

Mariana González · Silvia Susana Reyes ·
Andrea Rodrigo · Max Silberztein (Eds.)

Formalizing Natural Languages: Applications to Natural Language Processing and Digital Humanities

16th International Conference, NooJ 2022
Rosario, Argentina, June 14–16, 2022
Revised Selected Papers

Analyzing Political Discourse: Finding the Frames for Guilt and Responsibility	127
<i>Krešimir Šojat and Kristina Kocijan</i>	
Creation of Parallel Medical and Social Domains Corpora for the Machine Translation and Speech Synthesis Systems	139
<i>Mikita Suprunchuk, Nastassia Yarash, Yuras Hetsevich, Valery Varanovich, Siarhey Gaidurau, Yauheniya Zianouka, and Palina Sakava</i>	
Creation of a Legal Domain Corpus for the Belarusian Module in NooJ: Texts, Dictionaries, Grammars	151
<i>Valery Varanovich, Mikita Suprunchuk, Yauheniya Zianouka, Tsimafei Prakapenka, Anna Dolgova, and Yuras Hetsevich</i>	
Natural Language Processing Applications	
Construction of an Educational Game “CONJ_NOOJ”	165
<i>Héla Fehri and Nizar Jarray</i>	
Annotation of Procedural Questions in Standard Arabic Using Syntactic Grammars	178
<i>Essia Bessaies, Slim Mesfar, and Henda Ben Ghzela</i>	
Integrated NooJ Environment for Arabic Linguistic Disambiguation Improvement Using MWEs	189
<i>Dhekra Najar, Slim Mesfar, and Henda Ben Ghezela</i>	
The Digital Text Workshop Cloud, New Solutions for Super Calculation Environments	202
<i>Ilaria Veronesi, Rita Bucciarelli, Francesco Saverio Tortoriello, Andrea Rodrigo, Marianna Greco, Colomba La Ragione, and Javier Julian Enriquez</i>	
Author Index	215

Creation of a Legal Domain Corpus for the Belarusian Module in NooJ: Texts, Dictionaries, Grammars

Valery Varanovich^{1(✉)}, Mikita Suprunchuk², Yauheniya Zianouka³,
Tsimafei Prakapenka³, Anna Dolgova², and Yuras Hetsevich³

¹ Belarusian State University, Nezavisimosti Av. 4, Minsk, Belarus
ssrlab221@gmail.com

² Minsk State Linguistic University, Zakharov Str. 21, Minsk, Belarus

³ United Institute of Informatics Problems, Surhanava Str. 6, Minsk, Belarus
yuras.hetsevich@newman.bas-net.by

Abstract. This paper represents the procedure of creating a legal domain corpus in NooJ. It illustrates a primary analysis of the corpus including automatic text processing, emphasizing its main linguistic peculiarities, and analyzing prosodic characteristics of legal texts which is necessary for syntagmatic delimitation. It also depicts the compilation composition of three dictionaries in NooJ format: general, legal, and Belarusian-English-Russian dictionaries.

Keywords: Automatic language processing · Corpus of a legal domain · Dictionary · Law code · Syntactic grammar · Belarusian · Russian · English

1 Introduction

The current language situation in the Republic of Belarus is characterized primarily as state bilingualism. But despite the state bilingualism, the vast majority of legislative documents are implemented only in Russian. Thus, of the 26 codes of the Republic of Belarus, the texts of which are presented on the National Legal Internet Portal pravo.by 25 are officially adopted in Russian and only one in Belarusian.

To handle the question of translating legislative documents into Belarusian, Speech Synthesis and Recognition Laboratory of United Institute of Informatics Problems of National Academy of Sciences of Belarus in cooperation with specialists from the Faculty of Social and Cultural Communications of Belarusian State University have translated all codes of the Republic of Belarus into the Belarusian language using automatic services of corpus.by [1].

