A Diachronic Treebank of Russian Spanning More Than a Thousand Years

Aleksandrs Berdicevskis[1], Hanne Eckhoff[2]

[1]Språkbanken (The Swedish Language Bank), Universify of Gothenburg
[2]Faculty of Medieval and Modern Languages, University of Oxford
aleksandrs.berdicevskis@gu.se, hanne.eckhoff@mod-langs.ox.ac.uk

Abstract

We describe the Tromsø Old Russian and Old Church Slavonic Treebank (TOROT) that spans from the earliest Old Church Slavonic to modern Russian texts, covering more than a thousand years of continuous language history. We focus on the latest additions to the treebank, first of all, the modern subcorpus that was created by a high-quality conversion of the existing treebank of contemporary standard Russian (SynTagRus).

Keywords: Russian, Slavonic, Old Church Slavonic, treebank, diachrony, conversion

1. Introduction

The Tromsø Old Russian and OCS Treebank (TOROT, Eckhoff and Berdicevskis 2015) has been available in various releases since its beginnings in 2013, and its East Slavonic part now contains approximately 230K words.[1] This paper describes the TOROT 20200116 release,[2] which adds a conversion of the SynTagRus treebank. Former TOROT releases consisted of Old East Slavonic and Middle Russian texts from the 11th–19th century, covering a broad range of text types and genres such as chronicles, lives of saints, charters, birchbark letters, personal correspondence and tales. Now it also includes contemporary standard Russian. The new release thus turns it into a treebank covering every attested stage of Russian (contemporary standard Russian, Middle Russian, Old East Slavonic) and Old Church Slavonic, spanning more than a thousand years.

This paper is structured as follows: Section 2 discusses the TOROT treebank design. Section 3 is a description of SynTagRus. Section 4 describes the conversion, Section 5 concludes.

For detailed examples, we will be using tables with a format similar to simplified CONLL: id, Russian form, English gloss, head, relation. For SynTagRus relations, we will be using our translations of their original Russian names into English.

2. TOROT

2.1 The PROIEL dependency format

TOROT is a part of the PROIEL family of ancient Indo-European treebanks (Eckhoff et al. 2018) and is annotated using the PROIEL enhanced dependency format (Haug and Jøhndal 2008). The format was designed specifically for the needs of ancient Indo-European languages, which are characterised by rich inflection and relatively free word order.

The scheme differs from more classical approaches to dependency grammar, such as the Prague Dependency Treebank scheme (Hajič et al. 2018), in three main ways: use of secondary dependencies, limited use of empty nodes and a richer inventory of syntactic relation labels. Secondary dependencies are used to indicate external subjects, for example in control structures, and to indicate shared dependents, for example in structures with coordinated verbs. Empty nodes are allowed to give more information on elliptic structures and asyndetic coordination. The empty nodes are limited to verbs and conjunctions, the scheme does not allow empty nominal nodes. Finally, the range of relation labels is expanded to differentiate between types of arguments (OBJects, OBLiques, passive AGents) and between types of adnominal dependents (ATR for attributes, NARG for adnominal arguments, PART for adnominal partitives).

(1) *se slyšavše torci ubojašasja proběgoša*
'having heard this the Turks became afraid and ran away'
(Primary Chronicle, Laurentian manuscript, 163.4–5)

Example (1) and Figure 1 demonstrate several of these features: This Old East Slavonic sentence has an asyndetic coordination of the two predicates *ubojašasja* 'became afraid' and *proběgoša* 'ran away', both in the aorist 3rd person plural. This coordination is indicated by the empty conjunction node directly under the root, which is labeled PRED and has the two predicates as its daughters, also labeled PRED. The two coordinated predicates also share a subject, *torci* 'Turks', indicated by a secondary dependency labeled SUB. The first of the predicates has a conjunct participle daughter, *slyšavše* 'having heard', labeled XADV, a relation used for adverbial modifiers with external subjects. In this instance, the external subject is the same as that of the head verb, *torcy*. This is indicated by a secondary dependency labeled XSUB ('external subject'), which allows us to check if the conjunct participle has the expected gender, number and case agreement with the external subject. In this case it does (masculine nominative plural).

