

Bulgarian **R**esource **G**rammar – **E**fficient and **R**ealistic (BURGER)

Version 0.2



Technical Report

Petya Osenova

LingoLab, CSLI
Stanford

2010

Table of Contents

0. Introduction.....	3
I The Matrix Language Design and Bulgarian Specific Features.....	4
II The General Lexical Types and their features.....	6
1. Nouns.....	6
2. Adjectives.....	9
3. Adverbs.....	10
4. Prepositions.....	10
5. Numerals.....	11
6. Complementizers and Subordinators.....	11
7. Particles.....	13
8. Pronouns.....	14
8.1 Personal Pronouns.....	16
8.2 Possessive Pronouns.....	16
8.3 Reflexives.....	17
8.4 Determiner Pronouns.....	17
8.5 Relatives.....	18
9. Verbs.....	18
9.1 Main Verbs.....	20
9.2 Verb-Clitic Clusters.....	22
9.3 Auxiliaries and Modals.....	23
9.4 Copula.....	24
9.5 Impersonal Verbs.....	25
9.6 Participles, Gerund and Deverbals.....	25
10. Substantivization and Nominalization.....	26
11. Passivization and Deverbal Nouns.....	27
III Phrase Types.....	28
IV Lexicon.....	29
V Punctuation.....	30
VI Conclusion.....	31
Appendix 1: Lexicon types.....	33
Appendix 2:	44
Testset in Bulgarian with 212 sentences (194 positive and 19 negative)	
I Phenomena.....	44
II Specificities of translated testset examples.....	45
III Index of the testset examples (as in the testset).....	45
IV What do the negative examples check?.....	50
References.....	51

0. Introduction

BURGER was started as a customized matrix grammar for Bulgarian in January 2010 and as part of my 5-month Fulbright project at CSLI, Stanford. The project would not have been possible without the strong support of my host professor Ivan Sag, the patience and great collaboration of my grammar mentor and co-worker Dan Flickinger, my own enthusiasm (despite my little initial knowledge in LKB and Matrix world matters), Kiril Simov's strong belief that I can make it, and last, but not least – the Fulbright Programme via its local branch in Bulgaria, and its main headquarters in the USA.

To the best of our knowledge, this is the first quite exhaustive implementation of a grammar for a Slavic language in the family of Matrix grammars.

Matrix is intended as a typological core initiating the grammar writing about a specific language. Its purpose is also to ensure a basis for comparison among various language grammars. Although it supplies the skeleton of the grammar and is based on the experience with several languages, a number of changes had to be made, which override the Matrix suggestions, and which are required by the Bulgarian data.

Additionally to the Matrix, two other sources were used for the BURGER development.

1. the documentation on the already constructed grammars for other languages (English, Spanish, Portuguese, Modern Greek, German, etc.)¹, as well as the grammars themselves (considered as best practices) and 2. regular professional consultations and collaboration with Dan Flickinger – the developer of the English grammar (ERG).

At the moment, the grammar exists in two variants:

- as a 'toy grammar', which concentrated on the syntactic constructions coverage, and morphology has been added just to supply the necessary morphosyntactic data (**Version 0.2**)
- as a full 'morphologically-augmented grammar', which incorporates most of the verbal morphology for Bulgarian (for all personal verbs, partly for the impersonal and auxiliaries) - adjusted to the current syntactic coverage of the 'toy grammar' (**Version 0.3**).

Our aim is to continue developing further the 'morphologically-augmented grammar'. Despite the fact that the current report concentrates on Version 0.2, some brief notes come in order with respect to Version 0.3. The two grammars show comparable coverage results. They differ only in the verbal type hierarchy organization, which affects also the specificity of the inflectional types and rules, as well as the verb entries in the lexicon. Additionally, the initial loading of the rich grammar takes more time due to the big number of rules.

Bulgarian is a rich-inflected language, but in contrast to other Slavic languages, its richness lies in the verbal system, rather than in the nominal one. Thus, the transfer of the verbal inflections from the pre-existing morphological dictionary (Popov et. al 2003) resulted into over 2600 rules. At the moment, the 'morphologically-augmented grammar' transfer contains only the full morphology of the personal verbs. However, the addition of impersonal verbs as well as the other POS would not be a problematic or time-

¹ For this technical report I profited mostly from the Technical Reports on Portuguese and Modern Greek.

consuming procedure in the established setting. The idea behind the full morphological verb design is as follows:

- since Bulgarian verbs show a lot of alternations and irregularities across their grammatical categories (conjugation, tense, aspect, finite vs. infinite forms, other synthetic grammatical categories, such as imperative, etc.), the full paradigms per conjugation in the lexical types (Version 0.2 strategy) were abandoned as a generalization opportunity. Instead, each verb type was viewed as a combination of the appropriate subparadigms from the given morphological and/or lexical categories.
- the set of the respective subparadigms per category was attached to each verb in the lexicon. Thus, the lexicon was also “dressed” with the morphologically specific information for the distinct verbs. An example follows below for the verb *чета* (read):

```
cheta := v_np_i_le &  
[ STEM < "чета" >,  
  SYNSEM [LKEYS.KEYREL.PRED "чета_v_rel",  
    LOCAL.CAT.HEAD.MCLASS [FIN-PRESENT finite-present-013,  
      FIN-AORIST finite-aorist-028,  
      PART-AORIST participle-aorist-092,  
      FIN-IMPERF finite-imperf-014,  
      PART-IMPERF participle-imperf-015,  
      PART-PRESENT participle-present-006,  
      GERUND gerund-a-ejki,  
      PART-PASSIVE participle-passive-022,  
      FIN-IMPERATIVE imperative-012] ] ].
```

Since the integration of the developed ‘toy grammar’ syntactic types within the morphology-rich setting was done in the last days of my stay at Stanford, this report concentrates on the grammar world of Version 0.2. The morphology there is presented in the common way – lexical entries specify the lemma characteristics. Then, the rules change them into the respective wordforms. The lexicon remains clean. Needless to say, the generalizations are not inflection and category exhaustive.

However, the reader should keep in mind that all the phenomena are the same in both versions except for the verb presentations (hierarchy, rules, types). For example, in Version 0.2 the valency types are more numerous, but the lexicon is cleaner. In version 0.3 it is the opposite – smaller number of types, but morphologically “dressed” lexicon. Also, the test set is the same. The coverage results are comparable. In the future developments other changes in the morphological organization are expected.

I. The Matrix language design and Bulgarian specific features

The Matrix grammar promotes a semantic approach to the description of a language. For example, the verbs, the adjectives, the adverbs and the prepositions are events. Bulgarian, however, lacks a grammar, which describes all the phenomena at the semantic-syntactic interface. The existing research is mainly focused on the morphology and syntax.

Matrix provides the general ‘universal-oriented’ concept on the language phenomena and set of templates to follow and develop further. Ideally, one should only inherit from

Matrix, without changing it. In reality, each language challenges and is challenged by Matrix (see the discussions for Portuguese, Spanish, Modern Greek, etc.).

On the one hand, Matrix predefines some phenomena too strictly, on the other - gives possibilities for generalizations. All this is inevitable, since the ideal granularity between specificity and universality is difficult to establish. For that reason, the various language-driven grammar developments for each distinct language seems the most adequate strategy for the further improvements of Matrix being a universally-intended grammar core.

The Bulgarian rich morphology always seems to challenge the semantic approach. Thus, the adjectives, participles, numerals happen to have morphologically definite forms, while the definiteness marker is not a semantic property of these categories. For that reason, the most important thing in the grammar was to keep Syntactic and Semantic features separate (for example, agreement - in accordance with the ideas in Kathol 1997). In this way, the definiteness is operated either via the **MOD** feature, or via nominalization rules. In the first case (via the **MOD** feature), the event selects for a semantically definite [**SYNSEM.LOCAL.HOOK.INDEX.DEF+**], but morphologically indefinite noun [**LOCAL.AGR.DEF-**]. Bulgarian systematically differentiates among the semantic contribution and the morphological marking of the same category. This fact posited some difficulties in the starting design, since the same category had to be approached as semantic vs. morphological, and thus has to live in two different places in the structure (see above the example for **DEF**; also the tense can be accessed syntactically [**SYNSEM.LOCAL.CAT.HEAD.TAM.TENSE** *tense*] or semantically [**SYNSEM.LOCAL.CONT.HOOK.INDEX.E.TENSE** *tense*]). These issues have been taken into account in the current version.

Very often, the generalization cannot be kept at higher levels, because of the variety in the morphosyntactic behaviour types within the Bulgarian constructions.

Some changes, made in Matrix, are listed below:

- The determiner needs more than one relation (for cases like *whose*).
- New head types had to be incorporated into the disjunctive head types.
- The index of the modifying adjectives and adverbs, and their **MOD** index should not be structure-shared.
- The modifier phrase should allow also **QUE** non-empty modifier daughters.
- The optional-complement phrase had to allow also non-event head-daughters.
- The subject-head rule had to allow also interrogative pronouns as their subjects (i.e. **QUE** *1-dlist*).

Despite these changes, the aim of BURGER was to remain as close as possible to the current linguistic design of the Matrix Grammar.

This version 0.2 differs from version 0.1 in the following ways: it still covers the Matrix multilingual testset, but also has negative examples included as well as some probe sentences from BulTreeBank. The sentence final punctuation is added. Some phenomena are re-designed, and ambiguity is reduced slightly (see the Appendix with the Bulgarian sentences).

The grammar contains the following language specific files:

- a hierarchy of general lexical types (**bulgarian.tdl**)
- specific lexical types (**bg-le-types.tdl**)
- phrasal types (**bg-phrase-types.tdl**)
- inflectional types (**bg-infl-types.tdl**)
- inflectional rules (**irules.tdl**)
- lexical rules (**lrules.tdl**)
- lexicon (**lexicon.tdl**)
- labels (**labels.tdl**)
- root (**root.tdl**)

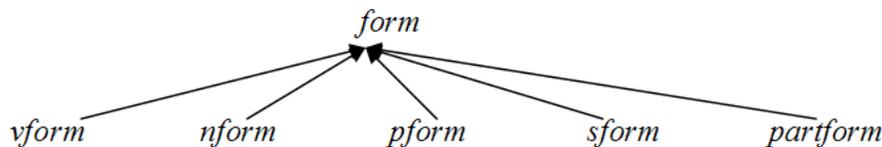
In the next chapters the basic types, features and principles behind the design will be presented.

II. The General lexical types and their features

There are features for the types in the grammar, which are defined globally. Such a feature is **FORM**, which is assigned to the type *head*.

head := [**FORM** *form*].

The hierarchy follows below:



The type *form* has the following subtypes depending on the *head* type: verbal form subtypes (*vform*), preposition form subtypes (*pform*), nominal form subtypes (*nform*), subordinator form subtypes (*sform*), particle form subtypes (*partform*).

The further subtype distinctions will be presented in the respective sections.

Another example for a feature, assigned globally, is the **SF** (sentence illocutionary force). It is assigned to event or referent indices: *event-or-ref-index* := [**SF** *iforce*].

1. Nouns

The nouns are traditionally divided into common and proper ones. Their head is a non-pronominal noun (*nonpro-noun-hd*), since the distinction is made between pronoun and non-pronoun heads (for more details see the Section on Pronouns). Most nouns have an empty **MOD** feature, because they usually are not modifiers. The nouns have constant **PERSON** value (*3rd*).

The **AGR**(eement) feature is used to encode the **PNG** (*person, number, gender*) values:

png := [**GENDER** *gender*,
 NUMBER *number*,
 PERSON *person*].

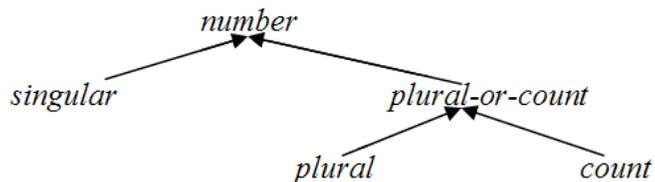
It also accommodates the feature **DEF**(inite), which is reserved for the definite nouns. In contrast to English, Bulgarian definite article is a phrasal affix in an NP, and also - part of the nominal word. The feature **DEF** is boolean, and it is separated from **PNG**, because nouns agree with their determiners or modifiers in **PNG**, but never in **DEF**.

AGR [**PNG** *png*,
 DEF *bool*]

The semantic contribution of the **DEF** feature is attributed only to referential indices, but its morphological contribution is allowed also for the events (such as, adjectives, participles, numerals).

Each noun belongs to one of the three gender categories: *masculine*, *feminine* or *neuter*. Thus, the **GENDER** feature has these three values. (The so-called ‘common-gender’ ones are not included in this version²).

The **NUMBER** feature has the following structure:



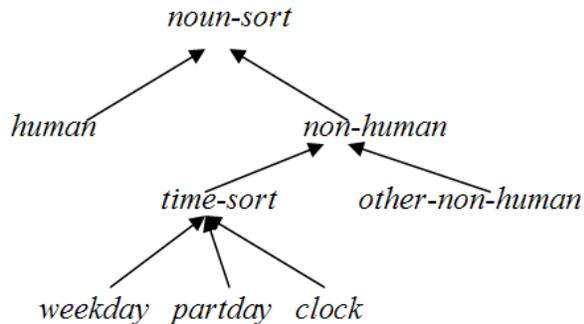
Some nouns (*masculine, non-human*) in a special position (after numerals) take the count form. The count form is a special plural form, remnant of the old dual form. In the same context, all other nouns take their normal plural form. For that reason, the type *plural-or-count* was introduced. For example, *десет стола* (‘ten chairs-count’, ten chairs) or *десет лъва* (‘ten lions-count’ ten lions), but *десет студенти* (‘ten students-masc, plural’, ten students), *десет жени* (‘ten women-fem, plural’, ten women), *десет деца* (‘ten children-neut, plural’, ten children).

Since the nouns, taking the count form are not only bound grammatically to one member of **GENDER**, namely *masculine*, but are also restricted by a non-grammatical property (being *non-human*), some fine-grained semantic categorization was incorporated. This increased the masculine types with two subtypes: that have the semantic **SORT** value *human* or that have the semantic **SORT** value *non-human*. However, the semantic sort categorizations had to go further, because of other specificities, which are discussed below. In future, even more semantic elaborations are expected to be required.

² These are nouns of the type *роднина* (relative), which might be feminine or masculine: *един роднина* (one relative-masc) or *една роднина* (one relative-fem).

Although most nouns are not modifiers, i.e. have an empty **MOD** feature, some of them do modify events. These are ‘time nouns’ which are part of modifying NPs of the type *всеки ден* ‘every day’, *тази вечер* ‘this evening’, *тази пролет* ‘this spring’, etc. Thus, additional partitioning within *non-human* type was introduced: *time-sort* vs. *other-non-human*. The *time-sort* was further subclassified into more specific sorts: *weekday*, *partday* and *clock*.

Here is the hierarchy:



There is one type - *basic-temporal-noun* – which generalizes over the non-empty **MOD** feature with respect to verb and noun heads. The noun, which participates in phrases, such as *всеки ден* ‘every day’ inherits directly from it with a restriction on verb heads only being modified. Thus, the value for the **SORT** is *time-sort*.

