

---

## ONE QUESTION, DIFFERENT ANNOTATION DEPTHS: A CASE STUDY IN EARLY SLAVIC\*

---

NILO PEDRAZZINI  
UNIVERSITY OF OXFORD &  
THE ALAN TURING INSTITUTE

**ABSTRACT** This paper addresses some of the challenges of carrying out corpus-based linguistic analyses on historical corpora of different sizes and annotation depths. Data from the TOROT Treebank is collected to carry out a case study on Early Slavic dative absolutes, showing the extent to which methodology and results may change depending on the amount of data and the levels of linguistic annotation available. The analysis indicates that deeply-annotated treebanks of limited size can be exploited to establish a solid guideline to analyze a phenomenon in shallowly-annotated corpora and even new, unannotated texts. This is particularly encouraging for historical languages, such as Early Slavic, showing very high diatopic and diachronic variation, which significantly undermines corpus-annotation automation and therefore calls for alternative strategies to counteract data scarcity.

### 1 INTRODUCTION

The value of deeply-annotated treebanks for historical syntax can hardly be overstated, but corpus building is also an extremely time-consuming process. This calls for strategies to optimally exploit corpora with different annotation depths, in order to counteract data scarcity or reach representativeness. Gathering quantitative evidence for phenomena with clear ties to deeper levels of linguistic analysis, such as information structure and discourse, can be challenging if only a minor portion of the corpus is annotated beyond morphology and syntax.

Early Slavic (ca. 10th-16th century) is ideally placed, from both the lin-

---

\* This work was supported by the Economic and Social Research Council [grant number ES/P000649/1]. I am indebted to Hanne Eckhoff, for the constant support and detailed critical feedback. I am also grateful to Marieke Meelen for the useful comments and Mary MacRobert for the feedback on an earlier draft. All shortcomings and errors remain, of course, my own.

guistic and the corpus perspective, to be employed to investigate some of these challenges, in light of the high linguistic variation within its relatively small corpus and the great imbalance in annotation depths between individual texts in the existing treebanks. This paper will exploit Early Slavic data from the Tromsø Old Russian and Old Church Slavonic Treebank (TOROT; Eckhoff & Berdičevskis 2015) of the PROIEL family of ancient Indo-European treebanks (Haug & Jøhndal 2008) and carry out a case study on an Early Slavic construction whose distribution has often been deemed to be dependent on discourse structure: the dative absolute.

### 1.1 Case study: Early Slavic dative absolutes

Dative absolutes are one of two main types of adjunct participle clauses in Old Church Slavonic (OCS), the language of the earliest attested Slavic sources. They involve a dative participle whose subject, also in the dative, needs not be coreferential with any argument of the matrix clause, shown in (1).<sup>1</sup> They differ from conjunct participles, whose subject is functionally controlled by a matrix argument<sup>2</sup> (typically the subject), shown in (2)-(3):

- (1) *i sꝥchodeštemꝥ imꝥ sꝥ gory*  
 and come.DOWN.PTCP.PRS.DAT.PL he.DAT.PL from mountain.GEN.SG  
*zapovědě imꝥ isꝥ glę*  
 command.AOR.3SG he.DAT.PL Jesus.NOM say.PTCP.PRS.M.NOM.SG  
 ‘As they came down from the mountain, Jesus commanded them  
 saying: “...”’ (Matthew 17.9, 39045)
- (2) *azꝥ prišedꝥ iscěljo i*  
 1SG.NOM come.PTCP.PST.M.NOM.SG heal.PRS.1SG he.ACC.SG  
 ‘I will come and heal him’ (Matthew 8.7, 50710)
- (3) *vꝥ četvrtqjo že stražq nošti. ide*  
 in fourth.F.ACC.SG PTC watch.F.ACC.SG night.F.GEN.SG go.AOR.3SG

1 Translations of biblical passages are mostly adapted from the King James Version and the New King James Version. Translations of other texts are my own. Glossing and abbreviations follow the Leipzig Glossing Rules, with the addition of AOR = aorist, IMPF = imperfect, SUP = supine, and PTC = participle. After the citation, I provide the sentence ID of the examples in TOROT (e.g. ‘Matthew 17.9, 39045’). The digits following the citation can be entered in the search box in the corpus user interface (<https://nestor.uit.no>) to read the sentences in context.

2 Control constructions are those in which ‘either syntactic or lexical constraints require coreference between an argument of the matrix clause [controller] and an argument of a subordinate or modifying adjunct clause [controllee]’ (Dalrymple, Lowe & Mycock 2019: 543). In functional (as opposed to anaphoric) control the case of the controllee is determined by the grammatical function (e.g. subject, object, etc.) of the controller.

One question, different annotation depths

*kǝ nimǝ isǝ chodǝ po morju.*  
 to he.DAT.PL Jesus.NOM walk.PTCP.PRS.M.NOM.SG on sea.M.DAT.SG  
 ‘And in the fourth watch of the night Jesus went unto them, walking  
 on the sea’ (Matthew 14.25, 38895)

Several studies have suggested that the functions of the dative absolute are better understood at the discourse level (see [Worth 1994](#); [Collins 2004, 2011](#); [Sakharova 2010](#)). One of the widespread intuitions is that its unifying function across different syntactic configurations and semantic interpretations is a ‘backgrounding’ or ‘stage-setting’ one ([Worth 1994: 30](#); [Corin 1995: 259](#); [Collins 2011: 113](#)).

Similarly to a sequence of two main clauses, the discourse relation between participle clauses and their matrix clause is semantically underspecified (see [Stump 1985](#); [Kortmann 1991: §3-4](#); [Bary & Haug 2011: 2](#); [Fabricius-Hansen & Haug 2012b: 3](#); [Sæbø 2012](#)).<sup>3</sup> The most common interpretation of a dative absolute as a temporal subordinate is thus the result of preferred pragmatic reading, rather than compositional semantics. The three translations indicated for (4) are therefore all acceptable.

- (4) *is bo ukloni sǝ narodu*  
 Jesus.NOM PTC withdraw.3SG.AOR REFL people.M.DAT.SG  
*sǝštju na mǝštǝ*  
 be.PTCP.PRS.M.DAT.SG on place.N.LOC.SG  
 ‘Jesus had withdrawn, a crowd *being* in that place’  
 ‘Jesus had withdrawn, *while* the crowd was there.’  
 ‘Jesus had withdrawn [unnoticed], *since* there was a crowd in that  
 place’ (John 5.13, 51849)

According to [Bary & Haug \(2011\)](#), predicative participles fulfil one of three main discourse functions (labeled as FRAMES, INDEPENDENT RHEMES, ELABORATIONS), each entailing a particular temporal relation to the matrix clause and the discourse at large, thus behaving to a large extent like grammaticalized versions of discourse or rhetorical relations within dynamic semantics frameworks, such as classic ([Kamp & Reyle 1993](#), [Kamp, Genabith & Reyle 2011](#)) and Segmented ([Asher & Lascarides 2003](#)) Discourse Representation Theory. As the corpus evidence in [Haug \(2012\)](#) shows, each function is also more likely to surface in specific syntactic, semantic and information-structural configurations. [Bary & Haug’s \(2011\)](#) analysis is based on Ancient

<sup>3</sup> This is true unless the participle is overtly subordinated, which is a much less common occurrence (see, for example, (21) in Section 2), in which case the subordinating conjunction may specify the semantic relation between the participle and the matrix.

Greek, but we can easily exemplify the concept in a nutshell using Early Slavic participles:

- **FRAMES (5)** anchor the matrix event in time.
  - They are highly anaphorical: both the event they encode and the event participants are typically presupposed or picked from the previous discourse.
  - Their dominant verbal aspect within a corpus is likely to vary with genre: ‘both while *x*-ing and after *x*-ing are possible ways of linking to a previously mentioned or accessible event’ (Haug 2012: 311-312).
  - They are ‘predictable’ (Haug 2012) predications in the discourse in which they appear, which is expected to be reflected as low lexical variation.

- (5) *Vъchodeŝte*                      *že vъ domъ*                      *cěluite*  
 enter.PTCP.PRS.M.NOM.SG   PTC   in   house.M.ACC.SG   greet.IMP.2SG  
*i*  
 he.ACC.SG  
 ‘As you enter the home, give it your greeting’  
 (Matthew 10.12, 38615)

- **INDEPENDENT RHEMES (6)** are discourse-coordinated, therefore information-structurally very similar to matrix clauses.
  - They are temporally independent of the matrix verb.
  - In narrative contexts, like main verbs, they are more likely to be perfectives.
  - They are expected to encode new information: this could surface as overall high lexical variation among participles with this function.

- (6) *vъstavъ*                      *chodi*  
 rise.PTCP.PST.M.NOM.SG   walk.IMP.2SG  
 ‘Rise and walk’                      (Matthew 9.5, 50750, 60014)

- **ELABORATIONS (7)** add granularity to the semantics of the main event.
  - They are more likely to be imperfectives, since their function is to provide more information (‘elaborate’ on) the matrix event by describing “concomitant circumstances” that are cotermporal with the main event’ (Haug 2012: 312).

One question, different annotation depths

- They are expected to encode new information.
- They are more likely to follow the matrix clause, both because temporally dependent on the main verb and because of the general propensity to place non-topical and non-focused information in post-verbal position.

- (7) *ašte ty esi chz. r̥ci nam̥*  
 if 2SG.NOM be.PRES.2SG Christ.NOM say.IMP.2SG 1PL.DAT  
*ne obin̥q̥e s̥*  
 NEG conceal.PTCP.PRS.M.NOM.SG REFL  
 ‘If You are the Christ, tell us plainly’ (John 10.24, 63103)

Bary & Haug’s (2011) framework is mostly built on conjunct participles,<sup>4</sup> which begs the question of whether the construction fulfils all three functions with some frequency, similarly to conjunct participles. The intuition is that dative absolutes may be more frequently identified as FRAMES, especially in light of the long-standing identification of the construction as a ‘backgrounding’ device. To the best of my knowledge, no attempt has been made to validate these intuitions quantitatively by means of historical treebank data.

## 1.2 Aims and methodology

The functions of the dative absolute are evaluated in three datasets extracted from treebanks with different annotation depths. The concept of ‘depth’ in corpus annotation is to some extent relative to existing corpora in the same language, but dependency annotation is certainly a necessary, albeit not always sufficient condition for a corpus to be qualified as ‘deeply-annotated’. In this paper *deeply-annotated* shall be used to refer to corpora containing annotation spanning morpho-syntax, syntactic dependency, and information structure. *Shallowly-annotated* will be used to refer to corpora with only up to dependency annotation, namely the minimum requirement for a corpus to be classified as a treebank. As the term suggests, *strategically-annotated* treebanks refer to previously unannotated texts, which only receive targeted annotation for the constructions one is interested in.

This paper addresses the following questions:

- What are some of the advantages of building small, but deeply-annotated treebanks for historical languages with very high diatopic and diachronic variation (such as Early Slavic)?

<sup>4</sup> The authors only include one example containing an absolute participle, which is treated as a FRAME (see Bary & Haug 2011: 8).

- To what extent do methodology and results differ between deeply- and shallowly-annotated historical treebanks?

