

# Towards an HPSG Account of the Bulgarian Definite Article

ESSLLI Student Session 2000

Guergana D. Popova

Seminar Für Sprachwissenschaft

Universität Tübingen

E-mail: gpopova@sfs.nphil.uni-tuebingen.de

10.08.2000

In this paper I will examine the status of the definite article in Bulgarian and will offer a formal account of its morphology and syntax in the framework of HPSG. I will argue that the definite article /DA/ is best considered a **phrasal affix** and that it should be lexically attached to its host. I also show that the currently available architecture for morphology can, with some modification, generate the relevant word forms. The modifications made for morphology will then be used to account for the phrasal distribution of the definite article.

I will first discuss the status of the Bulgarian DA and will then present the proposed analysis.

## 1 The Status of the Bulgarian Definite Article

1.1 The article can appear on nouns, adjectives, numerals, full form of possessive pronouns and participles. In all cases it surfaces as a string which attaches itself phonetically (and orthographically) to the end of the word. The article agrees (with a few exceptions) with its host in gender and/or number.

1.2 The article for singular masculine nouns which end in a consonant or a glide is *-ät* (full form) and *-a* (short form) with allomorpha *-iat*, *-ia*, see ex. (1)

- (1) *motor* ('engine'), *motorät* ('the engine').

Singular masculine nouns which end in "a" or "ä" take the article *-ia*, those that end in "o" or "e" take the article *-to*, see (2)

- (2) *diado* ('grandfather'), *diadoto* ('the grandfather')

Feminine singular nouns take the article *-ta*.

- (3) *masa* ('table'), *masata* ('the table')

Singular neuter nouns have the article *-to*

- (4) *dete* ('child'), *deteto* ('the child')

Plural nouns take *-e* or *-ta*

- (5) *khora* ('people'), *khorata* ('the people').

1.3 Adjectives, numerals, participles and possessive pronouns take the same article as the noun they agree with, subject to the following condition: **there is at most one DA per noun phrase and that appears as an ending on the first element of the noun phrase.** Cf. examples below:

- (6) *Knigata e na masata.*  
Book-THE is on table-THE.  
'The book is on the table.'

- (7) *Novata kniga e na masata.*  
New-THE book is on table-THE.  
'The new book is on the table.'

- (8) *Hubavata nova kniga e na masata.*  
Nice-THE new book is on table-THE.  
'The nice new book is on the table.'

- (9) \*Nova knigata e na masata.  
New book-THE is on table-THE.  
‘The new book is on the table.’

#### 1.4 Definitions:

- **Clitics** are phonologically deficient elements which cannot bear independent stress and whose distribution is governed either by normal rules of syntax (*simple clitics*) or by special rules (*special clitics*). Clitics are postlexically attached.
- **Phrasal affixes** are lexically attached to their stems, but their distribution is determined with respect to complete phrases.

1.5 Applying to the DA the criteria for distinguishing between words and clitics/affixes as outlined in (Zwicky, 1985) proves its dependent status. I have added to them some of the results of the work of (Elson, 1976)

1.6 The DA forms a phonological word with its host and never takes independent stress. It can, though, sometimes attract word stress onto itself, mainly in the case of monosyllabic singular masculine nouns of Bulgarian origin and singular feminine nouns. The same observation leads Elson (1976) to conclude that the articles in Bulgarian “would not qualify as enclitics because they can occur stressed like ordinary syllabic suffixes while the undisputed clitics cannot”.

1.7 (Elson, 1976) gives the following phonological arguments for the affixal status of the DA:

- whereas final devoicing occurs in Bulgarian at word boundaries, and between a word and a following clitic, it does not occur between the article and its host;
- the sequences of liquid and schwa which undergo metathesis before inflectional affixes but not before clitics, metathesize before the DA;
- the sequences obstruent+schwa+sonorant which lose the schwa before inflection, can do so before the DA as well.

1.8 The DA can be subject to the laws of vowel harmony, thus for plural nouns if the stem ends in “e” or “u”, the article is *-te*, if the stem ends in “ia” or “a”, the article is *-ta*. Cf. Stoianov (1983) and the following

examples: *borove* (‘firs’), *borovete* (‘the firs’), but *patishita* (‘roads’), *patishitata* (‘the roads’).

1.9 The article is strictly ordered with respect to adjacent morphemes in a word. It always appears as the last morpheme, and is added to the stem subsequent to inflectional endings.