(1) Кучето лаеше всеки ден.
 dog-the was-barking-imperfective every day
The dog barked every day.

The same holds for the *weekday* type. Only the value for **SORT** is the more specific *weekday*, since it can be modified by parts of the day, for example.

(2) Браун пристигна във вторник сутринта.
 Brown arrived in Tuesday morning-the
Brown arrived on Tuesday morning.

The *partday* noun in general might inherit directly from the *basic-temporal-noun*. However, a special type is provided for cases, in which it modifies other time nouns (such as, the *Tuesday morning* case). At the moment, this type is kept for restricting this noun-noun modification. The distinction is also justified by the differing selectional properties of the verbs and the nouns. In contrast, the type *basic-clock-noun* inherits from the *basic-common noun*, since it has an empty **MOD** value.

The appositions, such as *числото пет* ‘the number five’, are treated as compound phrases. The head-daughter (usually the first noun in the compounding) establishes a *compound-relation*, in which the head noun is bound as its first argument and the non-head noun - as its second argument. The head daughter is constrained by having a **HEAD**

value *compound*, and a **FORM** value *compound-init*. In Bulgarian there are, in fact, two morphological compound types for the word order *head-nonhead*. The first is when the head noun is or can be definite – *числото пет* ('number-the five', the number five), *планетата Земя* ('planet-the Earth' the Earth). This type is currently implemented. The second is when the head noun is always indefinite – *ляля Мария* (Aunt Maria), *улица "Луна"* ('street "Willow"', Willow Str.), *река Марица* ('river Maritsa', the Maritsa river). This second type needs another rule to be added. Also, the word order type *non-head-head* has not been implemented yet. It refers to phrases, in which the first noun is a loan from English (in most cases), and the second is a domestic word (*бизнес среща* 'business meeting').

The common nouns are further subclassified into intransitive and transitive - complement-taking - NP, PP or clauses.

The relational nouns are treated semantically as nouns with an additional argument (*noun-arg1-relation*), and syntactically – as taking PP complements.

(3) Снимката на Абрамс пристигна.

Picture-the of Abrams arrived.

The Abrams's picture arrived/The picture of Abrams arrived.

This analysis acknowledges the relational nature of the noun, but does not delve into the more specific semantic ambiguities or distinctions, such as possession and the subsumed relations (part-whole, kinship, etc.).

The proper nouns have the relation *named-relation*. They are only intransitive. The proper nouns at the moment do not undergo further subclassifications as person, location, organization, etc. This is left for the next stages of the grammar development.

Also, elaboration on more specific properties of noun heads will be added later (count vs. mass nouns; deverbal nouns, which do not undergo the typical derivation process by the suffix *-не* on the imperfective verbs; further types of noun compounds (as mentioned above), etc.)

2. Adjectives

In general, the adjectives might be intersective or scopal. The typical adjective is viewed as an intersective modifier, which selects for a non-definite noun, while structure-sharing all **PNG** characteristics with it. The comparative adjectival forms are introduced in a distinct type, which adds a *comparative-relation* to the inherent *event-relation*. In the canonical ordering, all adjectives appear before the nominal head.

The adjectives, which appear as second predications to non-auxiliary verbs, are treated as scopal. For example:

(4) Котката се преследваше сама.

Cat-the(fem,sg) se-refl was-chasing-imperf alone(fem,sg)

The cat chased herself.

The external argument (**XARG**) of *сама* (alone, by herself) should be structure-shared with the subject of the verb. Also, the subject and the adjective structure-share their **PNG** characteristics.

As it was mentioned above, the morphological behaviour of the adjectives conflicts its event-based semantic status. Thus, both contributions were separated with the help of the **AGR** feature being licensed at two places: morphological and semantic. Let us take as an example a definite adjective, which always selects for a non-definite noun: *старата котка* ('old-the-fem,sg cat-fem,sg', the old cat). Thus, when having a definite article, the adjective is ensured to select for a semantically definite noun via its **MOD** feature: [**SYNSEM.LOCAL.CONT.HOOK.INDEX.DEF+**]. In this way, the semantics contribution is handled without destroying the morphological behaviour of the modifier.

3. Adverbs

The adverbs follow the Matrix differentiation into intersective (*сега* (now), *тихо* (quietly)) and scopal ones (*вероятно* (*probably*)). Here the general adverbs cross-classify with the pronominal adverbs. They all share the same head (*adv*), but the pronominal ones also show a special value of the **FORM** feature (*interrog-reg*), and inherit additionally from the *pro-adv* type. The proliferation of types, which sometimes equal the lexical entry (such as, *almost-adv-le* or *kolko-noun-adverb-le*) is due to the fact that some adverbs modify various heads, which leads to a necessity for further constraints over the phrases these heads project.

(The hierarchy of heads for nonpronominal and pronominal nouns, adjectives and adverbs is given in the Section on Pronouns for consistency).

4. Prepositions

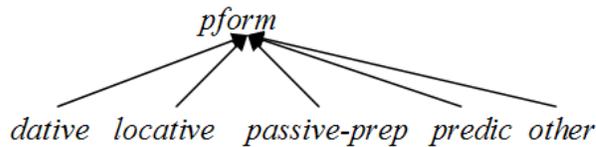
The prepositions are called adpositions (in the lines of Matrix). Their head value is *adp(osition)* and they obligatorily take complements. Also, they always precede their complements. They are divided into two subtypes: modifying prepositions, which have semantics, and non-modifying ones, which have no semantics, and are selected as complements by various heads. The first type behaves uniformly, while the semantically vacuous types exhibit some specificities. Here is an example for a modifying preposition:

(5) В градината имаше котки.

In garden-the had-imperfective cats

There were cats in the garden.

Concerning the semantically vacuous prepositions, there is a special type for the dative preposition (*на Иван* ('to Ivan')), which is used for verbs, taking dative obliques (such as *давам* (give)). There is a passive preposition (*от Иван* 'by Ivan'), which ensures the mapping from the subject to the corresponding oblique PP. There are also some other subtypes: predicative (*смятам някого за някакъв* 'I consider someone for something') and locative (for verbs like *сложих нещо в джоба* 'I put something in the pocket'). The classification is open to further refinements and additions. Here is the hierarchy of the **FORM** values, which are of type *pform*:



The hierarchy is flat, and thus the value *other* is assumed by default for the modifying prepositions.

5. Numerals

Numerals have the value of **HEAD** *num*. Otherwise, they are viewed as adjective-like intersective modifiers. The cardinal numerals from number 2 up via their **MOD** feature select for a *plural-or-count* noun. The numeral ‘one’ has a distinct type, since it shares its **PNG** characteristics with its head. In order to distinguish among number 1 and the rest, a value *num-sort* was introduced for the semantic feature **SORT**. The number 1 is assigned a value *one-sort*, while the other cardinals receive the value *several-sort*.

The type for ordinal numerals at the moment only specifies the structure-sharing of the **PNG** features between the numeral and its head. However, it will be elaborated in the next version.

There is a special type for the clock times. It is necessary, because the clock numeral modifies the masculine noun *час* (‘hour’) only in its definite form, thus showing idiosyncrasy.

(6) Абрамс пристигна в три часа.
 Abrams arrived in three hour-masc,def
Abrams arrived at three o'clock.

The words *много* (‘many’, ‘much’) and *малко* (‘little’, ‘few’) are also treated as numerals. They receive the **SORT** value *several-sort*. However, they have more types depending on the modified head (*verb*, *noun*, *adverb/adjective*). For example, when modifying a noun, they select its plural form and are in a preposition. Also, in this case the words can have definite forms (*многомо жени* (‘many-the women’, the many women), *малкомто мъже* (‘few-the men’, the few men); when modifying an adjective or an adverb, they are always in preposition:

(7) Абрамс беше много стар.
 Abrams was very old
Abrams was very old.

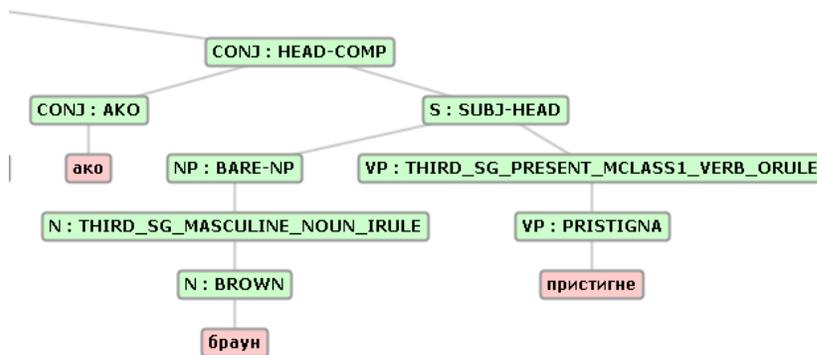
When modifying a verb, there are no ordering restrictions (no examples in the present Testset).

6. Complementizers and Subordinators

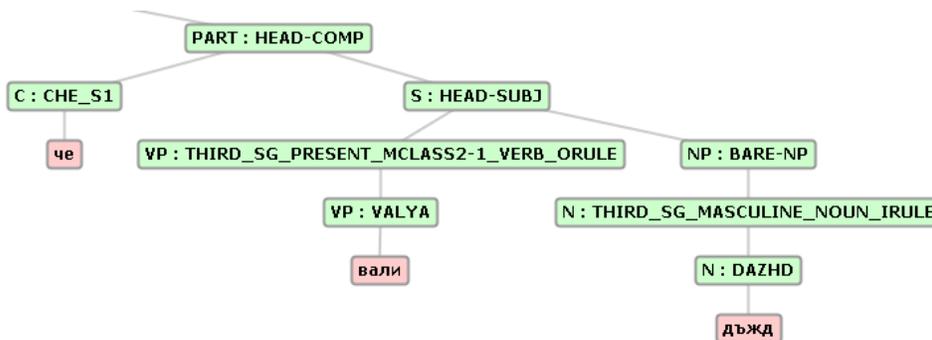
The complementizers and subordinators follow strictly the Matrix grammar specifications. Both types are used for convenience, since subordinators have a non-empty **MOD** feature (being modifiers), while the complementizers have an empty **MOD** feature (being clausal complements). Thus, the latter are subcategorized by the head.

Both – complementizers and subordinators select for their clauses as their complements. Two trees are given below, which are visualized in the CLaRK system. In picture [1] a clause excerpt is presented (‘if Brown arrives’), in which the subordinator *if* selects for a clausal complement. In picture (2) the clause excerpt (‘that rains rain’) presents a complementizer, which selects for a clausal complement:

[1]



[2]



There is a special type for the interrogative complementizers (like *дали* (whether)), since they are particles and propagate the interrogative value of the feature **SF ques**. These complementizers have the head *partcomp*, which inherits from both heads - *parthead* and *comp*(lementizer).

In the Bulgarian linguistic literature, it is common a complementizer role to be assigned also to the infinitival *да* ‘to’ construction in some cases. However, in our analysis this form is treated only as a nonfinite verbal form, which takes a lexical verb and inherits all its complements (see more in the Section on Verbs). The interrogative pronouns, which also happen to be in the complementizer position, are treated differently,

since they have varying grammatical roles within the clause. In this way, the complementizer type is reserved only for *че* (that) and *дали* (whether).

The subordinators modify the main clause as scopal modifiers. The subtypes follow the names of the specific subordinators, which are presented as values of the feature **FORM**: *ако* (if), *за да* (in order to), etc. The *че* (that) can serve not only as a complementizer, but also as a subordinator. For that reason, there are two distinct types in the lexicon.

7. Particles

The particles have the value of the **HEAD** *parthead*. In Bulgarian grammars the ‘se’ form is differently viewed as part of the pronominal pronoun system (reflexive pronoun):

(8) Тя се вижда. = Тя вижда себе си.
She ‘se’-short reflexive sees = She sees ‘sebe si’-full reflexive
She sees herself.

As a formation particle for the passive voice system:

(9) Кучето се преследва от Браун.
Dog-the ‘se’ chase-3rd person,sg from Brown
The dog is chased by Brown.

And as part of the lexical meaning of a verb:

(10) Той се учуди.
He ‘se’-reflexive wondered.
He wondered/He was amazed.

In our design this separating function-based strategy is not kept. All usages are handled by one type – namely, the accusative clitic pronoun type. Thus, more information is provided in the Section on the Pronouns and Verb-Clitic Clusters.

Another group of particles are the ‘question polar particles’: *ли* (li) and *нали* (nali). The first one is the neutral *yes-no* particle, while the second is the marked one, which seeks for the hearer’s confirmation on some fact (similar in meaning to the tag questions in English). They are treated as intersective modifiers. The proliferation of subtypes comes from the fact that *li*-particle can focalize NPs, VPs or whole sentences, while the *nali*-particle modifies only sentences or VPs. Compare the following examples where examples (11ab) refer to the neutral polar particle, while (12abc) refer to the tag-like particle:

(11a) [Кучето изляя] **ли**? or Кучето [изляя **ли**]?
Dog-the barked ‘li’-polar particle?
Did the dog BARK?

(11b) [Кучето **ли**] изляя?

Dog-the ‘li’-polar particle barked?
 Did THE DOG bark?

(12a) Кучето изляя, **нали**?
 Dog-the barked, ‘nali’-tag particle?
 The dog barked, didn’t it?

(12b) **Нали** кучето изляя?
 ‘Nali’-tag particle dog-the barked?
 The dog barked, didn’t it?

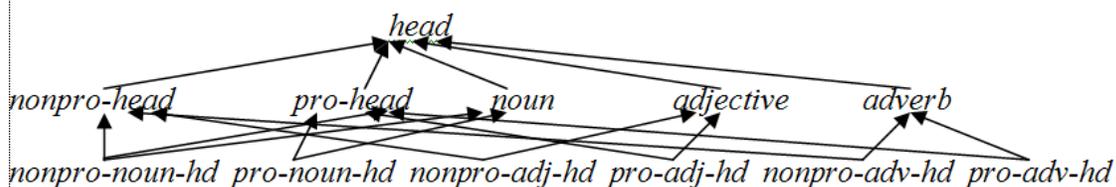
(12c) Кучето [**нали** изляя]?
 Dog-the ‘nali’-tag particle barked?
 The dog barked, didn’t it?

8. Pronouns

In Bulgarian, there are 9 or 10 pronoun types, depending on whether the reflexive is viewed as one distinct category or as part of either of both - personal and possessive pronouns. Here we adopted the second strategy. Thus, the types are as follows: *personal, possessive, personal-reflexive, possessive-reflexive, demonstrative, indefinite, interrogative, relative, negative, and collective*.

The pronouns are a non-homogenous group with respect to function, morphosyntactic behaviour among the pronoun classes and within them. Thus, the design here is more complex.

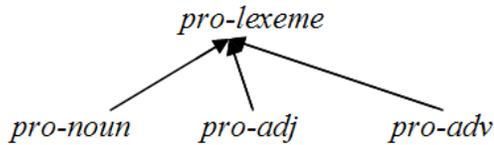
Since pronouns cross-classify functionally with other POS, typically: *nouns, adjectives* and *adverbs*, this fact reflected the head type hierarchy (presented below). Two further subtypes of head were introduced – *pro-head* (for pronominal lexemes) and *nonpro-head* (for non-pronominal lexemes). Then, the heads inherit either from *pro-head*, or from *nonpro-head*, and from either of the types: *noun, adjective* or *adverb*.