The authors of the article have set several tasks to solve this problem. The first one is to collect all legislative codes of the Republic of Belarus in the Belarusian language for creating a unified text corpus. There are some examples of such corpora in the

Y. Hetsevich—The authors would like to thank Zmicer Dzenisiuk for his help during the preparation of the article.

world [2–4]. For this, it is very important to perform a primary analysis of the legal domain corpus: to find out the main linguistic peculiarities of this kind of corpus in comparison with the Belarusian literary corpus. Secondly, different types of dictionaries (Belarusian-Russian, Belarusian-English, Belarusian-English-Russian dictionaries) are planned to be composed. This question is very topical for the Belarusian language since there are very few translated Belarusian-foreign and foreign-Belarusian dictionaries of legal terminology (now we are aware of six dictionaries that are different in their merits and significance for the ordering and development of Belarusian legal terminology). The last but not the least task is to develop special morphological and syntactic grammars for further prosodic analysis of legal texts [5, 6]. Automatic syntagmatic delimitation is still not solved for the Belarusian language, either. Therefore, developed NooJ grammars [7] will assist the process of creating a system of prosodic marks (including punctuation and intonation marks) and further automatic segmentation of Belarusian texts of the legal domain [8].

2 Compilation of a Legal Domain Corpus

To create a unified corpus of a legal domain, the specialists have translated all law codes from the Russian language into Belarusian one by one almost independently (Fig. 1).

Назва тэксту і крыніца	Пераклады тэксту	Спеўнікі і дадатковыя натрыпты
<u>Банковскі кодэкс Рэспублікі Беларусь</u>	<u>БАНКАЎСКІ КОДЭКС РЭСПУБЛІКІ БЕЛАРУСЬ</u> (апрацаваны метадыкай вычыты неафіцыйны пераклад тэксту)	
<u>Бюджэтны кодэкс Рэспублікі Беларусь</u>	<u>БЮДЖЭТНЫ КОДЭКС РЭСПУБЛІКІ БЕЛАРУСЬ</u> (апрацаваны метадыкай вычыты неафіцыйны пераклад тэксту)	<u>Спеўнікі зямлі</u>
<u>Военны кодэкс Рэспублікі Беларусь</u>	<u>ВОЕННЫ КОДЭКС РЭСПУБЛІКІ БЕЛАРУСЬ</u> (апрацаваны метадыкай вычыты неафіцыйны пераклад тэксту)	
<u>Воздушны кодэкс Рэспублікі Беларусь</u>	<u>ПАВЕТРАНЫ КОДЭКС РЭСПУБЛІКІ БЕЛАРУСЬ</u> (апрацаваны метадыкай вычыты неафіцыйны пераклад тэксту)	

Fig. 1. Law codes in Russian and their translation into Belarusian on the page of the Speech Synthesis and Recognition Laboratory (ssrlab.by).

The translation of each law code undergoes the following stages:

1. The usage of a machine translation system.
2. Proofreading of the obtained translation by the software presented on the corpus.by platform includes:

- spelling checking,
- analysis of statistical and supplemental information about symbols,
- analysis of statistics on the use of arbitrary symbolic sequences in an electronic text (counting the frequency of word forms),
- search and correction of errors in the spelling of the Belarusian letters “У” and “Ў”,
- recognition and selection of homographs in the text.

Manual text proofreading and correction of errors was accomplished sequentially by three editors: the first two were qualified linguists, the third was a professional lawyer. In the process of proofreading, a dictionary of the most frequent (regular) substitutions of words and phrases was created. Such a contextual dictionary is also planned to be used when training a machine translation system to improve the quality of the translated text.

Figure 2 below shows an example of the Civil Code of the Republic of Belarus in the Belarusian language obtained by semi-automatic translation from Russian.

ГЛАВА 1 ГРАМАДЗЯНСКАЕ ЗАКАНАДАЎСТВА

Артыкул 1. Адносіны, якія рэгулююцца грамадзянскім заканадаўствам

1. Грамадзянскае заканадаўства вызначае прававое становішча ўдзельнікаў грамадзянскага абароту, падставы ўзнікнення і парадак ажыццяўлення права ўласнасці і іншых рэчавых правоў, правоў на вынікі інтэлектуальнай дзейнасці, рэгулюе адносіны паміж асобамі, якія ажыццяўляюць прадпрымальніцкую дзейнасць, ці з іх удзелам, дагаворныя і іншыя абавязацельствы, а таксама іншыя маёмасныя і звязаныя з імі асабістыя немаёмасныя адносіны.

Прадпрымальніцкая дзейнасць – гэта самастойная дзейнасць юрыдычных і фізічных асоб, якая ажыццяўляецца імі ў грамадзянскім абарачэнні ад свайго імя, на сваю рызыку і пад сваю маёмасную адказнасць і накіраваная на сістэматычнае атрыманне прыбытку ад карыстання маёмасцю, продажу рэчаў, вырабленых, перапрацаваных ці набытых указанымі асобамі для продажу, а таксама ад выканання работ ці аказання паслуг, калі гэтыя работы ці паслугі прызначаюцца для рэалізацыі іншым асобам і не выкарыстоўваюцца для ўласнага ўжытку.

