Formal Representation of Bulgarian Pronouns

VELISLAVA STOYKOVA
Institute for Bulgarian Language - IBL
Bulgarian Academy of Sciences - BAS
52, Shipchensky proh. str. bl. 17, 1113 Sofia
BULGARIA
vstoykova@yahoo.com

Abstract: The paper*presents a computationally tractable application of Bulgarian pronouns representation using Universal Networking Language (UNL) formalism. It analyses grammar features of pronouns and offers a model of their formal representation based on incorporation of grammar, semantic and lexical knowledge by using standard UNL knowledge representation mechanisms. Also a comparison of different approaches to formal representation of Bulgarian possessive and reflexive-possessive pronouns inflectional morphology is offered. The interpretation is based on detailed analysis of grammar features and semantics of possessive and reflexive-possessive pronouns, and related formal representations are based on the use of semantic networks. The problem is interpreted as a grammar knowledge representation task.The UNL interpretation presents both morphological and syntactic knowledge and offers multilingual web-based application which can be further developed and elaborated.

Key–Words: Natural Language Processing Systems, Semantic Networks, Knowledge Representation.

1 Introduction

Natural Language Processing (NLP) systems use two general approaches to both speech [20] and text processing - statistically-based and rule-based. However, for both of them representing word [9] and its related grammar, lexical and contextual features is a central task. Thus, statistical approaches have been successfully used for NLP Neural Networks applications [4], whereas rule-based approaches were used for NLP knowledge representation tasks.

The semantic networks are widely used knowledge representation formalism. They offer grammar knowledge hierarchical semantic representation of both inflectional and conceptual knowledge by using mostly rule-based encodings. At the same time, different rule-based applications offer different techniques for encoding of almost all grammar features, including different encodings of one and the same grammar feature in similar semantic networks formalisms.

Further, we are going to analyse formal representation of Bulgarian pronouns using semantic networks interpretation in UNL framework. We also will compare two applications of inflectional morphology of Bulgarian possessive and reflexive-possessive pronouns represented in semantic networks – using DATR language for lexical knowledge representation, and using UNL [16].

*The research is supported by the project BG051PO001-3.3-05/0001 "Science and Business" of the Ministry of Education, Youth and Science of Bulgaria.

2 Pronouns

Pronouns exist in almost all European languages and are traditionally analysed as different part-of-speech with their own grammar features, semantics and specific usage. The main function of pronouns is to substitute another words or class of words, and with respect to that they have related grammar features. Pronouns' grammar features are language-specific, however, there are common view that pronouns play significant role in text mantaining its semantic relations and coherence. Also, some types of pronouns share semantics of deixis.

Nevertheless, the main function of pronouns is to relate words in the sentence or in the text so that, they can be formally presented using various types of formalisms.

2.1 Bulgarian Pronouns

The Bulgarian pronouns are traditionally described with respect to their semantics, grammar features, functions and usage. Generally, their semantics is connected to substitute, determine, relate, and agree both with other words in the sentence or within the whole text [12].

So, the grammar features of pronouns are connected to their semantics and functions and are highly dependent on agreement. In general, the pronouns have grammar features of person, number, and gender (for some also case, definiteness, etc.) These features can be successfuly used for formal representations of pronouns with respect to computational applications.

Thus, the existing formal semantic networks interpretation of possessive and reflexive-possessive pronouns using DATR language for lexical knowledge representation is based on the use of that features [18]. The formal interpretation of Bulgarian pronouns at the text level are presented in [5].

The Bulgarian pronouns are of following types: personal, reflexive, possessive, reflexive-possessive, demonstrative, relative, distributive, interrogative, indefinite, and negative. In our interpretation, we offer a systematic account for all types. Further, we are going to analyse and present semantic and grammar assumptions for related formal representation of Bulgarian pronouns [17].

