

Historical Corpora of Bulgarian Language and Second Position Markers

Tsvetana Dimitrova
Institute for Bulgarian Language
Bulgarian Academy of Sciences
cvetana@dcl.bas.bg

Andrej Bojadžiev
Faculty of Slavic Studies
Sofia University
aboy@uni-sofia.bg

Abstract

This paper demonstrates how historical corpora can be used in researching language phenomena. We exemplify the advantages and disadvantages through exploring three of the available corpora that contain textual sources of Old and Middle Bulgarian language to shed light on some aspects of the development of two words of ambiguous class. We discuss their behaviour to outline certain conditions for diachronic change they have undergone. The three corpora are accessible online (and offline – for downloading search results, xml files, etc.).

1. Introduction

This paper presents part of an ongoing work on the historical evolvement of clausal second position clitics and the clitic cluster in Bulgarian which attempts at explaining the conditions for the placement and movement of clitics and clitic-like elements towards the second position in the phrase and/or clause (it is the position immediately after the first emphatic (strong or stressed) syntactic constituent – the so-called Second Wackernagel position where reflexive, discourse, interrogative, and pronominal clitics can be found in different periods in the history of Bulgarian language). In this paper, we discuss the behaviour of two words – *bo* (*bo* “for, then”) and *oybo* (*oubo* “then, indeed, therefore”) that are often found in second position, in the context of methodological issues in development of historical corpora.

In the next section, we present the three corpora we have used for our study with a brief overview of their characteristics. In section 3., we discuss a couple of practical issues in dealing with historical corpora. Section 4. contains an empirical study of the two words that are often classified as conjunctions or particles in the traditional literature with an outline of the conditions when the research has to employ the data from historical corpora available.

2. The Corpora

We started our study by excerpting data from three corpora with Old Church Slavonic/Old Bulgarian texts. They are representative of the textual collections available nowadays for linguists to work with. The first – PROIEL corpus¹ – contains annotated texts without considering the variation in data, redactions, and transparent access to parallel data (the corpus contains parallel texts but they have been used for automatic and semi-automatic annotation and texts are not readily available in parallel). The second – Old Church Slavonic subcorpus in the TITUS database² – gives parallelized texts but they have been lemmatised only; parallel data involves the gospel text. The third corpus – the Historical Corpus of Bulgarian Language³ – has been developed for a couple of years to give access to an impressive electronic collection of texts – broad and diverse, although lacking transparent annotation so far.

¹ http://foni.uio.no:3000/users/sign_in

² <http://titus.uni-frankfurt.de/indexe.htm>

³ <http://histdict.uni-sofia.bg/textcorpus/list>

The PROIEL corpus has been developed at the University of Oslo since 2008. The corpus contains the gospel text from *Codex Marianus* (following the edition of Vatroslav Jagić, cf. Jagić, 1883), parts of the gospel text according to *Codex Zographensis* (again following the Jagić's edition, cf. Jagić, 1879) that is missing in *Codex Marianus* (Matthew 1:1 – 1:27) and texts from *Codex Suprasliensis* (this part of the corpus is still under preparation, and not all texts from *Codex Suprasliensis* are included and annotated; here, we use only the available texts⁴). Although the texts are annotated (normalized wordform, lemma, part-of-speech, and applicable morphological information, plus syntactic annotation and attempt at information structure annotation), there is no readily usable marking of corresponding passages across languages and texts. We have isolated the patterns (syntactic, with respect to word ordering and right and left adjoined constituents) that we are interested in for the discussion in Section 3. However, texts are translations, so the access to sources pertaining to different redactions and/or translations, is needed to support the comparative research across texts and language phenomena (as shown by example (4) in 4.1., there are well known differences between the texts according to different manuscripts).

The TITUS corpus gives a valuable access to aligned and parallelized texts albeit not annotated with morphological and/or syntactic information. However, they are lemmatized and it is easy to search for different inflectional and orthographic forms. Access to parallel texts with corresponding passages across texts and in comparison to Greek New Testament (NT) is easy although it does not resolve the issue of handy access to different sources within the Byzantine tradition. There is no marking of the common passages across texts either (quotations, idiomatic constructions, etc.).

