morpheme [6]. It is different for genders however, for masculine gender two types of definite morphemes exist – to determine a defined in a different way entities, which have two phonetic alternations, respectively. For feminine and for neuter gender only one definite morpheme exists, respectively. For plural, two definite morphemes are used depending on the ending vocal of the main plural form.

possessed		POSSESSOR								
number	gender	singular					plural			definiteness
		person								
		1 st pers.	2 nd pers.	3 rd person			1 st pers.	2 nd pers.	3 rd pers.	
		gender								
		male	female	neuter						
sg.	male	moj	tvoj	negov	nein	negov	nash	vash	tehen	undefined
sg.	female	moja	tvoja	negova	nejna	negova	nasha	vasha	tjahna	
sg.	neuter	moe	tvoe	negovo	nejno	negovo	nashe	vashe	tjahno	
pl.		moi	tvoi	negovi	nejni	negovi	nashi	vashi	tehni	
sg.	male	moja	tvoja	negovija	nejnija	negovija	nashija	vashija	tehnija	defined
sg.	male	mojat	tvojat	negovijat	nejnijat	negovijat	nashijat	vashijat	tehnijat	
sg.	female	mojata	tvojata	negovata	nejnata	negovata	nashata	vashata	tjahnata	
sg.	neuter	moeto	tvoeto	negovoto	nejnoto	negovoto	nasheto	vasheto	tjahnoto	
pl.		moite	tvoite	negovite	nejnite	negovite	nashite	vashite	tehnite	

Fig. 1. The Bulgarian possessive pronouns inflection table representation.

The features of gender and number of definite article are different from the gender and number features of possessive pronouns, themselves. The former are inflectional whereas the later are not inflectional, even both they can express agreement and both they are significant for part-of-speech parsing. The relation between them is important starting point for our interpretation. The inflectional forms of possessive pronouns are presented at Fig 1.

Our approach to inflectional morphology of possessive pronouns is based on the encoding already published in [9]. It presents an inheritance semantic network consisting of different inflectional hierarchical type nodes of rule-based formal grammar and a lexical database (the pronouns). The particular queries to be evaluated are related inflected forms. It also offers an account of sound alternations.

The interpretation is based on adjectives' encoding [8] and takes as a starting point linguistic motivation, in particular, the priority of one or another grammar feature. Thus, it uses definite article's grammar features of gender and number to trigger the value of related inflected forms. It also is similar to the approach presented in [10, 11] to possessive pronouns formal representation using frameworks of Universal Networking Language [12] which present another type of semantic network.

Further, we are going to extend that approach by relating inflectional rules and inflectional type hierarchy to grammar features of possessive pronouns themselves.

3.1 The inflectional type hierarchy

The DATR analysis of possessive pronouns starts with node DET which defines all inflecting morphemes for definite article as follows: ¹

¹ Here and elsewhere in the description we use Latin alphabet to present morphemes instead of Cyrillic. Because of the mismatching between both some typically Bulgarian phonological alternations are marked by two letters, whereas in Cyrillic alphabet they are marked by one.

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.
```

The basic node AdjG defines grammar rules for generating all inflected forms for the feature of gender, number, and definiteness and is exactly the same as for the adjectives.

AdjG:

```
<sing undef masc> == "<root>"
<sing undef femn> == "<root gend>" _a
<sing undef neut> == "<root gend>" _o
<sing def_2 masc> == "<plur undef masc>" DET
<sing def_1 masc> == "<plur undef masc>" DET
<sing def_1> == "<sing undef>" DET
<plur undef> == "<root gend>" _i
<plur def_1> == "<plur undef>" DET.
```

Node Adj inherits all grammar rules from node AdjG but employs grammar rules for generating inflected forms for the feature of comparison of degree. It also is the same as for the adjectives. We employ it to make encoding as short as possible regardless the fact that possessive pronouns generates compared inflected forms very rare. The pronouns 'negov' (his) and 'nein' (her) uses this type of inflectional rules.

Adj:

