



**University of  
Zurich** UZH

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2020

---

## **Definiteness markers in the Life of St Petka**

Šimko, Ivan

DOI: <https://doi.org/10.1515/slav-2020-0014>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-188027>

Journal Article

Published Version

Originally published at:

Šimko, Ivan (2020). Definiteness markers in the Life of St Petka. *Zeitschrift für Slawistik*, 65(2):272-307.

DOI: <https://doi.org/10.1515/slav-2020-0014>

Ivan Šimko\*

## Definiteness markers in the *Life of St Petka*

<https://doi.org/10.1515/slav-2020-0014>

**Abstract:** The following paper addresses the problem of the diachronic development of definiteness markers in Bulgarian as attested in literature. As the previous research focused on the postponed article as the main definiteness marker in present-day Bulgarian, our study compares its use with that of other markers, namely the long-form adjective and the adnominal demonstrative pronoun. The traditional view sees a steady convergence between the literary norm and the dialectal use of the article. To challenge this view, we observe the frequency of use of all definiteness markers on an empirical basis. Our study uses a sample containing six versions of the hagiography of St Petka of Tarnovo, also known as St Parascheva of Epibates. The text for this study is provided by a historical corpus of pre-standardized Balkan Slavic literature, which is being constructed to enable us comparisons between the dialectal and diachronic data. The samples are chosen to minimize the effects of genre, structure and content differences on the results. The frequency of individual types of definiteness markers in respective versions of the text shows that the development of the damaskini language was not only different in respect to dialectal developments in the area, but in some respects even contrary to them.

**Keywords:** Balkan Slavic, Damaskini, definite article, Church Slavonic, historical linguistics

---

**Note:** This study has been written within the framework of the project ‘Ill-bred sons’, family and friends: tracing the multiple affiliations of Balkan Slavic, led by Prof. Barbara Sonnenhauser of the University of Zurich, funded by the Swiss National Science Foundation (SNSF project grant IZRPO \\_177557/1), to whom I would like to express my thanks for support. I would also like to express my thanks to Prof. Jürgen Fuchsbauer of the University of Innsbruck for providing necessary consultations and materials and to Mag. Teodora Vuković (working in SNSF/ Era.Net RUS project grant 100015\\_176378/1) for help with the IT aspect of this study.

---

**\*Corresponding author: Ivan Šimko**, University of Zurich, Slavic Seminary, Plattenstrasse 43, 8032 Zurich, Switzerland, E-Mail: [ivan.simko@uzh.ch](mailto:ivan.simko@uzh.ch)

## 1 Introduction

The postponed definite article is a characteristic feature of the Bulgarian language. It is one of its most prominent Balkan areal features: similar constructions may be found in Albanian, Greek, Romanian and neighboring Slavic varieties as well. As such, it has been a subject to many studies from the earliest times (Leake 1814: 380; Schleicher 1983 [1850]: 210) until today (Haarmann 1976: 85; Friedman 2006: 661; Lindstedt 2014: 168). An expression of definiteness with a special morphological marker can only rarely be found in Slavic languages outside of the Balkans (Miletić 1889; Hansen 2004; Mendoza 2011). The process of its emergence is well studied (e.g. Conev 1934; Mayer 1988; Mladenova 2007), but its relation to older definiteness markers, such as the distinction between short- and long-form adjectives (Flier 1974: 65; Lunt 2001: 142) and the adnominal demonstrative pronoun (Miletić 1923: 28), still leaves some questions open.

In this paper, we would like to discuss these questions with the help of empirical methods. Due to the character of the early Modern Bulgarian literature, it is possible to compare linguistic features in parallel corpora of texts recurring in this tradition. One such text is the hagiography of St Petka, which has served as a testbed for a new, reusable digital corpus of pre-standardized Balkan Slavic varieties currently in development. This tool enables us to study the full variety of the morphological realization of respective definiteness markers.

This article begins (2.) with a short description of the transitional period between the fall of Tarnovo and the standardization of the modern literary Bulgarian in the 19th century, as well as of the literary sources preserved from this era. In the next chapters, we shall briefly discuss (3.) the concept of definiteness and (4.) its morphosyntactic expressions. Following these theoretical sketches, we shall (5.1) introduce a sample from our corpus, (5.2) explain the way our corpus deals with language variation between our sources, and finally (5.3) compare the frequencies of particular types of respective definiteness markers in them.

## 2 Pre-standardized Bulgarian

The language of the Old Church Slavonic sources (10th–12th century) is often described to be close to a vernacular spoken in the southern half of the Slavic-speaking area at the time of Constantine and Methodius (Lunt 2001: 4; Trunte 2005: xi; Dimitrova-Vulchanova & Vulchanov 2009: 91). However, sources from this period contain mostly sacred texts. The texts tried to differ minimally from originals. Later writers, especially from the Tarnovo school in the second half of the 14th century, systematically avoided some recent changes. Only due to irregular devia-

tions from this norm (i.e. “mistakes”) we get a glimpse at the changes in the local vernacular. Generally, the gap between the literary Church Slavonic and spoken Balkan Slavic varieties was growing.

The following three centuries represent a transitional stage in the history of Balkan Slavic. The orthography of the Tarnovo school is slowly replaced by that of the Resava school, which loosens some archaic conventions<sup>1</sup>. Shortly thereafter, in 1568, Damascenus Stoudites publishes his *Thesaurus*, a collection of homilies for religious and moral education of the lay people. Providing eclectic contents in an accessible language, the book gains popularity among both Greek and Slavic clergy. It is soon translated into Church Slavonic and later, within a generation, into the Bulgarian vernacular. These texts develop into collections of varying contents and the Damascenus’ name becomes eponymous with this new literary tradition, the *damaskini*.

These two events were a key to the development of modern literary Bulgarian. The technology of print was virtually inaccessible in the Ottoman-controlled territories: *Abagar* (1651), the first book in modern Bulgarian, was printed in Rome, and also the original Stoudites’ *Thesaurus* was published in Venice<sup>2</sup>. The *damaskini* still bear characteristics of medieval literature. Although based on printed sources, they all are unique manuscripts, *ad hoc* recensions of previous models. Contents of most *damaskini* collections, especially later ones, are different. A moderately educated scribe could be the editor and the compiler of his *damaskin* as well.

*Damaskini* sources are not written in a unified language. They often contain older texts in Church Slavonic, alongside those translated into a new literary language *na narodnoj osnove* (‘on a vernacular basis’; cf. Demina 1985: 31). These texts are usually authored by Damascenus himself. This language, termed differently from source to source as a “Bulgarian” or a “commoner” language, differs from source to source as well. One of the more voluminous of such sources, the Tixonravov *damaskin* (Demina 1972), contains translations into at least two different dialects. There are no systematic attempts at codification of Bulgarian vernaculars until the 19th century. Only then a centralized literary norm, supported by print, mass education and spread of secular literature, could start to take shape

---

1 The most striking difference between the two schools is the writing of some archaic letters. Characters for former nasal vowels *а* (*ę*) and *я* (*ǫ*) are replaced by *е* (*e*) and *ѣ* (*u*) in the Resava orthography, writing e.g. *jezikъ* or *jezykъ* ‘tongue’, *ruka* ‘hand’, instead of former *jezykъ*, *rъka*. The nasals were kept in the Tarnovo orthography, but subject to different rules. Tarnovo orthography remained in use in Moldavia, and it is preserved in some Russian sources as well (Velčeva 1966: 111; Velčeva & Mladenova 2013: 101).

2 Many reprints of *Thesaurus* were published in Venice as well (e.g. *Thēsaurus* 1799).

and hold. This all makes the damaskini an invaluable source for diachronic studies, but also not an easily accessible one.

### 3 Concept of Definiteness

Damaskini were used to demonstrate the developments of the definiteness markers very soon after their discovery by modern scholars in the late 19th century (Miletič 1903: xlii; Conev 1906: 510). Recently, they were used by Mladenova, who compared the use of definite article in multiple texts, including the *Life of St Petka* (Mladenova 2007: 258), across a number of damaskini sources and her own translations. In the terminology of these scholars, the reader is usually presented with the description of *členna forma* ('articled form'), but not always with the description of the function expressed by the article, i.e. by the concept of definiteness.

Since the earliest theoretical works on the topic (e.g. Momčilov 1868: 29; Miletič 1889: 1<sup>3</sup>), most linguists agree that the Bulgarian (or, more generally, Balkan Slavic) definite article had developed from earlier postponed demonstrative pronouns. From the aspect of morphology, the article is realized as a suffix (MASC.SG Bulg. *-ъt/-jat*, Mac. *-ot*, FEM.SG *-ta*, NEUT.SG *-to*, PL *-te*, NEUT.PL *-ta*) in most paradigms. A similar development of demonstratives can be found in other languages marking definiteness as well (Lyons 1999: 325). Both modern demonstratives and articles refer to identifiable objects: the hearer should be able to identify their referents (Mladenova 2007: 3). The question should then be: what is their difference?

According to Lyons (1999: 17), demonstrative pronouns are non-inclusive determiners. While definiteness markers refer to the "totality of the objects or mass in the context which satisfy the description" (Lyons 1999: 11), a demonstrative limits the choice. Present-day standard Bulgarian reflects this choice in the way of the following examples by Mladenova (2007: 185):

- |     |   |           |                    |               |                 |
|-----|---|-----------|--------------------|---------------|-----------------|
| (1) | <i>Eto</i>                              | <i>gi</i> | <i>podarъcite.</i> | <i>Gledaj</i> | <i>kuklata!</i> |
|     | here                                    | they.ACC  | gift.PL.DEF        | look.2SG.IMP  | doll.DEF        |
|     | 'Here are the gifts. Look at the doll!' |           |                    |               |                 |

<sup>3</sup> Earliest grammarians avoid discussions about the function or origin of the Bulgarian article. Although Neophyte (Rilski 1835: 43) likely understood how the Greek article was used, as he considered it to be an equivalent of the Bulgarian one, he does not describe it. Bogorov (Andreov 1844: 12) describes the function laconically *da si otredimъ razumотъ* ('to focus our mind').

- (2) *Eto gi podarǎcite. Gledaj тази kukla!*  
 here they.ACC gift.PL.DEF look.2SG.IMP that doll

According to Mladenova, the first example could work only in a situation if there is one single doll among the gifts. In such a situation, the choice between both expressions would be free; if there were more dolls among the presents, only (2) would work. Similarly *podarǎcite* (and not *\*tezi podarǎci*) refers to the bunch of presents as a whole, and not only a specific part of them, including the (or that) doll. This is a case, which Mladenova (2007: 60) describes as “deictic definiteness”. The problem is, that such references in written texts, both contemporary and old, are too rare. Thus, we cannot conclude much about the diachronic development of the function of these two markers on their basis alone.

Another function of Bulgarian definiteness markers is to express anaphoric reference (Mladenova 2007: 99). Objects, which have already been mentioned within the text, are regularly marked by articles. Such situations, at least in comparison to deictic definiteness, are abundant in the literary sources, ancient and modern as well. A modern Bulgarian example of such “anaphoric definiteness” can be observed on *ženite* ‘the women’ in the following passage from Jordan Jovkov’s short story *Šibil*:

- (3) *toj vidja, če dolu po pǎtja idat ženi.*  
 he see.3SG.AOR that below on path.DEF.OBJ go.3PL.PRS woman.PL  
 [...] *Ženite se pokazaxa na zavoja*  
 woman.PL.DEF REFL show.3PL.AOR at turn.DEF.OBJ  
 ‘He saw women going on the path below. [...] The women showed themselves at a turn.’