The third corpus – the Historical Corpus of Bulgarian Language (HCBL) – gives access to a great variety of texts (104 as of June 2014), some of which are of very late dating. The still missing annotation makes comparative research a bit complicated but the collection is extremely valuable because it covers texts according to manuscripts (and not editions), some rare and very interesting non-canonical texts, and late developments. This corpus is open-ended in the sense that non-canonical and non-literary materials can be added such as inscriptions, dialect data, databases of toponyms, personal names, etc.

Conditions	PROIEL	TITUS	HCBL
Metadata	Bibliographic reference to the edition only	Mirroring reference to the editions of the manuscripts	Reference to the manuscripts
Access to source	No	No	No
Annotation	Morphological, syntactic, lexical	Lemmatized only	No
Parallel data	No	Yes (no marking of parallel passages, citations, etc.)	No
Search engine	Yes	Yes	No
Text diversity	No	No	Extensive time period and genres

Table 1: Summary of the most important characteristics of the three corpora.

3. Practical Issues

Historical language study relies almost exclusively on written data as there are no sources that are more reliable for this research purpose. Corpus data is the empirical basis for diachronic linguistics, and by analysing it, we build hypotheses about linguistic processes within or outside a particular linguistic theory.

⁴ *Codex Suprasliensis* is included as part of the work in the UNESCO-funded project *The Tenth Century Cyrillic Manuscript Codex Suprasliensis* that aimed at digitizing this largest Old Church Slavonic manuscript.

As historical linguists do not have ready and non-compromised access to balanced corpora with well described sources covering entire periods, diverse content and genres, they often search for open-ended databases to collect materials they need. In this context, the notion of corpus may need broadening to cover different resources such as electronic text collections, editions, linguistic atlases, and dictionaries (Kytö, 2011). The Historical Corpus of Bulgarian Language is the only one among the three corpora used for our research that contains texts of diverse time periods and genres. However, it is still neither a corpus because it lacks annotation and metadata, nor a database because it is not really searchable. Therefore, here we define it as an open-ended e-text collection.

The trend, though, makes even harder to collect and align the materials to extract and observe the data because if we aim at studying the language system and its change in time (Mair, 2008), we need to take into account the linguistic phenomena as attested over time. Thus, although we may not be interested in the history of individual texts as instances of the output of the language system, we still have to take into account textual history (and the history of sources) to interpret the data we collect and analyze.

Moreover, if researchers do not have access to thoroughly described and annotated textual data, they may make use of design and arrangement of the data in a way that will rely on already available knowledge (reflected in traditional grammars and dictionaries, already annotated corpora, dialect atlases, and other handy data collections). One such approach involves heuristic alignment of historical texts with contemporary editions and/or translations of the same texts or editions of other texts that are readily available. For example, the TITUS database offers a parallel view of Old Church Slavonic NT text according to different manuscripts (*Codex Marianus*, *Codex Zographensis*, *Codex Assemanius*, and *Codex Sabbae*), Greek NT and Modern Russian NT translation. This parallel view is a fantastic tool for studying parallel constructions and specific phenomena.

In the next section, we will employ the three corpora for a field study on behavior of two words attested as early as the period of the earliest sources and preserved in some contemporary dialects. While summarizing our findings, we will sketch out the specifics of the three corpora.