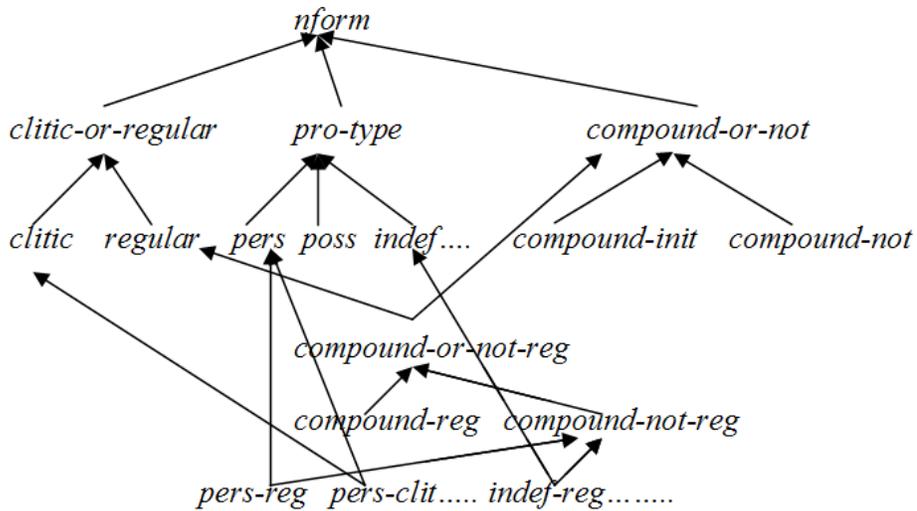
In this way, the noun, adjective and adverb heads are subdivided into pronominal and non-pronominal ones. Then, a type *pro-lexeme* was introduced, which has *pro-head* as its value for **HEAD**.

Although in Bulgarian (similarly to Portuguese) only pronouns are marked with syntactic case, the feature is assigned to the noun. In this way, the case is used not only as a morphological marker, but also at the syntactic level of the grammatical roles.

The pro-lexeme type is further subdivided into three types:



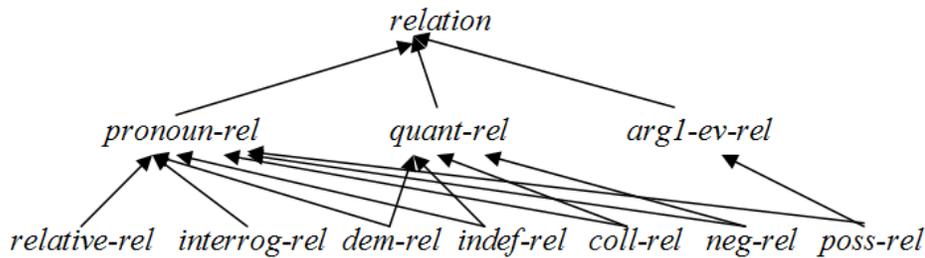
Additionally, the values of the *nform* feature are as follows:



The idea behind this classification is that, on the one hand, the type of the pronoun has to be encoded as a value of the feature **FORM**. At the same time, the fact has to be reflected that the non-nominative personal and the possessive pronouns can have either regular, or clitic forms. The other pronominal types have only regular forms. However, this hierarchy has also to interact with the *compound-or-not* classification of nouns in order some clashes of features to be avoided. For that reason, the typology became more complex. Thus, the types further cross-classify into more specific types: *pers-reg* (personal pronouns with regular form), *pers-clit* (personal pronouns with clitic form). Note that not all the hierarchy is given for space and clarity reasons. Note also that the reflexivity is actually handled at the semantic **SORT** level instead of at the feature **FORM**.

For the pronouns, the feature **SEMSORT** has value *pro-sort*. It is further subdivided into *norm-pro* and *refl-pro*. In this way, the personal and possessive reflexive types are contrasted to the remaining non-reflexive types.

The hierarchy of pronominal relations is presented on the next scheme:



The picture is simplified at the level below the *relation* type, and also with respect to the full relation names. As it can be seen, the *relative* and *interrogative* relations inherit only from the *pronoun* one. The relations of the determiner pronoun classes inherit also from the *quant-relation*. The *possessive-relation* inherits from *pronoun* and *arg1-ev-rel*.

With respect to the ordering of the pre-modifiers and determiners, at the moment some restrictions are added, such as the adjective or adjective-like modifiers cannot select for definite nouns or NPs. However, more work is needed for the elaboration of the exact element positions within the NP (Osenova 2009).

8.1 Personal Pronouns

The personal pronouns have the following categories: *person*, *case*, *number*. Only 3rd person forms have also the morphological category *gender*. The personal pronouns distinguish among clitic and regular forms in accusative and dative cases. The nominative forms have only regular forms. Additionally, they have reflexive counterparts, which do not have a nominative case. They differ from the non-reflexive ones by the value of the feature **SORT** *refl-pro*.

The clitics are viewed as lexical projections of the head (i.e. operated by special rules), while the regular forms are treated as head arguments (complements) (i.e. operated by head-complement principles). The mechanism of handling the clitics is explained in the Section on Verb-Clitic Clusters. Also, the clitic doubling phenomenon in Bulgarian is analyzed in this version.

8.2 Possessive Pronouns

The possessive pronouns have the following categories: *syntactic gender*, *possessor-gender*, *person*, *number*. They also distinguish among clitics and regular forms, on the one hand, and, reflexive and non-reflexive forms, on the other. In contrast to the personal pronouns, possessives inherit from *pro-adj* type, since they behave morphosyntactically as adjectives. In order to capture the information about the possessor-gender, a two-argument *possessive-relation* was introduced. Its **ARG1** is the index of the modified noun, i.e. the syntactic gender. Its **ARG2** is its semantic index, i.e. the possessor-gender.

Regular possessive forms canonically come before the noun head – *моя книга* (my book). The possessive clitics canonically come after a definite head – *книгата ми* ('book-the my-clitic', my book). The grammar parses also examples in which the possessive clitic takes the second position after a definite adjective within the NP: *старото ми куче* ('beautiful-the my-possessive clitic dog', my beautiful dog).

However, the idiosyncrasy is not handled in this version. This means that cases, in which non-definite nouns can be modified by a possessive clitic (such as the kinship notions – *майка ми* ('mother my-clitic', my mother)), as well as possessive clitics, which are raised in a preverbal position (*взех му палтото* 'took-I his-clitic coat-the', I took his coat) are not considered yet (more information on the clitic doubling and possessive raising phenomena can be found in (Avgustinova 2007), although there are differences in the grammar modeling).

8.3 Reflexives

The personal and possessive reflexives are handles respectively. At the moment, the personal dative *си* 'si' is not present. Only *се* 'se' is analyzed. The same holds for the possessive *си* 'si'. It should be noted that reflexivity in Matrix can be considered only on the syntactic level. The semantics is not reflected.

8.4 Determiner Pronouns

The following types of pronouns are analysed as determiners: demonstratives, indefinites, collectives, negatives, interrogatives. Two major points are worth mentioning here. Since the head is *det*, these lexical types are underspecified with respect to their functional part-of-speech, but they inherit from their genealogical supertype, which introduces the distinct for the pronoun class relations, and provides the specific value of the feature **FORM**. Also, while the entity pronouns behave like determiners, it is not the case for the rest of the class members (Needless to say, the pronominal adverbs keep only their derivation from the respective pronominalclass, but they are never treated as determiners. At the moment, only the interrogative ones are introduced).

For example, the demonstrative *този* (this-masc,sg), *тази* (this-fem,sg), *това* (this-neut,sg), *тези* (these-pl) is the most left element in the NP group, while for *такъв* (such-masc,sg) and its wordforms this is not the case. It can come after the form *един* 'one', for example:

(13) Видях една такава голяма котка.

Saw-I one such big cat.

I saw such a big cat.

Another illustration is the indefinite pronoun *нещо* (something-neut,sg), which can have a demonstrative determiner – *това нещо* (this-neut,sg something-neut,sg). The situation seems to be more uniform with respect to the negatives and the collectives. The interrogatives also mark the left boundary of the NP, hence they are treated as determiners. However, they are a special case, because they introduce the additional *interrogative-relation* on the top of the single *quantitative relation*. The possessive interrogative pronoun even adds one more relation – the *possessive* one. Interestingly, the negative particle *нищо* (neither.., (nor)) is treated as a determiner also. The polar negative particle *не* (no) is treated as a special kind of a verb (see the Section on Verbs).

Thus, some alternative non-determiner analyses, which inherit only from the pronoun subtypes, are introduced for the appointed special cases. This task needs further elaboration towards the full implementation of the pronoun class members.

8.5 The Relatives

Together with the interrogatives, this pronoun type introduces a non-empty **NON-LOCAL** value for the respective feature (for the interrogatives it is the feature **QUE**, while for the relatives it is the feature **REL**). The relative pronoun itself has an empty **MOD** value, and no semantic relation, which to contribute. The challenge is in the Relative Clause type, which will be presented later. Thus, here the Matrix grammar approach is more or less construction-based.

9. Verbs

The verb head is assigned the following set of features:

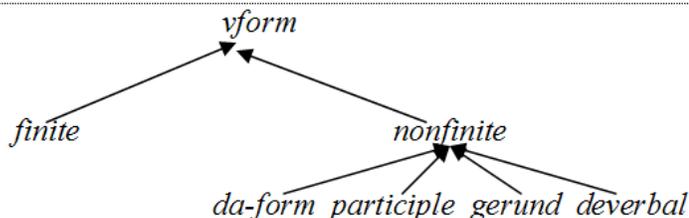
```
verb :+ [ AUX bool,  
         TAM tam,  
         FORM vform,  
         MCLASS mclass,  
         VOICE voice ].
```

The first one (**AUX**) shows whether the verb is auxiliary, or not. The **TAM** feature is used as in Matrix (**TENSE** *tense*, **ASPECT** *aspect* and **MOOD** *mood*). **FORM** has value of type *vform*, and has to do with being a finite or a non-finite form. The **MCLASS** (stands for *morphological class*) shows the division of Bulgarian verbs into 3 conjugation classes: first, second and third. The **VOICE** feature takes values of active or passive voice. Let us discuss these features in turn.

The [**AUX** +] is reserved for: the verb *съм* ('to be'), which participates in the formation of the analytical tenses; the 'da-construction' and *ще* ('will') marker for future tense. It also applies to the modal verbs.

The second feature needs more explanations. The values of **TENSE** handle only the synthetic forms: *present*, *aoarist* and *imperfect*. When *ще* (will) takes a finite present verb form, it also supplies the value *future* for the **TENSE** feature. There is a supertype *aoarist-or-imperfect*, since systematically 3rd conjugation verb forms coincide (except for 2 and 3 person, singular). The **ASPECT** feature is used for encoding the two types of morphological aspect: *perfective* and *imperfective* (*дам* ('give-perf')/*давам* ('give-imperf')). However, they have a supertype *perfective-or-imperfective*, because some verbs of foreign origin are considered biaspectual. The values of **MOOD** are *indicative*, *imperative* and *conditional*, although *conditional* applies to analytical forms, and *indicative* to both – analytical and synthetic forms.

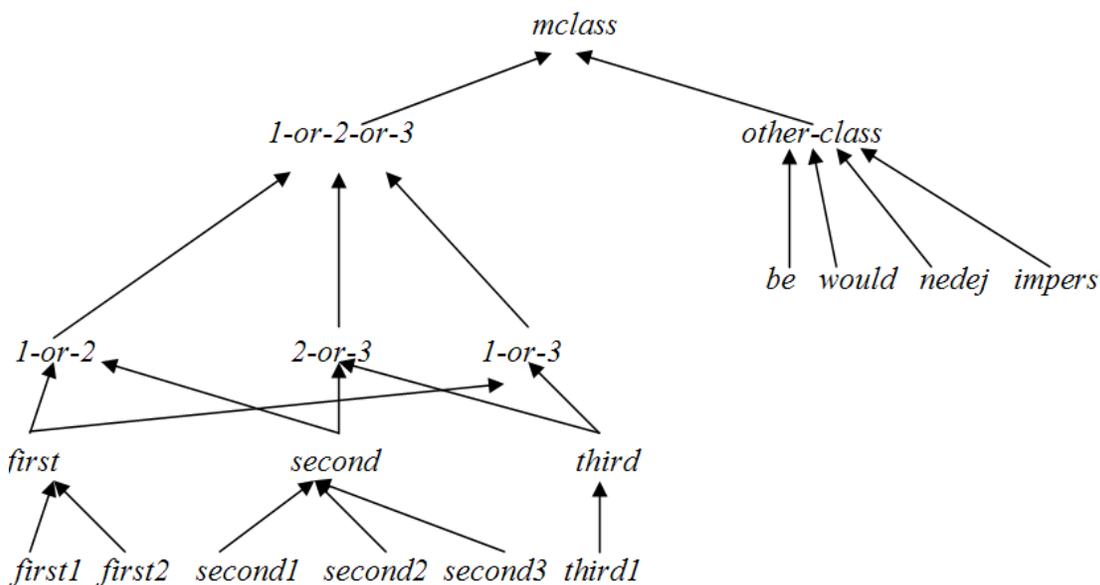
The third feature from the above list (**FORM**) specifies whether the verb is finite or non-finite. Here is the hierarchy:



Note that *da-form* is put under the nonfinite type as inheriting from the old Bulgarian infinitive. At the same time, the verb that comes after it, is finite. Under *gerund* the verbal adverbial is assumed. On the other hand, *deverbal* stands for the verbal noun.

The **MCLASS** feature generalizes over the three conjugations as well as some idiosyncratic cases. The hierarchy follows below. For the three conjugation classes the idea is to keep the possibilities for a generalization over certain categories (for example, imperative), and at the same time to have most specific types. The type *other-class* includes the conjugation of the verb *съм* ‘to be’, *щях* ‘would’, *недей* (nedej, ‘do not’, which is an alternative to the negative imperative) and *impers*. The type *finite* has two subtypes (not in the picture, because not used) – *personal-v* and *impersonal-v*. However, at the moment this distinction is kept in a different way via the **MCLASS** values.

As it was mentioned in the Introduction, this classification is far from exhaustive. It only demonstrates the complexity of the Bulgarian verbal system. A full implementation was done through the transfer of the patterns and rule from the Bulgarian Spelling dictionary. It follows a slightly different hierarchy and principles, which have to be incorporated into this version.



The feature **VOICE** reflects the active or passive voices. Additionally to the head feature **VOICE**, a semantic counterpart had to be added (**SVOICE**) in order this information to be accessible at the semantic level.

The verbs have the most complex subtype structure with respect to intransitivity, transitivity, ditransitivity and even more-argument-transitivity.

We should note that transitivity in our interpretations does not equal the subcategorization of accusative objects. It is rather understood in broader sense as complementation. Thus, the transitive types take NPs, PPs or clauses as their complements. The verbs always cross-classify with the perfective or the imperfective type lexeme.

9.1 Main Verbs

The types of the verbs are many, since their selectional properties are rich. The word order among the subject and the complements is considered free.