Section 2 exploits OCS data from the *Codex Marianus*, which contains the deepest annotation in TOROT. Dative absolutes from this dataset are analyzed in contrast with conjunct participles to be able to identify the functions of dative absolutes under Bary & Haug’s (2011) framework. Section 3 deals with shallowly-annotated treebanks, exploiting mixed (O)CS and Old East Slavic (OES) data. Section 4 uses a strategically-annotated Middle Bulgarian text. Final remarks follow in the conclusion.<sup>5</sup>

## 2 DATIVE ABSOLUTE IN DEEPLY-ANNOTATED TREEBANKS

From the *Codex Marianus* all potential dative absolutes (183) and conjunct participles (1512) were extracted. The constructions were then analyzed across variables which could indirectly capture Bary & Haug’s (2011) three participial functions. Loosely following Haug (2012), the main focus was on:

- the relative order of participles and matrix verb; aspect distribution (Section 2.1);
- the properties of the subjects involved (Section 2.2);
- the lexical variation among participles across different sentence configurations (Section 2.3).

### 2.1 Position and aspect distribution

As the frequencies in Table 1 show, there is a very strong tendency for dative absolutes to occur to the left of the main verb, while that is somewhat less the case for conjunct participles. The difference in proportions is in fact statistically significant.<sup>6</sup>

---

	Pre-matrix	Post-matrix
<b>Absolute</b>	93.4% (171)	6.6% (12)
<b>Conjunct</b>	66.6% (1047)	33.4% (525)

---

**Table 1** Matrix/participle relative order.

---

<sup>5</sup> All datasets, both raw and adjusted, as well as the R and Python scripts used in this paper are openly available at <https://doi.org/10.6084/m9.figshare.12894035.v1>.

<sup>6</sup>  $p < 0.001$  (Fisher’s exact test, two-tailed).

Only two dative absolutes to the left of the main verb are preceded by sister-node clauses<sup>7</sup> (Luke 2.42-43 and Luke 3.21), which means that the vast majority of all dative absolutes in the dataset are sentence-initial. As claimed in Bary & Haug (2011) and Haug (2012), framing participles should always be found in the leftmost position, since they set the stage and provide temporal anchoring for the whole sentence. FRAMES can in fact be treated as fronted adjuncts, that is, as information-structurally marked material in topicalized position (Haug 2012: 307).

Like *frame setters* in Chafe (1976: 51) and Krifka (2007: 45-48), the concept of FRAME in Bary & Haug (2011) is tightly connected to that of aboutness (or sentence) topic. As Fabricius-Hansen & Haug (2012a) explain, frame setters and aboutness topics are similar in that ‘they both refer to entities ... that are already established in, or easily inferable from, the preceding discourse’. Framing participles typically refer to a preceding event, as in (9), where the event providing the topic time for the whole sentence is Jesus casting out a demon. In other instances, they refer to bridging events linking to other events by means of ‘motion from one scene to another or perception of some previous action’ (Haug 2012: 299), as in (10), where the temporal anchor for the whole sentence is set by Jesus and the disciples moving – with the presupposition that they have set off from a previous setting.

- (9) *i izgъnanu běsu proglā*  
and cast.out.PTCP.PST.PASS.M.DAT.SG demon.M.DAT.SG speak.AOR.3SG  
*němy*  
mute.M.NOM.SG  
‘And when the demon was cast out, the mute spoke’  
(Matthew 9.33, 38591)

- (10) *i sъchodeštemъ imъ sъ gory*  
and come.down.PTCP.PRS.DAT.PL he.DAT.PL from mountain.GEN.SG  
*zapovědě imъ isъ glę*  
command.AOR.3SG he.DAT.PL Jesus.NOM say.PTCP.PRS.M.NOM.SG

<sup>7</sup> These do not include the second dative absolute in coordinated (or juxtaposed) dative absolutes modifying the same matrix, which are treated as sentence-initial, as in (8).

- (8) *i sęštju emu vъ vitanii. vъ domu simona*  
and be.PTCP.PRS.M.DAT.SG he.DAT.SG in Bethany in house Simon.GEN  
*prokaženaago. vъzležeštju emu pride žena*  
leper.GEN.SG sit.PTCP.PRS.M.DAT.SG he.DAT.SG come.AOR.3SG woman.NOM.SG  
‘And being in Bethany in the house of Simon the leper, as he sat at meat, there came a woman’  
(Mark 14.3)

‘And as they came down from the mountain, Jesus commanded them, saying’  
(Matthew 17.9, 39045)

Likewise, the conjunct participle in (11) refers to a preceding event in the discourse (the *Miraculous Catch of Fish* in Luke 5.1-7), while that in (12) anchors the topic time for the instruction contained in the matrix, by referring to the time frame within which its interpretation holds (‘when praying’).

- (11) *viděvъ*                      *že*   *simonъ*   *petrъ.*   *pripade*   *kъ*  
see.PTCP.PST.M.NOM.SG   PTC   Simon.NOM   Peter.NOM   fall.down.3SG   to  
*kolěnoma*   *isvama*                      *glę*  
knee.DAT.DU   Jesus.ADJ.DAT.DU   say.PTCP.PRS.M.NOM.SG  
‘When Simon Peter saw (that), he fell down at Jesus’ knees’  
(Luke 5.8, 51286)

- (12) *molęšte*                      *že*   *sę*   *ne*   *lichō*   *glte.*                      *ěkože*  
pray.PTCP.PRS.NOM.PL   PTC   REFL   NEG   vainly   speak.IMP.2PL   as  
*i*   *języčъnici*  
also   heathen.NOM.PL  
‘And when you pray, do not use vain repetitions as the heathen do’  
(Matthew 6.7, 51169)

Although FRAMES should generally be looked for to the left of the matrix verb, not all sentence-initial participles necessarily serve that function. However, there are already strong reasons to expect dative absolutes to be most typically framing participles, since those in sentence-initial position far outnumber those in any other configuration. As a result, it is also hard to formulate hypotheses on the differences between pre- and post-matrix absolutes. This is the case with aspect<sup>8</sup> distribution: the percentage of imperfective and perfective dative absolutes to the right of the matrix (Tables 2-3) is based on 12 occurrences only, which are arguably too few to form any hypothesis about differences in aspect distribution between pre- and post-matrix dative absolutes:

<sup>8</sup> OCS and Greek participles have been shown to largely function in the same way, namely, that so-called ‘present’ and ‘past’ participles express aspect, not relative tense (Eckhoff & Haug 2015; Kamphuis 2020). I will therefore talk about the ‘aspect’ of participles and refer to ‘past participles’ and ‘present participles’ as ‘perfective’ and ‘imperfective’ participles respectively.



	Imperfectives	Perfectives
<b>Absolute</b>	59.6% (102)	40.4% (69)
<b>Conjunct</b>	11.9% (125)	88.1% (922)

**Table 2** Pre-matrix participles: aspect distribution.

	Imperfectives	Perfectives
<b>Absolute</b>	50% (6)	50% (6)
<b>Conjunct</b>	89.7% (471)	10.3% (54)

**Table 3** Post-matrix participles: aspect distribution.

Although imperfectives appear to be somewhat more frequent, the difference in aspect distribution among pre-matrix dative absolutes is not statistically significant.<sup>9</sup> Aspect variation in dative absolutes fits the idea that the dominant aspect of FRAMES is expected to vary with the genre of the text and the discourse at large, rather than the syntactic configuration. Conjunct participles, on the other hand, show clear correlations, with a much stronger tendency for participles to the left of the main verb to be perfectives and for those to the right to be imperfectives.<sup>10</sup> This is in line with Haug’s (2012) Ancient Greek data, showing that participles to the right of the matrix are more likely to be ELABORATIONS while those to the left INDEPENDENT RHEMES. The former normally result in a complex rheme, where the event described by the elaborating participle adds granularity to the lexical meaning of the matrix verb, as shown in (13):

- (13) *i vōstrōzaachō učenici ego*  
 and pluck.IMPF.3PL disciple.NOM.PL he.GEN.SG  
*klasy. i ēdēachō istirajōšte*  
 head.of.grain.ACC.PL and eat.IMPF.3PL rub.PTCP.PRS.NOM.PL  
*rōkama*  
 hand.INS.DU  
 ‘And his disciples plucked the heads of grain and ate them, rubbing  
 them in their hands’ (Luke 6.1, 40221)

As pointed out by Haug (2012: 311), while the perfective is generally the dominant aspect in narrative contexts, this is particularly true of the simple nar-

<sup>9</sup>  $p = 0.553$  (Fisher’s exact test, two-tailed).

<sup>10</sup>  $p < 0.001$  (Fisher’s exact test, two-tailed).

rative style of the New Testament. The preponderance of perfective INDEPENDENT RHEMES is thus as unsurprising as that of perfective main verbs. Stacked perfective INDEPENDENT RHEMES always induce narrative progression (Bary & Haug 2011: 15-16), as in (14). Although much less frequent, even when imperfective, they are equivalent to main verbs from the information-structural perspective, which is often reflected in their English translation, as shown in (15):

- (14) *i abie tekъ edinъ otъ*  
 and immediately run.PTCP.PST.M.NOM.SG one.M.NOM.SG from  
*nichъ. i priemъ gobъ.*  
 he.GEN.PL and take.PTCP.PST.M.NOM.SG sponge.ACC.SG  
*isplъnъ ocъta. i vъznezъ na*  
 fill.PTCP.PST.M.NOM.SG vinegar.GEN.SG and put.PTCP.PST.M.NOM.SG on  
*trъstъ. napačše i*  
 reed.ACC.SG give.to.drink.IMPF.3SG he.ACC.SG  
 ‘Immediately one of them ran and took a sponge, filled it with sour  
 wine and put it on a reed, and offered it to him to drink’  
 (Matthew 27.48, 39789)

- (15) *nъ i psi prichodešte oblizaachъ*  
 but also dog.NOM.PL come.PTCP.PRS.NOM.PL lick.IMPF.3PL  
*gnoi ego*  
 wound.ACC.PL he.GEN.SG  
 ‘Even the dogs would come and lick his sores’ (Luke 16.21, 58438)

On the one hand, on the basis of only two parameters (verbal aspect and position relative to the matrix clause), it was possible to make a clear-cut distinction between dative absolutes as a whole and post-matrix conjunct participles, which are thus most typically expected to be ELABORATIONS. Prototypical FRAMES have, as it were, a diametrically opposite status to that of prototypical ELABORATIONS, both in terms of surface realization (the latter more likely post-matrix, the former most likely sentence-initial) and of information status (ELABORATIONS adding new information, FRAMES topicalizing a mentioned or accessible event). On the other hand, pre-matrix conjunct and absolute participles are still, to some extent, potentially competing FRAMES.

## 2.2 Subjects

Examples like (1) and (5) above are discursively very similar. The main difference appears to be syntactic: the subject of conjunct participles is co-

One question, different annotation depths

indexed with that of the matrix clause, whereas the subject of dative absolutes is not. However, it has often been noted that in both Early Slavic (Collins 2004, 2011) and Ancient Greek (Fuller 2006; Haug 2011) subject co-indexation between absolutes and matrix clauses is in fact attested (16),<sup>11</sup> despite traditional descriptive grammars treating them as ‘exceptions’.