The enhanced dependency scheme thus allows for annotation that preserves a maximum of structural information without making undue assumptions about the phrase structure of ancient languages with limited and skewed attestation.

---

[1] Word count here is not the same as token count (which is more traditional in NLP), since word count does not include punctuation marks.

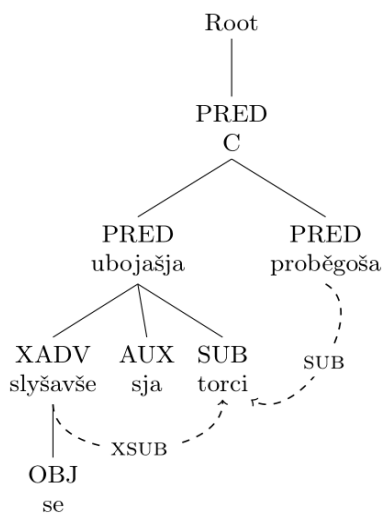[2] https://github.com/torottreebank/treebank-releases/releases/tag/20200116

Figure 1: Use of secondary dependencies and empty nodes in TOROT.

In a diachronic treebank choices must be made in order to capture syntactic change – structures in the earliest texts may differ considerably from corresponding structures in the latest texts, and it can be challenging to annotate texts from transitional periods. One way of seeing it is that there is a choice between a conservative approach (do not abandon the original structure until it is absolutely necessary) and a modernising approach (change the analysis of the structure at the earliest possible moment). While there are arguments in favour of both approaches, the TOROT syntactic annotation consistently sticks to the conservative principle. The same is the case for lemmatisation (maximally conservative spellings are chosen) and morphological analysis (morphological forms are taken at face value for as long as possible).

Figure 1 can again serve as an illustration. Over time the Old East Slavonic conjunct participles are reanalysed as non-inflecting gerunds, with a form which is in most cases identical with the old masculine nominative singular form of the participle. The example in Figure 1 thus serves as evidence that this has not happened yet: we observe agreement in the masculine nominative plural. We have adopted a maximally conservative annotation policy. Only in the SynTagRus conversion do we allow such examples to be taken as gerunds in the morphological annotation. In all pre-modern texts they are taken to be participles, and their morphology is taken at face value.

2.2 Workflow and text processing

The treebank is largely manually annotated, aided by automatic preprocessing (lemmatisation, statistical morphological tagging). Annotators work in an online annotation environment (for further description see Eckhoff et al. 2018: section 2) where they (1) adjust sentence division and tokenisation, (2) correct the

automatic lemmatisation and morphological annotation and (3) manually add syntactic annotation aided by rule-based guesses. All sentences are then manually inspected and corrected by a reviewer before they can be released.

The extent to which sentence division and tokenisation has to be adjusted depends on the source of the text. Whenever possible, already existing digitisations are used, but only if we deem them to be very faithful to the manuscript, with easily discernible editorial corrections, if any. For the Primary Chronicle (Figure 1), for example, we use the e-PVL.[3] If there is no such text available to us, we digitise texts ourselves, usually from manuscript facsimile. This means that in most cases our texts come with original mediaeval punctuation, which in general divides the texts into smaller units than the sentence. When texts are processed for import, they are split into sentences at full stops, colons and a few other punctuation marks. The annotators therefore very frequently adjust sentence boundaries as a part of the annotation workflow.

As for tokenisation, mediaeval Slavonic manuscripts are generally written in *scripta continua*, i.e. without word division. This means that any editor must make a decision on tokenisation, and they are often at odds with each other. When we digitise texts ourselves, we split the text into words according to the TOROT principles, for instance, the reflexive marker *sja* is always taken to be an independent word, separated from the verb with a whitespace. Many editions have a different policy, taking the verb and reflexive marker to be a single word whenever the reflexive marker appears postverbally. This is, for example, the policy in the e-PVL.