4. Empirical Study

Our empirical study covers the words *бо*⁵ (*bo* “for, then”) and *ойбо* (*oubo* “then, indeed, therefore”), with additional notes on *убо* (*ibo* “because”) – the origin of all of them can be traced to *бо*. *Bo* and *ойбо* are predominantly found in the second clausal and/or phrasal position after (prosodically and syntactically) strong constituent (in the Second Wackernagel position or 2P). The first strong constituent can be a wh-word in complementizer function such as *къмо* (*kăto* “who”), *чъмо* (*chăto* “what”), etc., including a prepositional phrase with a wh-word such as *но чъмо* (*po chăto* “why”). The strong constituent (verb, noun, adjective, adverb) in the first position can be preceded by a conjunction or a subjunction, negation particle *не* (*ne* “not”), and/or followed by the reflexive particle *са* (*sen* “self”), discourse particle *же* (*zhe*), pronominal clitics such as *ма* (*ten* “you-ACC,Sg”⁶), *му* (*ti* “you-DAT,Sg”), etc. These are mostly prosodically weak constituents – proclitics or enclitics (depending on whether the strong constituent is after or before them). In section 4.1., we discuss our observations on an annotated corpus (PROIEL), with additional data from the parallel texts included in TITUS. For further analysis, we need the Greek correspondences but parallel and comparable corpora of these sources are not readily available (and annotated). Therefore, we need to look further into traditional critical editions to extract the information about the Greek equivalents (Nestle-Aland, 2013).

4.1. Earlier Texts

In this section, we will present our observations on the earlier texts that are part of the PROIEL corpus with some raw and inconclusive numbers (instances of both *бо* and *ойбо* in the two large annotated textual segments of *Codex Marianus* and *Codex Suprasliensis* – respectively, Cod. Mar. and Cod. Supr.).

⁵ As the words will be repeated in the next pages, the transliteration will not be repeated and translation is to be given only to differentiate specific meanings in appropriate discussion passages.

⁶ The following abbreviations and conventional labels are used in the paper: ACC – accusative; DAT – Dative; GEN – Genitive; Sg – Singular; Pl – Plural; FUT – Future tense form; CL – clitic; QuCL – interrogative clitic; Pron – pronoun; PP – prepositional phrase.

Overall, *Codex Marianus* attests for 172 instances of *ουβο* and 343 of *βο*, and the texts of the *Codex Suprasliensis* included in PROIEL contain 272 instances of *ουβο* and 442 of *βο*.

Conditions	Cod.Mar.	Cod.Supr.	Cod.Mar.	Cod.Supr.
	<i>βο</i>	<i>βο</i>	<i>ουβο</i>	<i>ουβο</i>
After wh-pronoun (incl. wh in PP)	9	12	47	27
After a verb (incl. <i>быти</i> (<i>byti</i> “be”))	136	128	32	62
After a noun (incl. pronoun, etc.)	109	170	27	50
After any constituent followed by <i>же</i>	0	0	10	11
After any constituent followed by <i>с.а</i>	0	0	8	11
After any constituent followed by weak pronoun	0	0	4	7
After any constituent followed by <i>ли</i> (<i>li</i> – interrogative particle)	0	0	2	2
After <i>иже</i> (<i>izhe</i> “who/what”)	18	12	3	6
After <i>аще</i> (<i>ashte</i> “if”)	3	21	16	9
Before <i>же</i>	0	0	0	0
Before <i>с.а</i>	13	24	0	0
Before a weak pronoun	21	21	0	0
Before <i>ли</i>	0	0	0	0
Before <i>аще</i>	10	2	0	0

Table 2: Positions of *βο* and *ουβο* after and before other constituents as attested in *Codex Marianus*, and the texts from *Codex Suprasliensis* (in the annotated texts in PROIEL)

Originally, *βο* was a particle for emphasis and verification (Sławski, 1974: 285–286) of the preceding constituent – the emphasized word (often syntactically focused constituent). In the data, *βο* is almost exclusively preceded by only one constituent, except for *другъ къ другоу* (*drug kă drugou* “one another”), and the preceding constituent can be preceded only by a preposition or a negation (*не* “not”, *ни* “neither”). Other syntactically weak constituents such as *с.а* and pronominal clitics are placed after it.

