The Definite Article of Bulgarian Adjectives and Numerals in DATR

Velislava Stoykova

Institute of the Bulgarian Language, Bulgarian Academy of Sciences,
52, Shipchensky proh. str., bl. 17, 1113 Sofia, Bulgaria
vili1@bas.bg

Abstract. The paper describes the inflectional feature of definiteness in Bulgarian language. It takes into account the specific morphosyntactic properties of the adjectives and the numerals, and argues for a related computational approach of interpretation. The DATR language for lexical knowledge presentation is accepted as a framework for constructing a model of Bulgarian adjectives and numerals inflectional morphology. The particular type hierarchies are defined, and the basic architecture of the model is explained in terms of its linguistic motivation.

1 Introduction

The most interesting grammar feature of standard Bulgarian language in which it differs with all Slavic languages is the feature of definiteness. Taken together with the fact that Bulgarian language does not use cases to express syntactic structures, it requires a specific sequence of tasks to be resolved for natural language processing applications.

It is important to note that the definite article in Bulgarian is an ending morpheme. This fact gives a priority to morphological interpretations instead of syntactic ones. At the level of syntax, the definite article shows the subject (when it is not a proper name) since Bulgarian language uses a relatively free word order.

So, giving a detailed formal morphological interpretation of the definite article is an essential preliminary condition for part-of-speech parsing of Bulgarian.

2 The Semantics of Definiteness in Bulgarian

According to the traditional academic descriptive grammar works, the semantics of definiteness in standard Bulgarian language is expressed in three ways: lexical, morphological, and syntactic.

The lexical way of expressing definiteness is closely connected with the lexical semantics of a particular lexeme, so that, it would not be of semantic value for our formal interpretation. It is because of the fact that most of the formal grammar interpretations are supposed to deal with one-sense lexeme, which is the simplest possible type of lexeme.
At the syntactic level, the definiteness in Bulgarian could express various types of semantic relationships like a case (to show subject), part-of-whole, deixis etc. Some authors tend to analyse this feature as a way to compensate for the lack of case declensions.

It is important to note that instead of the very many existing descriptive grammar works devoted to the problem, the semantics of definiteness in Bulgarian seems still not a sufficiently well studied phenomenon.

2.1 The Formal Morphological Markers of the Definite Article

As it has been pointed, the syntactic function of definiteness in Bulgarian is expressed by a formal morphological marker which is an ending morpheme \[7\]. It is different for the genders, however, for the masculine gender two types of definite morphemes exist – to determine a defined in a different way entities, which have two phonetic variants, respectively.

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

The following part-of-speech in Bulgarian take a definite article: nouns, adjectives, numerals (both cardinals and ordinals), possessive pronouns, and reflexive-possessive pronoun.

The definite morphemes are the same for all part-of-speech, however, in the further description we are going to analyze the inflectional morphology simply for the adjectives and the numerals.

3 Some Preliminaries for Possible Formal Interpretation

As it has been pointed, the very complicated semantics of the feature of definiteness in Bulgarian requires a related specific formal approach of interpretation. In fact, it should consists of two embedded in one system stages.

The first one is a formal interpretation of the inflectional morphology, and the second should give a syntactic interpretation of the first. The first stage requires to apply standard computational morphology approaches like a rule-based concatenation of morphemes and construction of relevant rules for their combinations, whereas the second requires an effective syntactic processing of various generated inflecting forms.

Obviously, such a semantically complicated model could hardly be found among the existing formal models like HPSG, for example, especially designed mostly for syntactic processing and normally producing only one type of inflecting forms.

We considered the DATR language for lexical knowledge presentation as a suitable formal framework for modeling the inflectional morphology of Bulgarian definite article.