Mladenova considers the use of the definiteness marker for anaphoric reference to be optional. Anaphoric references with some stylistical, “emotional, ironic or critical” marking would be marked by demonstrative pronouns instead of the definite article (Mladenova 2007: 99). In the following passage from Anton Strašimirov’s story *Na nivata* (‘On the croft’), the author seems to express some amazement about the fact, that an intelligent child has to plow for his drunk father (again!), instead of going to school:

- (4) *Na osta sedeše – s osten v rǎkata – dete.*  
 on axis.DEF.OBJ sit.3SG.IMP with goad in hand.DEF child  
 [...] *A tova dete beše svetlo i vǎzpriemčivo*  
 but that child be.3SG.IMP bright and intelligent  
 ‘A child sat on the axis, with a goad in his hands. [...] But that child was bright and intelligent.’

Such constructions can be found in the damaskini sources as well. According to Miletič (1923: 28), the postponed article was freely interchangeable with adnominal demonstrative pronouns, which he interprets as a *prědpostavenъ членъ* ('prepositioned article'). His examples include mostly anaphoric references, such as "the letter" in the following passage from the *Slovo o čudesěxъ presvetie vladyčice* ('Homily about the miracles of Our Most Holy Lady'):

(5)	<i>Tъkvózi</i>	<i>ľžóvno</i>	<i>pisánie</i>	<i>napísa</i>	<i>onázi</i>			
	such.NOM.SG.DEM	false.NOM.SG	letter	write.3SG.AOR	DEM.NOM.SG.DEM			
	<i>ľstíva</i>	<i>carica. (...)</i>	<i>i</i>	<i>zaněsi</i>	<i>unúj</i>			
	cunning.NOM.SG	queen	and	bring.3SG.AOR	DEM.NOM.SG.DEM			
	<i>rukopísánie</i>	<i>ľstívo</i>	<i>na</i>	<i>gospodína</i>	<i>svoého</i>	<i>i</i>	<i>dádi</i>	
	letter	cunningly	to	lord.GEN.SG	REPL.GEN.SG	and	give.3SG.AOR	
	<i>rukopísánieto</i>	<i>da</i>	<i>gó</i>	<i>pričeté</i>				
	letter.DEM	to	it.ACC	read.3SG.PRS				

'Such a false letter was written by that cunning queen. [...] and then she brought the letter cunningly to her lord and gave (him) the letter to read'

*Svištov damaskin* (Miletič 1923: 159)

There seems to be no stylistic difference between the *unúj rukopísánie* and *rukopísánieto*, as Mladenova sees in the modern usage of these markers. Another difference between the present-day and the language of *Svištov damaskin* is the deictic type of the demonstrative pronoun. Demonstratives used to mark anaphoric references in present-day Bulgarian are usually of the type with an unmarked deixis (*tozi, taja, tova*), as seen in the example (4), while the older sources prefer distal pronouns (*onzi, onuj, onaja* etc.)<sup>4</sup>. Velčeva (1964: 167) argues, that the use of distal pronouns has been influenced by their use in the Tarnovo Church Slavonic literary tradition (which avoided the demonstrative suffixes altogether) and ultimately by its translations of the Greek pronoun *ekēinos*. The parallel use of these two competing markers for the anaphoric reference is common to various texts in sources studied by Miletič, which otherwise differ in many other linguistic and philological respects<sup>5</sup>. Thus, it is plausible to say that this phenomenon was common to a wider dialectal area, and not a feature of the scribe's individual style.

<sup>4</sup> Attested form *unúj* is a phonetic variant of *onuj*, with a shift in the unaccented vowel (cf. also *zaněsi* instead of expected \**zanese* 'she brought' or *dádi* instead of \**dade* 'she gave').

<sup>5</sup> Miletič refers to two damaskini from various eras and translation circles: *Koprivštica damaskin*, an older (17th cent.) manuscript, reflecting the early Sredna Gora circle (Petkanova-Toteva 1965: 63; or "type I Modern Bulgarian" damaskini, according to Demina 1968: 54; edition by Miletič 1908) and *Svištov damaskin*, a newer (dated 1753) one, based on a translation from the Northeastern Bulgaria

The third function of the definite article is to mark generally known objects (Lyons 1999: 4). In the following passage from Mladenova (2007: 20), it is clear who “the president” is:

- (6) *Sam prezident* *dojde* *na pogrebenieto.*  
 self president.DEF come.3SG.AOR to funeral.DEF  
 'The president himself came to the funeral.'

According to Lunt (2001: 142), the expression of definiteness for “generally known objects” can be found in Old Church Slavonic sources already, and likely even earlier. Greek noun phrases with definite articles marking such objects were translated using a specific morphological marker – the long-form adjective. Such is the case of “the Kingdom of Heaven” in the following passage from the Gospel of Matthew (Mt 5:3), as found in the *Codex Zographensis* (*Zogr.*) from the 10th–11th century<sup>6</sup>:

- (7) *autōn estin hē basileía tōn ouranōn*  
 their.GEN.PL be.3SG.PRS def.NOM.SG kingdom.NOM.SG DEF.GEN.PL heaven.GEN.PL  
*těxъ estъ cъso nъskoe*  
 they.GEN.PL be.3SG.PRS kingdom.NOM.SG heavenly.NOM.SG.LF  
 'for theirs is the Kingdom of Heaven'

The use of long-forms was not limited to this function. In the oldest sources, the long-form can express deictic and anaphoric references as well, as in the case of *bouię (děvbi)* ‘the foolish (maidens)’ in the following passage (Mt 25: 2–3) from *Zogr.*:

- (8) *pětъ že bě otъ níxъ boui [...]*  
 five REL be.3SG.AOR of they.GEN.PL foolish.NOM.PL

(Petkanova-Toteva 1965: 241; or “type IV”). He was already aware of the relations between the *Koprivštica*, *Trojan*, *Tixonravov* and *Ljubljana damaskin*, *Koprivštica d.* being considered the oldest among them (Miletič 1903: xxi).

Mladenova (2007) refers usually to *Tixonravov damaskin*, whose origins are similar to those of *Koprivštica d.* (see below), but also to later damaskins of Hilandar monastery, *Kotel*, *Žeravna*, *Svištov* (the only “type IV” damaskin among her sources) and others.

<sup>6</sup> For Greek Bible citations we are using the Perseus digital library – <http://www.perseus.tufts.edu/hopper/>. Last accessed: 9 January 2020. For OCS citations (*Zogr.*, *Cod. Assemanius*, *Cod. Marianus*, *Savvina kn.*) we are using transcriptions from the TITUS project – <http://titus.uni-frankfurt.de/texte/etcs/slav/aksl/zograph/zogra.htm>. Last accessed: 9 January 2020.



<i>bouię</i>	<i>bo</i>	<i>priemъšę</i>	<i>svętilnikъi</i>	<i>svoę,</i>	<i>ne</i>
foolish.NOM.PL.LF	as	take.ACT.PTCP	candlestick.ACC.PL	REFL.POSS.ACC.PL	NEG
<i>vъzęšę</i>	<i>bo</i>	<i>sъ</i>	<i>sobojo</i>	<i>i</i>	<i>olęę</i>
take.3PL.AOR	as	with	REFL.INST	too	oil.GEN.SG

'Five of them were foolish [...] The foolish ones took their lamps but did not take any oil with them.'

There are other functions of the mentioned definiteness markers, similar in meaning to the previous three, too. They may refer to objects known from the context of the narrative like e.g. *pogrebenieto* 'the funeral' in example (6), or inferred knowledge, like *rъkata* 'the hand' (for children are expected to have hands) in example (4) or *pъtja* 'the path' (as robbers tend to observe paths) in example (3). These are the major components of the concept of definiteness in Bulgarian.

There is not only a rich variety of what we understand as definiteness here, but also of its morphological realization, especially from the diachronic point of view. In our examples, we could see postponed demonstratives, like *podarъcite* in (1), prepositioned demonstrative pronouns, like *tova dete* in (4) and long-form adjectives, like *bouię (dęvъi)* in (8). A fourth form is the “short” article (*-a/-ja*), reserved for masculine nomina in object position, seen e.g. in example (3) as *po pъtja* 'on the path' or *na zavoja* 'at the turn'. In the following example by Mladenova (2007: 70), the *avtobusъt* 'the bus' receives the full article, while *noštниja vlak* 'the night train', as the sentence's object, a short one. The difference is not marked on the FEM noun *vъzkata* 'the connection':

(9)	<i>Avtobusъt</i>	<i>pravi</i>	<i>vъzkata</i>	<i>s</i>	<i>noštниja</i>	<i>vlak</i>
	bus.DEF	do.3SG.PRS	connection.DEF	with	night.DEF.OBJ	train
	'the bus does make a connection with the night train'					

Most grammarians of present-day Bulgarian (Andreov 1844: 12; Momčilov 1868: 29; Miletič 1887: 306; Conev 1934: 502; Mirčev 1978: 196; Mladenova 2007: 4, 70) use the term “article” (*člen* or *prilogъ*) only (excluding the extended demonstrative) for the suffixes based on postponed demonstratives (standard variants *-ъt/-jat, -ta, -to*, pl. *-te, -ta*) and their short variants (*-a/-ja*). These are attached to the leftmost element of the noun phrase. The functions of particular forms of definiteness expression changed in time, as we will try to explain in the following sections.

## 4 Expression of Definiteness

As we have seen, some definiteness marking can already be found in the Old Church Slavonic sources. But can we speak of a definite article in their time (10th–12th century) as well? Anaphoric references, generally known objects and to some extent objects mentioned previously in the narrative are nigh-consistently marked by long-form adjectives and short demonstrative pronouns with three deictic variants (e.g. *tb*, *sb*, *ontb* for MASC.NOM.SG). These are used sometimes alone, sometimes in combinations of both markers in the noun phrase. The pronoun often appears in place of Greek demonstrative *ekeĩnos*, such as in the following passage (Mt 24:48) attested in *Zogr.*:

- (10) *eàn dè éipē ho kakòs doũlos ekeĩnos*  
 if but say.3SG.AOR.SUBJ DEF.NOM.SG bad.NOM.SG servant.NOM.SG DEM.NOM.SG  
*ašte li rečety zblvi rabby tb*  
 if q say.3SG.PRS evil.NOM.SG.LF servant.NOM.SGDEM.NOM.SG  
 'if that evil servant shall say'

Constructions such as example (10) led some scholars to presume that the present-day definite article has already developed in Old Bulgarian (e.g. Mirčev 1978: 200), or even earlier (Miletič 1887, 1889). These markers begin quite early to interact with the preceding noun as enclitics. This can be observed in some sources on the preservations of the old MASC.NOM.SG ending. Traditionally, the ending was written as *-b*, although it had no phonetic value in Old Church Slavonic since at least the 10th century. As the demonstratives became part of the prosodic word domain, perceived as a part of the noun, the preceding *jer* of the ending was reinterpreted as “strong” according to Havlík’s law and thus reflected in Balkan Slavic dialects as a schwa, *-a* or *-o* after subsequent phonetic shifts<sup>7</sup>. Resulting *-ot* endings are reflected already in earliest literary sources. The *Codex Assemanius* (*Assem.*) renders the above passage from Matthew (24:48) in the following way:

7 Cf. Mirčev 1978: 200. According to Mirčev, the schwa and *-a* is more common in Eastern dialects, *-o* in Western ones. However, there is a group of Moesian dialects using *-o* as an article as well, in the area of Ixtiman (Mladenov 1963: 406).