The origin and clausal position of *βο* are parallel to the Greek *γάρ* (*gar* “for, indeed”) that was colloquially used to highlight the faculty or the property of something or someone. In the history of Bulgarian language, *βο* was gradually adopted for various functions, which, on the one hand, overlapped (partially or fully with the meaning of *ουβο*), and, on the other, were very close to those of *же* in its function of emphatic particle (there is no co-occurrence of *же* and *βο* alone – not as *никътоже βο*, *иже βο*, etc. - in the texts here). It was also adopted to function as a conjunction – in our data *βο* is found after the negation particle alone (without a preceding constituent). The conditions for the overlap depend on its position and function to emphasize the meaning of the preceding word (just like *же*), as: 1) a marker of cause or reason - “for” (introducing the reasoning); 2) a marker of clarification - “for, you see”; 3) a marker of inference - “certainly, by all means, so, then”.

The derivation variants of *βο* are many – *убо* (*ibo*, “for, because”), and *ουβο*, among others. They were often used in earlier Old Bulgarian texts to translate specific Greek constructions and are mostly calques (unlike *βο*). The following examples show co-occurrence of *βο* and *и* in the form of *ιβο* (phonetic variant of *убо* used to translate parallel constructions in Greek (with *και* (*kai* “and”) and *γάρ* (*gar* “for, indeed”; see also the occurrence of *и* in the meaning of “even, also” after *убо*), as in:

- (1) a. *ιβο* *υ* βεσβδα твоѣ авѣ та творить Cod. Mar. Mt. 26:73
indeed even speech your out you give⁷
και γαρ ή λαλιά σου δηλόν σε ποιει
- b. *ιβο* *υ* пси подь трапезоѣ ѣдаты Cod. Mar. Mk. 7:28
indeed and dogs under table eat
και γαρ τὰ κυνάρια ὑποκάτω τῆς τραπέζης ἐσθίουσιν⁸
- c. *ιβο* снъ ѿлвѣчскы не приде Cod. Mar. Mk. 10:45
indeed son human not come
και γαρ ὁ υἱὸς τοῦ ἀνθρώπου οὐκ ἦλθεν
- d. *ιβο* азъ ѿлвкъ есмь подь властелы оучинень Cod. Mar. Lk. 7:8
indeed I man am under authority appointed
και γαρ ἐγὼ ἄνθρωπός εἰμι ὑπὸ ἐξουσίαν τασσόμενος

The use of *ουβο* as particle for explanation and emphasis, if synonymous with *βο*, is considered the earliest (Tseytlin, 1994: 721–722). The further use of *ουβο* was dependent on its use after pronouns and pronominal adverbs, mainly in interrogative clauses (after a *wh*-word) – it is probably among its first functions as it is closest to the particle function (Tseytlin, 1994: 721–722).

- (2) a. οτѣ коудѣ *ουβο* имать плѣвель Cod. Mar. Mt. 13:27
from where then have weed
πόθεν οὖν ἔχει ζιζάνια
- b. Кто *ουβο* есть вѣрны рабъ и мѣдры. Cod. Mar. Mt. 24:45
who then be faithful servant and wise
Τίς ἄρα ἔστιν ὁ πιστός δοῦλος καὶ φρόνιμος,
- c. почто *ουβο* ὄсждаиши· ιέгоже богы не ὄсждаиеть·
why therefore judge whom God not judge
Τί τοίνυν κρίνεις ὄν ὁ Θεὸς κατακρίνει
Cod. Supr. 359:1 (PROIEL Supr. 31:147-148)
- d. бракъ *ουβο* готовъ есть Cod. Mar. Mt. 22:8
marriage truly ready be
Ὁ μὲν γάμος ἔτοιμός ἐστιν,

The corresponding Greek constituents vary a lot – *ἄρα* (*ara* “then”), *μὲν* (*men* “indeed”), *οὖν* (*un* “therefore”), *τοίνυν* (*toinun* “indeed, therefore”). The conjunction *οὖν* “then, therefore” is overwhelmingly placed in second position and is also found as *εἰ οὖν* (*ei un*) – *αἴτε ουβο* (*ashte oubō* “if then”). The adverb *μὲν* “indeed, truly” in (2d) occurs after the article in the NT Greek text while *ουβο* is in 2P.