3 Universal Networking Language

In the UNL approach, information conveyed by natural language is represented as a hypergraph composed of a set of directed binary labeled links (referred to as ”relations”) between nodes or hypernodes (the ”Universal Words”(UWs)), which stand for concepts. UWs can also be annotated with ”attributes” representing context information [19].

Universal Words (UWs) represent universal concepts and correspond to the nodes to be interlinked by ”relations” or modified by ”attributes” in a UNL graph. They can be associated to natural language open lexical categories (noun, verb, adjective and adverb). Additionally, UWs are organized in a hierarchy (the UNL Ontology), are defined in the UNL Knowledge Base and exemplified in the UNL Example Base, which are the lexical databases for UNL. As language-independent semantic units, UWs are equivalent to the sets of synonyms of a given language, approaching the concept of ”synset” used by the WordNet.

Attributes are arcs linking a node onto itself. In opposition to relations, they correspond to one-place predicates, i.e., function that take a single argument. In UNL, attributes have been normally used to represent information conveyed by natural language grammar categories (such as tense, mood, aspect, number, etc). Attributes are annotations made to nodes or hypernodes of a UNL hypergraph. They denote circumstances under which these nodes (or hypernodes) are used.

Attributes may convey three different kinds of information: (i) The information on the role of the node in the UNL graph, (ii) The information conveyed by bound morphemes and closed classes, such as affixes (gender, number, tense, aspect, mood, voice, etc), determiners (articles and demonstratives), etc., (iii) The information on the (external) context of the utterance. Attributes represent information that cannot be conveyed by UWs and relations.

Relations, are labeled arcs connecting a node to another node in a UNL graph. They correspond to two-place semantic predicates holding between two UWs. In UNL, relations have been normally used to represent semantic cases or thematic roles (such as agent, object, instrument, etc.) between UWs.

UNL-NL Grammars are sets of rules for translating UNL expressions into natural language (NL) sentences and vice-versa. They are normally unidirectional, i.e., enconversion grammar (NL-to-UNL) or deconversion grammar (UNL-to-NL), even though they share the same basic syntax.

In the UNL Grammar there are two basic types of rules: (i) Transformation rules - used to generate natural language sentences out of UNL graphs and vice-versa and (ii) Disambiguation rules - used to improve performance of transformation rules by constraining their applicability.

The UNL frameworks offer universal language-independent and open-source platform for multilingual applications [2]. Recently, lots of multilingual projects have been developed including ‘The Little Prince Project’ [7]. The UNL application for English language is developed but some applications for other languages like Russian [1] and Bulgarian [11, 14] are available as well.

4 Representing Bulgarian Pronouns in UNL

The UNL specifications [21] offer formal approach to represent Bulgarian pronouns accordingly to their respective grammar features. The representation scheme includes both lexical information presentation and related grammar rules. The UNL lexical information representation scheme allows the use of three types of dictionaries which list lexical entries and their related features.

4.1 Lexical Representation

UNL Dictionary includes UWs in form of flat list of alphabetically ordered UWs with their corresponding features. It also mantains a mechanism to present semantic features like hierachy, typology, synonymy,

etc. It is used for universal language-independent lexical semantic representation.

NL Dictionary is created for a particular language and includes flat list of natural language entries with their corresponding grammar features. It is language-specific and presents monolingual word entries which semantic representation scheme is structured to present related grammar features like morphology, syntax, etc.

UNL-NL Dictionary is a bilingual dictionary which links lexical entries of UNL Dictionary to entries of NL Dictionary through different linking mechanisms.

The UNL lexical representation scheme underlay various approaches with respect to types and structure of information that can be included. Some general issues are presented in [8].

4.2 Grammar Representation

The grammar rules include both morphological and syntactic information using the features of person, number and gender. They are capable to interpret inflectional features (like definiteness) as well. The inflection can be represented with respect to prefixes, suffixes, infixes, and to the sound alternations taking place during the process of inflection.