Fig. 2. A fragment of the civil code after semi-automatic translation.

After translating 17 law codes into Belarusian, a group of specialists compiled a unified corpus in NooJ format to create a dictionary of legal terms and expressions. Figure 3 shows the Belarusian corpus of a legal domain that consists of 17 law codes in NooJ format [9]. The total number of word tokens is 1,043,018.

File Name	Size	Last Modified
АБ ЭРКОНЕ І СІМІ	2552536	25.05.2021
БАНКАРСКИ КОДЭС РЭСПУБЛІКІ БЕЛАРУСЬ	2637397	25.05.2021
ВАДЖЭІТНЫ КОДЭС РЭСПУБЛІКІ БЕЛАРУСЬ	2335146	25.05.2021
ВОДНЫ КОДЭС РЭСПУБЛІКІ БЕЛАРУСЬ	770998	25.05.2021
ВЫДАРЧЫ КОДЭС РЭСПУБЛІКІ БЕЛАРУСЬ	2713848	25.05.2021
ГРАМАДЗІНСКІ КОДЭС РЭСПУБЛІКІ БЕЛАРУСЬ	7562913	25.05.2021
ЖЫТЛІНЫ КОДЭС РЭСПУБЛІКІ БЕЛАРУСЬ	3826297	25.05.2021
КОДЭС ГАНДЛІВАЎ А МАРАГІПАРСТВА РЭСПУБЛІКІ БЕЛАРУСЬ	1494740	25.05.2021
КОДЭС РЭСПУБЛІКІ БЕЛАРУСЬ АБ ЗІМНІ	1562630	25.05.2021
КОДЭС РЭСПУБЛІКІ БЕЛАРУСЬ АБ МЕТРАХ	1426401	25.05.2021
КОДЭС РЭСПУБЛІКІ БЕЛАРУСЬ АБ СУДОВЫМ ПАДЗЕ І СТАТУСЕ СУДДЗЬЯР	1374633	25.05.2021
КОДЭС УНІТРАНАЎ А ВОДНАГА ТРАНСПОРТУ РЭСПУБЛІКІ БЕЛАРУСЬ	1379173	25.05.2021
ЛІСНЫ КОДЭС РЭСПУБЛІКІ БЕЛАРУСЬ	1461709	25.05.2021
ПАВЕТНЫ КОДЭС РЭСПУБЛІКІ БЕЛАРУСЬ	1062929	25.05.2021
ПАДАТКОВЫ КОДЭС РЭСПУБЛІКІ БЕЛАРУСЬ	2728077	25.05.2021
ПРАЦЮНЫ КОДЭС РЭСПУБЛІКІ БЕЛАРУСЬ	2871299	25.05.2021
ПРАДСУДЛІНЫ ВЫДАРЧЫ КОДЭС РЭСПУБЛІКІ БЕЛАРУСЬ АБ АДМІНІСТРАЦЫЙНЫХ ПРАВАПАРУШЭННЯХ	2206412	25.05.2021

Fig. 3. The legal corpus in NooJ interface.

Characteristics of the corpus are as follows: about 1 million tokens; 731,584 word forms; 64,783 text units delimited by “\n”; 47,219 different annotations; 5,947,419 characters (136 diff); 4,911,529 letters (97 diff), 94,801 digits (10 diff), 724,456 blanks (3 diff), 216,633 other delimiters (26 diff).

The main linguistic peculiarities of the legal corpus are:

- There are more complex sentences than simple ones. Documents should be very clear and precise, that’s why reasons, explanations, and repetitions are often included in law articles. Some words are specific for this theme and should be added to the dictionary of NooJ [9].
- The noun is the most frequent part of speech.
- Some words are specific for this theme.
- In the Belarusian language in comparison with the Russian one, the number of words in texts is more but the number of characters is less: for example, in the Civil Code in Russian there are 129,348 word usages, 1,043,572 characters (excluding spaces), and in the Belarusian translation - 132,414 word usages, 1,015,631 characters. So, the average word length in the Russian text is 8.06 characters, in the Belarusian - 7.67 characters.
- Some grammatical forms are not recommended for usage in other styles but could be admitted in the official style, e. g. present participle (passive and active): *суд вырашае вынікаючыя*: РТСР.PRES.АКТ з гэтага пытанні ‘the court decides issues that result from it’ (Civil Code); *прадаставіць вызваляемае*: РТСР.PRES.PAS жылое памяшканне ‘to give a living space which is being cleared’ (Housing Code).