ουβο can be found (albeit sporadically) in the first clausal position – typical for subordinations and conjunctions (5 instances in Cod. Mar., and 2 in Cod. Supr.), and in the last position (as some adverbs, 1 in Cod. Mar., 2 in Cod. Supr.). *ουβο* is also found immediately after a weak constituent such as the conjunction *υ* (*i* “and”) and *δα* (*da* “to”). If there is another clitic, *ουβο* is usually found after it or after clitics in the clitic cluster (unlike *βο*). This means that it is placed (almost) exclusively after weak constituents such as *σα*, *μι*, *με* – (3a) and (3b), and pronominal clitics such as *τι* (“you-DAT”) and *μι* (“me-DAT”) – (3c).

⁷ Glosses are given only if there is no appropriate translation, i.e., *dogs* instead of *dog-PL*, but *Israel-DAT* (for the Dative form).

⁸ Nestle, Aland, 1979: 113, readings from various witnesses. The version of PROIEL follows Tischendorf, 1869: καὶ τὰ κυνάρια.

- (3) a. слышасте ли **оубо** Cod. Supr., 1, 3, 14a, 12 (27)
heard QuCL indeed
- b. състарѣвъ же са **оубо** Cod. Supr., 1, 16, 104b, 2 (208)
he became old DiscCL ReflCL indeed
- c. подобааше ти **оубо** Cod. Mar. Mt. 25:27
ἔδει σε **οὕν**
suited you-DAT indeed

There are isolated examples of immediate closeness to *оубо* and *бо* that can be interpreted as a result of an overlap in their functions. In TITUS, there is even a disagreement in translations in the parallel corpus (*бо оубо* in *Codex Marianus*, only *оубо* in *Codex Assemanius*, and *оудобъ* in *Codex Zographensis*).

- (4) a. ѣко **бо оубо** събираѣтъ плѣвелы. Cod. Mar. Mt. 13:40
as therefore is granted the weeds
ὡσπερ **οὕν** συλλέγεται τὰ ζιζάνια
- b. Ёкоже **оубо** плѣвел събираѣтъ са Cod. Assemanius Mt. 13:40
as therefore weeds granted
- c. ѣко **оудобъ** събираѣтъ плѣвелы Cod. Zogr. Mt. 13:40
as conveniently(?) granted weeds

The example with the variant readings in (4) shows that the correct interpretation of the language phenomena with respect to the language change requires access to parallel data.

4.2. Open-ended Text Collection

In this section, we discuss the additional data available through an open-ended text collection where we follow the changes in the phenomena. Sources are part of the Historical Corpus of Bulgarian Language which comprises diverse texts, with some very late ones such as *Damascenus Troianensis* (17th c.; NBKM № II, 11 or Kodov 88).

The raw statistics (without taking into account different meanings) shows interesting results with many later non-canonical sources exhibiting higher number for *оубо* and not for *бо* (in contrast to the earlier sources). The observations give a complex picture of the interplay between *бо* and *оубо*.

Source	бо	оубо
<i>Zlatoust of Jagić</i> (13 th c.; RNB, St. Petersburg, Q.п.I.56)	525	17
<i>Manasii Chronicle</i> (14 th c.; GIM, Moscow, Syn 38)	249	434
<i>Borili Regis Synodicum</i> (14 th c.; NBKM 289)	1 ⁹	37
<i>Codex of German</i> (14 th c.; Library of Romanian Patriarchy, №1)	486 ¹⁰	115
<i>Laudatio sanctae magnae martyris Dominicae</i> (1479; Rila Mon. 4/8, 603v-611v)	45	42
<i>Laudation sanctorum magnorum aequalium apostolic regum Constantini et Helenae</i> (1483; Rila Mon. 4/5, 424r-439r)	44	82
<i>Vita et acta sancti patris nostril Hilarionis episcopi ex Moglen</i> (1483; Rila Mon. 4/5, 161r-175r)	41	57

⁹ Co-occurring with *оубо*.