Sometimes the preserved *-o-* in similar positions is considered an older reflex of the Proto-Indo-European *o*-stem nom.sg ending *\*-os* (cf. Olander 2015: 3). In that case, the suffixation of postponed demonstratives would have proceeded even before the short *-u* (> OCS *-b*) was generalized as the nom.sg ending for the *o*-stem nouns.

- (11) *ašte li rečet*                    *zъlъi*                    *rabot*  
 if    Q    say.3SG.PRS            evil.NOM.SG.LF            servant.NOM.SG.DEM

Interactions similar to the one in the preceding example are not found between adjectives and demonstratives. The demonstrative was perceived either as a part of the following noun's prosodic word (i.e. it was a proclitic), or as a prosodic word on its own. Compare the same passage from the 11th century *Sava's Book* (*Sav.*):

- (12) *ašte li rečetъ*                    *zъlbi*                    *тъ*                    *rabъ.*  
 if    Q    say.3SG.PRS            evil.NOM.SG.LF            dem.NOM.SG            servant.NOM.SG

Later, the attached demonstrative with the preceding vowel was reinterpreted as a suffix, and as such would be attached to a preceding adjective. According to Conev (Conev 1906: 115) the earliest attestations of such a suffixed (or "articled") adjective can be found in the *Dobrejšo Gospel* (*Dobr.*) from the 13th century. The Mt 24:48 is rendered in this source in the following way:

- (13) *ašte li rečt*                    *zъlіotъ*                    *rabъ*  
 if    Q    say.3SG.PRS            evil.NOM.SG.LF.DEM            servant.NOM.SG

This evidence is often interpreted (Ilievski 1972: 67; Mirčev 1978: 202; Mladenova 2007: 4) as the final stage of the development of the article: as *-o* (as well as *-a* or schwa) had no other etymological justification at this position, it had to be perceived together with the demonstrative as a single, separate suffix. In this stage, its position was fixed to the leftmost element of the noun phrase. Thus, from the viewpoint of the said researchers, the Balkan Slavic definite article would have developed in the 13th century already<sup>8</sup>.

Such forms are not frequent in other sources from that time. The scribes were choosing from multiple options of expression of definiteness not only when translating Greek, but also for Church Slavonic. The old, unsuffixed long-form adjective did not fall out of use. Long-forms without such suffixes continue to mark definiteness by contrast to short-form adjectives still in the damaskini sources (Mladenova 2007: 339). For the orthography of the Tarnovo school, the suffixed

<sup>8</sup> Other attestations of short pronouns suffixed to adjectives are rare to find in other 13th century sources. Some sources are questionable. In the 16th century *Cserged Prayers*, a unique source from a Transylvanian village settled by fugitives from Western Bulgaria, we can find articulated adjectives (e.g. *nebeski-tui xori* 'of the heavenly host'). The relative isolation of the community from the rest of Bulgarian-speaking area makes the claim plausible, that such forms were present in the community's language in the time of their migration, i.e. in the 13th–14th century (Wahlström 2015: 39).

demonstratives of the *-tǔ* type were the less preferred, if not systematically avoided option. Some sources attest only other demonstratives (the types marked for deixis: i.e. *sǔ* and *onǔ*), while the word boundary between the pronoun and the substantive is orthographically marked by *jers* (as in the following example) or accents:

- (14) *eírǣskēn*      *ho*                      *basileũs*      *prǔs*      *tǔn*                      *hyiǔn*  
 say.3SG.AOR      DEF.NOM.SG      king.NOM.SG      to      DEF.ACC.SG      SON.ACC.SG  
*tǔn*                      *toũtu*  
 DEF.ACC.SG      this.GEN.SG
- rǣce*              *cǣb*              *ǔnǔ*              *kǔ*      *svoǣmu*              *sũu*  
 say.3SG.AOR      king.NOM.SG      that              to      refl.POSS.DAT      SON.DAT.SG
- 'the king said to his son' (Fuchsbauer 2010: 193)<sup>9</sup>

As already mentioned, the language of some damaskini sources knows a third option of expressing anaphoric definiteness as well: a prepositioned demonstrative pronoun (Miletič 1923: 28 f.; Velčeva & Mladenova 2013: 105). Unlike the postponed demonstratives, these are extended by a further demonstrative marker, as it can be observed in other Slavic languages as well<sup>10</sup>. Common variants of these are *-va*, *-j(a)*, *-zi*, as seen in previous examples, e.g. *tova* in (4), *tazi* in (2), *unũj* in (5)<sup>11</sup>. Their distribution was likely dialectal: the forms with the *-zi* suffix were originally used in Central and Moesian (Northeastern) dialects, while the form *-ja* was common in the Western area (Velčeva 1964: 173). The markers always remain uninflected, placed after the case ending, such as in *onǔmuzi*, here attested in the *Tixonravov damaskin*:

- (15) *togǣzi*      *rǣe*                      *sũy*                      *dimitrie*      *onǔmuzi*              *eǣkopu*  
 then.DEM      say.3SG.AOR      saint.NOM.SG.LF      Demetrius      DEM.DAT.SG.DEM      bishop.DAT.SG
- 'then St. Demetrius said to the bishop' (Demina 1972: 105)

<sup>9</sup> The example is from the 14th century Slavonic translation of *Dioptra*, an 11th century educatory poem by Philippos Monotropos of about 7000 verses. The translation shows no regular reflex for Greek articles (Fuchsbauer 2010: 177), the demonstratives of the *tǔ* type are not used at all.

<sup>10</sup> E.g. Cz. *tento* (Rejzek 2001: 691), Serb. *ovaj* (Skok 1971 II: 578), Russ. *etot* (Vasmer 1986 IV: 524) etc.

<sup>11</sup> Further examples of such demonstrative suffixes from contemporary dialects are presented in the Etymological Dictionary of BAN, e.g. *onǔto* (attested e.g. in the vicinity of today's Veliko Tarnovo), sometimes resulting a string of multiple suffixes, like *onǔzkana* (Vraca), *onovǣzika* (Dobruđža), *onũjnc* (Babica) etc. (Georgiev et al. IV 1995: 880 f.).

The oldest forms are those with the ending *-zi*, attested already since the 12th century (Mirčev 1978: 183), with the other markers following later<sup>12</sup>. These markings seemingly were not preferred in the 14th century Tarnovo orthography. Compare the following passage from the Euthymius' *Life of St Petka* (16), marking anaphoric relations with the emphatic particle *že* or demonstratives with a marked deixis (*smrada onogo*), with its translation from the 17th century (17), preferring the extended pronoun (*on'zi smrad*):

(16)	<i>Načeto</i>	<i>že</i>	<i>smradb</i>	<i>isxoditi</i>	<i>bezъměrnъ [...]</i>	<i>i</i>
	begin.3SG.AOR	EMPH	stench.NOM.SG	rise.INF	extreme.NOM.SG	and
	<i>stlbpniku,</i>	<i>ot</i>	<i>ježe</i>	<i>ne</i>	<i>mošti</i>	<i>nestъrpimago</i>
	stylite.DAT.SG	from	it.REL	NEG	can.INF	unbearable.GEN.SG
	<i>smrada</i>	<i>onogo</i>	<i>trъpěti (...)</i>			
	stench.GEN.SG	dem.GEN.SG	suffer.INF			

'An extreme stench began to rise (from the corpse). [...] (The) stylite, who couldn't bear that terrible stench (had to climb down from his pillar).' (Demina 1976: 33)

(17)	<i>počē</i>	<i>da</i>	<i>isxōdi</i>	<i>smrādъ</i>	<i>ot</i>	<i>njēgo</i>	<i>zlb [...]</i>
	begin.3SG.AOR	that	rise.3SG.PRS	stench.NOM.SG	from	it.GEN.SG <sup>13</sup>	bad.NOM.SG
	<i>i</i>	<i>on'zi</i>	<i>stlbpnikъ</i>	<i>vekie</i>	<i>ne</i>	<i>možē</i>	<i>da</i>
	and	DEM.NOM.SG.DEM	stylite.NOM.SG	more	NEG	can.3SG.PRS	to
	<i>trъpi</i>	<i>on'zi</i>	<i>smrādъ</i>				
	suffer.3SG.PRS	DEM.NOM.SG.DEM	stench.NOM.SG				

'A foul stench began to rise (from the corpse). [...] and that stylite couldn't bear that stench any more (and had to climb down from his pillar).'

*Tixonravov damaskin* (Demina 1972: 96)

<sup>12</sup> The etymology of the *-zi* suffix is uncertain. According to the BAN Dictionary (Georgiev et al. I 1971: 638) a pronoun *zi* alone is attested in some dialects, but the authors consider it a product of a secondary decomposition of the extended pronouns. A source could be the OCS nom.1sg pronoun *azъ*, which is in some later texts and dialects extended by a vowel (Mirčev 1978: 180). Thus the ending could be adopted for the demonstratives as well. *Wallachian Letters*, which attest *-zi* pronouns abundantly, use a nom.1sg pronoun form *aze* (along *ja* and *azъ*).

The *-j(a)* suffix follows the 3sg personal pronoun. Its hypothetical nominative *\*jb* is already attested in OCS only in compounds, such as the relative pronoun *iže* 'he that', 3sg personal pronoun *toi* 'he', or also the ending of the mentioned long-form adjectives. The form *-va* is likely based on an older demonstrative *ovъ*, which is still preserved in Macedonian and Serbian.

<sup>13</sup> The damaskini language did not distinguish between the genitive and accusative cases anymore. As the pronouns like *njēgo* or *-togo* ending etymologically correspond to Church Slavonic genitive (cf. Lunt 2001: 62), we mark it as such in the citations from damaskini sources. Cf. also below in section 4.2.

Other contemporary sources deviate from the Tarnovo school of Church Slavonic as well. Demonstrative pronouns with *-zi* suffixes are the preferred markers of anaphoric reference in the *Wallachian Letters* from the 15th century too<sup>14</sup>. One such instance is *togozi človeka* in the following letter (No. XVIII) from voivode Dan II (1420–1424, 1427–1431):

(18)	<i>ašte</i>	<i>kto</i>	<i>imat</i>	<i>něšto</i>	<i>dlbžno</i>	<i>vb</i>	<i>človeka</i>
	if	who	have.3SG.PRS	something	debt.ADV	in	human.GEN.SG
	<i>ot</i>	<i>tiqzi</i>	<i>zemlq,</i>	<i>a</i>	<i>on</i>	<i>da</i>	<i>si</i>
	from	DEM.ACC.SG.DEM	land.ACC.SG	and	he.NOM	to	REFL.DAT
	<i>naidet</i>	<i>togozi</i>	<i>človeka [...]</i>				
	find.3SG.PRS	DEM.GEN.SG.DEM	human.GEN.SG				

'If one has (a debtor) in this land, and he finds that (debtor) [...]'

So far it remains an open question, whether the long-form adjective (without the demonstrative suffix attached) was still viable in this function as well. It would be plausible to think otherwise: the anaphoric reference was marked at the end of the leftmost element of the noun phrase, being either the pronominal *-zi* extension, or the suffixed demonstrative. The literature starts to reflect these processes in the 15th century.