The related grammar features link Bulgarian pronouns to their correlates for different languages by using mostly grammar featutes of case (syntactic or morphological) or by agreement in gender and number (at morphological level). Generally, multilingual representation of pronouns allows development of both syntactic and morphological rules.

The UNL specifications [21] allow two types of transformation inflectional rules: (i) A-rules (affixation rules) apply over isolated word forms (as to generate possible inflections) and (ii) L-rules (linear rules) apply over lists of word forms (as to provide transformations in the surface structure). Affixation rules are used for adding morphemes to a given base form. They are used for generating inflections or derivations. There are two types of A-rules: (i) simple A-rules involve a single action (such as prefixation, suffixation, infixation and replacement), and (ii) complex A-rules involve more than one action (such as circumfixation).

There are four types of simple A-rules: (i) prefixation, for adding morphemes at the beginning of a base form, (ii) suffixation, for adding morphemes at the end of a base form, (iii) infixation, for adding morphemes to the middle of the base form, (iv) replacement, for changing the base form.

Further, we will analyse Bulgarian pronouns formal representation with respect to its lexical database

Figure 1: The structure of Bulgarian pronouns lexical database in UNL representation.

and NL dictionary language-specific feature presentation by comparing examples of personal, possessive and reflexive pronouns.

4.3 Bulgarian Pronouns Lexical Database

As analysed above, Bulgarian pronouns lexical database presents in its structure both grammar and semantic relations. It uses a model of collecting different types of information which can be further processed. The structure is presented at Fig. 1 and includes following types of information about pronominal lexical enrty: lemma, part-of-speech, lexical structure, gender, number, inflectional paradigm, subcategorization frame, etc. The above types of information present mostly structured grammar knowledge.

Semantic relations are presented by using type hierarchies, synonyms, etc. However, semantic relations can present grammar knowledge as well. For example, synonymic relations can link different (non-inflectional) case forms of a particular pronoun or lexical relations can link different inflected forms to a base form. Such approach is presented at Fig. 2 for Bulgarian personal pronoun ”az” (I) where a lexical entry ”az” is linked to its case forms in Bulgarian (”me”, ”men”, ”se”, ”sebe si”) and in English (I, me, my, mine, myself).

The lexical entry of Bulgarian personal pronoun ”az” (I) also includes grammar information about the type of pronoun - personal, the person - first (PER=1PS), and the number - singular. It also includes syntactic information given by POS=PPR and

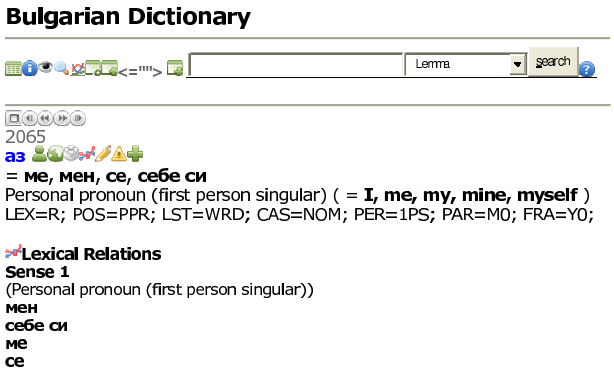

Figure 2: The Bulgarian personal pronoun ”az” (I) in UNL representation.

case information CAS=NOM which are used by transformation rules.

4.4 Morphological Representation

Generally, Bulgarian pronouns do not use inflection to realize their grammar function. However, possessive and reflexive-possessive pronouns have inflected forms for the features of gender, number and definiteness. The formal interpretation of possessive and reflexive-possessive pronouns made with DATR language for lexical knowledge representation [18] which uses also semantic networks is based on the idea of inheritance hierarchy of inflectional types most of which were already defined for adjectives.

Furher, we wll compare both semantic networks representations of inflectional morphology of Bulgarian possessive and reflexive-possessive pronouns – by using DATR language for lexical knowledge representation, and by using UNL [15].