1.10 The article has access to the internal morphological structure of the stem it attaches itself to. For masculine nouns ending in a consonant the article has two full and two short forms: *-ât*, *-iat* and *-a*, *-ia*. The forms *-ia* and *-iat* appear after consonants that are historically soft. Thus, the form corresponding to the noun *den* (‘day’) is *deniat* (‘the day’). Most often the soft allomorph will appear on nouns ending in the agentive suffix *-tel*, e.g. *uchi-tel-iat* (‘the teacher’). If a noun ends in “tel”, but this is not the suffix *-tel*, then the article will remain *-ât*, cf. *motel-ât* (‘the motel’)

1.11 The combination of an element X and a DA forms a single unit with respect to syntactic rules. Thus we cannot delete, replace with a pro-form or move only the element X in a sentence, without doing the same with the DA.

1.12 The above point out to an affixal nature for the Bulgarian DA. However, the DA in Bulgarian can appear only once in a given noun phrase. Therefore, (10) is grammatical, but (11) is not.

(10) Pârvata kâshita se pokaza na horizonta.  
First-DEF house REFL appeared on horizon-DEF.

‘The first house appeared on the horizon.’

(11) \*Pârvata kâshitata se pokaza na horizonta.

First-DEF house-DEF REFL appeared on horizon-DEF.

‘The first house appeared on the horizon.’

The presence of another determiner in a given NP, e.g. a demonstrative pronoun, blocks the use of a noun with a DA. Therefore, (12) is also ungrammatical. The demonstrative pronoun itself cannot bear the DA (cf. the ungrammaticality of (13)).

(12) \*Tazi káshtata e mnogo zdrava.  
This house-DEF is very sturdy.  
‘This house is very sturdy.’

(13) \*Tazita káshta e mnogo zdrava.  
This-DEF house is very sturdy.  
‘This house is very sturdy.’

## 2 Clitic or Affix?

Zwicky and Pullum (1983) give 6 criteria to distinguish between affixes and clitics.

2.1 “Clitics can exhibit a low degree of selection with respect to their hosts, while affixes exhibit a high degree of selection with respect to their stems”. The DA in Bulgarian seems to have a rather low degree of selection. It can attach itself to nouns, adjectives, participles, numerals, and some pronouns. Miller (1992) notes, however, that this criterion is compatible with both affixal and postlexical status.

2.2 “Arbitrary gaps in the set of combinations are more characteristic of affixed words than of clitic groups”.

The DA can attach itself to all nouns and noun modifiers (with some exceptions). The rules of combination with proper names are more complex. It could be argued though that the exceptions are not arbitrary, but are rather based on semantic considerations.

There is one arguably arbitrary gap in the paradigm. Short possessive pronouns in Bulgarian can only appear after definite nouns. See example (10).

(14) káshta\*(ta) mi  
house-\*(DEF) my  
‘my house’

There is a group of nouns, though, which can be followed with a short possessive pronoun, but without the DA. These form a natural class: they all refer to family relations. See (11) for an example.

(15) bashta(\*ta)mu  
father-\*(DEF) his  
‘his father’

This exception would point to the DA being an affix, rather than a clitic.

2.4 “Semantic idiosyncrasies are more characteristic of affixed words than of clitic groups”. The DA does not exhibit semantic idiosyncrasies.

2.5 “Syntactic rules can affect affixed words, but cannot affect clitic groups”. As we have already pointed out, the combination of host + DA is treated as a unit by syntactic operations which at no point make reference to its internal structure. This suggests that the DA is an affix rather than a clitic.

2.6 “Clitics can attach to material already containing clitics but affixes cannot”. The DA attaches itself to the stem after it has received its inflectional morphemes. Nothing else can intervene between the stem and the DA.

To the above Miller (1992) adds the criteria of coordination. In his analysis an item must be an affix if it cannot have wide scope over coordination, and if it must be repeated on each conjunct. Applying the coordination criteria to the Bulgarian DA, however, gives inconclusive results.

The phonological facts strongly suggest an affixal treatment for the DA. Its behaviour in syntax, however, makes it impossible to put it on an equal footing with inflectional affixes. The notion of **phrasal affix** seems to offer an appropriate categorization for the DA.

I will, therefore, give weight to the phonological facts and conclude that the definite article in Bulgarian is a phrasal affix and must be generated by morphology, with syntactic rules governing its distribution.

## 3 The Analysis

### 3.1 Morphology

I will assume here the HPSG architecture for morphology as first outlined by Höhle (1999) and elaborated on in (Sailer, 1999) and (Richter and Sailer,

to appear). Sailer’s remarks are largely based on work by (Reinhard, in progress).