¹⁰ With one co-occurrence: *яко бо обо и колико нж*.

<i>Vita et acta sancti patris nostril Ioannis in monte</i> (14 th c.; Zogr. Mon. 172 (olim 103 II g.6), 93r-104r)	30	81
<i>Vita et acta sanctae matris nostrae Parascevae</i> (14 th c.; Zogr. Mon. 172 (olim 103 II g.6, 93r-104r), 74r-82v)	36	38

Table 3: Occurrences of *бо* and *оубо* in later texts from the Historical Corpus of Bulgarian Language

In the latest source – *Damascenus Troianensis* – there are no instances of *оубо* and *бо*. Historical-apocalyptic literature consistently prefers *оубо* instead of *бо* in later texts. In *Homilia Hypatii Ephesiensis* there is only *бо* (disregarding the meaning), as in the following examples:

- (5) a. **бо** вь шесты днѣ се в'се бзѣ съдѣлавъ · послѣднею дѣло¹¹ ·
 God indeed on sixthday (in the name of the God made last thing)
- b. надь тѣми **бо** вьтораю смр'ть не имать власти
 over them-INST indeed second death not has power

The same is observed in *Visio Danielis propheti. De regibus. De novissimis diebus. De fine saeculi*:

- (6) и съразеть **бо** се бранию крѣпкою
 and (stroke down) indeed (with the fierce battle)

A possible explanation extends to postulated stylistic differences between *бо* и *оубо*. In *S. Methodii episcopi revelatione de regibus et novissimis diebus*, all 17 instances of *бо* are associated with different meanings; *оубо* is found only once but in the same discourse contexts as *бо* – in (7d) below, where we give the translation of the segment with the difference in the meaning between the two words.

- (7) a. рече **бо** бѣ Излю ·
 said then God Israel-DAT
- b. вь ти **бо** днѣ · боу(д)ть члѣвци ·
 in these then days be-FUT men
- c. творити **бо** нач'н]еть тѣ(д)а · знам[ения и] чюд[еса] многа
 create then start then signs and wonders many
- d. тог(д)а всѣке **бо** хетрости/!/ то дѣаволоу съкр[а]гѣють
 then every then skills Conj Demon-DAT go short of
- и не оуспѣють ничесоже сѣи **оубо** нечисти скврѣньни гноусни ѣзѣци
 and not succeed nothing-GEN this truly sinful unclean disgusting people
 “then every Devil's skills will disappear, and these all truly sinful unclean disgusting people will not succeed”

The observations are additionally hampered by the orthographic variants such as *бѣ* and *бо*; *оубо*, *8бо*, *8бѣ*, *оубѣ*, etc. Variation in graphics and the changes in lexical and morphological forms of the words are among the greatest obstacles to the annotation and structuring of these data.

Nowadays, *бо* can be found in most Slavic languages (Trubachev, 1975: 141–142). It has preserved its particle function, and keeps the second position. In Russian dialects, *бо* is synonymous with *же* as in:

¹¹ The examples are excerpted from the corpus so there is no reference to edition (<http://histdict.uni-sofia.bg/textcorpus/list>).

Садись бо, принеси бо “Take a sit *then*, bring along *then*“. If it is kept as a conjunction, it moves towards the first position in the clause as in the Russian Smolensk dialect Ня поїде, бо боїтся яго “(He) didn’t go *because* he is afraid of him“ (Filin, 1968: 34–35). The last example shows that бо has kept its unique syntactic function of connecting two clauses while it is placed in the second clause but not in the first position of the clause it introduces (unlike most conjunctions).