4.4.1 Linguistic and computational approaches to inflectional morphology

The traditional interpretation of inflectional morphology given at the descriptive academic grammar works is a presentation of tables. The tables consist of all possible inflected forms of a related word with respect to its subsequent grammar features. The standard computational approach to 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.

Natural Language Processing applications use different techniques to represent phonological, morphological and syntactic knowledge. The sound alternations influence the inflectional morphology and as a result, they form irregular word forms.

Thus, we have a rather unsystematically formed variety of regular and irregular word forms [13], which require a non-monotonic rule-based formal interpretation. Additional difficulties come from the fact that sound alternations can be occurred both in stems, prefixes, suffixes and also on their boundaries, which suggest extremely complicated solutions.

4.4.2 DATR language for lexical knowledge representation

The DATR language is a non-monotonic language for defining the inheritance networks through path / value equations [3]. 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. 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 these values is deterministic.

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

The DATR language has a lot of implementations however the analysed application was made by using QDATR 2.0 [22] (see related file `bul_det.dtr`). This PROLOG encoding uses Sussex DATR notation. DATR allows construction of various types of language models (language theories), and the implementation allows to process words in Cyrillic alphabet.

The detailed description of UNL knowledge representation scheme and its machanisms to present morphology by means of application of different types of rules was previously presented in secion 3.

4.4.3 The semantics and the grammar features of possessive pronouns

The semantics of the possessive pronouns in Bulgarian includes various relationships like: possession (depending whether it is an object or a subject of possession), part-of-whole, relational, etc. The main semantic relationship of the possession varies depending

whether it is referred to the possessor or to the thing being possessed [10].

Only the full forms of the possessive pronouns have inflection. The full forms of the possessive pronouns are: 'moj' (my), 'tvoj' (your), 'negov' (his), 'nein' (her), 'nash' (our), 'vash' (your), 'tehen' (their). They have the 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 the information about the possession at the syntactic level using agreement and is expressed by a formal morphological marker which is an ending morpheme. It is different for genders however, for the masculine gender two types of definite morphemes exist - to determine a defined in a different way entities, which have two phonetic alternations, respectively.

For the feminine and for the neuter gender only one definite morpheme exists, respectively. For the plural, two definite morphemes are used depending on the ending vocal of the main plural form.

The inflectional morphology table representation of possessive pronouns is given at Fig. 3

The features of gender and number of the definite article are different from the gender and number features of the possessive pronouns, themselves. The former are inflectional whereas the later are not inflectional, even both they can express agreement.

Thus, our task is to analyse related architecture and principles of rule-based interpretations of the possessive pronouns inflectional morphology using DATR language for lexical knowledge representation and UNL.

4.4.4 The DATR encoding of Bulgarian possessive pronouns inflectional morphology

The DATR encoding of inflectional morphology of Bulgarian possessive pronouns given at [18] presents an inheritance semantic network consisting of different inflectional type nodes, a rule-based formal grammar, and a lexical database (the pronouns). The particular queries to be evaluated are related inflected word forms. It also offers an account of sound alternations [13] and related visual representation using hierarchical graph models [6].

The interpretation is based on the adjectives encoding [14] and takes as a starting point linguistic motivation, in particular, the priority of one or another

possessed		POSSESSOR													definiteness		
number	gender	person													undefined		defined
		singular								plural							
		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		tehnija
sg.	female	moja	tvoja	negova		nejna		negova		nasha	vasha				tjahna		tehnijat
sg.	neuter	moe	tvoe	negovo		nejno		negovo		nashe	vashe				tjahno		tehnite
pl.		moi	tvoi	negovi		nejni		negovi		nashi	vashi				tehni		
sg.	male	moja	tvoja	negovija		nejnija		negovija		nashija	vashija				tehnija		
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		

Figure 3: The possessive pronouns inflection table representation.

grammar feature. Thus, the feature of gender is accepted as a specific trigger to change the values of the inflected forms for the features of number and definiteness.