- (16) *i*      *bystъ*                      *idoštemъ*                      *imъ.*                      *ištistišę*  
and happen.AOR.3SG go.PTCP.PRS.DAT.PL he.DAT.PL cleanse.AOR.3PL  
*sę*  
REFL  
‘And it came to pass that, as they went, they were cleansed’  
(Luke 17.14, 41043)

As we mentioned above, the concept of FRAME is closely connected to that of aboutness topic, as they both refer to entities that are already established in the preceding discourse. We can expect this to be reflected in the information status of subjects in framing constructions, which the information-structural annotation in PROIEL/TOROT can help capture. This level of annotation includes givenness tags on discourse referents<sup>12</sup> and anaphoric links from anaphors to antecedents. The treebank does not include annotation of complex information-structural categories such as topic and focus. A well-tested algorithm<sup>13</sup> is used to select potential topic candidates and to assign them a *topic score*, which measures their relative topicworthiness based on properties known to correlate with topichood cross-linguistically (Eckhoff 2018: 34).<sup>14</sup> This may help shed light on the differences between dative absolutes and conjunct participles in cases where they seem to functionally overlap, as we pointed out above. Dative absolutes should rank significantly higher than conjunct participles at involving subjects that are likely to be aboutness topics within their sentence. The box plot in Figure 1 shows the average topic score assigned to the (overt) subjects of dative absolutes and conjunct participles in TOROT. Since the main overlap is among dative absolutes and conjunct participles to the left of the matrix clause, only participles in that configuration have been included. Furthermore, since participles in sentence-initial posi-

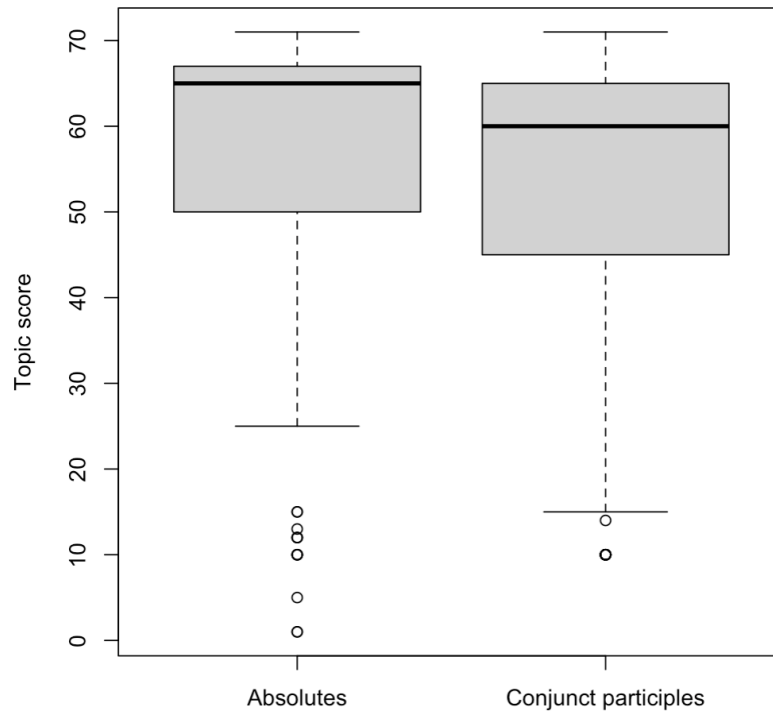
<sup>11</sup> Note that example (16) is parsed as [<sub>PRED</sub> *i* *bystъ* [<sub>COMP</sub> [<sub>ADV</sub> *idoštemъ* *imъ.*] *ištistišę* *sę*]].

<sup>12</sup> See Haug, Eckhoff & Welo (2014) for a detailed description of the givenness annotation scheme in PROIEL and its theoretical foundations.

<sup>13</sup> The algorithm was written by Dag Haug and Hanne Eckhoff and is described in great detail in Eckhoff (2018).

<sup>14</sup> These properties are: 1) givenness status 2) the number of mentions in the 30 preceding sentences 3) word order 4) realization (e.g. prodrops, personal pronouns, proper names) 5) syntactic relation 6) animacy 7) word order 8) the properties of the immediate antecedent and of the intervening competing candidates (Eckhoff 2018: 34).

tions are the hardest to disambiguate in terms of discourse function, which can typically be either *FRAMES* or *INDEPENDENT RHEMES*, only the leftmost conjunct participle is included from so-called ‘serial constructions’ (i.e. multiple stacked conjunct participles leading up to the matrix verb).



**Figure 1** Average topic score of subjects in sentence-initial dative absolutes and conjunct participles.

The median (the thick line inside the box in Figure 1) for dative absolutes is at 65, whereas the one for conjunct participles is at 60. The box itself represents the interquartile range of the distribution (i.e. the central 50% of the distribution). We can see that the median for dative absolutes coincides with the third quartile (the upper limit of the upper half of the box) of conjunct participles, indicating that the difference between the two distributions may be statistically significant. Among dative absolutes, subjects with a topic score below 25 are treated as outliers (the small circles below and above the whiskers), whereas with conjunct participles the lower limit is 15. A one-tailed Mann-Whitney *U*-test was performed and, on average, sentence-initial dative absolutes involve subjects with significantly higher topic scores than the subjects

of conjunct participles in the same position.<sup>15</sup>

Although these figures are still not helpful in disambiguating pre-matrix conjunct participles as INDEPENDENT RHEMES or FRAMES, they are, however, a further indication that dative absolutes involve subjects that are highly likely to represent aboutness topics in the sentences in which they appear: more likely, in fact, than conjunct participles in the same configuration.

The topic-like character of the dative absolute is also reflected in the position of its subjects, as shown in Table 4.

SV	VS	Null/NA
20.2% (37)	60.1% (110)	19.7% (36)

**Table 4** DA subject position.

21 of the 36 occurrences classified as ‘Null/NA’ in Table 4 are expressions of the type *pozdě byvǝʃu* ‘once it got late’, *byvǝʃju že dɔni* ‘when the day came’ or *večerou byvǝʃju* ‘when the evening came’. In PROIEL/TOROT these are all analyzed as impersonal temporal constructions consisting of a copula and a predicative adverb or noun, the latter agreeing with the participle. Although one could argue for a finer-grained distinction between personal and impersonal temporal constructions among these occurrences, from the discourse perspective these can all still be analyzed as generic stage-setters.

11 Null/NA occurrences come from coordinated dative absolutes sharing one subject. In those cases, the second conjunct is counted as null-subject.

Of the remaining four Null/NA occurrences only one example (17) is a potential null-subject proper:

- (17) *se izide sějɛi da sɛtɔ. i*  
 behold go.out.AOR.3SG sow.PTCP.PRS.M.NOM.SG to sow.SUP and  
*sějɔʃtumu. ova ubo padɔ pri*  
 sow.PTCP.PRS.M.DAT.SG some.NOM.PL then fall.AOR.3PL along  
*pɔti.*  
 way.LOC.SG  
 ‘Behold, a sower went out to sow. And as he sowed, some seed fell  
 by the wayside; and the birds came and devoured them’  
 (Matthew 13.4, 38776)

The example has, however, no real peculiarities from the information-structural or the discourse perspective: *sějɔʃtumu* ‘while sowing’ picks up its subject from the immediately preceding discourse (*izide* ‘(a sower) went out’),

<sup>15</sup>  $W=30806, p = 0.009$ .

in line with the anaphoric character of FRAMES. More interesting, perhaps, is that the participle appears in the long form,<sup>16</sup> which is notably rare among dative absolutes (Lindberg 2013: 37): the choice of the long-form might in fact have to do with having a null (hence definite, known, just mentioned) subject.

In the majority of dative absolutes the subject follows the participles: Haug (2012: 320) argues that the VS configuration largely has to do with the ‘re-instatement’ of old referents that had been out of the centre of attention in the preceding discourse and which framing participles help contextualize in the new discourse. It must be mentioned that, while the order of participle and subject in OCS largely follows the Greek originals, there are a few mismatches,<sup>17</sup> where there seems to be a slight tendency for OCS to be drawn towards the VS configuration, as shown in Table 5).

	OCS SV	OCS VS
Greek SV	33	7
Greek VS	3	80

**Table 5** Subjects of Greek/OCS absolutes by position.

Furthermore, there are 22 occurrences of dative absolutes with the VS configuration which do not translate a Greek genitive absolute, suggesting that the participle-subject order does reflect a discourse property of the construction regardless of the Greek originals.

As Table 6 shows, most realized subjects belong to categories typically encoding old or accessible referents:

<sup>16</sup> OCS adjectives (attributive and nominalized) and participles are inflected in either the ‘long’ or the ‘short’ form (sometimes referred to as ‘weak’ and ‘strong’ respectively, from the Germanic tradition). The difference is mostly one of definiteness: old or inferable nominal referents are marked on the adjective through the long form, whereas adjectives whose referent cannot be retrieved via context or world knowledge normally appear in the short form. In (17), the expected short form would have been *\*sějōštu*.

<sup>17</sup> The numbers refer only to the occurrences for which an OCS dative absolute corresponds to a genitive absolute, which is not always the case. They also do not include the temporal impersonal constructions mentioned above.

PoS	Frequency
<b>Personal pronouns</b>	50.3% (92)
<b>Common nouns</b>	20.2% (37)
<b>Proper nouns</b>	5.4% (10)
<b>Demonstrative pronouns</b>	2.2% (4)
<b>Indefinite pronouns</b>	2.2% (4)
Null/NA	19.7% (36)

**Table 6** Dative absolutes: subject parts of speech.

Only indefinite pronouns and common nouns can potentially introduce new referents: these together represent only 22.4% of all subjects, without, however, considering that several common nouns will also likely have old or accessible referents (e.g. *narod* ‘crowd, multitude’, *běs* ‘demon’, *gospod* ‘Lord’, *sloněce* ‘sun’).

As many as 59.9% of all overt subjects are a third person pronoun *\*i*, as shown in Table 7.<sup>18</sup>

Lemma	Frequency
<i>*i</i> ‘he, they’	59.9% (88)
<i>narod</i> ‘crowd, multitude’	3.4% (5)
<i>isus</i> ‘Jesus’	2.7% (4)
<i>sloněce</i> ‘sun’	2.2% (4)
<i>všs</i> ‘all’	2.0% (3)
<i>t</i> ‘that (one)’	2.0% (3)
<i>běs</i> ‘demon’	1.4% (2)
<i>gospod</i> ‘Lord’	1.4% (2)
<i>ljudije</i> ‘people’	1.4% (2)
<i>člověk</i> ‘person’	1.4% (2)

**Table 7** Dative absolutes: 10 most-frequent subject lemmas (out of total number of overt subjects).

It is hard, however, to fully appreciate the relevance of these figures on their own: because of the subject matter, the majority of third-person singulars refer to Jesus and third-person plurals to the Apostles. Only larger and more diverse datasets may be able to provide a reliable picture (see Section 3.2).

<sup>18</sup> The asterisk (here and below) is due to the fact the third-person personal pronoun is not attested in OCS in the nominative singular. Instead, demonstratives are generally used in the nominative (*t* ‘that one’, *on* ‘that one there’ and *s* ‘this one’).

### 2.3 Lexical variation

Lexical variation among participles in different syntactic configurations can also indirectly point to important functional differences.

As Haug (2012: 320) argues on the basis of Ancient Greek, FRAMES are assumed to be more ‘predictable’, since they are ‘presupposed (anaphoric or accommodated) and not explicitly asserted’, so that we expect to find less lexical variation among sentence-initial participles than in other positions. ELABORATIONS and INDEPENDENT RHEMES are instead expected to present more variation as they typically introduce new information. Since the post-matrix position is likely to be predominantly occupied by ELABORATIONS, we expect that position to have a very high degree of lexical variation. Pre-matrix conjunct participles are instead often ambiguous between FRAMES and INDEPENDENT RHEMES, so we expect a lower degree of lexical variation than among post-matrix participles. In Table 8, type-token ratio (TTR)<sup>19</sup> is used as a measure of overall lexical richness. The first column indicates how many lemmas belong to the 10 most-frequent lemmas in that configuration (i.e. whether we are dealing with few high-frequency lemmas).<sup>20</sup>

	10 most-frequent lemmas	TTR
<b>Pre-matrix</b>	42.1%	0.22
<b>Post-matrix</b>	29.7%	0.48

**Table 8** Lexical variation among participles by position relative to the matrix clause (regardless of participle type).