The results of primary corpus processing show the next data:

- Unknown words, nearly 2,660: *адклікальнага, адпаламу;*
- Unique word forms, nearly 365 words: *бенефіцыярны, сублізінг;*
- Abbreviations, nearly 20: *АН, ФСК*
- Spelling errors, nearly 50: *стацілі, праналежнасць;*

3 Creating Medical and Social Trilingual Dictionaries in NooJ

Text corpora are a valuable knowledge source about language peculiarities. They allow us to monitor facts and tendencies that are difficult to notice by common speaking or reading. Lexical and grammatical elements of the Belarusian NooJ module were used for the lexical and grammar marking of the law corpus which was created by the specialists of the laboratory. In this case, a number of words was not recognized. In the result of their detailed analysis a part of words was included in the general dictionary of the Belarusian NooJ module as they are considerably widespread: e.g. *ан'яненне* 'intoxication, drunkenness', *аўдыявізуальны* 'audiovisual', *гідрацыкл* 'hydrocycle', etc. Unrecognized rare words were collected into a specialized dictionary. It includes law terms *апекавальнік* 'guardian, caretaker', *аднісанне* 'refusal', *дыспаша* 'claims adjustment', *контрстайіны* 'demurrage'; other sphere terms *авердрафт* 'overdraft', *абвалоўка* 'mound', *дыспрозій* 'dysprosium', *сунесь* 'sandy loam' (according to the law codes themes); regular derivatives *сухавяршышны* 'stag-headed', *перапрамонтны* 'pre-repair', *непражыванне* 'non-residence (they are easily formed by the general word-building rules, so it is unreasonable to include them into the general dictionary); borrowings that are only being adapted into the Belarusian language *суббербоўт* 'sub-bareboat', *ф'ючарсны* 'futures', *эквайрынг* 'acquiring', etc.

After manual analysis of all unknown words, two groups of vocabulary were composed: words of legal terminology and general vocabulary. As the Belarusian NooJ module already had the general vocabulary "general_be.dic" it was decided to create an additional dictionary for general unknown words (nearly 150 word forms in initial form) and a law dictionary (nearly 200 forms).

The principle of creating these dictionaries is the same. All unknown words were originally processed by Word Paradigm Generator. This is an online service that generates the paradigm of a word. At the service input, a word or its word form is given. The service searches for a paradigm in the dictionary. If a ready-made paradigm is absent, the user receives a generated paradigm of the entered word or paradigms created on the basis of words similar in spelling. The resulting paradigms have an accent, indicated by the "+" symbol, and special tags separated from the word by an underscore "_". The main core of this program is processing words in NooJ format. It means that Word Paradigm Generator suggests appropriate variants for an unknown word without any manual correction except choosing the right paradigm. This is semi-automatic processing of unknown words.

The tool is free and available on Computational Platform for Electronic Text and Speech Processing corpus.by (<https://corpus.by/WordParadigmGenerator/?lang=en>).

Every initial form (lemma) from the list of unknown general words was assigned a morphological class. This class shows the flexion features, i. e. how the word changes. Then the law dictionary in the Belarusian language Law_codes_be.dic was compiled for NooJ format and now is available for further text processing with Belarusian texts of any domains. It is an addition to the main NooJ dictionary for the Belarusian language that has around 200 words in the initial form.

A trilingual dictionary is a perspective phenomenon in the field of Belarusian diplomacy, modern foreign policy, and international relations. The presence of a trilingual dictionary in everyday practice makes it possible to constantly replenish and adjust the lexical stock in three languages. The knowledge of Belarusian and Russian languages, and the spread of English as the “language of international integration” contributes to the creation of extensive interaction between the systems of the three languages in communication. This makes it necessary to compile trilingual dictionaries, especially of legal terminology in terms of this research. The presence of such a dictionary base will help to prevent various mistakes that are possible while translating law codes from Russian and mastering the vocabulary and grammatical structure of Belarusian, Russian and English languages. The use of a trilingual dictionary makes it possible to rely not only on the Russian language but also on other languages being studied, which is important for processing the material in our native language in other spheres of life. This is a good reason to create a Belarusian-Russian-English law dictionary from Law_codes_be.dic (Fig. 5).