The DATR account of Bulgarian inflectional morphology offers encoding which defines inflectional rules using semantic network structure. It starts with node DET consisting of all definite morphemes as follows[1]:

```
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.
```

Node AdjG was already defined for the adjectives and for the numerals, and defines related inflectional rules.

```
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 inflectional rules of node AdjG but it employs also the grammar rules for generating the forms for the feature of comparison of degree. The node is the same as for the adjectives.

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

Node Adj_2 also was defined for the adjectives and includes inflectional rules of the pronoun 'tehen' (their), which realizes 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.

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

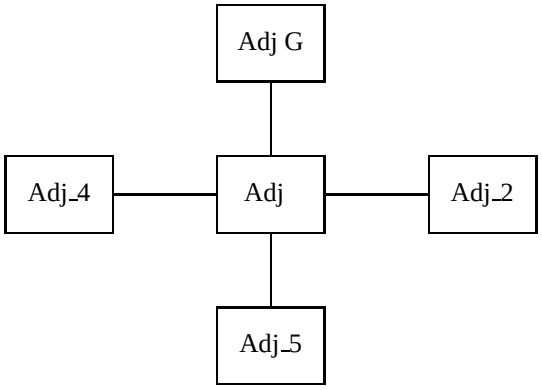

Figure 4: The possessive pronouns inflectional types hierarchy.

```
Adj_2:
      <> == Adj
      <plur undef> == "<root plur>" _i.
```

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

```
Adj_4:
<> == Adj
<sing undef neut> == "<root gend>" _e.
```

The new employed Adj_5 defines the inflectional rules for the pronouns 'moj' (my) and 'tvoj' (your). It evaluates the feminine and the neuter gender forms by using palatal morphemes, and generates the 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 encoding presents inflectional rules for generation of all related inflected forms of possessive pronouns. Node DET defines the definite inflectional morphemes and all other nodes define the inflectional rules for 4 related inflectional types which form inflectional types hierarchy presented at Fig. 4.

Thus, node Adj defines the rules for the pronouns 'negov' and 'nein'; node Adj_2 defines the rules for

the pronoun 'tehen'; node Adj_4 defines the rules for the pronouns 'nash' and 'vash', and node Adj_5 defines the inflectional rules for the pronouns 'moj' and 'tvoj'.

The pronouns are given as lexical database and are attached to the respective nodes which inflectional rules they use. The example pronoun 'moj' is given as attached to node Adj_5 and defined by its person, number, gender and inflectional roots.

```
Moj: <> == Adj_5
       <person> == one
       <number> == sing
       <gender> == masc
       <root> == moj
       <root gend> == mo.
```

Its generated inflected forms are as follows:

```
Moj: <sing undef masc> == moj.
Moj: <sing undef femn> == moja.
Moj: <sing undef neut> == moe.
Moj: <plur undef> == moi.
Moj: <sing def_1 masc> == mojat.
Moj: <sing def_2 masc> == moja.
Moj: <sing def_1 femn> == mojata.
Moj: <sing def_1 neut> == moeto.
Moj: <plur def_1> == moite.
```

The DATR representation of possessive pronouns uses inheritance hierarchical formal representation to interpret inflectional morphology rules and uses 4 inflectional rules most of which were defined for the adjectives. It accounts for the sound alternations and for irregular inflected forms. It also uses semantic hierarchical representation of the inflectional grammar features of gender, number and definiteness and concise encoding.

4.4.5 The UNL encoding of Bulgarian possessive pronouns inflectional morphology

In the UNL account of possessive pronouns presented in [14], the inflectional grammar features: gender, number and definiteness are accepted as a starting point of the encoding, and the inflectional rules are defined. The grammar features which are not inflectional (like person and non-inflectional gender) are presented as invariables (according to the UNL formalism definitions) and are included in the UNL dictionary database.

The UNL formal interpretation of possessive pronouns offers interpretation of inflectional morphology which uses A-rules. It also uses the idea of inflectional types and defines almost 5 inflectional types for generation of all inflected forms of possessive pronouns. The definitions of inflectional rules include interpretation also of sound alternations taking place during the process of inflection.

In further description, we are going to use the notation defined by the UNL specifications [21]. We are starting with the analysis of the inflectional rules for the possessive pronouns 'moj' (my) and 'tvoj' (your) which belong to one common inflectional type M165.