Some authors (Mladenov, 1941: 36) have stipulated that Bulgarian dialects keep traces of бо in бoедно (*boedno*), бoедна (*boedna*), бoедно (*boedno*) (with variants of бyд- (*bud-*), бaд- (*bad-*) in the Rhodope and Southern Bulgarian dialects) to be traced back to бо един, бо едно, бо една with the meaning of the indefinite pronoun някой (*nyakoju* “*somebody-M*“), някоя (*nyakoja* “*somebody-F*“), някое (*nyakoe* “*somebody-N*“), and sporadically can be interpreted as negative pronouns никой (*nikoju* “*nobody-M*“), никоя (*nikoja* “*nobody-F*“), никое (*nikoe* “*nobody-N*“) (Mirchev, 1932). However, the *Bulgarian Etymological Dictionary* suggests etymology from *любо едѣнь (BER, 1971). Бо can be found very later, although sporadically, as a conjunction in the meaning of “because“ (Ilchev, 1974: 37).

5. Closing Remarks

The discussion above shows that the benefits of a corpus study for an observation on the evolvement of language phenomena in context. However, neither available collection of historical texts of Bulgarian language offers working access to structured comprehensive data. The lack of context means that valuable linguistic information on syntax, for example, remains hidden which hampers the access to syntax-semantics information for the status of the markers we have studied in this paper.

The historical linguists interested in the history of Bulgarian still need structured resources with user-friendly marking (annotation) of the linguistic information, metadata (sources, dating, editions, etc.) and visualization and search interface to allow them to make use of valuable data.

Acknowledgements

The present paper was partially prepared within the project *Integrating New Practices and Knowledge in Undergraduate and Graduate Courses in Computational Linguistics* (BG051PO001-3.3.06-0022) implemented with the financial support of the Human Resources Development Operational Programme 2007 – 2013 co-financed by the European Social Fund of the European Union. The authors take full responsibility for the content of the present paper.

References

- Jagić, V. (1879). *Quattuor evangeliorum codex glagoliticus olim Zographensis nunc Petropolitanus*. Berlin.
- Jagić, V. (1883). *Quattuor Evangeliorum versionis palaeoslovenicae Codex Marianus Glagoliticus*. Saint Petersburg.
- Kytö, M. (2011). Corpora and Historical Linguistics. *Revista Brasileira de Linguística Aplicada*. 11(2): 417-457.
- Mair, C. (2008). Corpora and the Study of Recent Change in Language. In *Corpus Linguistics: an International Handbook*. Berlin/New York: Walter de Gruyter.
- Nestle-Aland (1979). *Greek-English New Testament*. 26th revised edition. Stuttgart: Deutsche Bibelgesellschaft.
- Nestle-Aland (2013). *Novum Testamentum Graece*. 28th revised edition. Stuttgart: Deutsche Bibelgesellschaft.
- Sławski, F. (1974). *Słownik prastowiański*. Tom I. A–B. Wrocław-Warszawa-Kraków-Gdańsk: Wydawnictwo Polskiej Akademii nauk.

- Tischendorf, C. v. (1869). *Novum Testamentum Graecae*. Editio octava critica maior. Leipzig: Giesecke & Devrient.
- BER, 1971: *Български етимологичен речник*. (1971). Том 1. Под ред. на Владимир Георгиев. София: Изд. на БАН.
- Пчев, 1974: Илчев, Ст. (1974). *Речник на редки, остарели и диалектни думи в литературата ни от XIX и XX век*. София: БАН.
- Mirchev, 1932: Mirchev, K. (1932). Източномакедонското и родопското “боедин”, словенското (n)obeden и западнославянското “žádný – žaden”. *Македонски преглед* 8 (2). 9–22.
- Mladenov, 1941: Младенов, Ст. (1941). *Етимологически и правописен речник на българския книжовен език*. София: Хр. Г. Данов.
- Trubachev, 1975: Трубачев, О. Н. (1975). *Этимологический словарь славянских языков (Праславянский лексический фонд)*. Вып. 2. Москва: Наука.
- Filin, 1968: Филин, Ф. (1968). *Словарь русских народных говоров*. Вып. 3. Москва: Наука.
- Tzeytlin, 1994: Цейтлин, Р., Вечерка, Р., Благова, Э. *Старославянский словарь (по рукописям X- XI веков)*. Москва: Русский язык.