There are intransitive verbs like *лая* (bark) or *вали* (rains), which select only for a subject.

There are transitive verbs, which select NPs (*зная* (know)), PPs (*харесам на някого* (to be likable to somebody)), or clauses like *вярвам, че* (believe that..). It should be noted that the verbs which have a reflexive particle (by its form, not by its meaning) *се* (se) as part of their lexical meaning, select for this ‘se’ as their first complement. For example, *чудя се* (to be puzzled). The mechanism here is more complex. Since we have one lexical type for ‘se’, here the regular form is used only for ensuring the clitic realization. It supplies an anti-synsem. Thus, it can never unify itself, but allows for the clitic mechanism as a handling one. This approach can be viewed as redundancy one from a certain point of view, but for the moment we consider it useful in the overall grammar architecture.

This fact, however, influences the arity of the types. Being intransitive by definition, such verbs become transitive by this analysis. Needless to say, the complementation of ‘se’ should be considered as an outlier being just a technical solution.

The transitive types include also the selection of a ‘da-construction’. Here two possibilities are taken into account. The first is when the main verb does not necessarily control the subject of the selected clause. For example:

(14) *Абрамс искаше Браун да лае.*
Abrams wanted Brown ‘da’ bark-1sg
Abrams wanted Brown to bark.

The second is when the main verb controls the subject. For example:

(15) *Абрамс продължаваше да лае.*
Abrams-3sg continued ‘da’ bark-3sg
Abrams kept barking.

Then, an additional restriction applies that **XARG** of the embedded verb is the same as the **XARG** of the main one.

A special case within the transitive verb group is the impersonal verb *изглежда* ‘seems’³, which selects for a clausal complement, but the subject belongs actually to the embedded clause verb. For example:

(16a) Абрамс изглежда лае.
Abrams seems barks
Abrams seems to bark.

(16b) Изглежда Абрамс лае.
Seems Abrams barks
It seems that Abrams barks.

The ditransitive types include the following subtypes:

The verb takes an NP and a PP as complements. The typical example is the verb *дам* (give). But here also apply verbs like *сложа нещо някъде* (to put something somewhere) or *смятам някого за някакъв* (to consider someone for something-adj).

There is a similar subtype, where the verb takes a PP and an AdjP. For example, *изглеждам на някого някакъв* (I seem to somebody something-adj).

Another subtype is the verb, which lexically selects for ‘se’ particle and also for a clause. As it was mentioned above, such a case normally should be considered as transitive, but technically it is ditransitive, and for that reason – described here. For example:

(17a) Абрамс се чудеше дали Браун лае.
Abrams ‘se’ wondered whether Brown barks
Abrams wondered whether Brown barked.

(17b) Абрамс се чудеше кое куче лаеше.
Abrams ‘se’ wondered which dog barked
Abrams wondered which dog barked.

The control structures here are of two types. The first one is the *subject-subject* identity. For example, the verb *обещая* (promise), which selects for a PP and a ‘da-clause’:

(18) Абрамс обеща на Браун да лае.
Abrams promised to Brown ‘da’ bark-3sg
Abrams promised Brown to bark.

In such constructions also ‘se’ can play the role of the first complement. For example, the verb *опитвам се* (try):

(19) Абрамс се опитваше да лае.
Abrams ‘se’ tried ‘da’ bark-3sg
Abrams tried to bark.

³ In fact, the behaviour of this verb turned out to be rather idiosyncratic. For that reason, three separate entries were added depending on its selectional properties and its position in the sentence. Here this issue is not commented in detail.

The second one is *object-subject* identity. For example, the verb *убежда* (persuade):

(20)⁴ Убедих Иван да дойде.
Persuaded-I Ivan 'da' come
I persuaded Ivan to come.

In the present grammar there is one verb with 5 arguments. This is *обзалагам се* (to bet). It selects for a subject and four complements ('se', PP, PP, clause):

(21) Абрамс се обзаложи с Браун на цигара, че е валило дъжд.
Abrams 'se' bet with Brown on cigarette that is rained
Abrams bet Brown a cigarette that it rained.

9.2 Verb-Clitic Clusters

Bulgarian has personal pronoun clitics which reserve the place of the complement (direct or indirect object). Similarly to Greek, it shows the phenomenon of clitic doubling. In contrast to Greek and Romance languages, we view the clitics here as such, not as affixes (Osenova and Simov 2007; Avgustinova (2007)). However, the general approach is the same. The clitic does not contribute its separate semantics, because it is not a full-fledged complement. Instead, the verb incorporates clitic's contribution in its own semantics. Thus, the personal pronoun clitic-lexemes have an empty relation list, while the regular pronoun forms have a pronoun relation.

In order to handle the clitics that combine with verbal heads, the following boolean features were introduced to the type *category*:

cat :+ [**CLINIT** bool,
 DOCLIT bool,
 IOCLIT bool].

The first one **CLINIT** (clitic initial) should say whether the clitic can or cannot precede its head. The second and third manage the presence of the clitics, respectively - **DOCLIT** (accusative direct object clitic) and **IOCLIT** (dative indirect object clitic).

Altogether 6 rules were designed: 3 for the ordering rule *clitic-head*, and 3 mirror ones for the *head-clitic* position. The idea is that the verb in this phrase has a lexical synsem (*lex-synsem*), and the same holds also for the resulting verb-clitic construction. The verb is the semantic head. The first set has to block ungrammatical expressions in the sentence-initial position, like *го видях (*him-clitic saw-I), *ме преследват (*me-clitic chase-they). The construction, being lexical, further allows the real complement(s) to come via the HEAD-COMP principle. In a preposition of these clusters, besides a complement, there can come also the subject or a modifier. Then, the ordering is allowed as grammatical. The restrictions are handled via the features **HC-INIT** and **HS-INIT**, introduced and discussed in the next Section.

The three mirror variants reflect the following situations:

⁴ The construction should parse, but there is no example sentence in the testset.

1. transitive verb which might take an accusative clitic (*виждам те* ‘see-I you-clitic’);
 2. ditransitive verb which might take dative and/or accusative clitic before that (*давам ти* ‘give-I you-dative-clitic’; *давам я* ‘give-I her-accusative-clitic’; *давам ти я* ‘give-I you-dative-clitic her-accusative-clitic’);
 3. transitive verb with dative clitic, which might also take other complements (*благодаря му за нещо* ‘thank-I him-dative-clitic for something’).
- The accusative clitic rule also handles the usages of the form ‘se’.

9.3 Auxiliaries and Modals

Auxiliaries participate in the formation of the traditionally recognized analytical tenses in Bulgarian. Additionally, the following word-order oriented features are introduced to the category:

cat :+ [HC-INIT bool,
HS-INIT bool].

The **HC-INIT** (head-complement initial) is responsible for the ordering of the complements, while **HS-INIT** (head-subject initial) is responsible for the ordering of the subject. The latter is inherited always from the first argument of a phrase up. Thus, at the sentential level it is ensured that ungrammatical structures like **съм тук* (*am-I here) are not allowed.

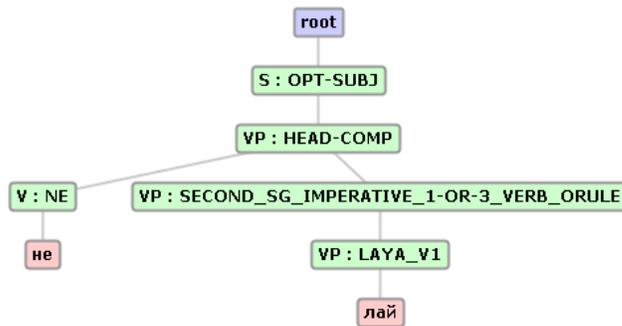
The auxiliaries include the verb *съм* (‘to be’), the future tense particle *ще* (will), the future tense in the past verb *цях* (would) and the nonfinite form *да* (to). All of these auxiliaries (with the exception of *would*) take as their first complement a lexical item. They inherit the complements of their first verbal lexical complement. The verb ‘to be’ takes an aorist participle, which is compatible with the respective active and passive forms of the participles. The *will* and *to* take finite verbs in present tense. However, the ‘to be’ auxiliary is splitted into two subtypes due to ordering reasons: present and non-present (for the moment, past). Thus, the present type blocks ungrammatical constructions with initial head: **съм разбрал* (‘am-I understood’, I have understood), **сте дошли* (‘are-you come-participle’, You have come). Such constraints do not apply to the other forms of the auxiliary (*бях разбрал* (‘was understood’, I had understood), *бяхте дошли* (‘were-you come-participle’, You had come)). The lexical verb complement can come either before, or after the head. Thus, **HC-INIT** feature here is inactive. However, the **HS-INIT** feature with value (-) says that the head cannot come before the subject (**е дошъл той* ‘is come he’, He has come).

The *will* and *to* types stipulate that **HC-INIT** has value (+), because their complement always comes after the head (*да дойда* (‘to come-I’, I to come), *ще дойда* (‘will come-I’, I will come) but not: **дойда да* (*come-I to), **дойда ще* (*come-I will)). *Will* additionally makes the tense *future*.

Would type takes *da-form* as its first complement, i.e. it selects for a phrasal synsem. It has its own conjugation.

It should be noted that the polar negative particle *не* (no) is also treated as an auxiliary. It always precedes its complement. The complement might be a distinct constituent or a clause. It introduces a *negation-relation*.

[3]



In picture [3] a simple example is given with the sentence: *не лай!* (do not bark!). The negative marker is a verb, which selects for a complement. Below the semantic presentation is also given (although some slight tuning is necessary).

[LTOP: *h1*

INDEX: *e2* [E.ASPECT: *imperfective* E.MOOD: *imperative* SF: *comm*]

RELS:<

[“negation_rel”

LBL: *h1*

ARG0: *e2*

ARG1: *h3*

[“лая_v_rel”

LBL: *h4*

ARG0: *e2*

ARG1: *x5* [x PNG.NUMBER *singular* PNG.PERSON *2nd* SF *comm*]

HCONS <*h3* qeq *h4*>]

The modals are treated as a special kind of auxiliaries. Similarly to *would*, they take a phrasal *da-form* (*мога да дойда* ‘can-I to come-I’, I can come). The imperative negative form *недей* (do not), which takes also *da-form* is analyzed as a modal verb, introducing a *negative-relation* (*недей да идваи* ‘do-not to come-you’, Do not come!).

9.4 Copula

The copula has the same orthographic form as the auxiliary verb *съм* ‘to be’. Thus, they both share the same conjugation class [MCLASS *be*]. It follows the same ordering restrictions, as above (the present form cannot be at the beginning of the sentence, while the past form can). Additionally, the types proliferate due to the specific interaction with the various complements – nouns, adjectives, PPs or adverbs. Although adjectives, adverbs and prepositions have *event* index, they cannot share the same generalized type. Adjectives structure-share their PNG characteristics with the copula’s XARG – the subject. The adverbs have to be restricted to intersective modifiers when taken as

complements. The common behaviour is that all these heads raise their semantic index to the copula, which is semantically vacuous itself.

The nouns, however, have a referential index. In this case, the copula behaves like a transitive verb, which selects for its complement. No index is raised from the noun complement up to the copula.

Altogether, 8 lexical types are introduced: two for present and past copula forms. Each of the two then is divided into four subtypes depending on the complement (present copula – noun; present copula – adjective; present copula – adverb; present copula - PP; past copula – noun; past copula – adjective; past copula – adverb; past copula - PP).

9.5 Impersonal Verbs

Impersonal verbs conform to their own class [**MCLASS** *impers*]. They can be non-modal and modal. They all have the morphosyntactic restriction of selecting for a pronominal subject in 3rd person, singular, neuter. The expletiveness is not marked in this version. From the non-modal ones the following types are included: *изглежда* (seem), *има* (there is) and *вали* (rains). The first one selects for a complementizer clause, the second one has as its first complement an NP, and the last is intransitive.

The modal verb is *трябва* (has to, ought to).

Having in mind the restrictions on the subject of the impersonal verb, the subject of the main verb in the complement clause might be extracted in front of the seem-verb (see examples (16ab)). In the first sentence the main verb subject is extracted in front of the seem-verb. In the second sentence there is no extraction, but just an optional subject of the seem-verb.

9.6 Participles, Gerund and Deverbals

These are nonfinite verbal forms, which are derived from the respective verbs.

It was necessary to distribute the information of the **MOD** feature at the correct level of granularity. Thus, the finite verb forms have empty **MOD** feature, since they do not modify anything. However, the participles (with the exception of one type only) can be modifiers. For that reason, the basic verb types do not have any opinion with respect to **MOD** feature. This split is introduced in the inflection types file. The participles have a non-empty **MOD** feature and hence - treated as intersective modifiers. The participles do not proliferate into subtypes. They are rather handled via compositional combinations of the values on the features **TENSE** and **SVOICE**. For example, the active present participle has [**TENSE** *present*] and [**SVOICE** *active*] (*четящ* ‘reading’); the active aorist participle has [**TENSE** *aorist*] and [**SVOICE** *active*] (*чел* ‘read’); the passive aorist participle has [**TENSE** *aorist*] and [**SVOICE** *passive*] (*четен* ‘read’). The imperfect active participle is not considered at the moment, since it does not participate in the analytical tense formation. Being verbal forms, participles have *event* index. However, they also can have morphologically definite forms (*четящият човек* (‘reading-the man’, the reading man); *пристигналата жена* (‘come-the woman’, the woman that has come); *намерената книга* (‘found-the book’, the found book)). Hence, similarly to the approach to the adjectives, the morphologically definite participles select for ‘semantically definite’ heads through their **MOD** feature:

MOD <[LOCAL.CONT.HOOK.INDEX.DEF +]>.

In this way, a compromise is made – an event can be morphologically marked by a definiteness marker, but semantically it cannot be definite. Semantically definite is only the nominal that is being modified.

The active present participles have the restriction that they can be derived only from imperfective verbs.

Gerunds behave like intersective adverbials. Their **XARG** is the same as the verb's subject. They can be derived only from imperfective verbs:

(22) Кучето пристигна лаейки.

Dog-the came barking

The dog came barking..

The deverbals in Bulgarian are nouns. Similarly to the active present participles and gerunds, they can be derived only from imperfective verbs:

(23) Преследването на котки е старо нещо.

Chasing-the of cats is old thing.

The chasing of the cats is an old thing.

The passive participles and the deverbals undergo further syntactic changes. The former becomes part of the passive rules, which swap accordingly the arguments of the input finite verb. The latter undergoes a nominalization changing its head category into a noun and its index into referential. These rules are presented in the respective sections below.

10. Substantivization and Nominalization

In (Osenova and Simov 2007) the substantivization is defined as the promotion of another nominal to a noun, while the nominalization – as the promotion of non-nominals or clauses into nouns. The current grammar version complies with these views.

In Bulgarian, the morphologically definite adjectives, numerals, possessive pronouns, participles can be used instead of the head noun. Thus, a rule was introduced to capture such cases:

(24) Моята лаеше.

My-the was-barking-imperfect tense

Mine was barking.

At the moment, we are cautious about allowing also non-definite forms to be substantivized. Only cardinal numerals are allowed for this option:

(25) Три лаят.