```
<> == AdjG
<compar> == po_ "<>"
<superl> == naj_ "<>".
```

Node Adj_2 is defined as it is employed for the adjectives and it presents inflectional rules of possessive pronoun 'tehen' (their), which realises two types of phonetic alternation during the process of inflection. At this node an additional inflectional base form <root plur> is introduced to account for the complexity.

Adj_2:

```
<> == Adj
<plur undef> == "<root plur>" _i.
```

Node Adj_4 is defined also for adjectives. It encodes inflectional rules of possessive pronouns 'nash' (our) and 'vash' (your) and inherits all grammar rules of node Adj. The new employed grammar rule changes inflectional morpheme for neuter gender into -e.

Adj_4:

```
<> == Adj
<sing undef neut> == "<root gend>" _e.
```

The new node Adj_5 is employed and it consists of inflectional rules for possessive pronouns 'moj' (my) and 'voj' (your). It defines feminine and neuter gender forms by using palatal morphemes, and generates defined inflected forms of masculine gender by using the <root gend> base.

Adj_5:

```
<> == Adj
<sing undef femn> == "<root gend>" _ja
<sing undef neut> == "<root gend>" _e
<sing def_1 masc> == "<root gend>" DET
<sing def_2 masc> == "<root gend>" DET.
```

The grammar features of gender and number of possessive pronouns themselves can be presented as relational information introducing node PossPron which consists of rules that relate lexical database information presented at the nodes of possessive pronouns themselves ('Moj', 'Tvoj', 'Negov', 'Nein', 'Nash', 'Vash' and 'TeHEN') to inflectional grammar rules nodes AdjG, Adj_2, Adj_4 and Adj_5. The node is as follows:

PossPron:

```
<first sing_p sing undef masc> == <sing undef masc> Moj
<first sing_p sing undef femn> == <sing undef femn> Moj
<first sing_p sing undef neut> == <sing undef neut> Moj
<first sing_p sing def_2 masc> == <sing def_2 masc> Moj
<first sing_p sing def_1 masc> == <sing def_1 masc> Moj
<first sing_p sing def_1> == <sing def_1> Moj
<first sing_p plur undef> == <plur undef> Moj
<first sing_p plur def_1> == <plur def_1> Moj
<second sing_p sing undef masc> == <sing undef masc> Tvoj
<second sing_p sing undef femn> == <sing undef femn> Tvoj
<second sing_p sing undef neut> == <sing undef neut> Tvoj
<second sing_p sing def_2 masc> == <sing def_2 masc> Tvoj
<second sing_p sing def_1 masc> == <sing def_1 masc> Tvoj
<second sing_p sing def_1> == <sing def_1> Tvoj
<second sing_p plur undef> == <plur undef> Tvoj
<second sing_p plur def_1> == <plur def_1> Tvoj
<third sing_p masc_p sing undef masc> == <sing undef masc> Negov
<third sing_p masc_p sing undef femn> == <sing undef femn> Negov
<third sing_p masc_p sing undef neut> == <sing undef neut> Negov
<third sing_p masc_p sing def_2 masc> == <sing def_2 masc> Negov
<third sing_p masc_p sing def_1 masc> == <sing def_1 masc> Negov
<third sing_p masc_p sing def_1> == <sing def_1> Negov
<third sing_p masc_p plur undef> == <plur undef> Negov
<third sing_p masc_p plur def_1> == <plur def_1> Negov
<third sing_p femn_p sing undef masc> == <sing undef masc> Nein
<third sing_p femn_p sing undef femn> == <sing undef femn> Nein
<third sing_p femn_p sing undef neut> == <sing undef neut> Nein
<third sing_p femn_p sing def_2 masc> == <sing def_2 masc> Nein
<third sing_p femn_p sing def_1 masc> == <sing def_1 masc> Nein
<third sing_p femn_p sing def_1> == <sing def_1> Nein
<third sing_p femn_p plur undef> == <plur undef> Nein
<third sing_p femn_p plur def_1> == <plur def_1> Nein
<first plur_p sing undef masc> == <sing undef masc> Nash
<first plur_p sing undef femn> == <sing undef femn> Nash
<first plur_p sing undef neut> == <sing undef neut> Nash
<first plur_p sing def_2 masc> == <sing def_2 masc> Nash
<first plur_p sing def_1 masc> == <sing def_1 masc> Nash
<first plur_p sing def_1> == <sing def_1> Nash
```