Much like in Greek (see the data in Haug (2012)), OCS participles to the right of the main verb show an overall higher degree of lexical variation than those to the left, as the difference in TTR suggests. The first column of Ta-

<sup>19</sup> Type-token ratio, a simple measure of lexical richness (a text’s vocabulary size divided by the text size), is used here to look at the proportion between number of occurrences and unique participle lemmas, as an additional criterion to weigh lexical variation among participles, especially given the difference in dataset size between the two constructions. The closer the TTR to 1, the higher the degree of lexical richness, thus, in our case, variation among participle lemmas.

<sup>20</sup> The numbers in Tables 8 and 9 do not include *glagolati* ‘speak, say’ (normally imperfective: *gl[agol]ję*) for post-matrix conjunct participles and *otvęštati* ‘answer’ (normally perfective: *otvęštavę*) for pre-matrix conjunct participles, since these would likely skew the frequencies. As noted by Haug (2012: 288) on the respective Greek forms (*legōn* ‘say’ and *apokritheis* ‘answer’), these work roughly like quotative particles, a usage sometimes considered a Semiticism (ibid.). *Glagolati* ‘talk, say’ in absolute constructions should arguably not be considered on a par with the same lemma in conjunct participles, since in dative absolutes it is generally used as an intransitive verb (‘talking’ rather than ‘saying’).



ble 8 indicates that the 10 most-frequent lemmas among pre-matrix participles account for 42.1% of all pre-matrix participles (regardless of type, that is, both absolutes and conjunct participles). The figure goes down to 29.7% for post-matrix participles. Since Haug (2012) focuses on conjunct participles, without providing separate figures for absolute constructions, it is also worthwhile to compare the degree of lexical variation among dative absolutes and conjunct participles in pre-matrix position alone (where there is most frequently functional overlap), as shown in Table 9.

	10 most-frequent lemmas	TTR
<b>Conjunct</b>	43.9%	0.22
<b>Absolute</b>	52.0%	0.44

**Table 9** Lexical variation among conjunct participles and dative absolutes in pre-matrix position.

The figures for conjunct participles are very similar to those seen in Table 8 for pre-matrix participles as a whole. On the other hand, the 10 most-frequent lemmas among pre-matrix dative absolutes account for 52% of all dative absolutes in that configuration (compared to 42.1% in Table 8). The TTR, however, is also much higher (0.47, compared to 0.22 in Table 8). While this might also be due to the difference in sample size (171 pre-matrix dative absolutes against 934 pre-matrix conjunct participles), these numbers may also indicate that dative absolutes tend to be used very frequently with very few lemmas. Only a much larger dataset may confirm whether this is the case.

## 2.4 Summary

Dative absolutes in the *Codex Marianus* were found to align strongly with the characteristics of Bary & Haug’s (2011) FRAMES regardless of context. On the other hand, conjunct participles appear to fulfil very different discourse functions depending on syntactic configuration and aspect. This aligns with Bary & Haug’s (2011) claim that predicative participles behave, at least from the temporal-semantic perspective, like grammaticalized versions of discourse relations. Information-structural annotation also indicated that the subjects of sentence-initial dative absolutes, unlike those of conjunct participles in the same position, are very likely topic candidates in the sentences in which they appear, which is a further indication of the identification of absolutes as FRAMES.

The next section replicates the analysis of dative absolutes carried out so far on a much larger and diverse dataset, leveraging these findings as a

blueprint to obtain further insights into the construction’s functions across different texts.

### 3 DATIVE ABSOLUTES IN SHALLOWLY-ANNOTATED TREEBANKS

A number of Early Slavic texts in TOROT currently have morphological and syntactic tagging, but no information-structural annotation.<sup>21</sup> Most of the frequencies presented in the previous paragraphs on the basis of the Gospels could, however, still be reproduced on this dataset. Once again, the focus was on:

- the relative order of dative absolutes and matrix verbs; aspect distribution (Section 3.1);
- the properties of realized subjects (Section 3.2);
- the lexical variation among dative absolutes in different subcorpora (Section 3.3).

#### 3.1 Position and aspect distribution

The same query used to extract dative absolutes from the *Codex Marianus* yielded no clear matrix clause for 17.2% of the occurrences (*NA* in Table 10).<sup>22</sup>

---

DA-matrix	Matrix-DA	NA
64.6% (902)	18.2% (254)	17.2% (241)

---

**Table 10** Dative absolute/matrix clause relative position.

---

Most occurrences counted under *NA* are dative absolutes connected to the matrix clause by a coordinating conjunction (*i* ‘and’ or *a* ‘but, whereas’). This

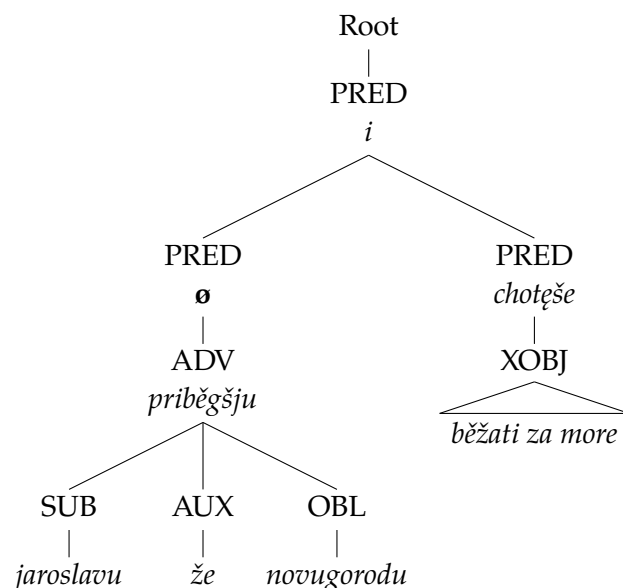
<sup>21</sup> This dataset contains mixed OCS, later Church Slavonic, OES and Middle Russian texts, from manuscripts of various provenances and dating. For OCS: *Codex Suprasliensis*, *Psalterium Sinaiticum* and *Kiev Missal*. For later Church Slavonic (or OCS texts read in later copies or heavily regional redactions): *Vita Constantini*, *Vita Methodii*, *Chrabr’s Treatise on the letters and Life of Petka Tarnovska* (Novaković). For OES: the *Primary Chronicle* from the *Codex Laurentianus* and (excerpts) from the *Codex Hypatianus*, *Suzdal Chronicle* (*Codex Laurentianus*), *Novgorod First Chronicle* (Synodal MS) and *Uspenskij sbornik*. Finally, for Middle Russian, the *Life of Sergij Radonezh*. Note that ‘(O)CS’ will be used to refer together to OCS and later Church Slavonic texts from this dataset (i.e. excluding the *Codex Marianus* from the first case study), unless otherwise specified.

<sup>22</sup> 1397 dative absolutes were extracted from this dataset through a query performed in June 2020.

One question, different annotation depths

is because in order to avoid direct coordination of dative absolutes with finite clauses, mostly for the sake of retrievability, the annotation convention in TOROT is to treat such examples as elliptical constructions. During their annotation, an empty verbal node is added above the absolute to stand in for the matrix clause, as Figure 2 and example (18) show.<sup>23</sup>

- (18) *jaroslavu že přiběgšju novugorodu*  
 Jaroslav.DAT PTC arrive-running.PTCP.PST.M.DAT.SG Novgorod.DAT  
*i chotěše běžati za more*  
 and want.IMPF.3SG escape.INF beyond sea.ACC  
 ‘When Jaroslav arrived to Novgorod in his flight, he planned to  
 escape overseas’ (PVL 143.18-19, 128835)



**Figure 2** The syntactic analysis of (18) in TOROT.

The remainder of the occurrences for which no information on the main verb was immediately available are absolutes lacking a clearly identifiable matrix clause, to which they could be attached in the same sentence. In many of these examples they introduce reported speech, a usage examined by Collins (2011: 113-122) in his broader analysis of ‘syntactically independent’ dative absolutes in Early Slavic. As Collins notes, when the wider discourse is taken

23 Abbreviations: PVL = *Primary Chronicle* (Codex Laurentianus); Supr. = *Codex Suprasliensis*.

- The subsequent discourse (19b)-(19c) reveals that (19a) is in fact part of a series of non-finite clauses leading up to a finite main clause (*reč* '(he) said') (19c). The order of the different types of clauses in (19) can be motivated at the level of discourse organisation, where the dative absolutes in (19a) still function as FRAMES beyond clause-level. As argued by Collins (2011: 126), 'the relation between the absolute and the unit (in some cases, larger than a

clause) to which it is most closely linked in semantic terms is not always subordination in the syntactic sense' and its function can be seen as 'signalling that the proposition that it expresses is secondary in its discourse context'.

Table 10 shows that 64.6% are found to the left of the matrix clause. While these clearly represent the majority of dative absolutes, there are as many as 18.2% of the occurrences following the matrix. This indicates that, outside the limited narrative styles represented by the Gospels, the matrix-dative absolute configuration is not as uncommon as the first dataset suggested.

A closer look at potential post-matrix dative absolutes reveals that some of these occurrences are either pre-matrix from the discourse perspective (e.g. *byst*-clauses<sup>24</sup> in (20)) or augmented with an overt subordinating conjunction, such as *jako* 'as, since' or *ašte li* 'if',<sup>25</sup> which, as the syntactic head of the participle, were extracted by the query as the potential matrix of the construction (thus resulting as post-matrix from the raw dataset). Some absolutes introduced by a subordinating conjunction do indeed follow the matrix (21), whereas others precede it (22):

- (20) *i* *byst* *besědujōštēma* *ima* *i*  
 and happen.AOR.3SG converse.PTCP.PRS.DAT.DU he.DAT.3DU and  
*sъпраšajōštēma* *sę* *čto* *byvōšee* *i* *čto*  
 discuss.PTCP.PRS.DAT.DU REFL what be.PTCP.PST.N.ACC.SG and what  
*bōdōšteje* *i* *sam* *sps*  
 be.PTCP.FUT.N.ACC.SG and self.NOM saviour.NOM  
*približ* *sę* *iděše* *s* *nima*  
 approach.PTCP.PST.M.NOM.SG REFL go.IMP.F.3SG with he.INS.3DU  
 'And it came to pass that, as they were conversing and discussing  
 what had been and what would be, the Saviour himself approached

24 *Byst*-clauses are sentences introduced by *byst* 'it came to pass, it happened'. Like the main Ancient Greek counterpart *egeneto*, as well as analogous forms in several other historical Indo-European languages (e.g. Latin *factum est*, Old English *gelamp*, Gothic *warþ*), these have relatively rigid discourse structures, where 'it came to pass' works as a discourse marker, followed by a background-foreground pair (see Brinton 1996: 134, Traugott & Dasher 2002: 52). In the OCS example (20), the background is given by the whole phrase from *besědujōštēma* 'conversing' to *bōdōšteje* 'being (future)', while the foreground is everything following *bōdōšteje* 'being (future)'. Despite the conjunction between background and foreground in (20), the background is still forward-looking (in SDRT-terms; see Asher, Prévot & Vieu 2007: 6-14) from the discourse perspective.