Period	Word count	
(Old) Church Slavonic	138,851	
Old East Slavonic	139,180	
Middle Russian	92,555	
Standard Contemporary Russian	in total	860,720
	full annotation	784,361
	morph. annotation (a draft tree present)	30,130
	morph. annotation (no tree present)	46,229

Table 1: Word count by language/period in the 20200116 release. See section 4.3 for more details about the modern subcorpus and the sentences without syntactic trees.

In our text preprocessing routine we split the text into words on the basis of whitespaces. When we annotate texts with word division policies different from the TOROT ones, the annotators must therefore often retokenise. This is illustrated in Figure 1: In the running text from the e-PVL, *ubojašasja* '(they) became afraid' is a single token. In our dependency analysis, however, *ubojaša* and *sja* are separate tokens and separate nodes in the dependency analysis.

Table 1 contains word counts by period for the 20200116 release.

2.3   Release formats

TOROT is released on Github in native XML and CONLL-X format. From January 2020 these releases include the SynTagRus conversion. A subset (currently about 150K words) of the East Slavonic part of TOROT has been available as a converted Universal Dependencies treebank[4] since May 2019. The treebank can also be browsed in the Syntacticus treebank facility.[5]

3.   SynTagRus

SynTagRus (Boguslavsky 2014) is the largest existing treebank of contemporary standard Russian. The treebank was annotated using ETAP-3 (Apresjan et al. 2003), a rule-based processing system, all the sentences were manually checked by human annotators. SynTagRus is being constantly expanded, its latest version contains around 70K sentences and more than 1M words. In TOROT, an older (2014) and smaller (860K words originally) version is presented. We corrected some minor annotation errors in the original treebank manually before conversion.

The SynTagRus annotation scheme is based on the Meaning–Text model (Mel'čuk 1995), which makes it strongly dependent on lexical semantics and gives it a highly granular argument structure representation based on ranked valencies.

Several conversions of SynTagRus into various other formats have been reported (Penn style: Luu, Malamud and Xue 2016; Stanford style: Lipenkova and Souček 2014; HPSG derivation trees: Avgustinova and Zhang 2010; Prague style: Mareček and Kljueva 2009). Most importantly, it has been converted to Universal Dependencies (UD) and is now part of the UD collection (Droganova, Lyashevskaya and Zeman 2018).

Even though TOROT is now also available in UD format, the present conversion is not redundant. UD is less informative than both formats (which means information from both treebanks is being lost). Furthermore, TOROT is converted to fit the conversion of the other treebanks from the PROIEL family, while SynTagRus is not. There is thus a considerable annotation "dialect" difference.

4.   Conversion

4.1   Morphological conversion

By *morphological annotation* we mean here part of speech (POS), lemma and morphological features. Most of the morphological conversion was straightforward, with the exception of cases when SynTagRus annotation is less granular than TOROT, which are summarised in Table 2.

We recovered missing information mostly by using handmade lexical lists (e.g. of those SynTagRus nouns that should become personal pronouns in TOROT). In some cases, when detailed syntactic information is necessary to determine the TOROT POS (e.g. relative adverb), morphological annotation was corrected during the syntactic conversion.

POS	TOROT	SynTagRus
Pronouns	seven different POS	absent (labelled as either nouns or adjectives)
Numerals	cardinal and ordinal	single POS
Subjunctions	different from conjunctions	same POS as conjunctions
Nouns	common and proper	single POS
Verbs	three voices (active, passive, middle)	two voices (active, passive)
Adverbs	interrogative, relative and "usual" adverbs	adverbs and particles

Table 2: Major morphological discrepancies between TOROT and SynTagRus

Proper nouns were identified through capitalisation. If a noun appeared sentence-initially, the decision was made based on how often it is capitalised in other positions.

All active-voice verbs ending in the reflexive suffix *-sja* were labelled as middle-voice, which probably results in a slight overgeneration of middle voice.

In addition, SynTagRus treats verbal aspect as a grammatical category, while TOROT considers an imperfective verb and its perfective counterpart to be two different lemmas. This was solved by making information about aspect (*.pf* or *.ipf*) part of the lemma, e.g. VZJAT'.pf 'take (perfective)' vs. BRAT'.ipf 'take (imperfective)'.