Three bark

Three bark.

The determiners also can be used as nouns depending on their type, form and the context (*Всеки казва* (Everybody says (that)), *Някои лаят* (Some bark), *Този не го познавам*

(‘This not him-accusative clitic know-I’, I do not know this (man))). However, the determiners and numerals can optionally have a partitive complement PP (*Три от кучетата лаят* (‘Three of dogs-the bark’, Three of the dogs bark), *Някои от студентите са тук* (‘Some of students-the are here’, Some of the students are here)). For these cases, a *partitive-relation* is introduced at both type specifications – determiner and cardinal numeral.

A nominalization rule is applied to sentences, in which the complementizer clause becomes a subject of the main verb (*Очевидно е, че кучето лае* (‘Obvious is that dog-the barks’, It is obvious that the dog barks)). The nominalized subject is always neuter, singular and nominative. Also, a *nominalize-relation* is introduced.

Since such sentences presuppose the implicit demonstrative: *това* (this), the CP-to-NP rule has a non-empty SPR list in order to be selected by the determiner.

11. Passivization and Deverbal Nouns

In Bulgarian, there are two passive paradigms – the auxiliary ‘to be’ in the respective tense plus the passive participle of the main verb, and the verb in the respective tense plus ‘se’. Examples follow for both cases:

(26) Кучето беше преследвано от Браун.

Dog-the-*neut,sg* was chased-*neut,sg* from Brown

The dog was chased by Brown.

(27) Кучето се преследваше от Браун.

Dog-the-*neut,sg* ‘se’ chased-*imperfect tense* from Brown

The dog was chased by Brown.

Thus, a general rule was introduced first, which takes care of swapping the NP complement of the active verb into its subject. Then two separate rules followed both paradigms. However, the inflection strategy is different in each case. The derived passive participial remained uninflected in order to be forced to undergo the passive rule. The ‘se’ passive takes inflected verbs and then the passive rule is applied. In both cases, the only complement of the passivized verb the PP, which is structure-shared with the active verb’s subject, marked with the preposition *om* (by). The other complements are left underspecified.

The deverbals are also considered a nonfinite type. They are then projected into nouns via a special nominalizing rule, which introduces a *nominalize-relation*. The PP phrase, headed by the passive preposition *om* (by) is treated as a complement only. All the other prepositional phrases are considered modifiers. The strategy is similar to Abney’s ‘islands of certainty’ in the chunk parsing – only the sure case is assigned a complement grammatical role, but this is precise. In the rest of the cases, inevitably, some information is lost, but also errors are avoided. To sum up, 3 cases altogether are predicted when nominalizing the deverbal form:

1. The subject is always projected into a complement *om* PP-phrase (by) – only if present in the sentence.

(28)⁵ Преследването на кучето **от** Браун беше смешно.
Chasing-the of dog-the **from** Brown was funny.
*The chasing of the dog **by** Brown was funny.*

2. The noun complement is also projected to a PP, but left underspecified as argument in the noun construction, i.e. considered modifier in the nominal environment. (see example (23))

3. The non-changing PPs in the projection are left unchanged and treated again as modifiers.

(29)⁶ Идването на Иван **в** Чикаго ме разстрои.
Coming-the of Ivan **in** Chicago me-acc,clitic upset.
*Ivan's coming **to** Chicago upset me.*

III. Phrase Types

Some of the phrase types have already been mentioned in the sections, dedicated to the various POS heads. Here the rest of the rules are described.

The word order seems significant to the modifier, complement, subject types as well as their heads in Bulgarian. Thus, always head initial and head final rules are stipulated. Only in some cases, there are specific constraints in their realizations.

Following Matrix, two basic rules on Modification are introduced: *head-modifier* and *modifier-head*. Also, the ordering rules apply distinctly to the intersective and the scopal modifiers. Thus, the rule for the intersective *modifier-head* phrase had to be split in two. For example, the adjectives and numerals canonically precede their noun heads, while PPs can come in either order with respect to their heads.

(30a) Имаше котки в градината.
Had-impers cats in garden-the
There were cats in the garden.

(30b) В градината имаше котки.
In garden-the had-impers cats
In the garden there were cats.

Then, the rule for *head-modifier* had also to be split into two subrules. This had to allow for time nouns to modify verbal or noun heads. (see examples (1) and (2))

Concerning complementation, the following has been done:

For the *clitic-head* or *head-clitic* lexical phrases, special rules have been introduced. The rule takes a word, which is of type *clitic*, but uninflected, i.e. [INFLECTED -]. This idiosyncratic approach is justified in ensuring that the heads of these lexical clusters are taking only clitics, not other words. Additionally to the transitive *verb-clitic cluster*, also a ditransitive rule was added. It takes care of cases, such as: *дадох му книга* ('gave-I

⁵ Not in the test set.

⁶ Not in the test set.

him-clitic book’, I gave him a book) and **му дадох книга* (‘him-clitic gave-I book’, To him I gave a book).

Also, rules for allowing the free order of both full-fledged complements were added. For example:

(31a)⁷ Той даде книгата на Иван.
He gave book-the to Ivan
He gave the book to Ivan.

(31b)⁸ Той даде на Иван книгата.
He gave to Ivan book-the
He gave Ivan the book.

These rules together with head-complement rules also ensure the analyses of clitic doubling:

(32) Котката го преследваше него.
Cat-the him-acc,clitic chased-imperfect tense him-acc,complement
The cat chased him.

Bulgarian is a grammatically-encoded pro-drop language. Thus, it makes use of *optional-subject-phrase*. However, the complements also might be omitted in the discourse. Thus, the *optional-complement-phrase* is also active in the grammar. However, as it was already mentioned, the complements have a relatively free order with respect to their head. Thus, in the latter rule the value of the feature **HS-INIT** (head-subject initial) is inherited at the phrase level from the head daughter.

The **Head-Spec** phrase follows straightforwardly the Matrix grammar specification. Bulgarian has incorporated its definite article as a morphological part of the word, and also as a part of the most left modifier in an NP. The indefinite article (which is still a debatable issue in Bulgarian grammar) might be expressed by a zero morpheme (i.e. the wordform itself), or by the numeral *един* ‘one’. Thus, the *bare-NP-phrase* rule is very much appreciated as a projection of the nominal heads.

The **Relative-Clause** rule via its **MOD** feature modifies its noun head. Note that this is the Relative clause, not the relative pronoun, that exhibits the **MOD** feature.

The **Head-Filler phrase** at the moment overgenerates. A split between relative-clause mechanism and verb extraction is considered.

The extraction is limited to nouns. PPs are not currently handled.

IV. The Lexicon

The idea is, similarly to the English grammar (ERG), that the last leaf type, which appears in the lexicon, presents explicitly the main syntactic information for this type in a globally ordered way. In this approach, the lexicon resembles a positionally organized

⁷ The pattern is in the testset, but not this very sentence.

⁸ The pattern is in the testset, but not this very sentence.

POS tagset. However, additionally to the main morphosyntactic characteristics, it also adds valence features (intransitivity, transitivity, types of the transitivity, etc.)

Let us consider, for example, the following mnemonic name:

n_ _cm_le

It means noun, intransitive, common, masculine. The ending 'le' means the last type (leaf) under the type hierarchy.

The lexicon consists of lexemes, whose identifiers are transliterated in Latin, and whose **STEM** is in Bulgarian. The entry consists of also a semantic predicate, whose beginning is in Bulgarian. The idea is to keep the lexicon as simple as possible. For that reason, only the relation is introduced, where available. See the next example for the noun *мъж* 'man':

(1)

```
mazh_n1 := n_ _cm_le &  
[ STEM < "мъж" >,  
  SYNSEM.LKEYS.KEYREL.PRED "мъж_n_1_rel" ].
```

A decision was made that the lexical aspect of the verb will be preserved in the **STEM**, but the aspect pairs will receive one common name predicate. Perfective was chosen for the absolute pairs (*изляя* – *излайвам* (bark)), and imperfective for pairs with empty prefix (*чудя се* – *почудя се* (wonder)). See the following example for the verbs 'bark-perfective' and 'bark-imperfective'. The **STEM** values differ, but the **PRED** values are identical:

(2)

```
izlaya := v_ _p1_le &  
[ STEM < "изляя" >,  
  SYNSEM.LKEYS.KEYREL.PRED "изляя_v_rel" ].
```

```
izlayvam := v_ _i3_le &  
[ STEM < "излайвам" >,  
  SYNSEM.LKEYS.KEYREL.PRED "изляя_v_rel" ].
```

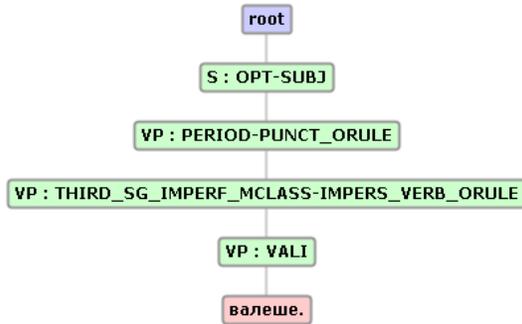
It should be noted that the morphologically rich BG grammar the verb lexicon is more complicated in its design, since it includes also the respective subparadigms.

V. Punctuation

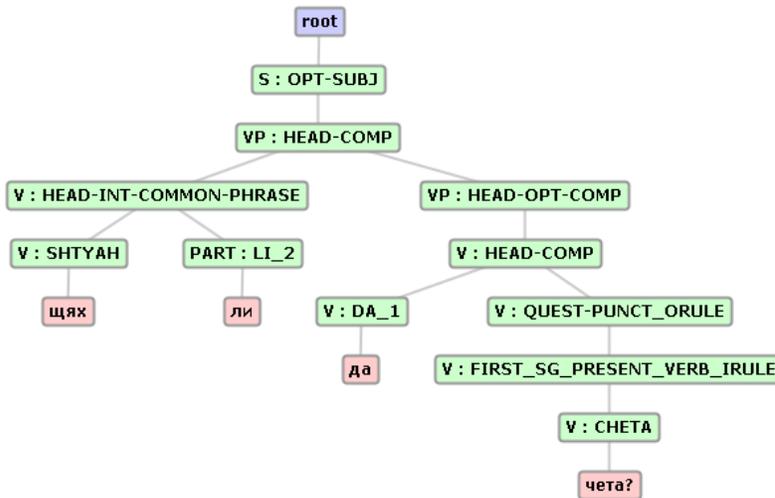
Only three types of the final punctuation have been added: period, question mark and exclamation mark. The principles of affixation were followed, which are presented in English and Portuguese grammars. Below two examples are presented from the

visualizations in CLaRK system. In picture [4] the sentence: *It rained* is presented, while in picture [5] – the sentence *Would I read?*:

[4]



[5]



VI. Conclusion

The present grammar is a second version, but still considered as a baseline for the further elaborations.

The evaluation of this grammar version was done with the system [tdsb] (Oepen 2001). Comparison is made between the coverage and performance of Version 0.1 and Version 0.2 (although the time span between the two was 1 month only).

The coverage of Version 0.1 is as follows:

COVERAGE Version 0.1						
total items	positive items	word string	lexical items	distinct analyses	total results	overall coverage
180	180	4.07	10.42	4.52	180	100 %

The coverage of Version 0.2 is as follows:

COVERAGE Version 0.2						
total items	positive items	word string	lexical items	distinct analyses	total results	overall coverage
213	194	4.10	10.36	3.73	194	100 %

In Version 0.2 the number of sentences is bigger, and includes also some negative examples. The word string slightly rises. The lexicon is still not very big, although there are few additional examples to the testset. The ambiguity was slightly decreased.

The number of distinct analyses is still higher than required for a good precision. Thus, the directions to go are: more lexical items, reduction of as many undesired analyses as possible, and further incorporation of the syntactic phenomena (plus more elaborated punctuation).

The performance metrics for Version 0.1 are as follows:

PERFORMANCE						
items	etasks	filter	edges	first	total	space
180	4718	64.0	107	0.10	0.10	85657

The performance metrics for Version 0.2 are as follows:

PERFORMANCE						
items	etasks	filter	edges	first	total	space
213	2026	66.6	68	0.05	0.04	76563

The processing time is reduced as well as the memory load. However, a conclusion to make is that the lexical and the structural ambiguities have to be addressed properly in the next phases.

APPENDIX 1

Lexicon Types

Noun lexicon types

n_ _cm-hum_le – noun, intransitive, common, masculine, human
n_ _cm-non-hum_le – noun, intransitive, common, masculine, non-human
n_np_cm-hum_le – noun, transitive (np), common, masculine, human
n_np_cm-non-hum_le – noun, transitive (np), common, masculine, non-human
n_pp_cm-hum_le – noun, transitive (pp), common, masculine, human
n_pp_cm-non-hum_le – noun, transitive (pp), common, masculine, non-human
n_cl_cm-hum_le – noun, transitive (clause), common, masculine, human
n_cl_cm-non-hum_le – noun, transitive (clause), common, masculine, non-human
n_ _cm-temporal_le – noun, intransitive, common, masculine, temporal, non-human
n_ _cm-week-day_le – noun, intransitive, common, masculine, weekday, non-human
n_ _cf-part-day_le – noun, intransitive, common, feminine, partday, non-human
n_ _cm-clock_le – noun, intransitive, common, masculine, clock, non-human

n_ _cf_le – noun, intransitive, common, feminine
n_np_cf_le – noun, transitive (np), common, feminine
n_pp_cf_le – noun, transitive (pp), common, feminine
n_cl_cf_le – noun, transitive (clause), common, feminine

n_ _cn_le – noun, intransitive, common, neuter
n_np_cn_le – noun, transitive (np), common, neuter
n_pp_cn_le – noun, transitive (pp), common, neuter
n_cl_cn_le – noun, transitive (clause), common, neuter

n_np-compound_cn_le – noun, transitive (np), common, neuter, compound

n_ _pm_le – noun, intransitive, proper, masculine
n_ _pf_le – noun, intransitive, proper, feminine
n_ _pn_le – noun, intransitive, proper, neuter

Adjective lexicon types

a_ _le – adjective, intransitive
a_pp_le – adjective, transitive (pp)
a_cl_le – adjective, transitive (clause)
a_ _compar_le – adjective, intransitive, comparative
a_ _pred_le – adjective, intransitive, predicative

Adverb lexicon types

int-adverb-le – intersective, adverb
scopal-adverb-le – scopal, adverb

ques-adverb-le – intersective, adverb, interrogative pronoun
kolko-noun-adverb-le – intersective, adverb, interrogative pronoun (how many),
modifies nouns
kolko-other-adverb-le - intersective, adverb, interrogative pronoun (how many),
modifies other heads
almost-adv-le - intersective, adverb, almost

Numeral lexicon types

cardinal-adj-clock-le – cardinal, numeral, clock times
cardinal-one-adj-le - cardinal, numeral, one
cardinal-adj-le – cardinal, numeral, from 3 up
ordinal-adj-le – ordinal, numeral
many-few-adj-verb-le – many or few, modifying verbs
many-few-adj-noun-le – many or few, modifying nouns
many-few-adj-adj-le – many or few, modifying adjectives and adverbs

Complementizer lexicon types

che-comp-le – complemenizer ‘che’ (that)
dali-comp-le – complementizer ‘dali’ (whether)