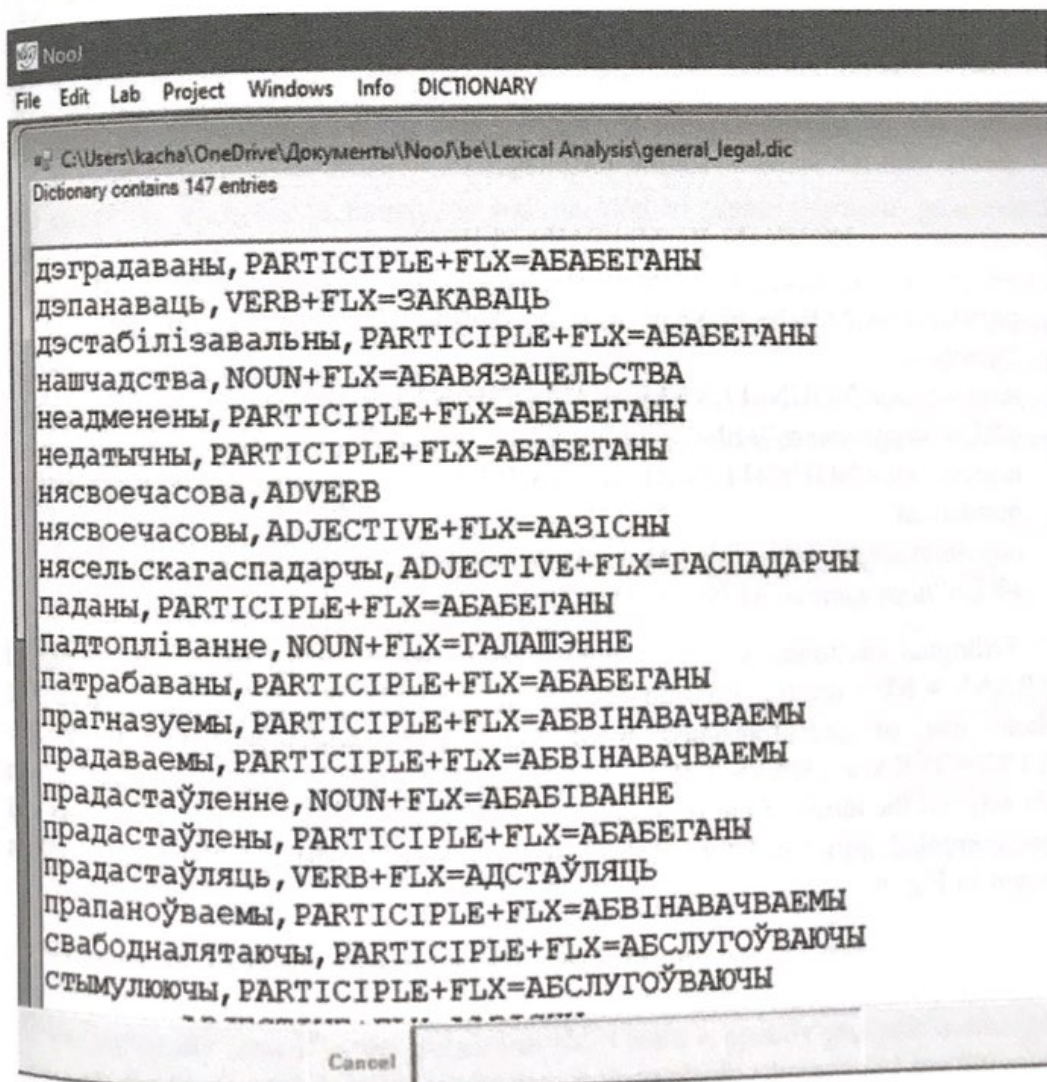


Fig. 5. Belarusian law dictionary (Law_general_be.dic) in NooJ format.