In rare cases when necessary morphological information about certain categories could not be retrieved, the respective categories (e.g. gender) were marked as unspecified.

4.2   Syntactic conversion

SynTagRus has 67 syntactic relations, while TOROT has only 24 (of which two are used only to label secondary dependencies). This means that the conversion in the opposite direction would have been much more complicated.

Below we describe the most prominent non-isomorphisms between the two formats.

Empty nodes.

In SynTagRus, the so-called "phantom" nodes are used to represent "syntactically disjointed" constructions, which usually means gapping. Phantoms can be of any POS, cf. example (2) as represented in Table 3 and Table 4.

(2) *zavist' uznavaema kak zavist', žalost' — kak žalost'*
'envy is recognisable as envy, pity as pity'
(*Nauka i žizn'* 11, 2015).

---

[4] https://github.com/UniversalDependencies/UD_Old_Russian-TOROT

[5] syntacticus.org

Node	Form	Gloss	Head	Rel
1	zavist'	envy	2	predic
2	uznavaema	recognisable	0	root
3	kak	as	2	compar
4	zavist'	envy	3	compar-conj
5	žalost'	pity	6	predic
6	FANTOM	(recognisable)	2	sent-coord
7	kak	as	6	compar
8	žalost'	pity	7	compar-conj

Table 3: Example (2) in the SynTagRus format.

Node	Form	Gloss	Head	Rel
1	zavist'	envy	2	sub
	Secondary dependency:		3	xsub
2		VERB	6	pred
3	uznavaema	recognisable	2	xobj
4	kak	as	2	adv
5	zavist'	envy	4	sub
6		CONJ	0	pred
7	žalost'	pity	8	sub
	Secondary dependency:		9	xsub
8		VERB	6	pred
9	kak	as	8	xobj
10	žalost'	pity	9	sub

Table 4: Example (2) in the TOROT format.

In TOROT, the empty nodes are limited to verbs and conjunctions. Every sentence has to be headed by a verb, which means an empty node must always be inserted when a verb is omitted. Verb ellipsis is a frequent phenomenon, especially in contemporary standard Russian where the copular 'be' is almost never used in present tense. In SynTagRus, however, there are no such restrictions, and no empty node will be inserted.

The conversion is performed as follows. SynTagRus phantoms that are verbs are straightforwardly converted to TOROT empty nodes. SynTagRus phantoms that are not verbs (e.g. 'envy' in example (2)) are deleted, and their dependent is promoted into their place, which is the TOROT policy for dealing with ellipsis. If there are several dependents, the first one (in linear order) is promoted, which may potentially yield errors.

If a sentence is not headed by a verb or if relations that are restricted to verbs in TOROT have a different head in SynTagRus, an empty verb is inserted and the necessary reattachments are made.

The requirement to always have a verbal head in a clause, which works reasonably well for older texts, sometimes makes the trees in contemporary standard Russian unnecessary complex, e.g. in case of elliptical sentences like 'fine' or parenthetical constructions like 'of course'. However, they are still annotated consistently and can be easily queried.

Empty conjunctions are straightforwardly inserted in cases of asyndetic coordination, e.g. node 6 in Table 3 (see more in **Coordination**).

Coordination. According to Popel et al.'s (2013) classification, SynTagRus and TOROT approaches to coordination are variants of resp. Moscow-style and Prague-style. In SynTagRus, the first conjunct is the head, the conjuction (if present) is its dependent (via the COORD relation or SENT-COORD for sentential coordination), the second conjunct is a dependent on a conjunction (via the COORD-CONJ relation), see Figure 2.
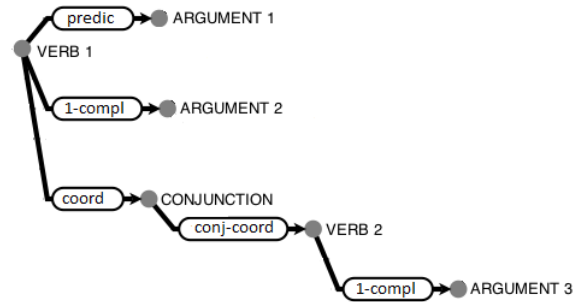