The postponed article was accepted into the literary language only later, in the damaskini period. The use of the article remained a controversial topic for early grammarians of the 19th century. In his *Life of St Alexius*, Konstantin Ognjanovič has argued against the use of the *městoimennija částicy* ('pronominal particles') *ta*, *to* or *te*, which he noticed to be pronounced after nouns and adjectives, as these are not used in any other Slavic language. The grammar would become “more beautiful” without them (Ognjanovič 1833: 62). In fact, these “particles” are found in Ognjanovič’s text as well, and thus they were likely deeply entrenched in the vernacular. They can be found in FEM.SG, NEUT.SG and in plurals, written separately from the noun or adjective:

<sup>14</sup> Seemingly conform to the Tarnovo orthographic rules, a postpositional *tb* is not (nor any other such pronoun) attested. This opens the question, whether the extended pronoun did not appear in the East as the new definiteness, replacing the distinction of short- and long-form adjectives, in contrast to the early article based on short pronoun, appearing in the West. Such a rough localization (based on a limited number of sources, e.g. *Cserged Prayers* and *Dobrejšo Gospel* for the West, *Wallachian Letters* and Tarnovo literature for the East) is speculative.

- (19) *I sedéxa vsička tã sédmica*  
 and sit.3PL.IMPF whole.FEM.SG DEM.FEM.SG week  
 'And they sat for the whole week' (Ognjanovič 1833: 60)

A Bulgarophone reader may wonder: how does he render the masculine article? In the Bulgarian literature of the early 19th century, the full MASC.SG.DEF particle *-bt* is not very common; “short” articles like *-a* or *-ja* are used instead. This morphem is still used in the present-day literary Bulgarian, but only in the position of a syntactic object, as already mentioned in the example (9). In the early 19th century, the traditional NOM.SG case marker *-b* remained in use only if the noun was not definite. The short article could be used for nouns in the subject position too, like *diavola* ‘the devil’ in the following passage from the 1856 edition of Sophronius’ *Life of St Petka*:

- (20) *diavola sja prestruvaše na razny zvěrove*  
 devil.DEF ref.ACC transform.3SG.IMPF to various.PL beast.PL  
 ‘The devil transformed himself to various beasts’ (Sofronii 1856: 257)

Other sources sometimes use a form of the short article *-o*. Unlike the ending *-a* or *-ja*, it does not have a parallel in Church Slavonic inflection paradigms:

- (21) *móžete da go izvadite o(t) stlpo*  
 can.2PL.PRS to it.ACC remove.2PL.PRS from pillar.DEF  
 ‘Could you remove (the body) from the pillar?’  
*Pop Punčov sbornik* (1796: 76r)

Both markers are today attested as dialectal variants of the definite article. They often appear alongside each other in damaskini texts, sometimes alongside old case markers too. The short article visually resembles the old GEN/ACC.SG case marker. It is written either as *-a*, or as *-ja* after palatalized consonants; the latter variant appears as *-ę* (Ⓐ) or *-ě* (Ⓑ) in older texts. But unlike the case marker, it is placed at the end of the leftmost element of a definite noun phrase only. The rest of the noun phrase remains unsuffixed:

- (22) *Tamo živějala na zimnyę mrazb*  
 there live.PTCP.PST.FEM.SG at winterly.DEF frost.NOM.SG  
 ‘There she lived in winterly frost’ (Sofronii 1856: 257)

The conventions concerning the use of short and full articles in literature developed in spite of the dialectal developments (Conev 1934: 510; Mayer 1988: 61;

Mladenova 2007: 70). Articles still appear beside the old cases in the early 19th century. In earlier damaskini sources, at least three distinctive case markers (NOM, GEN/ACC, DAT) frequently appear in the nominal MASC paradigm and two (NOM, ACC) in the FEM paradigm. Neuters have only a single, general case (Wahlström 2015: 81). Nominal inflection is often reflected in the article, the suffixed demonstrative, too. Frequently attested, although with a differing degree of regularity, are the MASC.GEN/ACC.SG ending *-a-togo* (example 23) or *-a-tok*, MASC.DAT.SG ending *-u-to-mu* (24) and FEM.ACC.SG ending *-b-tb* or *-u-tu* (25)<sup>15</sup>:

- (23) *filosófitē*                      *poslúšaxa*                      *cáreťtogo*  
 philosopher.NOM.PL.DEM.PL                      obey.3PL.AOR                      emperor.GEN.SG.DEM.GEN.SG  
 'the philosophers obeyed the emperor'  
*Svišťtov damaskin* (Miletič 1923: 17)
- (24) *padà*                      *ljúti*                      *zmeì*      *pod*      *krákata*  
 fall.3SG.PRS                      fierce.NOM.SG.LF                      dragon      under      leg.PL.DEM.PL  
*konjútomu*  
 horse.DAT.SG.DEM.DAT.SG  
 'the fierce dragon fell under the horse's legs'  
*Tixonravov damaskin* (Demina 1972: 318)
- (25) *póče*                      *avrám̃b*                      *da ljúbi*                      *róbinjutu*  
 start.3SG.AOR                      Abram.NOM.SG                      to love.3SG.PRS                      servant.ACC.SG.DEM.ACC.SG  
 'Abram started to love the servant'  
*Pop Punčov sbornik* (1796: 20r)

In these instances, the morphosyntactic case appears alongside the definiteness marker. Although a case may graphically resemble the short article, they can be distinguished by their distribution. The definiteness is marked usually only once per noun phrase, while the case is marked on all elements of it, including the article itself (*prévijatogo*)<sup>16</sup>:

<sup>15</sup> The marking of the FEM.ACC.SG (including masculine *a*-stem nouns like *bašta* 'father', *voivoda* 'duke' etc.) is different between the particular manuscripts. The variant *-u-tu* is common in Pop Punčo's texts (example 25). *Svišťtov damaskin* marks a distinctive ACC.SG only when the ending is stressed and unpalatalized, e.g. *ot bašt̃b si Adáma* 'by his father Adam' (Miletič 1923: 27), but *calúna májka si* 'she kissed her mother'. These differences seem to reflect Moesian dialectal features.

<sup>16</sup> It is remarkable that while the usual MASC.NOM.SG.def adjective is based on the long-form, the oblique cases were based on a short-form. *Svišťtov damaskin* has only one instance of an articulated



- (26) *ne*                                    *ště*                                    *da*                                    *zěme*                                    *drúgigo*                                    *vrěxú*  
 NEG                                    want.3SG.PRS                                    to                                    take.3SG.PRS                                    second.GEN.SG.LF                                    over  
*prěvíjatogo*                                    *sy*                                    *stopána*  
 first.GEN.SG.DEM.GEN.SG                                    REFL.DAT                                    husband.GEN.SG  
 'after her first husband, she didn't take another one'  
*Svištov damaskin* (Miletič 1923: 121)

Such systems containing both case inflection and definiteness marking on leftmost noun phrase elements are attested in present-day dialects, but only on the periphery of the Balkan Slavic area, e.g. in the areas of Timok, Albania and Rhodopes, the latter reflected in the following example from the village of Tixomir:

- (27) *na visóketok*                                    *čiléka*  
 of tall.GEN.SG.DEM.GEN.SG                                    human.GEN.SG  
 'of the tall man' (Wahlström 2015: 44)

The most present-day Balkan Slavic dialects, as well as both Bulgarian and Macedonian standards<sup>17</sup>, do not use this kind of markers. A system with a single MASC.SG article was already generalized in most dialects. Some of these systems further simplified the MASC.SG article by removing the final consonant, etymologically corresponding to the Church Slavonic demonstrative. One such dialect is described already by Neophyte of Rila in his *Grammar*:

- (28) *pokróvo*                                    *na-*                                    *vysókio*                                    *dómь*  
 roof.NOM.SG.DEF                                    POSS                                    tall.NOM.SG.DEF                                    house.NOM.SG  
 'the roof of the tall house' (Rilski 1835: 59)

It can be said the MASC.SG etymologically corresponds in these systems to the old nominative marker rather than to the demonstrative pronoun. Its dialectal vari-

---

oblique long-form, the comparative *pómalkigotogo* (Miletič 1923: 115). Unarticled adjectives in oblique cases show in most damaskini sources (excluding Pop Punčo) long-forms only.

<sup>17</sup> A curious exception among the literary standards is the recent codification of Pomak, based on southern Rhodopean dialects, which is taught at some schools in Greece. The nominal inflection paradigm in the textbook includes distinct case markings for both noun and article, which can also be marked for deixis, e.g. NOM.SG.PROX *zhón-a-sa* 'this woman', ACC.SG.PROX *zhón-o-so* (Kokkas 2006: 72; hyphens added). The article is suffixed to the leftmost element of the noun phrase, of course, e.g. *Claúdius primázava na skřito Danítsk-a-nek král-e* 'Claudius secretly kills the Danish king' (Kokkas 2006: 109).

ants mostly reflect the local phonetic reflex of an Old Bulgarian strong *jer* (Mirčev 1978: 127).

To sum up, the expression of definiteness in Bulgarian developed differently in the literature and the vernacular. In the earliest (OCS) literary sources, postponed short demonstratives (*тѣ*, *сѣ* or *онѣ* in MASC.SG) were used as markers of anaphoric relations and general knowledge. Although avoided by the Tarnovo school of the 14th century (at least the *тѣ* type), they reappear again in the 17th century damaskini sources (using only the *тѣ* type by then), systematically suffixed to the leftmost element of the noun phrase, reflecting the phrase's inflection as well. Damaskini sources use abundantly the extended pronoun (*tozi*, *onja* etc.), expressing anaphoric definiteness as well. A distinction between the short- and long-form adjectives is preserved up to the 19th century, at least in the nominative case. In this time, case marking in literature is reduced to the distinction of syntactic subjects and objects in the MASC.SG.DEF paradigm.

The dialects developed differently. The suffixation of the short demonstrative (all three variants at least in certain Macedonian, Rhodopean and Torlak areas<sup>18</sup>, *тѣ* in the rest) to the first element of the noun phrase (as reflected in rare cases like the *збѣиотѣ рабѣ*) and the emergence of extended pronouns (attested e.g. in *Wallachian Letters*) were early changes, happening before the said attestations in the literature, i.e. before the 15th century. They likely replaced the distinction of short- and long-form adjectives as a (anaphoric or general) definiteness marker, short-forms disappearing from the dialects (maybe excluding Northwest) in all non-nominative cases.

After these changes, the nominal inflection was simplified, first preserving the case markers used in damaskini (examples 23–26). This affected all definiteness markers – suffixed and extended pronouns alike. The rests of case nominal inflection vanished in most dialects in the 16th–19th century era<sup>19</sup>. If we follow the explanation of Conev (1934: 510), every such simplification was followed by a general-

**18** The dating of the emergence of articles marked for deixis is controversial. Usually unsuffixed (i.e. not showing phonetical interaction and fixation to the first element of the noun phrase) short demonstratives with deixis marking can be found in Church Slavonic texts, e.g. *сѣрѣ онѣ* in the example (14) or *сврѣда онѣ* in (16). It is a matter of controversy whether the modern articles of this type continue the Old Bulgarian markers (e.g. Ilievski 1972: 72), or not (Šaur 1975: 239), or if they can be described as “articles” at all (Topolińska 2006). However, such forms are not attested in the analyzed texts.

**19** Cf. Wahlström (2015: 81). Rests of case inflection could be still found in dialects outside the Rhodopes, Timok and Albania up to the recent times. The Bulgarian atlas of dialects attests the use of a specific oblique case for masculine proper names and family relations, e.g. in Orizare (Black Sea coast): *as pǎvna dǎd-a Néd'a* 'I remember uncle Nedjo' (*Atlas* 1964 I: 116; hyphens added). The case ending *-a* for such words was phonetically different from the short article (*-ѣ* in Orizare), and it was used in many dialects with the full article as well.

ization of the nominative or oblique definite ending. Some of the dialects further shortened the suffix in MASC.SG paradigm afterwards. The last change – reintroduction of case marking of MASC.SG nomina (excluding the proper names and family relations) by the difference of full and short articles – unifies the literary and spoken language, and thus we get the present-day, “standardized” article.