```
MCL&PST&DEF:="j">"ja";
MCL&PST&DEF:="j">"jat";
FEM&PST:="j">"ja";
FEM&PST&DEF:="j">"jata";
NEU&PST:="j">"e";
NEU&PST&DEF:="j">"eto";
PLR&PST:="j">"i";
PLR&PST&DEF:="j">"ite";
```

The possessive pronoun 'negov' (his) does not realize phonetic alternations during the process of inflection. Its inflectional grammar rules (type M167) define the masculine, feminine, neuter and plural undefined and defined inflected word forms by attachment of the gender, plural or definite morphemes to the base word form.

```
MCL&PST&DEF:=0>"ija";
MCL&PST&DEF:=0>"ijat";
FEM&PST:=0>"a";
FEM&PST&DEF:=0>"ata";
NEU&PST:=0>"o";
NEU&PST&DEF:=0>"oto";
PLR&PST:=0>"i";
PLR&PST&DEF:=0>"ite";
```

The possessive pronoun 'nein' (her) realizes one phonetic alternation (the transition of "i" into "j") during the process of inflection. The inflectional rules (type M168) for masculine, feminine, neuter and plural undefined and defined inflected word forms are as follows:

```
MCL&PST&DEF:="in">"jn",0>"ija";
MCL&PST&DEF:="in">"jn">"",0>"ijat";
FEM&PST:="">"in">"jn",0>"a";
```

```
FEM&PST&DEF:="in">"jn",0>"ata";
NEU&PST:="in">"jn",0>"o";
NEU&PST&DEF:="in">"jn",0>"oto";
PLR&PST:="in">"jn",0>"i";
PLR&PST&DEF:="in">"jn",0>"ite";
```

The possessive pronouns 'nash' (our) and 'vash' (your) does not realize phonetic alternations during the process of inflection (type M166). It uses the same inflectional rules as for the pronoun 'negov' (his). The only difference is that the inflectional rule for neuter gender undefined and defined forms are as follows:

```
NEU&PST:=0>"e";
NEU&PST&DEF:=0>"eto";
```

The most complicated inflectional rules (type M169) are for he pronoun 'tehen' (their) which realizes two phonetic alternations during the process of inflection only for the feminine and neuter gender (which are interpreted by applying the rules for replacement), and one phonetic alternation for the masculine and plural inflected forms.