Figure 2: Coordination in SynTagRus (reproduced with permission from Berdicevskis and Eckhoff 2015)

In TOROT, the conjunction is the head (a null conjunction is inserted in case of asyndetic coordination) and all the conjuncts are its dependents, no special relation is used, see Figure 3.
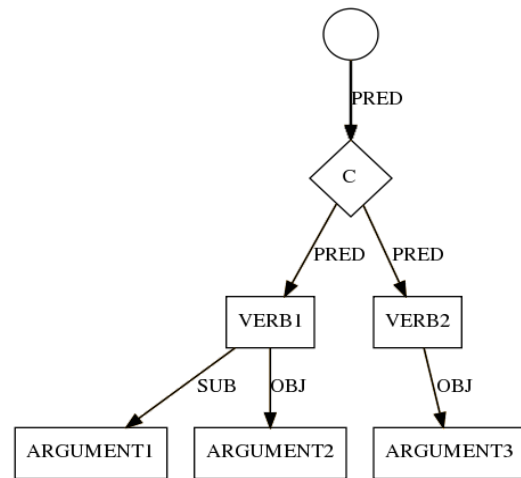

Figure 3: Coordination in TOROT (reproduced with permission from Berdicevskis and Eckhoff 2015)

The SynTagRus approach enables simpler syntactic queries, whereas the TOROT approach makes it possible to render complicated stacked structures better. In addition, TOROT uses secondary dependencies to indicate predicate identity (in case of verb ellipsis) and shared dependents (see also Figure 1).

The conversion algorithm handles coordination well, apart from rare cases of several entangled coordinated structures. Berdicevskis & Eckhoff (2015) devised a method for inserting secondary dependencies for shared verb arguments, which achieves near-ceiling performance for subjects (which are often shared), but not for other

arguments. The method, however, is not yet implemented in the conversion.

Secondary dependencies. As has already been mentioned, secondary dependencies are used in TOROT not only in coordinated structures, but also to indicate external subjects, for example in control and raising structures (see e.g. relations from nodes 3 and 9 to resp. 1 and 7 in Table 3). External subjects are not marked in any way in SynTagRus, but we devised a simple heuristic, which chooses either the subject or the object of the head verb depending on the syntactic structure of the sentence, and which performs very well (see Section 4.4).

Lexical semantics. SynTagRus heavily relies on lexico-semantic properties of words (an extensive handmade dictionary of Russian was used by the rule-based parser during the automatic pre-annotation of the treebank).

For instance, the verb argument representation is highly granular, but does not include much information on the morphosyntactic category of the argument itself. While the 1-COMPL ('first complement') relation is typically used for direct objects, with the verb 'to live' it is used for PPs and adverbs denoting locations ('to live in Norway', 'to live here' etc.). The relative rank of an argument in a valency frame, not its form, decides what relation label it gets.

In this case, we largely relied on nominal case to map the SynTagRus relations onto the TOROT ones (SUBJ, OBJ, OBL and ADV).

Another example concerns adnominal arguments, i.e. cases when a noun function as an argument of another noun (often, but not necessarily deverbal), e.g. *the catching of fish*. Both schemes recognise adnominal arguments, SynTagRus uses a granular valency-based representation, while TOROT uses a single NARG relation. The problem is that some of the nouns that can have arguments in SynTagRus are not allowed that in TOROT (an attributive relation has to be used instead). To filter them, we manually created an extensive lexical list.

4.3 Metadata

The original SynTagRus files usually contain information about the title of the text, its author and publication place, which is preserved after the conversion (resp. "title", "author" and "printed-text-place"). Internal metadata (the annotator, the editor, the date of adding the text) are not preserved. Unfortunately, the original metadata do not have a "date" or "year" field to indicate when the text was written. For periodical articles this information is sometimes present in the "printed-text-place" field or in the "title" field, but not in a consistent format.