Subordinator lexicon types

che-sub-le – subordinator ‘che’ (that)
ako-le - subordinator ‘ako’ (if)
zada-le - subordinator ‘za da’ (in order to)

Particle lexicon types

se-le – reflexive particle
polar-ques-part-le – polar particle ‘li’ as modifier of NPs and VPs or sentences
polar-quest-verb-part-le - polar particle ‘li’ as modifier of lexical verbs
polar-mod-ques-part-le –polar particle ‘nali’ as modifier of a sentence

Pronoun lexicon types

Personal pronouns

nom-pers-pro-noun-le – personal, nominative, regular
acc-pers-pro-noun-reg-le – personal, accusative, regular
acc-pers-pro-noun-clit-le - personal, accusative, clitic
dat-pers-pro-noun-clit-le - personal, dative, clitic
reg-pers-refl-pro-noun-le – personal, accusative, reflexive, regular
pers-refl-acc-clitic-lexeme-le - personal, accusative, reflexive, clitic
pers-refl-dat-clitic-lexeme-le - personal, dative, reflexive, clitic

Possessive pronouns

poss-1p-le – possessive, 1st person, regular
poss-2p-le - possessive, 2nd person, regular
poss-3pm-le - possessive, 3rd person, masculine, regular
poss-3pf-le - possessive, 3rd person, feminine, regular
poss-3pn-le - possessive, 3rd person, neuter, regular
poss-pro-adj-clit-le - possessive, clitic
poss-refl-pro-adj-reg-le – possessive, reflexive, regular
poss-refl-pro-adj-clit-le - possessive, reflexive, clitic

(Determiner) pronouns

demonstrative-entity-le – demonstrative determiner
demonstrative-property-le – demonstrative pronoun for properties
indefinite-entity-le – indefinite determiner
indefinite-neshto-le – indefinite pronoun, neuter
collective-entity-le – collective determiner
negative-entity-le – negative determiner
question-pro-noun-le – interrogative determiner
question-poss-pro-noun-le – interrogative possessive determiner
relative-pro-noun-le – relative pronoun
nito-le – negative particle determiner

Verb lexicon types⁹

Impersonal types

v_-_imp_i_le – verb, intransitive, impersonal, imperfective
v_seem_le – verb, seem (transitive, takes a clause, imperfective)
v_there-is_le – verb, there-is (transitive, takes an NP, imperfective)
v_must_le – verb, must (transitive, takes da-form, imperfective)

Intransitive perfective types

v_-_p1_le – verb, intransitive, perfective, mclass 1
v_-_p1-1_le – verb, intransitive, perfective, mclass 1-1
v_-_p2-1_le - verb, intransitive, perfective, mclass 2-1
v_-_p2-2_le - verb, intransitive, perfective, mclass 2-2
v_-_p2-3_le - verb, intransitive, perfective, mclass 2-3
v_-_p3_le - verb, intransitive, perfective, mclass 3

Intransitive imperfective types

⁹ Note that on the one hand, some types are isolated just to cover the testset lexicon. On the other hand, there are also exhaustive types, which do not cover any verb in the present testset, but are envisaged to cover the new entries. Hence, some asymmetry might be observed within the distinct type groups.

v_ _i1_le - verb, intransitive, imperfective, mclass 1
v_ _i1-1_le - verb, intransitive, imperfective, mclass 1-1
v_ _i2-1_le - verb, intransitive, imperfective, mclass 2-1
v_ _i2-2_le - verb, intransitive, imperfective, mclass 2-2
v_ _i2-3_le - verb, intransitive, imperfective, mclass 2-3
v_ _i3_le - verb, intransitive, imperfective, mclass 3
v_ _i3-1_le - verb, intransitive, imperfective, mclass 3-1

Transitive perfective NP types

v_np_p1_le – verb, NP transitive, perfective, mclass 1
v_np_p1-1_le - verb, NP transitive, perfective, mclass 1-1
v_np_p2-1_le - verb, NP transitive, perfective, mclass 2-1
v_np_p2-2_le - verb, NP transitive, perfective, mclass 2-2
v_np_p2-3_le - verb, NP transitive, perfective, mclass 2-3
v_np_p3_le - verb, NP transitive, perfective, mclass 3
v_np_p3-1_le - verb, NP transitive, perfective, mclass 3-1

Transitive imperfective NP types

v_np_i1_le - verb, NP transitive, imperfective, mclass 1
v_np_i1-1_le - verb, NP transitive, imperfective, mclass 1-1
v_np_i2-1_le - verb, NP transitive, imperfective, mclass 2-1
v_np_i2-2_le - verb, NP transitive, imperfective, mclass 2-2
v_np_i2-3_le - verb, NP transitive, imperfective, mclass 2-3
v_np_i3_le - verb, NP transitive, imperfective, mclass 3
v_np_i3-1_le - verb, NP transitive, imperfective, mclass 3-1

Transitive perfective PP types

v_pp_p1_le - verb, PP transitive, perfective, mclass 1
v_pp_p1-1_le - verb, PP transitive, perfective, mclass 1-1
v_pp_p2-1_le - verb, PP transitive, perfective, mclass 2-1
v_pp_p2-2_le - verb, PP transitive, perfective, mclass 2-2
v_pp_p2-3_le - verb, PP transitive, perfective, mclass 2-3
v_pp_p3_le - verb, PP transitive, perfective, mclass 3
v_pp_p3-1_le - verb, PP transitive, perfective, mclass 3-1

Transitive imperfective PP types

v_pp_i1_le - verb, PP transitive, imperfective, mclass 1
v_pp_i1-1_le - verb, PP transitive, imperfective, mclass 1-1
v_pp_i2-1_le - verb, PP transitive, imperfective, mclass 2-1
v_pp_i2-2_le - verb, PP transitive, imperfective, mclass 2-2
v_pp_i2-3_le - verb, PP transitive, imperfective, mclass 2-3

v_pp_i3_le - verb, PP transitive, imperfective, mclass 3
v_pp_i3-1_le - verb, PP transitive, imperfective, mclass 3-1

Transitive perfective clauses with CHE (that) complemenizer types

v_che_p1_le - verb, CHE transitive, perfective, mclass 1
v_che_p1-1_le - verb, CHE transitive, perfective, mclass 1-1
v_che_p2-1_le - verb, CHE transitive, perfective, mclass 2-1
v_che_p2-2_le - verb, CHE transitive, perfective, mclass 2-2
v_che_p2-3_le - verb, CHE transitive, perfective, mclass 2-3
v_che_p3_le - verb, CHE transitive, perfective, mclass 3
v_che_p3-1_le verb, CHE transitive, perfective, mclass 3-1

Transitive imperfective clauses with CHE (that) complemenizer types

v_che_i1_le - verb, CHE transitive, imperfective, mclass 1
v_che_i1-1_le - verb, CHE transitive, imperfective, mclass 1-1
v_che_i2-1_le - verb, CHE transitive, imperfective, mclass 2-1
v_che_i2-2_le - verb, CHE transitive, imperfective, mclass 2-2
v_che_i2-3_le - verb, CHE transitive, imperfective, mclass 2-3
v_che_i3_le - verb, CHE transitive, imperfective, mclass 3
v_che_i3-1_le verb, CHE transitive, imperfective, mclass 3-1

Transitive perfective clauses with DALI (whether) complemenizer types

v_dali_p1_le - verb, DALI transitive, perfective, mclass 1
v_dali_p1-1_le - verb, DALI transitive, perfective, mclass 1-1
v_dali_p2-1_le - verb, DALI transitive, perfective, mclass 2-1
v_dali_p2-2_le - verb, DALI transitive, perfective, mclass 2-2
v_dali_p2-3_le - verb, DALI transitive, perfective, mclass 2-3
v_dali_p3_le - verb, DALI transitive, perfective, mclass 3
v_dali_p3-1_le - verb, DALI transitive, perfective, mclass 3-1

Transitive imperfective clauses with DALI (whether) complemenizer types

v_dali_i1_le - verb, DALI transitive, imperfective, mclass 1
v_dali_i1-1_le - verb, DALI transitive, imperfective, mclass 1-1
v_dali_i2-1_le - verb, DALI transitive, imperfective, mclass 2-1
v_dali_i2-2_le - verb, DALI transitive, imperfective, mclass 2-2
v_dali_i2-3_le - verb, DALI transitive, imperfective, mclass 2-3
v_dali_i3_le - verb, DALI transitive, imperfective, mclass 3
v_dali_i3-1_le - verb, DALI transitive, imperfective, mclass 3-1

Transitive perfective clauses with DA-FORM as CP complemenizer types (relaxed subject) (example - want)

v_da-cp_p1_le - verb, DA-FORM as CP, transitive, perfective, mclass 1
v_da-cp_p2-1_le - verb, DA-FORM as CP, transitive, perfective, mclass 2-1
v_da-cp_p2-2_le - verb, DA-FORM as CP, transitive, perfective, mclass 2-2
v_da-cp_p2-3_le - verb, DA-FORM as CP, transitive, perfective, mclass 2-3
v_da-cp_p3_le - verb, DA-FORM as CP, transitive, perfective, mclass 3
v_da-cp_p3-1_le - verb, DA-FORM as CP, transitive, perfective, mclass 3-1

Transitive imperfective clauses with DA-FORM as CP complemenizer types (relaxed subject)

v_da-cp_i1_le - verb, DA-FORM as CP, transitive, imperfective, mclass 1
v_da-cp_i2-1_le - verb, DA-FORM as CP, transitive, imperfective, mclass 2-1
v_da-cp_i2-2_le - verb, DA-FORM as CP, transitive, imperfective, mclass 2-2
v_da-cp_i2-3_le - verb, DA-FORM as CP, transitive, imperfective, mclass 2-3
v_da-cp_i3_le - verb, DA-FORM as CP, transitive, imperfective, mclass 3
v_da-cp_i3-1_le - verb, DA-FORM as CP, transitive, imperfective, mclass 3-1

Transitive perfective clauses with DA-FORM as VP complemenizer types (subject identity) (example – start, continue...)

v_da-vp_p1_le – verb, DA-FORM as VP, transitive, perfective, mclass 1
v_da-vp_p2-1_le – verb, DA-FORM as VP, transitive, perfective, mclass 2-1
v_da-vp_p2-2_le – verb, DA-FORM as VP, transitive, perfective, mclass 2-2
v_da-vp_p2-3_le – verb, DA-FORM as VP, transitive, perfective, mclass 2-3
v_da-vp_p3_le – verb, DA-FORM as VP, transitive, perfective, mclass 3
v_da-vp_p3-1_le - verb, DA-FORM as VP, transitive, perfective, mclass 3-1

Transitive imperfective clauses with DA-FORM as VP complemenizer types (subject identity) (example – start, continue...)

v_da-vp_i1_le – verb, DA-FORM as VP, transitive, imperfective, mclass 1
v_da-vp_i2-1_le – verb, DA-FORM as VP, transitive, imperfective, mclass 2-1
v_da-vp_i2-2_le – verb, DA-FORM as VP, transitive, imperfective, mclass 2-2
v_da-vp_i2-3_le – verb, DA-FORM as VP, transitive, imperfective, mclass 2-3
v_da-vp_i3_le – verb, DA-FORM as VP, transitive, imperfective, mclass 3
v_da-vp_i3-1_le - verb, DA-FORM as VP, transitive, imperfective, mclass 3-1

Transitive perfective SE verbs

v_se_p1_le – verb, SE complement, transitive, perfective, mclass 1
v_se_p2-1_le – verb, SE complement, transitive, perfective, mclass 2-1
v_se_p2-2_le – verb, SE complement, transitive, perfective, mclass 2-2
v_se_p2-3_le – verb, SE complement, transitive, perfective, mclass 2-3
v_se_p3_le – verb, SE complement, transitive, perfective, mclass 3
v_se_p3-1_le – verb, SE complement, transitive, perfective, mclass 3-1

Transitive imperfective SE verbs

v_se_i1_le – verb, SE complement, transitive, imperfective, mclass 1
v_se_i2-1_le – verb, SE complement, transitive, imperfective, mclass 2-1
v_se_i2-2_le – verb, SE complement, transitive, imperfective, mclass 2-2
v_se_i2-3_le – verb, SE complement, transitive, imperfective, mclass 2-3
v_se_i3_le – verb, SE complement, transitive, imperfective, mclass 3
v_se_i3-1_le – verb, SE complement, transitive, imperfective, mclass 3-1

Ditransitive perfective SE verbs with PP complement

v_se-pp_p1_le – verb, SE-PP complement, ditransitive, perfective, mclass 1
v_se-pp_p2-1_le – verb, SE-PP complement, ditransitive, perfective, mclass 2-1
v_se-pp_p2-2_le – verb, SE-PP complement, ditransitive, perfective, mclass 2-2
v_se-pp_p2-3_le – verb, SE-PP complement, ditransitive, perfective, mclass 2-3
v_se-pp_p3_le – verb, SE-PP complement, ditransitive, perfective, mclass 3
v_se-pp_p3-1_le – verb, SE-PP complement, ditransitive, perfective, mclass 3-1

Ditransitive imperfective SE verbs with PP complement

v_se-pp_i1_le – verb, SE-PP complement, ditransitive, imperfective, mclass 1
v_se-pp_i2-1_le – verb, SE-PP complement, ditransitive, imperfective, mclass 2-1
v_se-pp_i2-2_le – verb, SE-PP complement, ditransitive, imperfective, mclass 2-2
v_se-pp_i2-3_le – verb, SE-PP complement, ditransitive, imperfective, mclass 2-3
v_se-pp_i3_le – verb, SE-PP complement, ditransitive, imperfective, mclass 3
v_se-pp_i3-1_le – verb, SE-PP complement, ditransitive, imperfective, mclass 3-1

Ditransitive perfective verbs with DA-FORM as VP complement (subject-subject identity) (example – to promise someone to do something)

v_da1-vp_p1_le - verb, DA-FORM as VP, ditransitive, subj-subj, perfective, mclass 1
v_da1-vp_p1-1_le - verb, DA-FORM as VP, ditransitive, subj-subj, perfective, mclass 1-1
v_da1-vp_p1-1_dat-le - verb, DA-FORM as VP, ditransitive, subj-subj, perfective, mclass 1-1, preposition with empty semantics and MOD value
v_da1-vp_p2-1_le - verb, DA-FORM as VP, ditransitive, perfective, mclass 2-1
v_da1-vp_p2-2_le - verb, DA-FORM as VP, ditransitive, perfective, mclass 2-2
v_da1-vp_p2-3_le - verb, DA-FORM as VP, ditransitive, subj-subj, perfective, mclass 2-3
v_da1-vp_p3_le - verb, DA-FORM as VP, ditransitive, subj-subj, perfective, mclass 3
v_da1-vp_p3-1_le - verb, DA-FORM as VP, ditransitive, subj-subj, perfective, mclass 3-1