25 The occurrence of absolute constructions introduced by a connective is obviously not unheard of (see in particular Stump 1985, and König 1991 on English *with*-augmented absolutes). It would instead be worthwhile to investigate the distribution of such connectives: it has sometimes been suggested that the semantic interpretation of a dative absolute, while driven by pragmatic reasoning, may be influenced by a principle of iconicity, whereby pre-matrix absolutes can always receive a temporal reading, unlike post-matrix ones (Worth 1994: 34).

them and started walking with them' (Supr. 41, 168486)

- (21) *i plakati se nača popadŭja. jako mertou*  
 and cry.INF REFL begin.AOR.3SG priest's-wife.NOM as dead.DAT  
*suštju onomu*  
 be.PTCP.PRS.M.DA.SG that.DAT  
 'And the priest's wife started to weep, as if he were dead'  
 (PVL 261.15-16, 133479)

- (22) *ašte li sice. izvolšu bgu. ... da budet*  
 if Q thus desire.PTCP.PST.M.DAT.SG god.DAT ... VOL be.FUT.3SG  
*volę gnę.*  
 will.F.NOM.SG lord.ADJ.F.NOM.SG  
 'It thus God desires ... may Lord's will be done'  
 (Life of Sergij of Radonezh 13, 198411)

The query could have been changed to extract the actual position of overtly-coordinated, as well as 'empty-headed', dative absolutes. However, doing so would not have added much insight to the relative frequency of pre- and post-matrix dative absolutes. Furthermore, the analysis of the raw dataset allows us to identify 'non-canonical' configurations of the constructions more easily. Most importantly, seemingly independent dative absolutes should be treated separately, since they have long been an object of discussion (Worth 1994; Corin 1995; Collins 2004, 2011).

Concerning aspect distribution, as Table 11 shows, perfectives and imperfectives are almost equally frequent in the dataset (47.8% and 52.2% respectively). Among dative absolutes to the left of the matrix, the two aspects are nearly even, in line with the pattern found in the Gospels. This was explained, following Haug (2012: 311-312), by the fact that the dominant aspect of framing participles is expected to vary with genre – unlike conjunct participles, whose aspect was shown to strongly correlate with their position. At the same time, among dative absolutes to the right of the matrix, imperfectives are more frequent:

	Perfectives	Imperfectives
<b>DA-Matrix</b>	52.8% (476)	47.2% (426)
<b>Matrix-DA</b>	26% (66)	74% (188)
NA	52.3% (126)	47.7% (115)
<b>Tot.</b>	47.8% (668)	52.2% (729)

**Table 11** Dative absolutes: aspect distribution (row percentage).

One question, different annotation depths

These frequencies beg the question of whether post-matrix dative absolutes, being more frequently imperfectives, are in fact more likely to function as ELABORATIONS than FRAMES. Lexical variation among dative absolutes (see Section 3.3) in the two configurations may help clarify the issue.

### 3.2 *Properties of subjects*

As Table 12 shows, over half of the subjects are represented by parts of speech which are either inherently anaphoric (personal and demonstrative pronouns) or likely to encode old or accessible referents (proper nouns).

PoS	Frequency
Common nouns	41.2%
Personal pronouns	26.1%
Proper nouns	15%
Demonstrative pronouns	10.1%
Adjectives	3.1%
Indefinite pronouns	2.6%
Verbal nouns	0.8%
Cardinal numerals	0.6%
Ordinal numerals	0.2%
Interrogative pronouns	<0.1%
Relative pronouns	<0.1%
Possessive pronouns	<0.1%

**Table 12** Subject parts of speech.

In comparison to the Gospels, subject parts of speech appear to be more varied (12 parts of speech, against 5 in the Gospels). However, many of these occur relatively rarely and the only four standing out are common nouns, personal pronouns, proper nouns, and demonstratives. Furthermore, the picture becomes more similar to the one observed in the Gospels if we consider lexical variation among subjects. Table 13 shows how many subjects belong to the ten most-frequent lemmas in each subcorpus. The first row refers to the lexical variation that emerged from the entire dataset without normalizing the (O)CS and OES spellings, whereas the second row compares that result with the one that emerged after normalization. The differences in normalization between the (O)CS and the OES subcorpora in TOROT result in listing as two separate lemmas some lexical items, which should be regarded as one, as two separate lemmas. In our case, further normalization of the two subcorpora can only produce even lower overall variation, but it is still useful to be

aware of the extent to which our observations may be affected.<sup>26</sup> The following rows show the lexical variation in the individual subcorpora, as well as the one that emerged by integrating the Gospels dataset from Section 2, for ease of comparison.

Subcorpus	10 most-frequent lemmas	TTR
(O)CS + OES (non-normalized)	40.7%	0.28
(O)CS + OES (normalized)	44.2%	0.26
(O)CS	44.5%	0.36
OES	48.8%	0.26
(O)CS + OES + Gospels (normalized)	46%	0.24
(O)CS + Gospels	49.2%	0.31
Gospels	78.2%	0.27

**Table 13** Lexical variation of subjects by subcorpus.

Even without further normalizing the spellings we are able to gather that lexical variation is relatively low, in line with the dative absolutes from the previous dataset. As Table 13 shows, further normalization only yielded a slight decrease in variation (with 44.2% of subjects belonging to the ten most-frequent lemmas, compared to the 40.7% from the non-normalized dataset). By integrating the dative absolutes from the Gospels into the full dataset (and normalizing the spellings), the results are still very similar to the one we obtain without incorporating them (with 46% of the subjects belonging to the 10 most-frequent lemmas, compared to the 44.2% obtained from the dataset without the Gospels). By adding them only to the (O)CS subcorpus the figure that we obtain is somewhat higher (49.2% compared to the 44.5% from the (O)CS subcorpus without the Gospels). What is most indicative of the figures in Table 13, however, is that the TTR is consistently between 0.26 and 0.36, which is a sign of low lexical variation. This also suggests that the subcorpora used in this case study are likely to be representative of the features of subjects of dative absolutes as a whole.

Given the heterogeneity of the dataset in terms of genres and language

<sup>26</sup> Normalization of (O)CS and OES spellings in TOROT follows two different conventions, so that further normalization is needed to obtain a more accurate picture of lexical variation across all Early Slavic texts regardless of their Slavic variety. The main changes made are: 1) *je* and word-initial *a* were changed to *e* and *ja* respectively, 2) Cyrillic characters for nasal vowels were changed to the respective east Slavic outcome (*ę/ję* > *ja*, *ѣ/jѣ* > *ju*), 3) *ě* was changed to *e*, 4) multiple variants for *y* were eliminated (and all set to <ѣ>), 5) word-final *ii* and *yi* in OES texts were changed to *ь* and *ѣ* respectively, 6) *ТѣRT ТѣRT* > *TRѣT TRѣT*. The normalization routine can be found in the data depository at <https://doi.org/10.6084/m9.figshare.12894035.v1>.



varieties, it is useful to check whether there are differences between the types of subjects in the (O)CS and in the OES subcorpora. Table 14 shows the most-frequent subject lemmas by subcorpus.

Subcorpus	Lemmas
(O)CS + OES (non-normalized)	*i ‘he’, <i>bogъ</i> ‘God’, <i>onyi</i> ‘that one there’, <i>azъ</i> ‘I’, <i>dъnъ</i> ‘day’, <i>tyi</i> ‘that one (OES spelling)’, <i>kъnjazъ</i> ‘king’, <i>tъ</i> ‘that one ((O)CS spelling)’, <i>bratija</i> ‘community (of monks)’, <i>lěto</i> ‘year’
(O)CS + OES (normalized)	*i ‘he’, <i>onъ</i> ‘that one (there)’, <i>bogъ</i> ‘God’, <i>tъ</i> ‘that one’, <i>azъ</i> ‘I’, <i>kъnjazъ</i> ‘king’, <i>sъ</i> ‘this one’, <i>dъnъ</i> ‘day’, <i>bratija</i> ‘community (of monks)’, <i>lěto</i> ‘year’
(O)CS	*i ‘he’, <i>tъ</i> ‘that one’, <i>azъ</i> ‘I’, <i>sъ</i> ‘this one’, <i>svętъ</i> ‘saint’, <i>onъ</i> ‘that one (there)’, <i>ty</i> ‘you’, <i>lěto</i> ‘year’, <i>vъsъ</i> ‘all’, <i>gospodъ</i> ‘Lord’
OES	*i ‘he’, <i>onyi</i> ‘that one (there)’, <i>bogъ</i> ‘God’, <i>tyi</i> ‘that one’, <i>kъnjazъ</i> ‘king’, <i>azъ</i> ‘I’, <i>bratija</i> ‘community (of monks)’, <i>izъjaslavъ</i> ‘Izjaslav’, <i>dъnъ</i> ‘day’, <i>volodiměrъ</i> ‘Volodimir’

**Table 14** 10 most-frequent subject lemmas by subcorpus.

Although common nouns comprise >40% of all subject parts of speech, the most frequent of these encode old or accessible referents (e.g. *bogъ* ‘God’, *kъnjazъ* ‘king’, *bratija* ‘community (of monks)’, *gospodъ* ‘Lord’). Other lemmas are likely to encode prominent subjects in a limited part of the subcorpus. All occurrences of *svętъ* ‘saint’, for instance, are found in hagiographical passages contained in the *Codex Suprasliensis*, where they likely introduce a new event related to the life of a given saint, as in (23).

- (23) *vъšъdъšu*                      *že* *stuumu* *i* *stavъšu*  
 enter.PTCP.PST.M.DAT.SG    PTC   saint.DAT   and   stand.PTCP.PST.M.DAT.SG  
*na sōdišti·*    *rěšę*                      *kъ nemu* *vъsvi·*  
 on seat.LOC   say.AOR.3PL   to   he.DAT   mage.NOM.PL  
 ‘When the Saint came in and stood at the seat the mages told him:  
 “...”  
(Supr. 23, 138584)

The lack of information-structural annotation prevents us from validating these observations across the entirety of the treebank and for every lemma. If the goal was to check this information for individual lemmas, one could do so manually by looking at the context for every occurrence. However, especially with large treebanks, this process is often too time-consuming and therefore unfeasible. Furthermore, in order to check whether the observations made on the information status of a particular form reflect general characteristics of the construction, we would need to repeat the process for every relevant lemma. If information-structural annotation (which we defined as ‘deep’) was available for this treebank, as in the treebank used in Section 1, we would be able, for instance, to exploit the givenness status annotation, as well as the information on *relative saliency*, to compare the average pick-up rates and overall information status of the referents of the subjects in the (O)CS/OES dataset and in the *Codex Marianus*. Since the treebank used in this Section is only shallowly-annotated (i.e. it only contains up to syntactic dependency annotation), it would only be feasible to obtain this type of information for a small number of lemmas.

### 3.3 Lexical variation

Lexical variation among dative absolutes is generally low across all configurations. Overall, as Table 15 shows, 38.9% of them belong to the 10 most-frequent lemmas (with a total TTR of 0.30, which is a low TTR).

Position	10 most-frequent lemmas	TTR
<b>Pre-matrix</b>	40.6%	0.33
<b>Post-matrix</b>	40.9%	0.53
NA	44%	0.49
<b>Total</b>	38.9%	0.30

**Table 15** Lexical variation of participles by position (normalized dataset).

The higher TTR for post-matrix dative absolutes indicates that, overall, a higher number of lemmas are used in that position. However, the high percentage of dative absolutes belonging to the 10 most-frequent lemmas suggests that the pool of frequently used verbs is very limited - even more so than pre-matrix ones, which have a lower TTR (i.e. each lemma is relatively more evenly spread across all lemmas in that position).