In addition, the converter provides detailed information about possible and probable conversion errors. In TOROT, every sentence has a "status" field, which in the converted files is normally set to "annotated". If, however, the converter estimates the probability of a major error to be very high, it changes the status to "unannotated". Out of 59240 sentences, 3584 (76359 words) get this status. Of these, for 2194 sentences (46229 words) the resulting

syntactic structure cannot pass the TOROT validator due to some formal error, for these sentences the syntactic structure is completely erased. In other words, out of 3584 "unannotated" sentences 1390 (30130 words) actually do have a syntactic annotation (which is likely to contain at least one major error), while 2194 do not. The morphological annotation is always present.

The converter also outputs several lists of other potential errors (focusing on the errors in the verb argument structure), providing some guesses about the type and the probability of the error.

4.4 Evaluation

To evaluate the conversion quality, we drew a random sample of 100 sentences with the "annotated" status and checked the annotation manually. In the cases when several formal analyses can be argued acceptable, we tried to choose the one that is most compatible with the annotation in the other TOROT parts. The results are summarised in Table 5.

Metric	Score (%)
Labelled attachment score (LAS)	accuracy 93.0
Unlabelled attachment score (UAS)	accuracy 95.7
POS	accuracy 98.4
Morphology	accuracy 98.6
Secondary dependencies	precision 91.3 recall 97.7
Empty token insertion	accuracy 99.1

Table 5: Conversion quality.

When calculating the accuracy of morphological annotation, tags like "unspecified" were not considered errors. When evaluating the insertion of secondary dependencies, we took into account only those edges that we tried to insert, i.e. those to the external subjects in cases of control and raising, but not the predicate-identity and the shared-dependent edges.

Overall, the annotation quality is high. It is only marginally worse than that by human experts and probably better than what is currently achievable by automatic parsing from scratch. The state-of-the-art result of parsing SynTagRus in UD is LAS of 92.5% (Zeman et al. 2018), but the TOROT scheme is likely to yield worse results, not to mention that prior to the conversion, there have been no training data for contemporary standard Russian in this format.

5. Conclusions

We present a new release of TOROT, a diachronic treebank that now spans from Old Church Slavonic (250K words) to Old East Slavonic and Middle Russian (230K words) and contemporary standard Russian (780K words). The treebank is freely available for download and accessible for searching and browsing through online interfaces.

TOROT is already being used for diachronic studies on Slavonic morphology and syntax (see, for instance, Eckhoff 2018, Zanchi and Naccarato 2016, Mishina 2016; Berdicevskis 2015). Hopefully it will become even more useful with the addition of the modern subcorpus.

Given its size and annotation quality, the treebank can also be used to train machine-learning models for automatic annotation of both modern and historical Slavonic texts.

Acknowledgements

6. Bibliographical References

Apresjan, Ju., Boguslavsky, I., Iomdin, L., Lazursky, A., Sannikov, V., Sizov V, and Tsinman, L. ETAP-3 linguistic processor: A full-fledged NLP implementation of the MTT. In *Proceedings of the First International Conference on Meaning-Text Theory*, 279–288.

Avgustinova, T., and Zhang, Y. (2010). Conversion of a Russian dependency treebank into HPSG derivations. In *Ninth International Workshop on Treebanks and Linguistic Theories*, pp. 7–18.

Berdicevskis, A. (2015). Estimating grammeme redundancy by measuring their importance for syntactic parser performance. In *Proceedings of the Sixth Workshop on Cognitive Aspects of Computational Language Learning*, pp. 65–73.

Berdicevskis, A., and Eckhoff, H. 2015. Automatic identification of shared arguments in verbal coordinations. In *Computational linguistics and intellectual technologies. Papers from the annual international conference "Dialogue"*, 14, pp. 33–43.

Boguslavsky, I. (2014). SynTagRus–a Deeply Annotated Corpus of Russian. In P. Bluemental, I. Novakova & D. Siepmann (Eds.) *Les émotions dans le discours-Emotions in Discourse*, 367–380. New York: Peter Lang.