Ditransitive imperfective verbs with DA-FORM as VP complement (subject-subject identity) (example – to promise someone to do something)

v_da1-vp_i1_le - verb, DA-FORM as VP, ditransitive, subj-subj, imperfective, mclass 1

v_da1-vp_i2-1_le - verb, DA-FORM as VP, ditransitive, subj-subj, imperfective, mclass 2-1

v_da1-vp_i2-2_le - verb, DA-FORM as VP, ditransitive, subj-subj, imperfective, mclass 2-2

v_da1-vp_i2-3_le - verb, DA-FORM as VP, ditransitive, subj-subj, imperfective, mclass 2-3

v_da1-vp_i3_le - verb, DA-FORM as VP, ditransitive, subj-subj, imperfective, mclass 3

v_da1-vp_i3-1_le - verb, DA-FORM as VP, ditransitive, subj-subj, imperfective, mclass 3-1

Ditransitive perfective verbs with DA-FORM as VP complement (object-subject identity) (example – to persuade someone to do something)

v_da2-vp_p1_le - verb, DA-FORM as VP, ditransitive, obj-subj, perfective, mclass 1

v_da2-vp_p2-1_le - verb, DA-FORM as VP, ditransitive, obj-subj, perfective, mclass 2-1

v_da2-vp_p2-2_le - verb, DA-FORM as VP, ditransitive, obj-subj, perfective, mclass 2-2

v_da2-vp_p2-3_le - verb, DA-FORM as VP, ditransitive, obj-subj, perfective, mclass 2-3

v_da2-vp_p3_le - verb, DA-FORM as VP, ditransitive, obj-subj, perfective, mclass 3

v_da2-vp_p3-1_le - verb, DA-FORM as VP, ditransitive, obj-subj, perfective, mclass 3-1

Ditransitive imperfective verbs with DA-FORM as VP complement (object-subject identity) (example – to persuade someone to do something)

v_da2-vp_i1_le - verb, DA-FORM as VP, ditransitive, obj-subj, imperfective, mclass 1

v_da2-vp_i2-1_le - verb, DA-FORM as VP, ditransitive, obj-subj, imperfective, mclass 2-1

v_da2-vp_i2-2_le - verb, DA-FORM as VP, ditransitive, obj-subj, imperfective, mclass 2-2

v_da2-vp_i2-3_le - verb, DA-FORM as VP, ditransitive, obj-subj, imperfective, mclass 2-3

v_da2-vp_i3_le - verb, DA-FORM as VP, ditransitive, obj-subj, imperfective, mclass 3

v_da2-vp_i3-1_le - verb, DA-FORM as VP, ditransitive, obj-subj, imperfective, mclass 3-1

Ditransitive perfective SE verbs with DA-FORM as VP complement (subject identity)(example - *оnуmвaм ce da дoйдa* ('try-I se to come-I', I try to come))

v_se-da-vp_p1_le – verb SE, DA-FORM as VP, ditransitive, subj-subj, perfective, mclass 1

v_se-da-vp_p2-1_le - verb SE, DA-FORM as VP, ditransitive, subj-subj, perfective, mclass 2-1

v_se-da-vp_p2-2_le – verb SE, DA-FORM as VP, ditransitive, subj-subj, perfective, mclass 2-2

v_se-da-vp_p2-3_le – verb SE, DA-FORM as VP, ditransitive, subj-subj, perfective, mclass 2-3

v_se-da-vp_p3_le – verb SE, DA-FORM as VP, ditransitive, subj-subj, perfective, mclass 3

v_se-da-vp_p3-1_le – verb SE, DA-FORM as VP, ditransitive, subj-subj, perfective, mclass 3-1

Ditransitive imperfective SE verbs with DA-FORM as VP complement (subject identity)(example - *ονυμαι σε να δοῦδα* ('try-I se to come-I', I try to come))

v_se-da-vp_i1_le - verb SE, DA-FORM as VP, ditransitive, subj-subj, imperfective, mclass 1

v_se-da-vp_i2-1_le - verb SE, DA-FORM as VP, ditransitive, subj-subj, imperfective, mclass 2-1

v_se-da-vp_i2-2_le - verb SE, DA-FORM as VP, ditransitive, subj-subj, imperfective, mclass 2-2

v_se-da-vp_i2-3_le - verb SE, DA-FORM as VP, ditransitive, subj-subj, imperfective, mclass 2-3

v_se-da-vp_i3_le - verb SE, DA-FORM as VP, ditransitive, subj-subj, imperfective, mclass 3

v_se-da-vp_i3-1_le - verb SE, DA-FORM as VP, ditransitive, subj-subj, imperfective, mclass 3-1

Ditransitive perfective SE verbs with DALI (whether) as CP complement (example – *чудя се дали идва* ('wonder-I se whether comes-he/she/it', I wonder whether he/she/it comes))

v_se-dali_p1_le – verb SE, DALI as CP, ditransitive, perfective, mclass 1

v_se-dali_p2-1_le – verb SE, DALI as CP, ditransitive, perfective, mclass 2-1

v_se-dali_p2-2_le – verb SE, DALI as CP, ditransitive, perfective, mclass 2-2

v_se-dali_p2-3_le – verb SE, DALI as CP, ditransitive, perfective, mclass 2-3

v_se-dali_p3_le – verb SE, DALI as CP, ditransitive, perfective, mclass 3

v_se-dali_p3-1_le – verb SE, DALI as CP, ditransitive, perfective, mclass 3-1

Ditransitive imperfective SE verbs with DALI (whether) as CP complement (example – *чудя се дали идва* ('wonder-I se whether comes-he/she/it', I wonder whether he/she/it comes))

v_se-dali_i1_le – verb SE, DALI as CP, ditransitive, imperfective, mclass 1

v_se-dali_i2-1_le - verb SE, DALI as CP, ditransitive, imperfective, mclass 2-1

v_se-dali_i2-2_le - verb SE, DALI as CP, ditransitive, imperfective, mclass 2-2

v_se-dali_i2-3_le - verb SE, DALI as CP, ditransitive, imperfective, mclass 2-3

v_se-dali_i3_le - verb SE, DALI as CP, ditransitive, imperfective, mclass 3

v_se-dali_i3-1_le - verb SE, DALI as CP, ditransitive, imperfective, mclass 3-1

Ditransitive perfective SE verbs with INTERROGATIVE PRONOUN (who,...) as CP complement

(example – *чудя се кой идва* ('wonder-I se who comes', I wonder who comes))

v_se-ques_p1_le - verb SE, QUES WORD as CP, ditransitive, perfective, mclass 1
v_se-ques_p2-1_le - verb SE, QUES WORD as CP, ditransitive, perfective, mclass 2-1
v_se-ques_p2-2_le - verb SE, QUES WORD as CP, ditransitive, perfective, mclass 2-2
v_se-ques_p2-3_le - verb SE, QUES WORD as CP, ditransitive, perfective, mclass 2-3
v_se-ques_p3_le - verb SE, QUES WORD as CP, ditransitive, perfective, mclass 3
v_se-ques_p3-1_le - verb SE, QUES WORD as CP, ditransitive, perfective, mclass 3-1

Ditransitive imperfective SE verbs with INTERROGATIVE PRONOUN (who,...) as CP complement

(example – *чудя се кой идва* ('wonder-I se who comes', I wonder who comes))

v_se-ques_i1_le - verb SE, QUES WORD as CP, ditransitive, imperfective, mclass 1
v_se-ques_i2-1_le - verb SE, QUES WORD as CP, ditransitive, imperfective, mclass 2-1
v_se-ques_i2-2_le - verb SE, QUES WORD as CP, ditransitive, imperfective, mclass 2-2
v_se-ques_i2-3_le - verb SE, QUES WORD as CP, ditransitive, imperfective, mclass 2-3
v_se-ques_i3_le - verb SE, QUES WORD as CP, ditransitive, imperfective, mclass 3
v_se-ques_i3-1_le - verb SE, QUES WORD as CP, ditransitive, imperfective, mclass 3-1

Ditransitive perfective verbs with NP and PP as complements

(examples – *give, consider..*)

v_np-pp_p1_le - verb, NP-PP, ditransitive, perfective, mclass 1
v_np-pp_p2_le - verb, NP-PP, ditransitive, perfective, mclass 2
v_np-pp_p2-1_le - verb, NP-PP, ditransitive, perfective, mclass 2-1
v_np-pp_p2-2_le - verb, NP-PP, ditransitive, perfective, mclass 2-2
v_np-pp_p2-2_loc-le - verb, NP-PP, ditransitive, perfective, mclass 2-2, MOD empty PP, locative
v_np-pp_p2-3_le - verb, NP-PP, ditransitive, perfective, mclass 2-3
v_np-pp_p3_le - verb, NP-PP, ditransitive, perfective, mclass 3
v_np-pp_p3-1_le - verb, NP-PP, ditransitive, perfective, mclass 3-1
v_np-pp_p3-1-dat_le - verb, NP-PP, ditransitive, perfective, mclass 3-1, MOD empty PP, dative

Ditransitive imperfective verbs with NP and PP as complements

(examples – *give, consider..*)

v_np-pp_i1_le - verb, NP-PP, ditransitive, imperfective, mclass 1
v_np-pp_i2_le - verb, NP-PP, ditransitive, imperfective, mclass 2
v_np-pp_i2-1_le - verb, NP-PP, ditransitive, imperfective, mclass 2-1
v_np-pp_i2-2_le - verb, NP-PP, ditransitive, imperfective, mclass 2-2
v_np-pp_i2-3_le - verb, NP-PP, ditransitive, imperfective, mclass 2-3

v_np-pp_i3_le - verb, NP-PP, ditransitive, imperfective, mclass 3
v_np-pp_i3-dat_le – verb, NP-PP, ditransitive, imperfective, mclass 3, MOD empty PP, dative
v_np-pp_i3-pred_le - verb, NP-PP, ditransitive, imperfective, mclass 3, MOD empty PP, predicative
v_np-pp_i3-adj-pred_le - verb, NP-PP (with complement adjective), ditransitive, imperfective, mclass 3, MOD empty PP, predicative
v_np-pp_i3-1_le - verb, NP-PP, ditransitive, imperfective, mclass 3-1
v_np-pp_i3-1-dat_le - verb, NP-PP, ditransitive, imperfective, mclass 3-1, MOD empty PP, dative

Ditransitive imperfective verbs with AdjP and PP as complements
(examples – *изглеждам на някого някакъв* ('seem-I to someone some-adjective'))

v_adj-pp_i3_le - verb, AdjP-PP, ditransitive, imperfective, mclass 3, MOD empty PP

Four-transitive perfective verbs with SE, PP, PP and CL complements (bet)

v_se-2pp-comp_p2-2_le - verb, SE-PP-PP-CL, four-transitive, perfective, mclass 2-2

Auxiliary types

pres-be-le – auxiliary 'be', present tense
past-be-le – auxiliary 'be', past tense
will-le – auxiliary 'will'
da-le – auxiliary 'da'
would-le – auxiliary 'would'
ne-le – auxiliary 'ne'
nedej-aux-le – auxiliary 'nedej'
modal-can-le – modal 'can'

Copula types

Past tense

past-copula-prep-le – copula, PP complement, past tense
past-copula-adj-le – copula, Adjectival complement, past tense
past-copula-noun-le – copula, Noun complement, past tense
past-copula-adv-le – copula, Adverbial complement, past tense

Present tense

pres-copula-prep-le – copula, PP complement, present tense
pres-copula-adj-le – copula, Adjectival complement, present tense
pres-copula-noun-le - copula, Noun complement, present tense
pres-copula-adv-le - copula, Adverbial complement, present tense

APPENDIX 2: THE BULGARIAN VERSION OF THE MULTILINGUAL TESTSET

1. Phenomena representation

impersonal intransitive constructions (11)
impersonal transitive constructions (991, 801)
subject-verb (21, 141, 201, 211, 351)
inchoative (unaccusative) (31)
subject-verb-direct object (41, 151, 671, 681)
subject-verb-indirect object-direct object (51)
subject-verb-direct object-indirect object (61)
many-argument verb (NP bet [PP with somebody] [PP on something] [that P]) (71)
verbs or other head that take complementizer clauses (81, 291, 301, 551, 811, 821, 871, 931, 971)
subject control transitive constructions (91, 401, 661, 791)
object control transitive constructions (841)
non-control transitive constructions (101)
quantifiers and their scope (111, 121, 901, 1001,1011)
possessives (pronouns and PPs) (131, 611, 621)
reflexives (161)
substantivation (171, 181, 191, 221, 601)
partitive constructions (231, 591)
negation (241)
polar questions (251)
questions with interrogative pronouns (261, 271, 581,1022)
imperatives and negative imperatives (281, 1061)
relative clauses (311, 321, 561)
passives (331)
reduced relatives (341)
analytical tenses (361, 371, 381, 391)
modals (411)
modal + negation (421)
intersective adjective modification (431, 921)
intersective adverb modification (441, 471, 901, 911)
scopal adverb modification (451)
intersective PP modification (461, 481, 491, 651)
copula constructions (501, 511, 891)
participles as modifiers (521, 531)
deverbal nouns (541, 921)
nouns as modifiers (571, 961)
numerals (631, 641, 881, 891)
relational nouns (691)
adverbial resultatives (701)
subject-verb-direct object–locative complement (711)

if-clauses (721)
 coordination (731, 741, 751, 761)
 ‘question tags’ (771)
 sentential subject (781)
 equi (801)
 ‘in order to’ clauses (831)
 light verb + PP + adjective (851) (in either word order)
 light verb + NP + PP (with adjective complement) (861) (in either word order)
 dates (941)
 clock (951)
 comparative adjective modification (981)
 compounds (1021)
 object deletion (1041, 1051)
 gerund (1071)
 clitic doubling in BG (accusative – 821; dative - 831)

2. Some Specificities of the translated Bulgarian testset.

Here come some comments on the existing testset in English and some of its reflections in Bulgarian counterparts.

First of all, Bulgarian is a pro-drop language. Thus, it has always counterparts with null subjects. In the discourse, it can also omit its complements in many cases. Second, Bulgarian verbs encode aspect lexically. Thus, the English sentences often have been translated with verbs in both aspects (perfective or imperfective). In some sentences more Bulgarian verb synonyms have been provided to the English one. Next, Bulgarian has clitics. Thus, translations with a clitic and a full-fledged complement have been provided to the single English one, when appropriate. Bulgarian has a double negation mechanism (similarly to other Slavic languages) and in contrast to English. Bulgarian polar questions are formed with a special question particle, which has also a focalizing role. Bulgarian has only semantically restricted class of noun modifiers (time nouns). The modification is mostly done by the adjectives (*garden dog* (en) vs. *градинско куче* (bg, ‘garden-adjective dog’))¹⁰. Some alternations that are challenging for English, are not relevant for Bulgarian. For example: *Browne squeezed the cat in* and *Browne squeezed in the cat* are translated in the same way: *Браун вмъкна котката*. The same holds for the well-known give-alternation: *Abrams handed Browne the cigarette* and *Abrams handed the cigarette to Browne*. The Bulgarian translation just ‘swaps’ the complements, but does not change them: *Абрамс даде на Браун цигарата* and *Абрамс даде цигарата на Браун*.

3. The sentences with number markers as in the original testset¹¹

валеше (11)
 валеше дъжд (11)
 валя (11)

¹⁰ Please note that there is a big productive group with noun modifiers, which concerns the loans from English (*бизнес среща* ‘business meeting’). However, such cases are not present in the current testset.