Much like in the Gospels, the relative high frequency of *byti* ‘be’ is quite apparent, making up 16.9% of all dative absolutes in the corpus – nearly three times the next most-frequent lemma:

Lemma	Frequency
<i>byti</i> ‘be’	16.9% (236)
<i>přiti</i> ‘arrive’	5.9% (82)
<i>rešti</i> ‘say’	2.7% (38)
<i>glagolati</i> ‘speak, say’	2.5% (35)
<i>iti</i> ‘go’	2.5% (35)
<i>chotěti</i> ‘want’	2.5% (35)
<i>sěděti</i> ‘sit’	2% (28)
<i>žiti</i> ‘live’	1.4% (19)
<i>minŕti</i> ‘pass’	1.3% (18)
<i>stojati</i> ‘stand’	1.3% (18)

**Table 16** Dative absolutes: 10 most-frequent lemmas (normalized dataset).

The lemmas in Table 16 are all found with some frequency in both the (O)CS and the OES subcorpora, with the exception of *žiti* ‘live’, which is found only once in the (O)CS *Codex Suprasliensis*, the rest all occurring in the OES subcorpus. Lemmas denoting motion/position or speech, as well as existential *žiti* ‘live’ and *byti* ‘be’ could easily be considered to be framing devices on a par with the typical dative absolutes of the Gospels, translating into expressions of the type ‘as *x* was praying/speaking’, ‘after *x* arrived’, or ‘when *x* was still alive’.

Mental-state verbs such as *chotěti* ‘want’ and *věděti* ‘know’ (the latter not in Table 17, but occurring 16 times) involve a different reading than the typical temporal ones seen in the OCS dataset. This, however, does not necessarily entail overall different functions: as already mentioned, unless augmented with overt connectives (e.g. ‘because’, ‘when’, ‘while’), the semantics of participial adjuncts is generally underspecified, which is why a semantic framework integrating discourse relations was needed in the first place to fully account for dative absolutes. It is instead interesting to note that most occurrences of *věděti* ‘know’ (all imperfective) appear in a specific genre and language variety (the OES chronicles<sup>27</sup>), and that its subjects are often names of kings (i.e. *jaroslavъ* ‘Jaroslav’, *mъstislavъ* ‘Mstislav’, *volodimerъko* ‘Volodimirko’, *borisъ* ‘Boris’, *izjaslavъ* ‘Izjaslav’; see examples (24)-(25)). It is tempting to view this as a tendency for dative absolutes to occur in fixed or semi-fixed verb-subject combinations, especially considering that the common noun *kъnjazъ* ‘king’, despite being among the most-frequent lemmas in the subcorpus,

<sup>27</sup> Of the 16 total occurrences, 8 are found in the *Suzdal Chronicle*, 2 in the PVL, 1 in the *First Novgorod Chronicle*; outside the chronicles, 4 are in the *Uspenskij Sbornik* and only 1 in the *Codex Suprasliensis*.

is never found as the subject of *věděti* ‘know’.

- (24) *jaroslavu že ne věduštju ôttěně*  
 Jaroslav.DAT PTC NEG know.PTCP.PRS.M.DAT.SG father.ADJ.F.DAT.SG  
*smrti. varęzi bęchu mnozi ou*  
 death.F.DAT Varangian.NOM.PL be.IMPF.3PL much.NOM.PL at  
*jaroslava*  
 Jaroslav.GEN  
 ‘While Jaroslav had not heard of his father’s death yet, many  
 Varangians were under his command’ (PVL 140.16-18, 128721)

- (25) *volodimerku že togo ne věduštju so*  
 Volodimerko.DAT PTC that.GEN NEG know.PTCP.PRS.M.DAT.SG with  
*andrějem. i stasta u mičęska*  
 Andrej.INS and set.out.AOR.3DU at Mičesk.GEN  
 ‘Volodimerko and Andrej did not know that and went to Mičesk’  
 (Suzdal Chronicle, year 6658, 208359)

The remainder of the occurrences of *věděti* ‘know’ appear instead in a clearly fixed expression meaning ‘without anyone knowing’, ‘not known to anyone’, all in the OES subcorpus (as in (26)) except one example in *Suprasliensis* (27):

- (26) *ne věduštju nikomuže pridoch v*  
 NEG know.PTCP.PRS.M.DAT.SG nobody.DAT come.AOR.1SG in  
*pečeru.*  
 crypt.ACC  
 ‘Without anyone knowing I went to the crypt’ (PVL 210.3-4, 156216)

- (27) *i zatvori i vę chyzinę. pęti desęť ti*  
 and lock-up.AOR.3SG he.ACC in shack.LOC five.GEN ten.GEN PTC  
*pęti lętę sęšta. nikomuže inomu*  
 five.GEN year.GEN.PL be.PTCP.PRS.M.GEN.SG nobody.DAT.SG other.DAT  
*otę družiny vědęstu jako episkupę*  
 from company.GEN know.PTCP.PRS.M.DAT.SG that bishop.NOM.SG  
*jestę*  
 be.PRS.3SG  
 ‘And he locked him up in a shack when he was fifty-five, without  
 anyone else from the company knowing that he was a bishop’  
 (Supr. 25, 139641)

*Chotęti* ‘want’ on the other hand, is not always used in its lexical meaning of ‘wanting to’, but rather as a modal auxiliary translatable as ‘being about to’, as

in (28)-(29), thus not posing particular issues from the discourse perspective, where it can be interpreted as a generic FRAME (‘when *x* was about to do *y*, *z* happened’):

- (28) *na sud že jemu choteštu iti.*  
to judgement.ACC PTC he.DAT want.PTCP.PRS.M.DAT.SG go.INF  
*plaka se mati otročete sego.*  
cry.AOR.3SG REFL mother.NOM child.GEN this.GEN  
*glagoljušti*  
say.PTCP.PRS.F.NOM.SG  
‘But when he was ready to go for the Judgement, the mother of the child cried, saying: “...”’ (Vita Constantini 2, 216314)

- (29) *i chotěvъšu jemu po dvoju dъniju*  
and want.PTCP.PST.M.DAT.SG he.DAT on second.LOC day.LOC  
*otъplouti. javi se jemu vъ sně*  
sail.away.INF appear.AOR.3SG REFL he.DAT in dream.LOC  
*aggelъ gospodъnъ glagolę*  
angel.NOM lord.ADJ.NOM.SG say.PTCP.PRS.M.NOM.SG  
‘And as he was going to sail away on the second day, an angel of the Lord appeared to him in dream saying: “...”’ (Supr. 25, 139658)

One more evident pattern that arises from the dataset is the relatively larger number (compared to the Gospels) of non-finite constructions coordinated with a following finite clause, the former often involving a dative absolute. This has emerged indirectly through dependency information in the tree-bank, which revealed that a number of dative absolutes have another participle, rather than a finite verb, as their syntactic head, in many cases entailing that matrix of the dative absolute is actually an INDEPENDENT RHEME overtly coordinated with the finite clause, as in (30). In other cases, the dative absolute is headed in TOROT by an empty verbal node, suggesting that the absolute participle itself is coordinated with a following predication, as in (31).

- (30) *janevi že iduštju domovi. v druguju*  
Yan.DAT PTC go.PTCP.PRS.M.DAT.SG home.ADV in other.ACC  
*noštъ. medvědъ vъzlězъ*  
night.ACC bear.NOM attack.PTCP.PST.M.NOM.SG  
*ougryzъ eju i sněstъ*  
gnaw.PTCP.PST.M.NOM.SG he.3DU and eat.AOR.3SG  
‘While Yan was going home the next night, a bear attacked them, gnawed them and ate them up’ (PVL 178.17-19, 130173)

- (31) *približivŕšu*                      *že se s[vę]tuumu· i*  
 approach.PTCP.PST.M.DAT.SG PTC REFL saint.DAT and  
*sŕtvorivŕ*                      *christosovo*                      *znamenije na čelě*  
 make.PTCP.PST.M.NOM.SG Christ.ADJ.N.ACC sign.ACC on knee.LOC  
*svojemŕ vnide*                      *vŕ crkŕve*  
 his.LOC enter.AOR.3SG in church.ACC  
 ‘When the saint had drawn near and had crossed himself, on his  
 knees, he entered the church’ (Supr. 19, 137399)

(31) is particularly revealing, since it involves a dative absolute coordinated with a conjunct participle with which it shares the subject. While this absolute is unarguably a FRAME, the conjunct participle is ambiguous and could be interpreted as either a FRAME or an INDEPENDENT RHEME. Examples like (31) show that the functions of dative absolutes can be entirely consistent with their general identification as framing devices even when two of its most non-prototypical syntactic features (i.e. subject co-indexing with the matrix and coordination to a following predication) are present.

### 3.4 Summary

The analysis of the second dataset indicated that data emerged from small, but deeply-annotated treebanks can be used to make relatively safe predictions on much larger treebanks with shallower annotation. The second dataset lacked information-structural annotation, which would have been useful to directly compare the information status (e.g. givenness, topic score, relative saliency) of the subjects in dative absolutes to the left and right of the matrix. In particular, post-matrix dative absolutes were found to be much more common than in the previous dataset. While we were able to capture some of the features of post-matrix dative absolutes indirectly through lexical variation, annotation on information status could have provided further insights into how the construction differs across the different corpora.

The new data also suggested that, beside the pool of typically framing verbs common to most texts and, perhaps, historical stages, different genres and language varieties might involve different high-frequency verbs. In some cases these appear in fixed or semi-fixed expressions. The phenomenon thus lends itself well to genre-based distributional analyses, which is beyond the scope of this paper.<sup>28</sup> In Section 4 I extract and analyze dative absolutes from a completely new text by looking for the same variables which have so far proved most informative.

<sup>28</sup> See Kurešević (2006) for a study on dative absolutes using a genre-based approach, although not from a quantitative perspective.

#### 4 DATIVE ABSOLUTES IN STRATEGICALLY-ANNOTATED TEXTS

This final case study considers the extent to which corpora with considerably fewer and shallower levels of annotation can be exploited to investigate a discourse-driven syntactic phenomenon, particularly as a means of corroborating results emerging from treebanks with deeper annotation.

To this end, a previously unannotated text, the Slavonic *Story of Abraham of Qidun and his niece Mary*,<sup>29</sup> contained in the 14th-century Middle Bulgarian *Bdinski Sbornik* (Dujčev 1972; Scharpe & Vyncke 1973), was added to TOROT.<sup>30</sup> The text was automatically lemmatized, and part-of-speech and morphological tags were added using Scherrer, Rabus & Mocken's (2018) pre-modern Slavic CLSTM tagger.<sup>31</sup>

Before post-correction, 121 potential participles were extracted, 15 of which were identified as potential dative absolutes solely on the basis of morphological tagging. Of these, two turned out to be part of the same passive absolute construction (*prizvaně byvši* 'having been called'), where one of the dative participles (*byvši*) is the copula (i.e. two of the potential absolutes were actually one). One occurrence was found to be a finite verb (*nareku* 'I call, will call'), which the tagger confused with a dative participle, possibly because of the stem *narek-* (morphologically suggesting a past active participle) and the *-u* ending (suggesting a masculine or neuter dative). Finally, two of the potential occurrences were revealed to be regular dative participles agreeing with a preceding dative noun.

All remaining potential absolutes (11 occurrences) were confirmed as such, indicating that even automatic processing alone can be useful if the main goal is to detect and gather a large number of examples for a given phenomenon over a wide pool of texts.

In our case, the small amount of occurrences would allow us to closely analyze each dative absolute and the contexts in which they appear on a case by case basis. However, it is useful to look at the extent to which sheer automatic lemmatization, part-of-speech and morphological tagging can be used

29 For a detailed discussion of the text's relation to its Greek sources, including translation techniques in this and related hagiographies, see Stern (2013, 2015, 2016, 2018). I am grateful to Dieter Stern for pointing me to the relevant literature on the topic and for kindly sending me copies of the cited material.

30 The text in TOROT (<https://nestor.uit.no/sources/1009>) was adapted from the digital edition of the *Bdinski Sbornik* curated by David Birnbaum (<http://bdinski.obdurodon.org>).

31 I am grateful to Yves Scherrer for the support and for kindly providing an updated version of the model, including a conversion script to obtain the original tag format used in PROIEL-TOROT, which is now available in the model's repository (<https://github.com/yvesscherrer/lstmtagger>). The tagger is described in detail in Scherrer et al. (2018) and Scherrer & Rabus (2019).



to validate the results that emerged from the previous two datasets.