```
MCL&PST&DEF:="en">"ni",0>"ja";
MCL&PST&DEF:="en">"ni",0>"jat";
FEM&PST:="en">"na","e":"ja";
FEM&PST&DEF:="en">"na","e":"ja",0>"ta";
NEU&PST:="en">"no","e":"ja";
NEU&PST&DEF:="en">"no","e":"ja",0>"to";
PLR&PST:="en">"ni";
PLR&PST&DEF:="en">"ni",0>"te";
```

The UNL account of the possessive pronouns uses complex A-rules for adding, suffixation, prefixation and replacement. It defines 5 inflectional types (M165-M169) very similar to that of the adjectives. The encoding uses related rules definition to generate inflected forms and accounts for sound alternations and irregular inflected forms.

Fig. 5 shows types of information included in lexical entry of Bulgarian possessive pronoun "moj" (my) which contains both grammar and semantic features. It relates morphological information for inflection to its assigned inflectional type PAR=M165, and shows all generated pronominal inflected forms. The specification POS=SPR is used for syntactic grammar rules and links pronoun to its related multilingual semantic synonyms.

Figure 5: The Bulgarian possessive pronoun "moj" (my) in UNL representation.

4.4.6 The semantics and the grammar features of reflexive-possessive pronoun

The semantics of the reflexive-possessive pronoun combines the semantics of the possession relationship and that of the reflexivity. That means it expresses the 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, and agrees in gender and number).

The reflexive-possessive pronoun is one and it has a full and a short form, and both they can be used with respect to agreement. However, only its full form 'svoj' (-self) has the inflectional grammar features of gender, number, and definiteness (Fig. 6), which are similar to that of the adjectives and of the possessive pronouns.

The DATR formal account of reflexive-possessive pronoun inflectional morphology uses the inflectional rules already defined at node Adj_5 and the same principle as for the possessive pronouns (Fig. 7).

The UNL formal interpretation of reflexive-possessive pronoun is consistent with the encoding of the possessive pronouns 'moj' (my) and 'tvoj' (your) and is exactly the same. It uses the inflectional rules already defined at inflectional type M165 and accounts for all related inflected forms (Fig. 5).

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

Figure 6: The reflexive-possessive pronoun's inflection tabular representation.

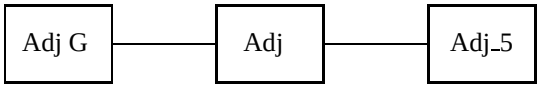

Figure 7: The reflexive-possessive pronoun inflectional types hierarchy.

4.4.7 Comparison between the formal representations

The above description presents encodings of inflectional morphology of Bulgarian possessive and reflexive-possessive pronouns using DATR language for lexical knowledge representation and using UNL. Both encodings define grammar rules for generation of all related inflected forms based on the use of inflectional grammar features of gender, number, and definiteness. For both interpretations, the pronouns are defined as lexical database at the dictionary by their base forms and their non-inflectional features of gender and person.

The DATR formal interpretation uses inheritance hierarchical representation of inflectional rules and concise encoding of 4 inflectional types, whereas UNL formal interpretation uses complex A-rules for adding, suffixation, prefixation and replacement and defines 5 inflectional types.

4.5 Syntax

The UNL interpretation of Bulgarian pronouns also uses transformation grammar rules to interpret syntactic grammar features of pronouns like case, definiteness, etc. The rules define syntactic function of a related pronoun and connect it to its lexical database entry by using specifiers POS=SPR, POS=PPR, etc. (Fig. 2 and Fig. 5). The above linking mechanism support multilingual representation with application to machine translation.

5 Conclusion

The analysed formal representation of Bulgarian pronouns in UNL frameworks uses semantic networks formalism to offer computationally tractable multilingual web-based application. It offers a model to present semantic and grammar features of pronouns in universal knowledge representation scheme which combines both grammar and lexical knowledge presentation.