Droganova, K., Lyashevskaya, O., and Zeman, D. (2018). Data conversion and consistency of monolingual corpora: Russian UD treebanks. In *Proceedings of the 17th International Workshop on Treebanks and Linguistic Theories (TLT 2018)*, pp. 52–65. Linköping University Electronic Press.

Eckhoff, H. M. (2018). A corpus approach to the history of Russian *po* delimitatives. *Diachronica* 35(3): 338–366.

Eckhoff, H. and Berdičevskis, B. (2015). Linguistics vs. digital editions: The Tromsø Old Russian and OCS Treebank. *Scripta & e-Scripta* 14–15.

Eckhoff, H. M., Bech, K., Bouma, G., Eide, K., Haug, D., Haugen, O. E. and Jøhndal, M. (2018). The PROIEL treebank family: a standard for early attestations of Indo-European languages. *Language Resources and Evaluation* 52:1

Hajič, J., Bejček, E., Bémová, A., Buráňová, E., Hajičová, E., Havelka, J., Homola, P., Kárník, J., Kettnerová, V., Klyueva, N., Kolářová, V., Kučová, L., Lopatková, M., Mikulová, M., Mírovský, J., Nedoluzhko, A., Pajas, P., Panevová, J., Poláková, L., Rysová, M., Sgall, P., Spoustová, J., Straňák, P., Synková, P., Ševčíková, M., Štěpánek, J., Urešová, Z., Vidová Hladká, B., Zeman, D., Zikánová, Š. and Žabokrtský, Z. (2018). *Prague Dependency Treebank 3.5.* Institute of Formal and Applied Linguistics, LINDAT/CLARIN, Charles University, LINDAT/CLARIN PID: http://hdl.handle.net/11234/1-2621.

Haug, D. T. T. and Jøhndal, M. L. (2008). Creating a Parallel Treebank of the Old Indo-European Bible Translations. In C. Sporleder & K. Ribarov (Eds.), *Proceedings of the Second Workshop on Language Technology for Cultural Heritage Data* (LaTeCH 2008), pp. 27–34.

Lipenkova, J., and Souček, M. (2014). Converting Russian dependency treebank to Stanford typed dependencies representation. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, volume 2: Short Papers, pp. 143-147.

Luu, A., Malamud, S. A., and Xue, N. (2016). Converting SynTagRus dependency treebank into Penn treebank style. In *Proceedings of the 10th Linguistic Annotation Workshop held in conjunction with ACL 2016 (LAW-X 2016)*, pp. 16–21.

Mareček, D., and Kljueva, N. (2009). Converting Russian Treebank SynTagRus into Praguian PDT Style. In *Proceedings of the Workshop on Multilingual Resources, Technologies and Evaluation for Central and Eastern European Languages*. Association for Computational Linguistics, pp. 26–31.

Mel'čuk, I. (1988). *Dependency Syntax: Theory and Practice*. State University of New York Press.

Mishina, E. (2016). Ob osobyh upotreblenijah form aorista i dejstvitel'nogo pričastija prošedšego vremeni v drevnerusskih letopisjah. In Moldovan, A. (Ed.) *Materialy mezhdunarodnoj naučnoj konferencii "Grammatičeskie processy i sistemy v sinhronii i diahronii"*, pp. 420–439. Russian Academy of Sciences.

Popel, M., Mareček, D., Štěpánek, J., Zeman, D., and Žabokrtský, Z. (2013). Coordination structures in dependency treebanks. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers), pp. 517–527.

Zanchi, C., and Naccarato, C. (2016). Multiple prefixation in Old Church Slavonic and Old Russian. In Konošenko, M., Ljutikova, E., & Zimmerling, A. *Tipologija morfosintaksičeskih parametrov,* pp. 359–390.

Zeman, D., Hajič, J., Popel, M., Potthast, M., Straka, M., Ginter, F., Nivre, J., Petrov, S. 2018. CoNLL 2018 shared task: multilingual parsing from raw text to universal dependencies. In Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies 2018, pp. 1–21.