```

<first plur_p plur undef>      == <plur undef> Nash
<first plur_p plur def_1>     == <plur def_1> Nash
<second plur_p sing undef masc> == <sing undef masc> Vash
<second plur_p sing undef femn> == <sing undef femn> Vash
<second plur_p sing undef neut> == <sing undef neut> Vash
<second plur_p sing def_2 masc> == <sing def_2 masc> Vash
<second plur_p sing def_1 masc> == <sing def_1 masc> Vash
<second plur_p sing def_1>     == <sing def_1> Vash
<second plur_p plur undef>     == <plur undef> Vash
<second plur_p plur def_1>     == <plur def_1> Vash
<third plur_p sing undef masc> == <sing undef masc> Tehen
<third plur_p sing undef femn> == <sing undef femn> Tehen
<third plur_p sing undef neut> == <sing undef neut> Tehen
<third plur_p sing def_2 masc> == <sing def_2 masc> Tehen
<third plur_p sing def_1 masc> == <sing def_1 masc> Tehen
<third plur_p sing def_1>     == <sing def_1> Tehen
<third plur_p plur undef>     == <plur undef> Tehen
<third plur_p plur def_1>     == <plur def_1> Tehen.

```

The new introduced paths <first>, <second> and <third> refer to the feature of person, paths <masc_p> and <femn_p> refer to the feature of gender, and paths <sing_p> and <plur_p> refer to the feature of number of possessive pronouns themselves. The generated inflected forms of possessive pronoun 'Negov' are given at the Appendix.

4 Bulgarian reflexive-possessive pronoun in DATR

The semantics of reflexive-possessive pronoun in Bulgarian language combines semantics of possession relationship and that of reflexivity. It expresses possession relationship between the possessor (defined by the subject in the sentence, and agreed with it in gender and number) and the thing being possessed (to which the pronoun is referred to, and agrees in gender and number). The reflexive-possessive pronoun is one and it has

number	singular			plural
gender	male	female	neuter	
<i>undefined</i>	svoj	svoja	svoe	svoi
<i>defined</i>	svoja(t)	svojata	svoeto	svoite
	svojat			

Fig. 2. The Bulgarian reflexive-possessive pronoun's inflection table representation.

full and short form, and both they can be used with respect to agreement. However, only its full form 'svoj' (-self) has inflectional grammar features of person, gender, number, and definiteness, which are similar to that of adjectives and of possessive pronouns. Its inflected forms are given at Fig. 2.

4.1 The inflectional type hierarchy

The DATR formal account of reflexive-possessive pronoun inflectional morphology [9] uses inflectional rules already defined at node `Adj_5` and is part of it. It also uses the same principle as for possessive pronouns. The entire interpretation is extended with rules which relate reflexive-possessive pronoun to its grammar features of person and gender (the person and gender of the subject in the sentence which are reflexive) and the gender and number grammar features of its definite article (the gender and number of the complement in the sentence). The node `RefPoss` is introduced which is as follows:

```
RefPoss:
  <sing undef masc> == <sing undef masc> Svoj
  <sing undef femn> == <sing undef femn> Svoj
  <sing undef neut> == <sing undef neut> Svoj
  <sing def_2 masc> == <sing def_2 masc> Svoj
  <sing def_1 masc> == <sing def_1 masc> Svoj
  <sing def_1>      == <sing def_1> Svoj
  <plur undef>     == <plur undef> Svoj
  <plur def_1>    == <plur def_1> Svoj.
```

The encoding differ with that for possessive pronouns 'moj' and 'tvoj' because the features of person and gender of pronoun 'svoj' are reflexive and should be defined by the person and gender of the subject in the sentence. Thus, entire interpretation gives a relation also to syntactic interpretation for part-of-speech analysis. The related generated inflected forms of reflexive-possessive pronoun 'svoj' are given at the Appendix.

5 Conclusion

The entire formal semantic network interpretation of inflectional morphology of possessive and reflexive-possessive pronouns in Bulgarian language uses non-monotonic approach to account for both regular and irregular word forms by introducing different nodes with related inflectional rules. It also uses the features of gender and number as a trigger to change the values of related inflected forms and a relational database approach to relate the grammar features of definite article to grammar features of pronouns themselves allowing further syntactic interpretation.

Thus, it is good as a future works to extend that approach by enlarging the application offering syntactic rules.

References

1. Corbett, G. and Fraser, N. 1993. Network Morphology: a DATR account of Russian nominal inflection. *Journal of Linguistics*, 29, 113–142.
2. Evans, R. and Gazdar, G. 1989a. Inference in DATR. *Fourth Conference of the European Chapter of the Association for Computational Linguistics*, 66–71.
3. Evans, R. and Gazdar, G. 1989b. The semantics of DATR. In Anthony G. Cohn, ed. *Proceedings of the Seventh Conference of the Society for the Study of Artificial Intelligence and Simulation of Behaviour*, London: Pitman/Morgan Kaufmann, 79–87.
4. Evans, R. and Gazdar, G. 1996. DATR: A language for lexical knowledge representation. *Computational Linguistics*, 22(2):167–216.

5. Czuba, K. 1994. The DATR Web Pages at Sussex. http://www.cogs.susx.ac.uk/lab/nlp/datr/datnode49, file polish_n.dtr.
6. Gramatika na suvremennia bulgarski knizoven ezik. 1983. *Morphologia*, tom. 2. (in Bulgarian)
7. Nicolova, R. 1986 The Bulgarian Pronouns Nauka i izkustvo, Sofia, (in Bulgarian)
8. Stoykova, V. 2004. The definite article of bulgarian adjectives and numerals in DATR. In C. Bussler and D. Fensel, eds. *Artificial Intelligence: Methodology, Systems, and Applications. LNAI 3192*, Springer-Verlag, 256–266.
9. Stoykova, V. 2010 Bulgarian possessive and reflexive-possessive pronouns in DATR. In R. Trappl ed. *Cybernetics and Systems 2010.*, Vienna: Austrian Society for Cybernetic Studies, 426–432.
10. Stoykova, V. 2012 The Inflectional Morphology of Bulgarian Possessive and Reflexive-possessive Pronouns in Universal Networking Language. *Procedia Technology*, vol. 1, Elsevier, 400–406.
11. Stoykova, V. 2013 Representation of Possessive Pronouns in Universal Networking Language. In L. Iliadis, H. Papadopoulos, and C. Jayne eds. *Proceedings from EANN 2013, Part II, Communications in Computer and Information Science (CCIS)*, vol. 384, Springer-Verlag, 129–137.
12. Uchida, H. 2005 Universal Networking Language. UNDL Foundation
13. The DATR Web Pages at Sussex. 1997. <http://www.cogs.susx.ac.uk/lab/nlp/datr/>

Appendix:

Negov:

```
<> == Adj
<root> == negov
<root_gend> == negov.
```

```
PossPron: <third sing_p masc_p sing undef masc> == negov.
PossPron: <third sing_p masc_p sing undef femn> == negova.
PossPron: <third sing_p masc_p sing undef neut> == negovo.
PossPron: <third sing_p masc_p sing def_2 masc> == negovija.
PossPron: <third sing_p masc_p sing def_1 masc> == negovijat.
PossPron: <third sing_p masc_p sing def_1 femn> == negovata.
PossPron: <third sing_p masc_p sing def_1 neut> == negovoto.
PossPron: <third sing_p masc_p plur undef> == negovi.
PossPron: <third sing_p masc_p plur def_1> == negovite.
```

```
Svoj: <> == Adj_5
      <root> == svoj
      <root_gend> == svo.
```

```
RefPoss: <sing undef masc> == svoj.
RefPoss: <sing undef femn> == svoja.
RefPoss: <sing undef neut> == svoe.
RefPoss: <sing def_2 masc> == svoja.
RefPoss: <sing def_1 masc> == svojat.
RefPoss: <sing def_1 femn> == svojata.
RefPoss: <sing def_1 neut> == svoeto.
RefPoss: <plur undef> == svoi.
RefPoss: <plur def_1> == svoite.
```