It uses semantic frameworks which incorporate and link morphological and syntactic rules, and allows multilingual application for machine translation. Also, the comparison between two pronouns' formal representations using two types of semantic networks are discussed - by using DATR language for lexical knowledge representation and by using UNL.

The application is open for further development by incorporating new grammar rules, enlarging lexical database or to test and develop new models.

References:

[1] I. Boguslavsky, Some Lexical Issues of UNL, In J. Cardeosa, A. Gelbukh, E. Tovar, eds., *Universal Network Language: Advances in Theory and Applications. Research on Computing Science* 12, 2005, pp. 101–108.

[2] Ch. Boitet, I. Boguslavskij and J. Cardenosa, An Evaluation of UNL Usability for High Quality Multilingualization and Projections for a Future UNL++ Language, In A. Gelbukh, ed. *Proceedings of CICLing, Lecture Notes in Computer Sciences* 4394, 2007, pp. 361–373, Springer-Verlag.

[3] R. Evans and G. Gazdar, DATR: A Language for Lexical Knowledge Representation, *Computational Linguistics* 22(2),1986, pp. 167–216.

[4] N. Gomez, L. Mingo, J. Bobadilla, F. Serradilla and J. Calvo Manzano, Particle Swarm Optimization Models Applied to Neural Networks Using the R Language, *WSEAS Transactions on Systems*, issue 2, vol. 9, 2010, pp. 192–202, WSEAS Press.

[5] K. Ilieva, *Pronouns and text*, 1985, Bulgarian Academy of Sciences Publishing house (in Bulgarian).

[6] V. Kasyanov, Methods and Tools for Structural Information Visualization, *WSEAS Transactions on Systems*, issue 7, vol. 12, 2013, pp. 349–359, WSEAS Press.

[7] R. Martins, Le Petit Prince in UNL, *Proceedings from Language Resources Evaluation Conference 2011*, 2011, pp. 3201-3204.

[8] R. Martins, ed., *Lexical Issues of UNL: Universal Networking Language 2012 Panel*, 2013, Cambridge Scholars Publishing.

[9] F. Neri, Empirical Investigation of Word-of-mouth Phenomena in Markets: a Software Agent Approach, *WSEAS Transactions on Computers*, issue 8, vol. 4, 2005, pp. 987-994, WSEAS Press (Wisconsin, USA).

[10] R. Nicolova, *The Bulgarian Pronouns*, 1986, Nauka i izkustvo, Sofia (in Bulgarian).

[11] V. Noncheva, Y. Stancheva and V. Stoykova, The Little Prince project - Encoding of Bulgarian grammar, www.undl.org, 2011, UNDL Foundation.

[12] K. Popov and al., *The Grammar of Contemporary Bulgarian Language. Morphologia*, 2, 1986, Bulgarian Academy of Sciences Publishing house (in Bulgarian).

[13] V. Stoykova, Representing Lexical Knowledge for Bulgarian Inflectional Morphology in DATR, In N. Mastorakis, V. Mladenov, Z. Bojkovic and D. Simian, eds. *Latest Trends on Computers*, 2, 2010, pp. 612–616, WSEAS Press.

[14] V. Stoykova, The Inflectional Morphology of Bulgarian Possessive and Reflexive-possessive Pronouns in Universal Networking Language, *Procedia Technology*, 1, 2012, pp. 400–406, Elsevier.

[15] V. Stoykova, Visual Representation of Bulgarian Adjectives Inflectional Morphology, In M. Demiralp and A. Pisarchik, eds. *Advances in Applied Information Sciences*, 2012, pp. 91–96, WSEAS Press.

[16] V. Stoykova, Formal Representations of Bulgarian Possessive and Reflexive-possessive Pronouns, In O. Corbi, J. C. Metrolho, A. Lysko and R. Furferi, eds., *Recent Researches in Information Science and Applications*, 2013, pp. 144–149, WSEAS Press.

[17] V. Stoykova, Bulgarian Pronouns in Universal Networking Language, In O. Nakov, P. Borovska, A. Antonio, V. Mladenov, L. Zinchenko and A. Fuentes-Penna, eds., *Recent Advances in Computer Science*, 2013, pp. 142–145, WSEAS Press.

[18] V. Stoykova and Ch. Lozanov, Visual Semantic Network of Bulgarian Possessive and Reflexive-possessive Pronouns Inflection, In O. Nakov, P. Borovska, A. Antonio, V. Mladenov, L. Zinchenko and A. Fuentes-Penna, eds., *Recent Advances in Computer Science*, 2013, pp. 138–141, WSEAS Press.

[19] H. Uchida, M. Zhu and T. Della Senta, *Universal Networking Language*, 2005, UNDL Foundation.

[20] M. Voznak, Non-Intrusive Speech Quality Assessment in Simplified E-Model, *WSEAS Transactions on Systems*, issue 8, vol. 11, 2012, pp. 315-325, WSEAS Press.

[21] UNDL Foundation, 2011, URL http://www.undl.org.

[22] The DATR Web Pages at Sussex, 1997, http://www.cogs.susx.ac.uk/lab/nlp/datr.