Several predictions were made based on the patterns noticed so far. In particular, despite their small number, the new occurrences were expected to be homogeneous with respect to the properties of their subjects, since that is where the most consistent features of dative absolutes have emerged. On the other hand, precisely because there are few occurrences, none of these were expected to be found to the right of the matrix. Even though in our second (larger) case study the post-matrix position was not as uncommon as in the *Codex Marianus*, most dative absolutes were still found sentence-initially.

Any prediction made on aspectual distinctions would be ill-informed: even if the relative frequency of perfectives and imperfectives resembled the one found for absolutes in the previous datasets,<sup>32</sup> the small size of the dataset would not allow us to draw any conclusion regarding a different role for verbal aspect than the one which has surfaced hitherto.

Extracting a window of 3 tokens on each side of the dative participle was enough to identify each subject, considering that in the previous two datasets it typically immediately followed or preceded the participle.

The properties of the subjects in the dataset were found to be fully consistent with those seen in the previous case studies. The majority of them are again represented by third-person pronouns (6 occurrences), as in (32).<sup>33</sup>

- (32) *potomъ sědeštu jemu na odrě, reč kb*  
 after sit.PTCP.PRS.M.DAT.SG he.DAT on bed.LOC say.AOR.3SG to  
*njemu*  
 he.DAT

‘After that, as he was was sitting on the bed, she said to him: “...”

(*Bdinski Sbornik*, 10v: 13, 286459)

Time-denoting nouns, *dnъ* ‘day’ and *leto* ‘year’ (one occurrence each), could be seen as a separate group, appearing in constructions functionally similar to the widespread temporal ones seen in the Gospels (discussed in Section 2.2). A particularly clear echo of the latter can be detected in (33), where *utro* ‘morning’ would be analyzed in TOROT as a predicative complement, consistently with similar examples headed by *byti* ‘be’ (e.g. *pozdě byvŕšu* ‘once it got late’, Matthew 14.23 38892).

<sup>32</sup> Both aspects are represented to some extent (7 perfectives, 4 imperfectives).

<sup>33</sup> Citation indicates ‘folio number: line number’ (e.g. *Bdinski Sbornik* 13r: 10), following the manuscript’s digital edition ([bdinski.obdurodon.org](http://bdinski.obdurodon.org)). This information can also be retrieved directly through the TOROT Treebank in each individual token’s page (e.g. by clicking on *utru*, which redirects to <https://nestor.uit.no/tokens/3286275>, the folio and line numbers are found in the ‘Foreign IDs’ box).



One question, different annotation depths

- (33) *outrou že byvšu reč kь nei*  
 morning.DAT PTC be.PTCP.PST.N.DAT.SG say.AOR.3SG to she.DAT  
 ‘When morning came, he said to her: “...”’  
 (Bdinski Sbornik 13r: 10, 286531)

The remaining subjects, *vьsb* ‘all’, *sь* ‘this’ and *otьcb* ‘father’ (one occurrence each), widely represented as the subjects of dative absolutes subjects across all texts, hardly pose any difficulty in terms of functional continuity with the previous datasets. Finally, all 10 overt subjects follow the participle, which, as we already commented on, can be explained as part of their framing function, whereby old referents get reinstated as shown in Table 17.

Subject	Participle	Position
*i (3SG)	<i>prizvati</i> ‘summon’	VS
*i (3SG)	<i>sěděti</i> ‘sit’	VS
*i (3DU)	<i>sěděti</i> ‘sit’	VS
*i (3DU)	<i>pěti</i> ‘sing’	VS
*i (3SG)	<i>lobyzati</i> ‘kiss’	VS
*i (3DU)	<i>vьniti</i> ‘enter’	VS
<i>sь</i> ‘this’	<i>prilučiti (se)</i> ‘happen’	VS
<i>otьcb</i> ‘father’	<i>umřeti</i> ‘die’	VS
<i>lěto</i> ‘year’	<i>končati</i> ‘finish’	VS
<i>dьnь</i> ‘day’	<i>pomesti</i> ‘change, replace’	VS
(impersonal)	<i>byti</i> ‘be’	-

**Table 17** Summary of subject-participle lemma combination.

Although the occurrences are too few to make any generalization regarding lexical variation, it is still possible to comment on the types of event they seemingly encode. Lemmas denoting movement or position, constituting the majority in the previous two case studies, are also represented in the new datasets, as in (32) above and in (34):

- (34) *i vьsedьšima že ima. vidě odrь*  
 and enter.PTCP.PST.DAT.DU PTC he.DAT.DU see.AOR.3SG bed.ACC  
*vysokь nastlanь*  
 large.ACC spread.PTCP.PST.PASS.ACC.SG  
 ‘As they entered, he saw a large bed made up’  
 (Bdinski Sbornik, 10r: 4, 286436)

When taken out of context, some of the lemmas instead hardly appear as predictable predications from the OCS standpoint (35).

- (35) *ona že vstavši i objetъ i*  
 she PTC stand-up.PTCP.PST.F.NOM.SG and embrace.AOR.3SG he.ACC  
*i načetъ lobizati po vii jeho.*  
 and start.AOR.3SG kiss.INF on neck.LOC he.GEN  
*lobyzajušti že jee. vъzvонja ubo vonjeju*  
 kiss.PTCP.PRS.F.DAT.SG PTC she.GEN exhale.AOR.3SG then smell.INS  
*črъnorizъčъskoju tělo jeho*  
 monk.ADJ.INS body.NOM he.GEN  
 ‘She stood up, and embraced him, and started kissing his neck. As  
 she was kissing him, the smell of monasticism issued from his body’  
 (Bdinski Sbornik 8r: 5, 286385)

Most of these, however, do not present new information but refer back to an event already introduced in the previous discourse. The dative absolute’s lemma *lobyzati* ‘kiss’ in (35), for example, had already occurred as part of the main predication in the previous sentence (*načetъ lobizati* ‘(he) started kissing’), and gets reinstated by the dative absolute not as a main event but as a stage-setter for a following new event.

It is tempting to see this as a general feature of the phenomenon, whereby dative absolutes are likely to be used to introduce a topic shift. Manual checking of the preceding and following predications in the datasets did in fact suggest that the construction typically frames a following new event by using the predication of the preceding main clause as the new topic time.

In order to check this more systematically, sentence division was fixed,<sup>34</sup> the automatic morphological analysis was spot-corrected and syntactic annotation was added to all sentences with a dative absolute, as well as to the preceding and following two clauses (expecting this to be sufficient to detect patterns). By correcting sentence division, all dative absolutes were confirmed to be sentence-initial, hence preceding the matrix clause, as we predicted. ‘Strategic’ syntactic annotation allowed us to extract all the subjects and the main verbs of the surrounding sentences. For most of the occurrences, either one of two patterns emerged: for some, the dative absolute’s lemma

<sup>34</sup> Sentence division is performed automatically in the pre-processing stage on the basis of punctuation marks. The punctuation system of Early Slavic, however, does not follow modern principles. Full stops <.> and middle dots <.> very often separate portions of text smaller than a sentence, so that during text segmentation dependent clauses frequently get separated from their matrix. When adding syntactic annotation, correct sentence division is restored by attaching any dependent clause to its matrix.

appeared as the main verb in one of the previous two sentences, as in (35) above and (36), for others the subject of the the dative absolute appeared in the previous sentence as a focal<sup>35</sup> element (37)-(38):

- (36) *prizovi mi ju da se poveselju dn̄s*  
 summon.IMP 1SG.DAT she.ACC so-that REFL enjoy.PRS.1SG today  
*s njeju. zělo bo jako slyšal̄b jesm o*  
 with she.INS much for that hear.PTCP.RES.M be.PRS.1SG about  
*njei godě mi jest̄b.*  
 she.LOC pleasant.ADV 1SG.DAT be.PRS.3SG  
*prizvaně že byvši ei.*  
 summon.PTCP.PST.PASS.F.DAT.SG PTC be.PTCP.PST.F.DAT.SG she.DAT  
*pride k njemu*  
 come.AOR.3SG to he.DAT  
 ‘‘Summon her, so that I can enjoy myself with her today. From what  
 I’ve heard of her, I am much attracted by her’’. After being  
 summoned, she came to him.’ (Bdinski Sbornik, 7v: 11, 286377)

- (37) *k lět, čr̄n̄oriz̄b̄stvova s̄b nim̄b jako*  
 twenty year.GEN.PL live-monastically.AOR.3SG with he.INS like  
*agnica čstaa, i golubyca neskvr̄nnaa. i*  
 lamb.NOM pure.NOM and dove.NOM innocent.NOM and  
*končavšu se dvadeset̄nu lētu vr̄meni. i*  
 end.PTCP.PST.N.DAT.SG REFL twentieth.N.DAT year.DAT time.GEN and  
*neistov̄stvo s̄tvori dijavol̄b na nju i s̄t̄b*  
 fury.ACC make.AOR.3SG devil.NOM on she.ACC and snare.ACC  
*poleče da ulovit̄b ju.*  
 set-up.AOR.3SG to catch.PRS.3SG she.ACC  
 ‘For twenty years she lived monastically after him, like a chaste  
 lamb, and an innocent dove. When twenty years had elapsed, the  
 Devil turned his fury on her and set up nets to ensnare her’  
 (Bdinski Sbornik, 2v: 15, 286249)

- (38) *iměše blaženy s̄b brata po pl̄ti.*  
 have.IMP.F.3SG blessed.NOM this.NOM brother.ACC by flesh.DAT

<sup>35</sup> The information-structural notion of focus is intended here, along with Halliday (1967: 204) as ‘a message block which he [the speaker] wishes to be interpreted as informative. What is focal is “new” information; not in the sense that it cannot have been previously mentioned, although it is often the case that it has not been, but in the sense that the speaker presents it as not being recoverable from the preceding discourse’. In this sense, it overlaps with the notion of ‘rheme’ in other terminologies.

*imušta*                      *dvěstere*                      *edinočedju.*  
 have.PTCP.PRS.M.ACC.SG daughter.ACC only.ACC  
*umeršu*                      *že* *ocju*                      *jee.*                      *prěbys*                      *že*  
 die.PTCP.PST.M.DAT.SG PTC father.DAT she.GEN remain.AOR.3SG PTC  
*junotka*    *sir*  
 girl.NOM orphan.NOM  
 ‘The blessed one [Abraham] had a brother who had an only  
 daughter [Mary]. When the father [of Mary] died, the girl remained  
 orphan.’ (Bdinski Sbornik, 1v: 4, 286216)

In (36), the dative absolute *prizvane že byvoši ei* ‘after she was called’ picks up the lemma of the imperative *prizovi ju* ‘call her’. In (37), *20 let* ‘for 20 years’ is also focussed (arguably implying ‘for as long as 20 years’), whereas it then reappears as the subject of the dative absolute with a frame-setting function (*končavšu se dvadesetnu letu vremeni* ‘after 20 years had passed’). In (38), *ocju* ‘father’, the dative absolute’s subject, and *brata* ‘brother’, the object of the previous sentence, are actually the same referent, but the different lemma is due to a topic-change, from the referent’s brother (Abraham) to the referent’s daughter (Mary).

Dative absolutes in this dataset clearly function as topic-shifters, which was already highlighted in the previous two case studies as a function consistent with Bary & Haug’s (2011) FRAMES. Throughout the dataset, the construction was shown to systematically refer back to a preceding event as either concluded or reduced to the role of background for a new foregrounded event. This has either emerged from the predicates themselves, whereby a previous main predication appears in the following sentence(s) as the dative absolute’s lemma, or from the subject of the dative absolute referring back to event participants that were focussed in the previous propositions, but are changed to topics in the following discourse.