All terms of the Belarusian law dictionary Law_codes_be.dic were correlated with their Russian equivalents from the parallel corpus of a legal domain. The next step was their translation into English which was done by linguists manually. The specialists faced the following difficulties and noticed the following features of trilingual translation into English:

1. some of the Belarusian-Russian law terms can have more than one terminological equivalent with different meanings in the English language:

падведамаснасць, NOUN+FLX=АСКОМІСТАЦЦЬ+TRANS+RU="подведемственность"+TRANS+EN="subordinate"

падведамаснасць, NOUN+FLX=АСКОМІСТАЦЦЬ+TRANS+RU="подведемственность"+TRANS+EN="jurisdiction"

2. two different Belarusian-Russian nouns (of masculine and feminine gender) are translated by the same word equivalent in English:

сужонка, NOUN+FLX=АБАДПРАНКА+TRANS+RU="супруга"+TRANS+EN="marital partner"

сужэнец, NOUN+FLX=АБАДПРАНЕЦ+TRANS+RU="супруг"+TRANS+EN="marital partner"

3. there are Belarusian-Russian law terms which can be translated by several equivalents with the same or similar meaning:

паручальнік, NOUN+FLX=АБАГАЧАЛЬНІК2+RU="поручитель"+EN="fidejussor"

паручальнік, NOUN+FLX=АБАГАЧАЛЬНІК2+RU="поручитель"+EN="surety"

паручальнік, NOUN+FLX=АБАГАЧАЛЬНІК2+RU="поручитель"+EN="guarantor"

паручальнік, NOUN+FLX=АБАГАЧАЛЬНІК2+RU="поручитель"+EN="bail bondsman"

паручальнік, NOUN+FLX=АБАГАЧАЛЬНІК2+RU="поручитель"+EN="adpromissor"

Trilingual dictionaries can be composed in NooJ format by special command "TRANS + RU" for the Russian equivalent and "TRANS + EN" for English. The whole line of one translated word looks like "Абвалоўка, NOUN + FLX = ЛЕЧІЧОЎКА + TRANS + RU = "обваловка" + TRANS + EN = "mound". In this way, all the terms of the legal domain were processed manually by translators and then compiled into NooJ format. A new dictionary consists of 200 notions and is shown in Fig. 6.

- Да таго часу, пакуль палажэнні часткі першай гэтага пункта не дазваляюць вызначыць статус фізічнай асобы, асоба прызнаецца падатковым рэзідэнтам Рэспублікі Беларусь у бягучым каляндарным годзе, калі яна фактычна знаходзілася на тэрыторыі Рэспублікі Беларусь больш за сто восемдзсят тры дні ў папярэднім каляндарным годзе, а ў выпадку, калі фізічная асоба фактычна знаходзілася за граніцамі тэрыторыі Рэспублікі Беларусь сто восемдзсят тры дні і больш у папярэднім каляндарным годзе, яна не прызнаецца падатковым рэзідэнтам Рэспублікі Беларусь у бягучым каляндарным годзе. (Tax Code).
2. Simple sentences with homogeneous parts of the sentence, and the series of these homogeneous parts can be very common (up to 8–10). For example:
 Не дапускаецца ўстанаўленне падаткаў, збораў (пошлін) і льгот па іх выплаце, якія наносзяць урон нацыянальнай бяспецы Рэспублікі Беларусь, яе тэрытарыяльнай цэласнасці, палітычнай і фінансавай стабільнасці, у тым ліку тых, якія абмяжоўваюць свабоднае перамяшчэнне фізічных асоб, перамяшчэнне тавараў (работ, паслуг) ці фінансавых сродкаў у межах тэрыторыі Рэспублікі Беларусь ці тых, якія ствараюць у парушэнне Канстытуцыі Рэспублікі Беларусь і прынятых у адпаведнасці з ёй заканадаўчых актаў іншыя перашкоды для ажыццяўлення прадпрымальніцкай і іншай дзейнасці арганізацый і фізічных асоб, акрамя забароненай заканадаўчымі актамі. (Tax Code).
 3. The presence of numerous passive structures. For example:
 Замежныя (міжнародныя) назіральнікі акрэдытуюцца Цэнтральнай камісіяй. Пасяджэнні Цэнтральнай камісіі склікаюцца па меры неабходнасці Старшынёй Цэнтральнай камісіі.....скрынкі для галасавання правяраюцца, пламбіруюцца або апячатваюцца старшынёй участкавай камісіі... Праатакол падпісваецца старшынёй або намеснікам старшыні і членам камісіі. (Electios Code).
 4. The string of nouns, mainly in the genitive case, i.e. the use of the noun combinations in the genitive case.
 Тэрмін паўнамоцтваў назіральніка пачынаецца з дня яго акрэдытацыі і заканчваецца ў дзень устанаўлення адпаведнай камісіяй вынікаў падліку галасоў, вынікаў выбараў, рэферэндуму, адклікання дэпутата, члена Савета Рэспублікі. (Elections Code)
 Прадаўжэнне тэрміну карыстання нетрамі для здабычы карысных выкапняў, выкарыстання геатэрмальных рэсурсаў нетраў праводзіцца ў парадку, прадагледжаным артыкуламі 33 і 35 гэтага Кодэкса для давання горных адводаў (Mineral Resources Code).