Sailer introduces recursive structure into morphology and assumes the sorts *basic-morph* and *complex-morph* within the following sort hierarchy (Sailer, 1999):

```

sign PHON phonsstring
  SYNSEM synsem
  :
morph M-SUBCAT list(synsem)
  M-SPEC list(morph)
  :
basic-morph M-DTRS m-const-struct
complex-morph

```

Sailer (1999) outlines a word principle that ensures that we do not have to give redundant descriptions of lexemes and their word forms.

I will adopt these assumptions and use inflectional mechanisms to build word forms of nouns, adjectives, participles, etc., with the definite article attached to them. In the current setup the lexicon consists of basic morphs.

The definite article *-ât* will then have an entry as in Fig 1  
 The description of the stem “stol” follows the account of nouns in (Pollard and Sag, 1994), except that it also bears the feature DEF (with a negative value), is a morph and has an AGR feature. See Fig. 2.  
 Additional Principles:

- **The Definiteness Principle.** If at least one of the daughters of a phrase or of a complex-morph is marked DEF+, then the mother node also bears the marking DEF+. Otherwise it is DEF-.
- **QSTORE principle** The QSTORE in phrases and complex morphs is the set union of the QSTOREs of the daughters. (This is very much like the Quantifier Inheritance Principle of (Pollard and Sag, 1994, p.48); but is used here for morphology as well).

The description of the resulting complex-morph is given on Figure (3).

### 3.2 Syntax

One can ensure that the DA will only appear on the first element of a NP by using the DEF feature introduced on the level of morphology.

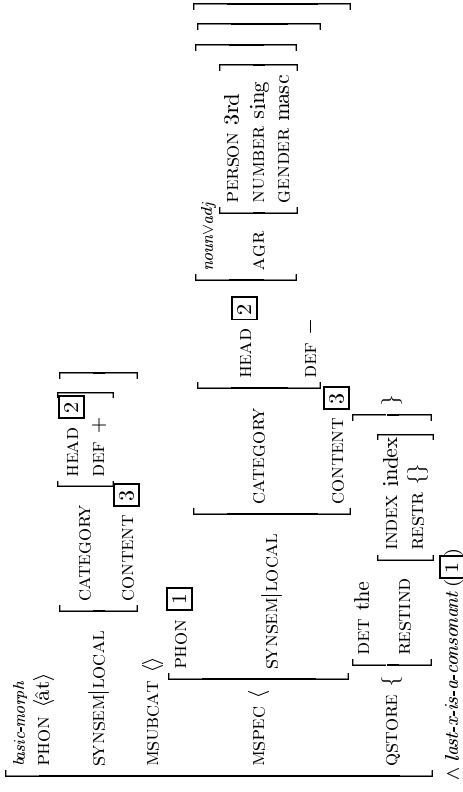


Figure 1: An entry for the DA *-ât*

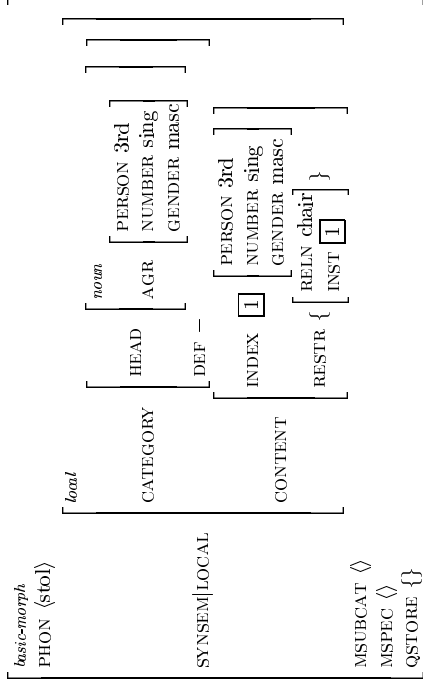


Figure 2: A lexical entry of the stem *stol*

It will be necessary to stipulate that modifiers, specifiers, etc., which normally precede the noun, can only select (via MOD or SPEC) for N with  $[DEF-]$ .

This will mean that contructions like (16) are correctly predicted to be grammatical, whereas the ungrammatical (17), (18), and the like are ruled out.

- (16)  ${}_{DEF+}$ [Novata  ${}_{DEF-}$ [interesna  ${}_{DEF-}$ [bálgarska  ${}_{DEF-}$ [kniga]]]].  
New-THE interesting Bulgarian book.  
'The new interesting Bulgarian book.'
- (17)  $*{}_{DEF+}$ [Nova  ${}_{DEF+}$ [interesnata  ${}_{DEF-}$ [bálgarska  ${}_{DEF-}$ [kniga]]]].  
New interesting-THE Bulgarian book.  
'The new interesting Bulgarian book.'
- (18)  $*{}_{DEF+}$ [Nova  ${}_{DEF+}$ [interesna  ${}_{DEF+}$ [bálgarska  ${}_{DEF+}$ [knigata]]]].  
New interesting Bulgarian book-THE.  
'The new interesting Bulgarian book.'