## 5 CONCLUSIONS

This paper explored some of the advantages and limitations of exploiting treebanks with different annotation depths when investigating syntactic phenomena whose distribution appears to be driven by discourse structure. As a case study, we tested the widespread intuition that Early Slavic dative absolutes are typically used as framing (stage-setting, backgrounding) devices. Deeply-annotated treebanks, despite small in size, proved very useful for establishing a solid blueprint to guide the analysis in much larger but shallowly-parsed corpora. Strategically-annotated treebanks, however, also proved to be useful tools. By means of morphological pre-processing alone it was pos-

sible to extract a number of potential occurrences of the relevant construction. This method could be replicated on a greater number of texts at once, in order to develop preliminary intuitions on a phenomenon before deciding whether they are worth further annotation. The last case study showed that strategically-annotated treebanks can certainly be used as a means of corroborating patterns emerging from fully annotated treebanks.

One of the main advantages of having syntactic-dependency annotation spanning beyond the relevant constructions lies in the possibility to compare potentially competing constructions by grammatical function. The properties of subjects, for instance, gave interesting insights into the functions of dative absolutes. Particularly for texts which contain an overall limited number of event participants, as is the case with the *Story of Abraham and Mary His Niece*, information-structural annotation would be particularly useful to be able to assess anaphoric phenomena systematically in larger stretches of discourse, rather than relying on a case-by-case approach when checking the relation between referents across sentences.

On an optimistic note, the transition from deeply-annotated treebanks to strategically-annotated ones, in decreasing depth of annotation, did not result in the need of drastically altering our approach. Thanks to its several levels of annotation, the patterns emerged from the OCS deeply-annotated treebank were robust enough to establish the core functions of the construction. This indicates that even on the basis of translations alone - which is necessarily the case for the earliest stages of Slavic - and of datasets of limited size, historical treebanks can be used to test hypotheses and provide valuable insights on the syntax of the relevant language, provided they are annotated across multiple levels of linguistic analysis. Deep annotation of small treebanks can therefore be useful to first test an hypothesis, before investing time in deep annotation of large corpora. Deeply-annotated treebanks can be exploited to make informed predictions on a given construction in new texts, using those as a relatively solid guideline to analyze its behaviour in larger, but shallowly annotated corpora. This is particularly encouraging for the study of the syntax of Early Slavic and, similarly, of other historical languages for which corpus-building automation is particularly difficult because of very high diachronic and diatopic variation.

## REFERENCES

- Asher, Nicholas & Alex Lascarides. 2003. *Logics of conversation*. Cambridge: Cambridge University Press.
- Asher, Nicholas, Laurent Prévot & Laure Vieu. 2007. Setting the background in discourse. *Discours* 1. 1–29.

- Bary, Corien & Dag T.T. Haug. 2011. Temporal anaphora across and inside sentences: The function of participles. *Semantics and Pragmatics* 4. 1–56.
- Brinton, Laurel J. 1996. *Pragmatic markers in English: Grammaticalization and discourse functions*. Berlin-New York: De Gruyter.
- Chafe, Wallace L. 1976. Givenness, contrastiveness, definiteness, subjects, topics and point of view. In Charles N. Li (ed.), *Subject and Topic*, 27–55. New York: Academic Press.
- Collins, Daniel E. 2004. Distance, subjecthood, and the early Slavic dative absolute. *Scando-Slavica* 50. 165–181.
- Collins, Daniel E. 2011. The pragmatics of “unruly” dative absolutes in Early Slavic. *Oslo Studies in Language* 3(3). 103–130.
- Corin, Andrew R. 1995. The dative absolute in Old Church Slavonic and Old East Slavic. *Die Welt der Slaven* 40. 251–284.
- Dalrymple, Mary, John J. Lowe & Louise Mycock. 2019. *The Oxford Reference Guide to Lexical Functional Grammar*. Oxford: Oxford University Press Oxford.
- Dujčev, Ivan. 1972. *Bdinski zbornik. Ghent Slavonic Ms 408, A.D. 1360*. London: Variorum Reprints facsimile edn.
- Eckhoff, Hanne M. 2018. Quantifying syntactic influence: Word order, possession and definiteness in Old Church Slavonic and Greek. In Björn Hansen, Jasmina Grković-Major & Barbara Sonnenhauser (eds.), *Diachronic Slavonic syntax: The interplay between internal development, language contact and metalinguistic factors*, 29–62. Berlin-Boston: De Gruyter.
- Eckhoff, Hanne M. & Aleksandrs Berdičevskis. 2015. Linguistics vs. digital editions: The Tromsø Old Russian and OCS treebank. *Scripta & E-Scripta* 14-15. 9–25.
- Eckhoff, Hanne M. & Dag T.T. Haug. 2015. Aspect and prefixation in Old Church Slavonic. *Diachronica* 32(2).
- Fabricius-Hansen, Cathrine & Dag T.T. Haug. 2012a. Co-eventive adjuncts: Main issues and clarifications. In Cathrine Fabricius-Hansen & Dag T.T. Haug (eds.), *Big events and small clauses*, 21–54. Berlin-Boston: De Gruyter.
- Fabricius-Hansen, Cathrine & Dag T.T. Haug. 2012b. Introduction. In Cathrine Fabricius-Hansen & Dag T.T. Haug (eds.), *Big events and small clauses*, 1–20. Berlin-Boston: De Gruyter.
- Fuller, Lois K. 2006. The “genitive absolute” in New Testament/Hellenistic Greek: A proposal for clearer understanding. *JGRChJ* 3. 142–167.
- Halliday, Michael A. K. 1967. Notes on transitivity and theme in English: Part 2. *Journal of Linguistics* 3(2). 199–244.
- Haug, Dag & Marius L. Jøhndal. 2008. Creating a parallel treebank of the Old Indo-European Bible translations. In Caroline Sporleder, Kiril

- Ribarov, Antal van den Bosch, Milena P. Dobрева, Matthew James Driscoll, Claire Grover, Piroska Lendvai, Anke Luedeling & Marco Pas-sarotti (eds.), *Proceedings of the Language Technology for Cultural Her-itage Data Workshop (LaTeCH 2008)*, Marrakech, Morocco, 1 June 2008, 27–34. ELRA. [http://www.lrec-conf.org/proceedings/lrec2008/workshops/W22\\_Proceedings.pdf#page=31](http://www.lrec-conf.org/proceedings/lrec2008/workshops/W22_Proceedings.pdf#page=31).
- Haug, Dag T.T. 2011. Backward control in Ancient Greek – subsumption or linearization? In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG11 Conference*, 278–298. Stanford: CSLI Publications.
- Haug, Dag T.T. 2012. Open verb-based adjuncts in New Testament Greek and the Latin of the Vulgate. In Cathrine Fabricius-Hansen & Dag T.T. Haug (eds.), *Big events and small clauses*, 287–321. Berlin-Boston: De Gruyter.
- Haug, Dag T.T., Hanne M. Eckhoff & Eirik Welo. 2014. The theoretical founda-tions of givenness annotation. In Kristin Bech & Kristine Gunn Eide (eds.), *Information structure and syntax in Germanic and Romance languages*, 17–52. Amsterdam: John Benjamins.
- Kamp, Hans & Uwe Reyle. 1993. *From discourse to logic: Introduction to mod-elltheoretic semantics of natural language, formal logic and Discourse Represen-tation Theory*. Dordrecht: Springer.
- Kamp, Hans, Josef Genabith & Uwe Reyle. 2011. Discourse Representation Theory. In Dov M. Gabbay & Franz Günthner (eds.), *Handbook of Philo-sophical Logic*, vol. 15, 125–394. Dordrecht: Springer.
- Kamphuis, Jaap. 2020. *Verbal aspect in Old Church Slavonic. A corpus-based ap-proach*. Leiden: Brill-Rodopi.
- Kortmann, Bernd. 1991. *Free adjuncts and absolutes in English: Problems of con-trol and interpretation*. London-New York: Routledge.
- Krifka, Manfred. 2007. Basic notions of information structure. *Interdisciplinary Studies on Information Structure* 6. 13–56.
- Kurešević, Marina. 2006. Apsolutni dativ u srpskoj srednjovekovnoj pis-menosti. *Zbornik Matice srpske za filologiju i lingvistiku* 49(1). 35–112.
- König, Ekkehard. 1991. *The meaning of focus particles: A comparative perspective*. London-New York: Routledge.
- Lindberg, Rebecka. 2013. *Definiteness in Old Church Slavonic: A study of how long and short form in adjectives reflect information status*. University of Oslo MA thesis.
- Sakharova, Anna. 2010. K voprosu o diskursivnykh funkciyakh pričastnykh konstrukcij v russkoj letopisi. *Russian Linguistics* 34. 87–11.
- Scharpe, Jan L. & Frans Vyncke. 1973. *Bdinski Zbornik. An Old-Slavonic menologium of women saints. (Ghent University Library Ms. 408, A.D. 1360)*. Brugge: De Tempel.

- Scherrer, Yves & Achim Rabus. 2019. Neural morphosyntactic tagging for Rusyn. *Natural Language Engineering* 25(5). 633–650.
- Scherrer, Yves, Achim Rabus & Susanne Mocken. 2018. New developments in tagging pre-modern orthodox Slavic texts. *Scripta & E-Scripta* 18. 9–33.
- Stern, Dieter. 2013. Izučaja istočniki Bdinskogo sbornika: Slovo ob Avraamii Kidunskom i ego plemjannice Marii. *Starobolgarska Literatura* 74–91.
- Stern, Dieter. 2015. Studien zum Bdinskij Sbornik: die slavische Erzählung des Abraham von Qidun und seiner Nichte Maria und ihre griechische Vorlage. *Cyrrillomethodianum* 20. 257–287.
- Stern, Dieter. 2016. Translations as text redactions: Observations on hagiographic translations (taken from the Bdinski Sbornik). *Slavica Ambrosiana* 6. 23–41.
- Stern, Dieter. 2018. Copying Greek into Slavonic? The Slavonic branch of the Greek tradition of the life of Abraham of Qidun. In Lara Sels, Jürgen Fuchsbauer, Vittorio Tomelleri & Ilse De Vos (eds.), *Editing medieval texts from a different angle: Slavonic and multilingual traditions* Orientalia Lovaniensia Analecta (Vol. 276), 225–250. Leuven, Paris, Bristol/CT: Peeters.
- Stump, Gregory. 1985. *The semantic variability of absolute constructions*. Dordrecht: Springer.
- Sæbø, Kjell J. 2012. Adverbial clauses. In Claudia Maienborn, Klaus Heusinger & Paul Portner (eds.), *Semantics: an international handbook of natural language meaning*, vol. 2, 1420–1441. Berlin-Boston: Mouton de Gruyter.
- Traugott, Elisabeth C. & Richard B. Dasher. 2002. *Regularity in semantic change*. Cambridge: Cambridge University Press.
- Worth, Dean S. 1994. The dative absolute in the Primary Chronicle: Some observations. *Harvard Ukrainian Studies* 18. 29–46.

Nilo Pedrazzini  
 University of Oxford & The Alan Turing Institute  
 St Hugh's College  
 St Margaret's Road  
 Oxford  
 OX2 6LE  
[nilo.pedrazzini@ling-phil.ox.ac.uk](mailto:nilo.pedrazzini@ling-phil.ox.ac.uk) & [npedrazzini@turing.ac.uk](mailto:npedrazzini@turing.ac.uk)  
[npedrazzini.github.io](https://github.com/npedrazzini)