Some of these statements remain in need of further clarification. One question is the persisting use of long-form adjective without the article in literature. According to Mladenova (2007: 339) the long-forms were still “interacting” with the definite article. Damaskini literature attests both the old long-form without *-tv* suffix (as *carstvo nebeskoe* ‘Kingdom of Heaven’ in example 7, or *xytrost diavol'skaa* ‘Devilish cunning’ below in example 37) in such phrases. Was the distinction between short- and long-form adjectives still a productive one, or was it an archaism of the literary tradition? Did it still express (perhaps one specific concept of) definiteness in the vernaculars (as it still does in Serbian), or did it have another function?

Another question concerns the relation between the inflection and definiteness markers. The present-day standardized language, the literature and some dialects as well preserve a distinction between syntactic objects and subjects. For this reason Mayer (1988: 61) and, to some degree, Mladenova (2007: 74) posit a continuity between the damaskini practice and present-day literary norm in this respect. This can be seen in the pronominal paradigms, where some cases (GEN/ACC, DAT) are partly preserved among the demonstrative pronouns used as definiteness markers (e.g. *onómuzi episkopu* in example 15) as well. Thus we can look, whether the frequency of oblique cases in damaskini sources differs much from today’s use of short articles.

Our evidence for the aforementioned conclusions is based only on a few examples. It merely provides the earliest possible date for a change, and minimal, if any, information about the location. If a new phenomenon becomes frequent in a single text, it could reflect a specific author’s style or mechanic translation. If a rising frequency of a feature is observed in multiple texts from a specific period and location, it is more likely a reflex of a language change. Thus, after presenting some episodic attestations, we will test our hypotheses on a broader empirical basis.

## 5 Damaskini Corpus

To present a sound basis for such an evidence, a corpus of pre-standardized Balkan Slavic texts is being collected and digitalized. It is composed of texts from the period between the 16th and 19th century, mostly from present-day Bulgaria, including texts of various sorts, both of ecclesiastic (damaskini, Gospels, liturgic texts) and secular (dictionaries, chronicle of Paisius, an Alexander Romance) na-

ture. In future it could include sources from earlier periods, many of them already digitalized in the context of other projects.

The use of the different forms of articles was empirically analyzed before by Mladenova from two aspects. She compared the use of articles in 6 damaskini versions according to individual semantic classes of definiteness with the expected presence of articles in her own translation (“current norm”; cf. Mladenova 2007: 258). Mladenova compared the version from *Tixonravov damaskin* with her translation from the aspect of short articles, i.e. the rate of the total number of masc nouns with an *-a* ending and the nouns which would receive the short article according to the current norm (Mladenova 2007: 76). In comparison to this study, we will compare the use of various definiteness markers, both from morphosyntactic and orthographic point of view.

## 5.1 Sample

For the purposes of this study we have chosen the hagiography of St Petka of Tarnovo, also known as St Parascheva of Epibates, an ascetic saint of the 10th century. The hagiography was written in the 14th century by Patriarch Euthymius of Tarnovo himself. The text was a part of the organization of the cult of St Petka as the patron saint of the Bulgarian kingdom, and it became popular in the damaskini tradition as well. It is attested in many – at least 18 known to the author – collections, available in both Church Slavonic and vernacular, in both Cyrillic and Greek scripts, as well as in some modern prints. Six versions of the text were used in this study as a historical parallel subcorpus. As we are using variants of the same text, of the same topic, genre and structure, with a similar length and style, the impact of non-linguistic features on their linguistic differences is minimalized. The hagiographic genre is less influenced by conservatism, which characterizes biblical and liturgical texts<sup>20</sup>. As there are many versions available, our sample includes representations of individual traditions of transcripts, different dialectal areas and various periods.

The first version (a1) is contained in the *Tixonravov damaskin*, already mentioned as one of the most extensive, a well preserved, and well-studied manuscript of this tradition, written in 17th century by the Sredna Gora circle<sup>21</sup>. A related version (a2), possibly written in Trjavna few generations later, comes from the *Ljubljana damaskin*, one of the first such documents ever studied by modern

<sup>20</sup> For challenges concerning a diachronic parallel corpus of biblical texts, cf. Zeldes (2007: 31).

<sup>21</sup> Miletič (1903: xxi) locates the origin of the manuscript in Sopot. Currently preserved under the signature OR Φ.299 №702 in the Russian State Library in Moscow. For the purposes of this article, we have used the critical edition by Demina (Demina 1972: 94–98).

scholars<sup>22</sup>. Despite the temporal distance, these two versions differ only marginally from each other. They reflect the same recension of Euthymius' *Life of St Petka*, accustomed to the Central Bulgarian dialect of the early 17th century<sup>23</sup>.

Another version (b) is available in the *Berlin damaskin*, written in the beginning of the 19th century<sup>24</sup> likely in Pleven. This version is based on another recension than (a1) and (a2)<sup>25</sup>. While preserving the same narrative structure, its language reflects features of the Moesian dialects, showing phonetic shifts of unstressed vowels and many Turkish loanwords.

A shorter version (c) is available to us in the *Pop Punčov sbornik*, written in 1796 in Mokreš, a village in the Montana (former Kutlovica) Province<sup>26</sup>. The language of this collection shows features of the transitional Northwestern Bulgarian dialect (Šaur 1970). The sources of this *Sbornik* are eclectic and rich – aside from hagiographies, apocrypha and homilies it includes excerpts from Paisius' History, folk tales and even a few jokes. The hagiography (*Slovo Paraskevi*) seems to be based on a newer translation from the East Slavic area.

The versions (d1) and (d2) are available to us from the various prints of the *Nedělnik*, a collection of Sunday homilies by the Bishop Sophronius of Vraca. The author is a major figure in Bulgarian literature, being the author of at least one well-preserved damaskin from the 1760s and the first modern Bulgarian autobiography as well. The original version (d1) was printed in 1806 in Romania (Sofronii

---

22 Based on the critical edition by Argirov (1895: 550–556). The original is preserved under the signature NUK Cod. Kop. 21 in the National and University Library of Slovenia in Ljubljana. The facsimile is available at the Library's website: <http://old.nuk.uni-lj.si/kopitarjevazbirka/ImageBrowser.asp?Kodeks=23&Slika=1>. Last accessed: 16 January 2020. The differences between the versions (a1) and (a2) are mostly of a lexical nature (cf. Demina 1968: 168).

23 The whereabouts on all particular translations are according to personal communication with Jürgen Fuchsbauer. The structure of the texts (a1/a2) and (b) reflects the shortened edition of Euthymius' *Life of St Petka* preserved in Božidar Vuković's *Zbornik za putnike* (1536). Vuković's edition has been copied in a later Church Slavonic damaskin from Adžar (Demina 1968: 162). The other texts reflect a later Church Slavonic edition from the *Kniga žitij svjatyx* by Dmitrij Rostovskij (1689).

24 The original is available under the signature Slav. fol. 36 (f. 179r-185v) in Biblioteka Jagiellońska in Cracow. According to Conev (Conev 1937 XXXI: 3) a sidenote dates the manuscript to the year 1791. On the basis of the watermarks on the paper, Ciaramella (1996: 120) argues the damaskin could not have been written before 1803.

25 The text is not preserved in older sources as a whole. A fragment of the same edition of the *Life of St Petka* can be found in the damaskin No. 133 of the Church Archive in Sofia from the 18th century (cf. Sprostranov 1900: 211). Similarly as *Svištov damaskin*, these sources are classified by Demina (1968: 54) as "type IV" Modern Bulgarian damaskini.

26 The original *Sbornik* is preserved under the signature HBKM 693 in the National Library in Sofia. A digital facsimile is available at the Library's website, as well as in the World Digital Library: <https://www.wdl.org/en/item/10659/>. Last accessed: 16 January 2020.

1806: 273–276), containing Sophronius' own edition of the *Life* under the title *Žitie prěpodobnyę matere našęę Paraskėvi Tėrnovskię*. The *Nedělnik* was republished fifty years later in Novi Sad, edited by Pop Teodor Xrulev (Sofronii 1856: 256–258). The variant of the text (*Za světaę Paraskeva*) in this edition (d2) is different from the one in the original. The editor used an East Slavic source similar to the one used by Pop Punčo. The language is, as the editor claims, a *čisto bulgarskij jazыkъ*, representing a later stage of a literary norm.

Our sources have the same topic, but differ in length and details. In the longer sources, our selection included about 1000 words (a1: 1016 words, a2: 1021, b: 1043). Three shorter versions, one by Pop Punčo (566 words) and the two versions from our editions of *Nedělnik* (d1: 771, d2: 929) were used as a complete text. An overview of the sources is provided in the following Table 1:

**Table 1:** overview of our sources

	source	date	language variety	size (tokens)
a1	<i>Tixonravov damaskin</i>	17th century	Central	1016
a2	<i>Ljubljana damaskin</i>	18th century	Central	1021
b	<i>Berlin damaskin</i>	19th century	Moesian	1043
c	<i>Pop Punčov sbornik</i>	1796	Northwestern	566
d1	original <i>Nedělnik</i>	1806	Eastern, archaized	771
d2	later edition of <i>Nedělnik</i>	1856	Eastern	929

## 5.2 Annotation

The next step is to adapt the sample for research purposes. As our corpus includes pre-standardized texts from various places and times, we cannot always rely on methods used to construct and annotate corpora of standardized languages. Neither can we expect a unified grammar and vocabulary in them. This presents a problem for the annotation and normalization of the corpus, which do require some unified categories.

One such problem is the normalization of word forms. The tokens come in various forms both because of orthographic freedom in writing of some vowels and endings (e.g. *xýtros* 'cunning' in a1, *xítrostъ* in a2; cf. below example 37). To mark such tokens as realizations of the same lemma, we usually follow the dictionary made on the basis of *Tixonravov damaskin* (Demina et al. 2012). But still, this source lacks many newer loanwords (e.g. *arěsa* 'likes'), present in the versions (b) and (c), as well as archaic Slavonicisms and Russian words (e.g. *kózni*

'devices'), characteristic of our source (d1). To normalize these words, we added sources reflecting multiple diachronic and dialectal varieties, like the *Etymological dictionary* of BAN (Georgiev et al. 1971–2010) or Church Slavonic dictionaries (e.g. Cejtin et al. 1994).

Marking the morphology proves to be more challenging. The language of our sources was undergoing significant changes in this aspect, and the variation (especially definiteness and case marking) is the focus of this study. The morphological features are sometimes obscured by the orthographic norms (e.g. *-ę* could mark the short article, as *zimnyę mrazъ* in example 22, but also the GEN/ACC case, as *cъrętogo* in example 23). For the purposes of our corpus, the individual words are annotated using tags for morphological, as well as some semantical and syntactical features, e.g. proper nouns, animacy or syntactical position of pronouns. The tagging is based on the MultextEast 4 standard, roughly based on the specifications for (modern) Bulgarian<sup>27</sup>. An example from (d2):

- (29) *bogъ*                    *ja*                    *proslavi*  
 God.NOM.SG            she.ACC            beatify.3SG.AOR  
 'God beatified her'

token	lemma	tag
<i>bogъ</i>	bog	Ncmsnny
<i>ja</i>	tja	Pp3fsa-c-
<i>proslavi</i>	proslavja	Vmia3s-anp

Concerning the tag, the first (capital) letter reflects the grammatical category ("N" representing nouns, "P" pronouns, "V" verbs etc.), the consequent letters reflect features relevant to the category (e.g. the noun tag "N" is followed by a "c" for a common noun, an "m" for the gender, an "s" for the number, another "n" for the case etc.). Our texts, however, do not always conform to modern Bulgarian. Compare the endings of the word for 'pillar' (*stlbp*) in the following examples:

- (30) *môžete*            *da*    *go*    *izvadite*                    *o<sup>t</sup>*    *stlbpo*  
 can.2PL.PRS        to    it.ACC    remove.2PL.PRS        from    pillar.DEF  
 'Could you remove (the body) from the pillar?'  
*Pop Punčov sbornik* (1796: 76r)

<sup>27</sup> The original set of morphological tags can be viewed online at the MultextEast website: <http://nl.ijs.si/ME/V4/msd/html/msd-bg.html>. Last accessed: 16 January 2020. The adaptation for the damaskini sources can be seen in the appendix of this article.