¹¹ Note that the numbers do not follow always in the expected order.

валя дъжд (11)
абрамс лаеше (21)
абрамс излая (21)
прозорецът се отвори (31)
абрамс преследваше браун (41)
абрамс даде на браун цигарата (51)
абрамс даде цигарата на браун (61)
абрамс знаеше че вали дъжд (81)
абрамс знаеше че вали (81)
абрамс възнамеряваше да лае (91)
абрамс подаде на браун цигарата (51)
абрамс предаде на браун цигарата (51)
абрамс връчи на браун цигарата (51)
абрамс подаде цигарата на браун (61)
абрамс предаде цигарата на браун (61)
абрамс връчи цигарата на браун (61)
абрамс възнамеряваше да излае (91)
абрамс се канеше да лае (91)
абрамс се канеше да излае (91)
абрамс искаше браун да лае (101)
абрамс искаше браун да излае (101)
всяка котка лаеше (111)
всяка котка излая (111)
всяка котка преследваше някое куче (121)
моята котка лаеше (131)
моята котка излая (131)
котката ми лаеше (131)
котката ми излая (131)
лаеше (141)
излая (141)
то лаеше (141)
то излая (141)
котката го преследваше (151)
котката преследваше това (151)
котката преследваше себе си (161)
моята лаеше (181)
моята излая (181)
моего лаеше (181)
моего излая (181)
онова се отвори (191)
котките лаят (201)
тютюнът пристигна (211)
някои лаят (221)
някои от котките лаят (231)
нито една котка не лаеше (241)
нито една котка не излая (241)

кучето лаеше ли (251)
кучето излая ли (251)
лаеше ли кучето (251)
излая ли кучето (251)
кое куче лаеше (261)
кое куче излая (261)
никоя котка не лаеше (241)
никоя котка не излая (241)
чие куче лаеше (271)
чие куче излая (271)
преследвай браун (281)
абрамс се чудеше дали браун лае (301)
абрамс се чудеше дали браун излая (301)
абрамс се почуди дали браун лае (301)
абрамс се почуди дали браун излая (301)
абрамс се чудеше кое куче лаеше (291)
абрамс се чудеше кое куче излая (291)
абрамс се почуди кое куче лаеше (291)
абрамс се почуди кое куче излая (291)
кучето което браун преследваше лаеше (311)
кучето което браун преследваше излая (311)
кучето лае (351)
кучето е излаяло (361)
кучето е лаяло (361)
кучето лаеше (371)
кучето беше лаяло (381)
кучето ще лае (391)
кучето ще излае (391)
кучето се кани да лае (401)
кучето се кани да залае (401)
кучето се кани да излае (401)
кучето можеше да лае (411)
кучето можеше да излае (411)
кучето не можеше да лае (421)
кучето не можеше да излае (421)
старото куче лаеше (431)
старото куче излая (431)
кучето лаеше тихо (441)
кучето излая тихо (441)
кучето вероятно лаеше (451)
кучето вероятно излая (451)
кучето сигурно лаеше (451)
кучето сигурно излая (451)
кучето лаеше в градината (461)
кучето излая в градината (461)
кучето лае сега (471)

градинското куче лаеше (481)
кучето от градината лаеше (481)
кучето от тютюневата градина лаеше (491)
кучето от тютюневата градина излая (491)
котката е стара (501)
котката е в градината (511)
кога лаеше кучето (581)
кучето на браун лае (611)
двеста и двайсет кучета лаят (641)
двайсет и три кучета лаят (631)
абрамс пристигна с кола (651)
абрамс продължаваше да лае (661)
браун вмъкна котката (671, 681)
снимката на абрамс пристигна (691)
абрамс избърса масата до чисто (701)
абрамс сложи браун в градината (711)
кучето ще лае ако браун пристигне (721)
кучето ще излае ако браун пристигне (721)
абрамс и браун пристигнаха (731)
абрамс браун и кучето пристигнаха (741)
кучето пристигна и залая (751)
кучето пристигна и излая (751)
кучето пристигна и браун залая (761)
кучето пристигна и браун излая (761)
очевидно е че кучето лаеше (781)
очевидно е че кучето излая (781)
абрамс обеща на браун да лае (791)
абрамс вярва че браун лае (811)
абрамс го притесняваше че браун лае (821)
на абрамс му трябваша десет минути за да пристигне (831)
абрамс лаеше от десет до три (881)
абрамс беше много стар (891)
кучето лаеше нали (771)
кучето излая нали (771)
абрамс лаеше много тихо (911)
абрамс излая много тихо (911)
браун го притесняваше че абрамс преследва котки (821)
трети юни дойде (941)
онова куче преследваше браун (1001)
някой преследваше браун (1011)
дойде трети юни (941)
кучето преследвано от браун лаеше (341)
лаещото куче преследваше браун (521)
преследваните кучета лаят (531)
кучетата преследват всичко което лае (561)
това че котката преследва браун е старо нещо (551)

имаше котки в градината (991)
в градината имаше котки (991)
кучето беше преследвано от браун (331)
абрамс знаеше че е валяло (81)
кучето се преследваше от браун (331)
котката сама се преследваше (161)
котката се преследваше сама (161)
кучето лаеше всеки ден (571)
кучето излайваше всеки ден (571)
почти всяко куче лаеше (901)
преследването на котки е старо нещо (541)
по-щастливото куче преследваше браун (981)
три от кучетата лаят (591)
три лаят (601)
абрамс се обзаложи с браун на цигара че е валяло (71)
абрамс изглежда на браун стар (851)
браун смята абрамс за стар (861)
числото пет притеснява браун (1031)
не лай (1061)
кучето пристигна лаейки (1071)
недей да лаеш (1061)
абрамс остави на браун да лае (841)
кучето което трябва да е преследвано лае (321)
колко щастлив беше абрамс (1021)
браун пристигна във вторник сутринта (961)
брауновото лае (621)
брауновото преследване на котки притесни абрамс (921)
брауновото преследване на котки притесняваше абрамс (921)
абрамс пристигна в три и двайсет (951)
абрамс пристигна в три часа и двайсет (951)
абрамс пристигна в три и двайсет минути (951)
котките намериха начин да лаят (971)
на абрамс му хареса идеята че браун може да лае (871)
абрамс се опитваше (1051)
абрамс можеше (1041)
абрамс изглежда лае (801)
изглежда абрамс лае (801)
старата ми котка преследва кучето.
старата му котка лае.
старата ми щастлива котка лае.
котката го преследваше него.
котката го преследваше браун.
котката него го преследваше.
за щастие той не уби никого.
наистина това беше прилеп.
няма ли да вечеряте?

трябва ли да му кажа?
щях ли да чета?
кучето което трябва да се преследва лае.
купувам му къща.

Negative examples

Checking Agreement:

- *котките лае (subject-verb disagreement)
- *старата куче лаеше (adjective-noun disagreement)
- *котката е старо (predicative disagreement)
- *кучето които лае преследва браун (noun head and relative pronoun disagreement)

Checking Definiteness:

- *кучето преследва старата котката (double DEF + is not possible on head and modifier)
- *кучето преследва стара котката ([DEF +] not possible on a modified head)
- *всяка котката лаеше (determiners select for [DEF -] nouns)
- *старата ми щастливата котка лае (only the first adjective in an NP should have [DEF +])

Checking Subject Control:

- *абрамс се канеше да изляят (violation of the subject identity)

Checking ungrammatical word order:

- *ли излая кучето
- *кучето лае ще
- *кучето иска лае да
- *котката го него преследваше
- *е стара
- *сьм преследвал браун

Checking determiner modifier order:

- *стара тази котка лае

Checking selectional properties:

- *абрамс искаше браун излае
- *мога излая
- *абрамс се почуди браун излая

References

- Avgustinova 2006:** T. Avgustinova. A Functional Typology of Copular "be": Towards an HPSG Formalisation. *Proceedings of the 13th International Conference on Head-Driven Phrase Structure Grammar*. S. Muller (Editor), CSLI Publications.
- Avgustinova 2007:** T. Avgustinova. *Language Family Oriented Perspective in Multilingual Grammar Design*. Linguistik International: Band 17. Peter Lang - Europaischer Verlag der Wissenschaft. Frankfurt am Main, Berlin, Bern, Bruxelles, New York, Oxford, Wien.
- Avgustinova and Uszkoreit 2006:** T. Avgustinova and H. Uszkoreit. Shared and Non-Shared Grammar in Modelling Slavic Morphosyntax. *Readings in Multilinguality. Selected papers for young researchers in BIS-21++*. G. Angelova, K. Simov, and M. Slavcheva (Editors). Institute for Parallel Processing, Bulgarian Academy of Sciences. Sofia (Published by Incoma Ltd., Shoumen, Bulgaria).
- Avgustinova and Uszkoreit 2003:** T. Avgustinova and H. Uszkoreit. Towards a typology of agreement phenomena. *The Role of Agreement in Natural Language: TLS 5 Proceedings*. W. E. Griffin (Editor), Texas Linguistics Forum 53. Austin, TX 2003. Pages 167-180.
- Baldwin et. al. 2005:** T. Baldwin, John Beavers, E. Bender, D. Flickinger, Ara Kim and Stephan Oepen. Beauty and the Beast: What running a broad coverage precision grammar over the BNC taught us about the grammar — and the corpus. In Stephan Kepser and Marga Reis (eds.), *Linguistic Evidence: Empirical, Theoretical, and Computational Perspectives*, Berlin: Mouton de Gruyter.
- Bender, Flickinger and Oepen 2002:** E. Bender, D. Flickinger and S. Oepen. The Grammar Matrix: An Open-Source Starter-Kit for the Rapid Development of Cross-Linguistically Consistent Broad-Coverage Precision Grammars. Carroll, John, Nelleke Oostdijk, and Richard Sutcliffe, eds. *Proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics*. Taipei, Taiwan. pp. 8-14.
- Bender 2007(a):** E. Bender. Combining Research and Pedagogy in the Development of a Crosslinguistic Grammar Resource. In *Proceedings of the GEAF 2007 Workshop*. Stanford CA: CSLI Publications.
- Bender 2007(b):** E. Bender. Socially Meaningful Syntactic Variation in Sign-Based Grammar. In Adger, David and Trousdale, Graeme, eds. *English Language and Linguistics 11(2) Special Issue*.
- Bender 2008:** E. Bender. Evaluating a Crosslinguistic Grammar Resource: A Case Study of Wambaya. In *Proceedings of ACL08:HLT*, Columbus, OH.
- Bender et. al 2004:** E. Bender, D. Flickinger, J. Good and I. Sag, Ivan. Montage: Leveraging Advances in Grammar Engineering, Linguistic Ontologies, and Markup for the Documentation of Underdescribed Languages. In *Proceedings of the Workshop on First Steps for Language Documentation of Minority Languages: Computational Linguistic Tools for Morphology, Lexicon and Corpus Compilation, LREC 2004*, Lisbon, Portugal.
- Bender et. al 2007:** E. Bender, L. Poulson, S. Drellishak, and Chris Evans. Validation and Regression Testing for a Cross-linguistic Grammar Resource. In *Proceedings of the ACL 2007 Workshop on Deep Linguistic Processing*. pp.136--143.

- Copestake 2002:** Ann Copestake. *Implementing Typed Feature Structure Grammars*. Stanford, CA: CSLI Publications.
- Copestake et. al 2005:** A. Copestake, D. Flickinger, C. Pollard and I. Sag. Minimal Recursion Semantics: An Introduction. In *Research on Language and Computation* (2005) 3:281–332.
- Flickinger 2000:** D. Flickinger. On building a more efficient grammar by exploiting types. In: *Natural Language Engineering, 6 (1) (Special Issue on Efficient Processing with HPSG)*, pp. 15-28.
- Flickinger and Bender 2003:** Compositional semantics in a multilingual grammar resource. In E.M. Bender, D. Flickinger, F. Fouvry, and M. Siegel, editors, *Proceedings of the Workshop on Ideas and Strategies for Multilingual Grammar Development*, ESSLLI 2003, pages 33–42, Vienna, Austria.
- Ginzburg and Sag 2000:** G. Ginzburg and I. Sag. *Interrogative Investigations: The Form, Meaning and Use of English Interrogatives*. CSLI, Stanford.
- Hellan and Haugereid 2003:** L. Hellan and P. Haugereid. NorSource: An exercise in Matrix grammar-building design. In E.M. Bender, D. Flickinger, F. Fouvry, and M. Siegel, editors, *Proceedings of the Workshop on Ideas and Strategies for Multilingual Grammar Development*, ESSLLI 2003, pages 41–48, Vienna, Austria.
- Kathol 1997:** A. Kathol. Agreement and the Syntax-Morphology Interface in HPSG. In Robert Levine and Georgia Green (eds.) *Studies in Current Phrase Structure Grammar*. Cambridge University Press. pp. 223-274.
- Kordoni and Neu 2005:** V. Kordoni and J. Neu. Deep analysis of Modern Greek. In K-Y Su, J. Tsujii, and J-H Lee, editors, *Lecture Notes in Computer Science*, volume 3248, pages 674–683. Springer-Verlag, Berlin.
- Muller and Kasper 2000:** S. Muller and W. Kasper. HPSG analysis of German. In W. Wachslar (ed.), *Verbmobil. Foundations of speech-to-speech translation*. (Artificial Intelligence ed., pp. 238-253). Berlin, Germany: Springer.
- Nicholson et. al 2008:** Jeremy Nicholson, Valia Kordoni, Yi Zhang, Timothy Baldwin and Rebecca Dridan. Evaluating and Extending the Coverage of HPSG Grammars. In *Proceedings of LREC-2008*, May 28-30, 2008, Marrakech, Morocco.
- Osenova and Simov 2007:** *Formal grammar of Bulgarian*. IPP, BAS, Sofia. (in Bulgarian).
- Osenova 2009:** *The Noun Phrases in Bulgarian*. ETO Publishing House, Sofia.
- Oepen 2001:** Stephan Oepen. [incr tsdb()] – competence and performance laboratory. User manual. Technical Report, Saarland University, Saarbruecken, Germany.
- Oepen, Flickinger, Toutanova, and Manning 2002:** Stephan Oepen, Dan Flickinger, Kristina Toutanova, and Christopher D. Manning LinGO Redwoods. A Rich and Dynamic Treebank for HPSG. In *Proceedings of The First Workshop on Treebanks and Linguistic Theories (TLT 2002)*, Sozopol, Bulgaria.
- Penchev 1993:** Y. Penchev. *Bulgarian Syntax. Government and Binding*. Plovdiv. (in Bulgarian)
- Pollard and Sag 1994:** C. Pollard and I. Sag. *Head-driven Phrase Structure Grammar*. University of Chicago Press, Chicago, Illinois, USA.

- Popov et. al 2003:** D. Popov, K.Simov, Sv. Vidinska and P.Osenova. *Spelling Dictionary of Bulgarian language*. Nauka i izkustvo. Sofia, 2003. (in Bulgarian)
- Sag 2008:** I. Sag. Feature Geometry and Predictions of Locality. In Greville Corbett and Anna Kibort (eds.), *Proceedings of the Workshop on Features*. Oxford: Oxford University Press.