Other determiners, like the demonstrative pronoun, can be marked  $[DEF+]$  in the lexicon. Since they also come first in a NP, this will predict the fact that they cannot bear or cooccur with the DA. Cf. example (12).

## 4 Conclusion

This paper examined the status of the definite article in Bulgarian and offered an account of its morphology and syntax in the framework of HPSG. It showed that the DA is best considered a phrasal affix and should be lexically attached to its host. It then went on to show that the currently available architecture for morphology can, with some modification, generate the relevant wordforms. What is more, the modifications made for morphology can then be used to account for the phrasal distribution of the DA. I have left out some peculiarities of its distribution with respect to NPs containing simultaneous demonstrative pronouns and adjectives (ex. (19)), NPs with two nouns (ex. (20) and (21)), NPs with adverbs (ex. (22)). These topics are left for future research.

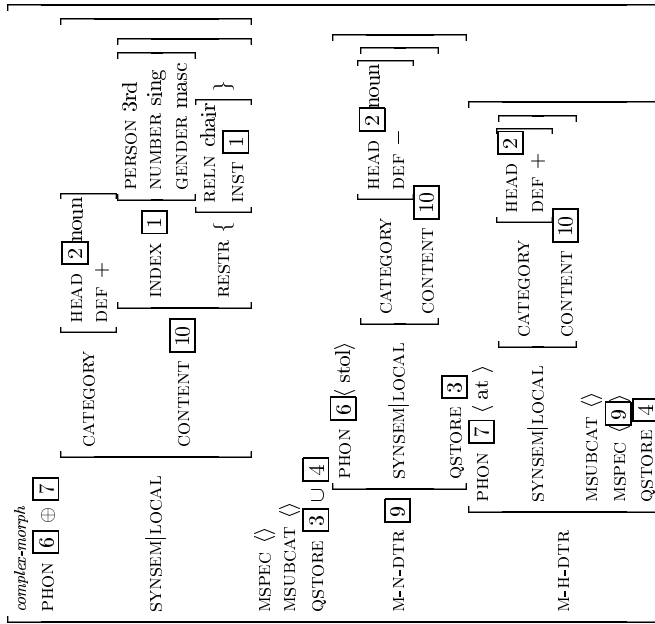


Figure 3: A description of the complex-morph *stóit*.

- Sailer, M. (1999). *The content of CONTENT*. PhD thesis, Tübingen University, Tübingen. Draft version of November 24, 1999.
- Stoianov, S., editor (1983). *Gramatika na sávmennáa búlgarski knižhoven ezik*. Izdatelstvo na búlgarskátata akademiia na naukite, Sofia.
- Zwicky, A. (1985). Clitics and particles. *Language*, **61**(2), 283-305.
- Zwicky, A. and Pullum, G. (1983). Cliticization vs. inflection: English *n't*. *Language*, **59**(3), 502-513.

- (19) Iskam тази, hubavata kola  
I-want this nice-THE car  
'I want this car, the nice one'
- (20) Asistent-rezhisorát priklínchi rechta si.  
Assistant director-THE ended speech his-REFL.  
'The assistant-director ended his speech'
- (21) Poetát revolútsioner zagina za svobodata.  
Poet revolutionary-THE died for freedom.  
'The poet revolutionary died for freedom'
- (22) Darvetata vliavo bialha izsecheni.  
Trees-THE [on the left]<sub>(adv)</sub> were cut.  
'The trees on the left were cut down'

## References

- Elson, M. J. (1976). The definite article in Bulgarian and Macedonian. *Slavic and East European Journal*, **20**(3), 273-279.
- Höhle, T. (1999). An architecture for phonology. In B. Borsley and A. Przepiórkowski, editors, *Slavic in Head-driven Phrase Structure Grammar*. CSLI Publications.
- Miller, P. (1992). *Clitics and Constituents in Phrase Structure Grammar*. New York, Garland.
- Pollard, C. J. and Sag, I. A. (1994). *Head-Driven Phrase Structure Grammar*. University of Chicago Press, Chicago.
- Reinhard, S. (in progress). -ung-Nominalisierungen und -ung-Komposita. Ein Fall für die Morphologie-Syntax-Semantik-Schnittstelle. PhD Thesis, Tübingen University.
- Richter, F. and Sailer, M. (to appear). On the left periphery of German finite sentences. To appear in a CSLI volume edited by Tibor Kiss and Detmar Meurers.