- (31) *povele*                      *da iskopajutъ*      *dupka*              *pri*              *stbpa*  
 command.3SG.AOR      to dig.3PL.PRS      hole.NOM.SG      next to      pillar.GEN.SG.DEM  
 '(the monk) commanded (them) to dig out a hole next to the pillar'  
*Pop Ioann's damaskin* (Angelov 1958: 102)

- (32) *podkàni*                      *se da slézi*                      *ot stól'potъ*  
 decide.3SG.AOR      REFL to descend.3SG.PRS      from pillar.DEF  
 '(the monk) decided to step down from the pillar'  
*Berlin damaskin* (182r)

Despite the three cited texts all stem from approximately the same time, they all use a different suffix on the 'pillar'-word. They show three variants of definiteness markers, one of them – Pop Ioann's (31) – using one resembling that of the present-day standardized language. Using the default Bulgarian MultextEast v4 tagset, the 'pillar'-word can be simply marked as definite, or specifically as a carrier of the short or of the full article. It is, however, questionable to tag the two non-standardized options (-o and -otъ) as equal to the standardized variants of the short (-a) and the full (-ъt) article, the first of them known from Neophyte's *Grammar*. The information about the realization (and hints about possible underlying morphological variety) of the article would be lost.

To overcome such instances of ambiguity in the corpus, we adapted the Bulgarian tagset to reflect both archaic and innovative features. These include nominal case markers, verbal infinitives and multiple options to mark the definiteness of adjectives. Using this adapted tagset the variants from examples (33–35) would be tagged as follows:

	<b>token</b>	<b>lemma</b>	<b>tag</b>
(33)	<i>stbpo</i>	stъp	Ncmsonn
(34)	<i>stbpa</i>	stъp	Ncmsgnn
(35)	<i>stól'potъ</i>	stъp	Ncmsoyн

The letter in the 5th position after the category tag “N” reflects the presence of the -t marker, not the concept of definiteness as such. We are strictly relying on morphological marking, leaving all semantic aspects to analyses based on this corpus. This has been partly reflected in the original MultextEast Bulgarian tagset as well, but only for masculine nouns: thus it would mark both tokens in (33) and (34) as a nominative with a short article (Ncmssnsn). The new set does not only distinguish such variants, but it is also able to mark proximal and distal articles used in dialects which know them. The difference between the tokens (33) and (34) is expressed in the 4th position, the “case marker”. Due to phonetic changes



and simplifications of inflection paradigms, the identification of case markers and the actual morphosyntactic relations understood as “cases” is not always clear. The tag refers to the graphic realization of the ending, not the function.

In ambiguous cases, etymology helps us set up a convention. For example, the difference between the *MASC.GEN.SG* and *MASC.ACC.SG* was already lost in Middle Bulgarian; in example (34) a *g* (by default a genitive marker) marks the suffix *-a*. A new option is the new “*o*-case”, which was added to mark the nominative form of the short article (33) and the full article with the “Macedonian” vowel reflex (35). By such a construction, the annotated part of the corpus does not merely help us to analyze its morphosyntactic features, but also makes it possible to compare them together to already existing corpora of Old Church Slavonic, Modern Bulgarian and related dialectal varieties.

### 5.3 Application

Using the new tagset we can unambiguously tag all of the variants of definiteness markers. Thus they can be automatically searched for in our corpus and we can assess the frequency of their use, as well as coincidences with other linguistic features. The particular variants of definiteness markers receive the following tags, with points representing tags irrelevant for the present study:

old long-form adjective	A . . . . l
extended demonstrative pronoun	Pd . . . . . a .
- <i>t</i> suffix (full article)	N . . . . y . , A . . . . y
<i>MASC</i> noun with an <i>-a</i> ending	N .msgn .
<i>MASC</i> noun with an <i>-o</i> ending	N .mso . .
<i>MASC</i> adjective with an <i>-ija</i> ending	A .msns
adj. pronoun with a <i>-t</i> suffix	Ps . . . . . ay , Px . . . . . ay
<i>MASC</i> pronoun with an <i>-ija</i> ending	P . . . . . as

To make the results comparable to the aforementioned studies by Mladenova (2007: 76, 258), and also between the respective versions of various length, we have added the percentual article rate: the rate of articted forms (including *-a* and *-o* endings in *MASC.SG*) among all nouns or adjectives in the text. This rate can be interpreted as the overall tendency of the scribe (or writer/editor) to employ these markers.

**Table 2:** definiteness and case markers in pre-standardized versions of the *Life of St Petka*

		a1	a2	b	c	d1	d2
<b>nouns</b>							
full article (-t suffix)	N . . . . y.	15	15	36 <sup>28</sup>	4	0	31
oblique MASC.SG	N.msn. . . <sup>29</sup>	21	23	21 <sup>30</sup>	12	28	19
-a endings	N.msg. . .	18	20	16	8	18	19
-o endings	N.mso. . .	0	0	3	1	0	0
datives	N. . . . d. . .	7	7	5	9	6	1
all non-vocative nouns	N. . . . v. . .	236	235	225	137	169	210
article rate		13.98 %	14.89 %	24.44 %	9.49 %	10.65 %	23.81 %
<b>adjectives</b>							
full article (-t suffix)	A . . . . y	2	2	14	0	0	18
old long-form	A . . . . l	26	31	24	13	67	8
-ija endings	A.msns	0	0	0	0	1	8
oblique cases	A . . . . n.	3	4	7	11	22	4
all adjectives	A . . . . .	70	72	91	46	86	78
article rate		2.86 %	2.78 %	15.38 %	–	1.16 %	33.33 %
<b>pronouns</b>							
poss. with a -t suffix	Ps . . . . . ay	3	3	11	0	0	4
refl. with a -t suffix	Px . . . . . ay	1	1	2	0	0	5
pron. with an -ija ending	P . . . . . as	0	0	0	0	0	1
ext. dem. pronouns	Pd . . . . . a.	15	15	7	19	15	11
<b>size (tokens)</b>		1016	1021	1043	566	771	929

<sup>28</sup> There is one instance of a postponed short pronoun written separately after the noun (*xodinie to*) in the *Berlin damaskin* as well. This instance is disregarded, as its state is ambiguous from a strict point of view.

<sup>29</sup> Vocatives (N.msv..) were subtracted from the results.

<sup>30</sup> The single attestation of an instrumental noun (*rédom* 'all around') is disregarded, as the word is used adverbially.

The *-t* suffix is quite frequent in the *Berlin damaskin* (b) and Xrulev's 1856 edition of the *Nedělnik* (d2), with a rise in use among adjectives. In other sources, it is only rarely to be found among the adjectives and adjectival pronouns. We can observe a gradual rise in the use of the *-t* suffix, both among nouns (between versions a1/a2 and b) and adjectives (between b and d2), as well as the abandonment of the old long-form adjective. In these cases, the literary language steadily follows the developments in the vernacular. The lack of these changes in Pop Punčo's (c) and Sophronius' (d1) versions seems to contradict the development, which can be seen in the present-day language or the (d2) version, which is arguably the most similar one to it.

Pop Punčo's version (c) does not attest a single case of an adjective or pronoun with a *-t* suffix. Although such constructions can be found in other texts of his collection (e.g. *tvojàta cřkva* 'your Church'; 1796: 36v), overall they are not frequent in this source. It is the only text containing both short-form adjectives in GEN.SG. Extended demonstratives, sometimes followed by a long-form (e.g. *onova štoe telo* 'that holy body'), are relatively more frequent than in other versions, considering the size (19 per 137 nouns, against 15 per 236 in the a1 version). Preference for this definiteness marking is a Northwestern dialectal feature. The replacement of extended pronouns by adjectives with a *-t* suffix likely had affected only Central dialects in Punčo's time<sup>31</sup>. The vernacular tendencies are summed up in Table 3.

**Table 3:** vernacular tendencies

	a1	a2	b	c	d1	d2
nominal article rate	13.98 %	14.89 %	24.44 %	9.49 %	10.65 %	23.81 %
adjectival article rate	2.86 %	2.78 %	15.38 %	–	1.16 %	33.33 %
short articles	18	20	19	9	19	28
relative to source size <sup>32</sup>	1.77 %	1.96 %	1.82 %	1.59 %	2.46 %	3.01 %
ext. dem. pronouns	15	15	7	19	15	11
relative to source size	1.48 %	1.47 %	0.67 %	3.36 %	1.95 %	1.18 %

<sup>31</sup> Adjectives with a *-t* suffix were absent in the version available in the damaskin of Pop Ioann from 1788 (Angelov 1958: 101 f.), based on a different, newer translation of the Euthymius' *Life of St Petka* from 1750s.

<sup>32</sup> Size is measured by the number of tokens.

Sophronius' original version (d1) scores high on the short article rate, but only due to old case markers. It does not only lack adjectival, but also nominal articles made by *-t* suffixes. Instead, he uses remarkably more often the old, unsuffixed long-form adjective. While articles are much more abundant in his autobiography (Sofronii 1976: 29 f.), Sophronius avoids the use of articles based on a suffixed *-t*. This may be a way to express a more formal register, staying close to the original. The vocabulary preserves many Slavonicisms (e.g. *xóče* 'wants', *pérvo obrázie* 'primordiality', *mnógo kráty* 'many times'), which can be found in the original (Rostovskij 1689) as well. The past tense is relatively often (64 times) expressed by *l*-participles, against 20 cases of imperfect and 11 aorists; in comparison, the version from the *Tixonravov damaskin* (a1) uses the *l*-participles only 12 times, against 45 imperfects and 54 aorists. The inflection is preserved in a more complex shape than in the damaskini version. In the feminine paradigm we can observe accusative (endings *-u/-ju*; 11 nouns, 1 adjective), dative (endings *-e/-i*; 13 nouns, 5 adjectives) and even instrumental (attested once) case endings. Yet we can find some dialectal influences as well. The adjectival short articles (*-ija*) are not attested in the earlier sources, like *vsýja* in the following passage:

- (36) *pomólila*                      *sja (...)*                      *za*                      *vsýja*                      *svét'*  
          pray.PST.PTCP                      REFL.ACC                      for                      whole.DEF                      world  
          'she prayed for the whole world'

The use of a long-form adjective is one of the major differences between Church Slavonic and Modern Bulgarian. The form is still preserved in Serbian, where it exists alongside the short-form, and also in North Slavic languages, where the short-form is generally lost. All four damaskini versions (a1, a2, b, c) still use both forms. Sophronius uses them even more frequently than in the older versions. Thus it still could be a productive definiteness marker, used at least for generally known objects (as e.g. *staa pét'ka* 'St Petka', *krásnyi rái* 'beautiful Paradise', *lukávŷ diávob* 'evil Devil') in the language used as a base for the translation. The orthographic tendencies are summed up in the following Table:

Table 4: orthographic tendencies

	a1	a2	b	c	d1	d2
MASC.SG syntactic object marking <sup>33</sup>	22	24	25	12	41	29
relative to source size	2.17 %	2.35 %	2.4 %	2.12 %	5.32 %	3.12 %
old long-forms	26	31	24	13	67	8
relative to source size	2.56 %	3.04 %	2.3 %	2.3 %	8.69 %	0.86 %

The data do not represent that much our present-day norm, but rather the tendencies toward the existing norms in the time of writing. The long-forms were a part of the orthographic norm adhered to by most damaskini, but they were not consistently used. This can be seen well in our oldest versions (a1) and (a2), which are otherwise nearly identical, as revealed by the comparison of corresponding passages:

- (37) a1: *síč'ka*      *xýtros*      ***diávol'skaa***      *isčéznuvaše*  
all.NOM.SG      cunning      devilish.LF      vanish.3SG.IMPF
- a2: *síč'ka*      *xítrostb*      *diávol'ska*      *izčéznuvaše*  
all.NOM.SG      cunning      devilish      vanish.3SG.IMPF
- 'all the devilish cunning was vanishing'

- (38) a1: *da*    *se*      *nasýti [...]*      *pěsni*      *aġġl'sky*  
to    REFL.ACC      satisfy.3SG.PRS      song.PL      angelic.PL
- a2: *da*    *sè*      *nasýti [...]*      *pěсны*      ***áġġb'skye***  
to    REFL.ACC      satisfy.3SG.PRS      song.PL      angelic.PL.LF
- 'to satisfy herself with [...] angelic songs'

Even if the translator may have known the function of the long-form, the scribes of the present damaskini sources did not. The long-form does not appear as the marker of anaphoric reference (like *bouię děvvi* in example 8) anymore; it was already replaced by the extended demonstrative in the Middle Bulgarian period.

The use of definiteness markers in the Berlin damaskin (b) is different: it uses articles on nouns with the same frequency as the modern text, but still preserves many old long-forms. In the exegetic part of this text, which is not attested in other damaskini (and not included in the present study), there is only one MASC.

<sup>33</sup> Contains all oblique cases (N.msn..., A.msn. minus vocatives) and short articles (-ija endings).

sg adjective with an article (*dobriat* 'the good') to be found against 15 long-forms. Unlike in case of Central Bulgarian damaskini, the feature was still liveable in the Moesian dialect of this source at least among adjectives<sup>34</sup>.

Finally, the language of the Xrulev's edition of Sophronius' text (d2) presents a sample closest both to the present standard and the vernacular, having the highest article rate and preserving only a handful of long-form adjectives. Furthermore, it is the only one attesting short articles on adjectival pronouns. The short article is consistently used in all positions in the masculine paradigm:

- (39) *kato gy pročulъ moldavskija knjazъ*  
 when they.ACC.PL hear.PST.PTCP Moldavian.DEF prince.NOM.SG  
*Ioannъ*  
 John.NOM.SG  
 'when the Prince John of Moldavia heard about them [...]'

## 6 Conclusion

The present study shows some differences between the use of definiteness markers in damaskini sources up to the 19th century and Modern Bulgarian. The article was still interchangeable with the other means of expression of definiteness, some emerging at the time (extended demonstrative), others being recessive (long-form adjective). Dialectal differences (e.g. between Pop Punčo and the Central Bulgarian damaskini) existed, but even greater differences are attested due to orthographic preferences (e.g. between respective editions of Sophronius' text d1 and d2). In other words, damaskini were not merely a "pre-standardized" literary tradition, they were the means of standardization. Sometimes at the cost of dialectal features, which were seen as "ugly".

Another point of this study was to demonstrate the capabilities of our corpus of pre-standardized Balkan Slavic literature. The digitalization of only a few carefully selected literary sources provides us with examples reflecting features of various dialects from both spatial and temporal perspective. It provides us with a new tool, by which we could step over the scope of one or two homilies in a single or two sources, and compare hypotheses across the borders of dialects, diachronic stages, and also scribal or orthographic traditions. In a further per-

<sup>34</sup> In fact, Mladenov (1963: 407) describes some dialects in the area of Ixtiman in the Northeastern Bulgaria, where both old long-forms and a forms with a short article (e.g. *golemij* 'the big' alongside *drugijo* 'the other', in Vodolej) can be found.

spective, in addition to a quantitative extension of the corpus, a semantic layer is being developed to distinguish syntactic relations, periphrastic constructions and changes in language register. In its present state, the corpus already enables us to browse through a large amount of texts and their linguistic features in a way unknown to the previous studies. This study barely exploits all the possibilities of research, which could be done with e.g. statistic or geographic methods. It is reusable and compatible with other means of corpus research.

Nevertheless, the study shows the persisting need of philological research in diachronic linguistics. The choice of sources is based on purely philological criteria. The classification and mutual relations between them is based on comparison of contents of both sources (damaskini) and particular texts (like *Life of St Petka*), on the variety of their original models, on particularities of their font, on side notes and whereabouts in the time of their discovery by modern scientists. The scarce information we have available about the authors, translators, scribes and also the historical context behind the very emergence of our sources comes mostly by this way as well. In short, the data extracted by a corpus study needs a check. The difference in use of particular features in text does not always reflect the frequency of use in speech. The perspectives of big data analysis and close reading are meant to support each other.

Sociolinguistic aspects are not to be ignored either. Damaskini sources present a material, which had a high subjective value for the contemporary readers (or, rather, the audience), who scarcely had any other means of literary education in the language of their community. This surely contributed to the preservation of archaisms in the text, in its orthography, in spite of the language change in dialects. By the means of our corpus, we can show this cultural phenomenon in numbers.

## Primary sources

- Angelov, Bonju 1958. *Iz starata bālgarska, ruska i srbska literatura. Kniga 1*. Sofia: BAN.
- Argirov, Stefan 1895. Ljubljanskijat bālgarski rākopis ot XVII vĕk. In *Sbornik za narodni umotvor-enija, nauka i knižnina*, 463–560. Sofia: Ministerstvoto na narodnoto prosvĕštenie.
- Conev, Benjo 1906. *Dobrĕjšovo četveroevangele: srĕdnobālgarski pametnik ot XIII vĕk* (= Bālgarski starini kn. I). Sofia: Ministerstvoto na narodnata prosvĕta.
- Demina, Evgenija I. 1972. *Tixonravovskij damaskin. Bolgarskij pamjatnik XVII v. Issledvanie i tekst*, 2. Sofia: BAN.
- Ilievski, Petar X. 1972. *Krninski damaskin*. Skopje: Institut za makedonski jazik “Krstе Misirkov”.
- Jovkov, Jordan 1976 [1927]. *Šibil*. <http://slovo.bg/showwork.php3?AuID=95&WorkID=2290&Lev el=2>. Last accessed: 1 April 2019.
- Kokkas, Nikolaos Th. 2004. *Uchem se pomatsko: Mathēmata pomakikēs glōssas*, tom 1. Xanthi: Politistiko Anaptyxiako Kentro Thrakēs.

- Miletič, Ljubomir 1908. *Koprivštenski damaskin: novobalgarski pametnik ot XVII vĕk* (= Bälgarski starini kn. II). Sofia: Arxeografskata komisija pri Ministerstvoto na narodnata prosvĕta.
- Miletič, Ljubomir 1923. *Svištovski damaskin: novobalgarski pametnik ot XVIII vĕk* (= Bälgarski starini kn. VII). Sofia: BAN.
- Momčilov, Ivan 1868. *Grammatika za novobalgarskija ezik*. Rusčuk: Pečatnicata na Dunavska oblast.
- Ognjanovič, Konstantin 1833. *Žitie stago Aleksia člvĕka bžija*. Budin-grad: Kralevsko Vseučilište.
- Rilski 1835 = Neofit sv. Rilski 1835. *Grammatika*. Kraguevac: Knjažesko-Serbskaa tipografia.
- Rostovskij 1689 = Dmitrij Tupalo ep. Rostovskij 1689. *Kniga žitij svjatyx*, 1. Kiev: Lavra Pečerskaja.
- Sofronii 1806 = Sofronij ep. Vračanskij 1806. *Kyriakodromion sireč Nedĕlnik – Poučenie*. Râmnicu: ep. Nektarij.
- Sofronii 1856 = Sofronij ep. Vračanskij 1856. *Evangelie poučitelno*. Ed. by Todor T. Xrulev. Novi Sad: Knigopečatnjata Dr. Dan. Medakov.
- Sofronii 1976 = Sofronij Vračanskij 1976. *Žizneopisanie*. Ed. by Nikolaj Dylevskij & Andrej N. Robinson. Leningrad: Nauka.
- Strašimirov, Anton 1999 [1892]. *Na nivata*. <http://www.slovo.bg/showwork.php3?AuID=196&WorKID=6137&Level=1>. Last accessed: 1 April 2019.
- Berlin damaskin*. Library of Jagellonian University. Cracow, Slav. fol 36.
- Thĕsauros 1799 = Damaskĕnos Stouditĕs 1799. Thĕsauros Damaskinou tou Hypodiakonou kai Stouditou tou Thessalonikeōs*. Enetia: Nikolaos Glykeos.
- Velčeva, Borjana & Olga M. Mladenova. 2013. *Loveški damaskin: novobalgarski pametnik ot XVII vek*. Sofia: Nacionalna biblioteka “Sv.Sv. Kiril i Metodij”.
- Vukovič, Božidar 1536. *Zbornik za putnike*. <http://digital.bms.rs/ebiblioteka/pageFlip/reader/index.php?type=publications&id=2269>. Last accessed: 1 April 2019.

## Secondary literature

- Andreov, Ivančo. 1844. *Pärvička bälgarska grammatika*. Bukurešt: Serd. K. Penkovič.
- Cejtlin, Ralja M., et al. 1994. *Staroslavjanskij slovar’: po rukopisjam X-XI vekov*. Moskva: Russkij jazyk.
- Ciaramella, Roberto 1996. Novi danni za Berlinskija damaskin. In *Palaeobulgarica* XX (3), 120–129.
- Conev 1937 XXXI = Conev, Benjo 1937. *Slavjanski rĕkopisi na berlinskata državna biblioteka* (= Sbornik na bälgarskata akademija na naukitĕ kn. XXXI), 1–79. Sofia: BAN.
- Demina, Evgenija I. 1968. *Tixonravovskij damaskin. Bolgarskij pamjatnik XVII v. Issledvanie i tekst*, 1. Sofia: BAN.
- Demina, Evgenija I. et al. 2012. *Rečnik na knižovnija bälgarski ezik na narodna osnova ot XVII vek*. Sofia: Valentin Trajanov.
- Dimitrova-Vulchanova, Mila 2009. Linearization Patterns in the Old Bulgarian Nominal Expression. In Dimitrova-Vulchanova, Mila & Olga M. Tomić, *Investigations in the Bulgarian and Macedonian Nominal Expression*, 74–94. Trondheim: Tapir.
- Flier, Michael S. 1974. *Aspects of Nominal Determination in Old Church Slavic*. The Hague: Mouton.
- Friedman, Victor A. 2006. Balkans as a Linguistic Area. In Brown, Keith (ed.), *Encyclopedia of Language and Linguistics*, 2<sup>nd</sup> ed., vol. 1, 657–672. Oxford: Elsevier.



- Fuchsbauer, Jürgen 2010. *Die Übertragung der Dioptra ins Slavische*. Wien: Universität Wien (Dissertation).
- Georgiev, Vladimir I. (ed.) 1972–2006. *Bălgarski etimologičen rečnik*, tom I–V. Sofia: BAN.
- Haarmann, Harald 1976. *Aspekte der Arealtypologie: die Problematik der europäischen Sprachbünde*. Tübingen: Narr.
- Hansen, Björn 2004. The life cycle of a definite marker: the development of the short and long form of the adjective in Russian, Old Church Slavonic and Serbian/Croatian. In *Zbornik za lingvistiku i filologiju XLVII* (1–2), 51–73.
- Leake, William M. 1814. *Researches in Greece*. London: John Boot.
- Lindstedt, Jouko 2014. Balkan Slavic and Balkan Romance: From congruence to convergence. In Besters-Dilger, Juliane et al., *Congruence in Contact-induced Language Change: Language Families, Typological Resemblance, and Perceived Similarity*, 168–183. Boston: De Gruyter.
- Lunt, Horace G. 2001. *Old Church Slavonic Grammar*, 7th ed. Berlin: Mouton de Gruyter.
- Lyons, Christopher. 1999. *Definiteness*. Cambridge: Cambridge Univ. Press.
- Mayer, Georg 1988. *The Definite Article in Contemporary Standard Bulgarian*. Wiesbaden: Otto Harrassowitz.
- Mendoza, Imke 2011. On the demonstrative pronouns and the question of the article in 17th century Russian. In *Russian Linguist* 35, 245–265.
- Miletič, Ljubomir 1887. Členā v bālgarskija ezik. In *Periodičesko spisanie*, kn. XXI i XXII, 305–331. Srēdec: Bālgarskoto knižovno družestvo.
- Miletič, Ljubomir [Miletič, L.] 1889. *O članu u bugarskom jeziku*. Zagreb: Tisak Dioničke Tisare.
- Mirčev, Kirill 1978. *Istoričeska gramatika na bālgarskija ezik*. Sofia: Nauka i izkustvo.
- Mladenov, Maksim Sl. 1963. Členuvani prilagatelni formi na -ij v severoiztočnite bālgarski govori. In *Bālgarski ezik XIII* kn. 4–5, 404–410.
- Mladenova, Olga M. 2007. *Definiteness in Bulgarian: Modelling the Process of Language Change* (= Trends in Linguistics 182). Berlin & New York: de Gruyter.
- Olander, Thomas 2012. Proto-Indo-European \*-os in Slavic. In *Russian Linguist* 36 (3), 319–341.
- Petkanova-Toteva, Donka 1965. *Damaskinite v bālgarskata literatura*. Sofia: BAN.
- Rejzek, Jiří 2001. *Český etymologický slovník*. Praha: Leda.
- Schleicher, August 1983 [1850]. *Die Sprachen Europas in systematischer Übersicht: linguistische Untersuchungen* (= Amsterdam Classics in Linguistics 4). Amsterdam & Philadelphia: John Benjamins BV.
- Skok, Petar 1971. *Etimologijski rječnik hrvatskoga ili srpskoga jezika*. Zagreb: JAZU.
- Sprostranov, Evtim 1900. *Opis na rākopisite v bibliotekata pri Sv. Sinod na Bālgarskata Cārkva v Sofia*. Sofia: Sv. Sinod.
- Šaur, Vladimír 1970. *Pop Punčov sbornik kak istočnik istoriko-dialektologičeskix issledovanij*. Praga: ČSAV.
- Šaur, Vladimír 1975. O významu zájmeného základu ov- a jeho vztahu k bulharskému a makedonskému členu. In *Slavia* 44 (3), 233–239.
- Todorov, Todor A. (ed.) 2002–2010. *Bālgarski etimologičen rečnik*, tom VI–VII. Sofia: Marin Drinov.
- Topolińska, Zuzana 2006. Trojniet člen – da ili ne. In *Južnoslovenski filolog LXII*, 7–15.
- Vasmer, Max 1986. *Étimologičeskij slovar' russkogo jazyka*. Perv. O. N. Trubačev. Moskva: Progress.
- Velčeva, Borjana 1964. Pokazatelni mestoimenija i narečija v novobālgarskite pametnici ot XVII i XVIII v. In *Izvestija na instituta za bālgarski ezik*, kn. X. Sofia: BAN.

- Velčeva, Borjana 1966. Norma i tradicija v bālgarskija knižoven ezik ot XVI-XVIII v. In *Bālgarski ezik* XVI kn. 2, 111–121.
- Wahlström, Max 2015. *The Loss of Case Inflection in Bulgarian and Macedonian* (= Slavica Hel-singiensia 47). Helsinki: University of Helsinki.
- Zeldes, Amir 2007. Machine Translation between Language States: Extracting Historical Gram-mar from a Parallel Diachronic Corpus of Polish. In *Proceedings of Corpus Linguistics*, 27–37.

## Appendix: Early Modern Bulgarian Tagset

Nouns		N	
1	type	c or p	semantic: common vs proper
2	gender	m, f, n	MASC, FEM, NEUT
3	number	s or p	SG, PL
4	case	n	nominative; inflection unmarked
		g	genitive; case marker <i>-a</i> in MASC/NEUT, <i>-i</i> in FEM.SG, zero marker in FEM.PL
		d	dative; <i>-u</i> in MASC/NEUT.SG, <i>-e</i> or <i>-ě</i> in FEM.SG, <i>-om/-am</i> in PL
		a	accusative (Resava style); <i>-u</i> in FEM.SG
		v	vocative; <i>-o</i> or <i>-e</i> , noun in appellative position
		l	locative; <i>-e</i> or <i>-ě</i> in MASC/NEUT.SG, <i>-ix</i> in PL
		i	instrumental; <i>-om/-am</i> in MASC/NEUT.SG, <i>-im</i> in PL
		o	<i>-o</i> in MASC.SG, <i>-b</i> or <i>-b</i> in FEM.SG (Tarnovo style)
5	definiteness	n	definiteness unmarked by a specific suffix
		y	<i>-t</i> suffix (full article)
		p	proximal deixis (e.g. <i>rabos</i> )
		d	distal deixis (e.g. <i>čovekon</i> )
6	animacy	y or n	semantic: animate vs inanimate

Adjectives		A	
1	type	c or p	semantic: common vs proper
2	gender	m, f, n	MASC, FEM, NEUT
3	number	s or p	SG, PL
4	case	n	nominative; inflection unmarked
		g	genitive; case marker <i>-a(go)</i> in MASC/NEUT.SG, <i>-ix</i> in PL
		d	dative; <i>-(m)u</i> in MASC/NEUT.SG, <i>-ei</i> in FEM.SG, <i>-im</i> in PL
		a	accusative; <i>-u</i> in FEM.SG
		l	locative; <i>-om</i> in MASC/NEUT.SG
		i	instrumental; <i>-im</i> in MASC/NEUT.SG, <i>-imi</i> in PL
		o	<i>-o</i> case ending (e.g. <i>zliotb</i> )
5	definiteness	n	definiteness unmarked by a specific suffix
		y	<i>-t</i> suffix (full article) on a long-form (e.g. <i>dobrijat</i> )
		p	proximal deixis
		d	distal deixis
		l	unsuffixed long-form (NOM.SG <i>-ii</i> , <i>-aa</i> or <i>-oe</i> , NOM.PL <i>-ie</i> )
		s	<i>-ija</i> suffix (short article; MASC only)
		t	<i>-t</i> suffix on a short form (e.g. <i>visokbt</i> )
Pronouns		P	
1	type	p	personal: e.g. <i>azb</i>
		d	demonstrative: <i>tozi</i> , <i>sii</i> , <i>onzi</i> , <i>togava</i> , <i>tolkova</i> , <i>takbv</i>
		i	indefinite: <i>njakbv</i> , <i>njakoi</i> , <i>edikoi</i> , <i>edin</i> , <i>vseki</i>
		s	possessive: <i>moi</i> , <i>negov</i> , <i>ego</i>
		q	interrogative: <i>koi</i> , <i>kbd</i> , <i>kakvo</i> , <i>kak</i>
		r	relative: <i>koito</i> , <i>deto</i> , <i>što</i> , <i>kbdeto</i> , <i>iže</i>
		x	reflexive: <i>svoi</i> , <i>se</i> , <i>sebe</i> , <i>si</i> , <i>sambsi</i>
		z	negative: <i>nikoi</i> , <i>nikakbv</i>
2	person	1, 2, 3	
3	gender	m, f, n	MASC, FEM, NEUT
4	number	s or p	SG, PL
5	case	n	nominative; inflection unmarked

		g	default oblique: e.g. <i>mene, nego, onogoz, sebě</i>
		d	dative; <i>nemu, njei, onomuva</i>
		a	accusative; FEM.SG <i>njea, onuju</i> and clitics: <i>se, me</i>
		l	rare; e.g. <i>niemъ, oněxъ</i>
		i	rare; e.g. <i>mnoju, nim</i>
6	owner number	s or p	for possessive pronouns only
7	owner gender	m, f, n	for possessive pronouns only
8	syntactic type	n	nominal, e.g. <i>azъ, nešto</i>
		a	adjectival, e.g. <i>moi, njakoi, on'zi</i>
		r	adverbial, e.g. <i>kogato, kъde, tolkova</i>
		c	short form pronouns, e.g. <i>go, si</i>
9	definiteness	n	definiteness unmarked by a specific suffix
		y	- <i>t</i> suffix (full article)
		p	proximal deixis, e.g. <i>się</i>
		d	distal deixis, e.g. <i>on'zi</i>
		s	- <i>ija</i> suffix (short article); MASC only, e.g. <i>negovija</i>
<b>Numerals</b>			
		M	
1	type	c, o, l	cardinal, ordinal, collective (e.g. <i>dva-ma</i> )
2	gender	m, f, n	MASC, FEM, NEUT
3	number	s or p	sg, pl
4	case	n	nominative; marked according to nominal/adj. endings
		g, d, a, l, i	
5	form	c	cyrillic (e.g. <i>*vi*</i> )
		d	digital (i.e. arabic, e.g. <i>12</i> )
		l	letter (e.g. <i>dvanadesętb</i> )
		r	roman (e.g. <i>XII</i> )
6	definiteness	n	marked according to adj. endings
		y, p, d, l, s, t	
<b>Verbs</b>			
		V	
1	type	m or a	main or auxilliary
2	form/mood	i	indicative
		m	imperative

		p	participle
		n	CS (or Serb.) full infinitive, e.g. <i>byti</i>
		s	short infinitive, e.g. <i>neděi se gnusi</i>
3	tense	p	present
		i	imperfect
		a	aorist
		o	CS present-tense endings (- <i>o/u</i> , - <i>ši</i> , - <i>tv</i> )
4	person	1, 2, 3	
5	number	s or p	SG, PL
6	gender	m, f, n	MASC, FEM, NEUT (applies only to participles)
7	voice	a	active (default tag and <i>l</i> -participles)
		p	passive ( <i>n</i> -participles)
		o	CS participles ending with - <i>y/-e(šti)</i> or - <i>tv(ši)</i>
		t	CS participles ending with - <i>tv</i> or - <i>mv</i>
8	negative	y or n	yes if <i>ne</i> -prefixed
9	aspect	i or p	imperfective (e.g. <i>pija</i> ) or perfective (e.g. <i>izpija</i> )
	Adverbs	R	e.g. <i>tvkmo</i>
	Prepositions	S	e.g. <i>na</i>
	Conjunctions	C	e.g. <i>ili</i>
	Particles	Q	e.g. <i>makarъ</i>
	Interjections	I	e.g. <i>aminъ</i>