5 Conclusion

The creation of a specialized corpus in NooJ processor is described in the article. A new law corpus contains 1,043,018 tokens. It includes legal texts in the Belarusian and Russian languages, which are law codes of the Republic of Belarus. One law code was officially introduced by the Parliament in the Belarusian language, the other 25 — in Russian. Laboratory scientists translated Russian-language law codes into

Belarusian. The translation quality was improved due to the combined machine-manual approach: texts were initially translated automatically, after that they were manually proofread and edited, then automatically proofread.

A general dictionary of the Belarusian NooJ module was complemented by new words in the process of corpus preparation; a specialized small dictionary of 200 words was compiled: a Belarusian-Russian-English list of law and professional terms. The dictionary is planned to be supplemented.

Syntactic grammars were constructed for machine translation. These grammars will help to divide long sentences without punctuation marks and provide a better reading of law texts. We also plan to find and describe syntactic and morphological markers and other words in which prosodic indicators influence the way we read law texts.

References

1. Computational platform for electronic text & speech processing. <http://www.corpus.by>. Accessed 01 Sep 2021
2. Byulaw. Law & Corpus Linguistics. <https://lawcorpus.byu.edu>. Accessed 01 Sep 2021
3. Monolingual Polish corpus in the law domain. <https://elrc-share.eu/repository/browse/monolingual-polish-corpus-in-the-law-domain>. Accessed 01 Sep 2021
4. British Law Report Corpus. <http://flax.nzdl.org/greenstone3/flax?a=fp&sa=collAbout&c=Blarc&if=>. Accessed 01 Sep 2021
5. Hetsevich, Y., Varanovich, V., Kachan, E., Lysy, S., Reentovich, I.: Semi-automatic part-of-speech annotating for belarusian dictionaries enrichment in NooJ. In: Barone, L., Monteleone, M., Silberstein, M. (eds.): Automatic Processing of Natural-Language Electronic Texts with NooJ: 10th International Conference, NooJ 2016, České Budějovice, June 9–11, 2016, Revised selected papers, pp. 101–111. Springer International Publishing (2016). https://doi.org/10.1007/978-3-319-55002-2_9
6. Hetsevich, Y., Kirdun, A.A.: Vykarystannie sistem mashynnaga pierakladu i sintezu mawliennia dla zabiespiachennia dastupnastsi zakanadawchych tekstaw na roznykh movakh u Respublitsy Bielarus. = Гецавіч, Ю.С., Кірду́н, А.А.: Выкарыстанне сістэм машыннага перакладу і сінтэзу маўлення для забеспячэння даступнасці заканадаўчых тэкстаў на розных мовах у Рэспубліцы Беларусь. In: Коваленко, Е. И. (ред.). Информационные технологии и право (Правовая информатизация – 2018): материалы VI Междунар. науч.-практ. конф., Минск, 17 мая 2018 г., с. 123–128. Минск (2018)
7. NooJ: A Linguistic Development Environment. <http://www.nooj4nlp.org/>. Accessed 01 Sep 2021
8. Udaskanaliennie pratsy awtamatyzyzavanykh sistem pa tekstakh jurydychnaj tematyki = Удасканаленне працы аўтаматызаваных сістэм па тэкстах юрыдычнай тэматыкі. <https://ssrllab.by/7804>. Accessed 01 Sep 2021
9. Reentovich, I., et al.: The First One-Million Corpus for the Belarusian NooJ Module. In: Okrut, T., Hetsevich, Y., Silberstein, M., Stanislavenka, H. (eds.): Automatic Processing of Natural-Language Electronic Texts with NooJ. In: 9th International Conference, NooJ 2015, Minsk, Belarus, June 11–13, 2015, Revised selected papers, pp. 3–15. Springer International Publishing (2016). <https://doi.org/10.1007/978-3-319-55002-2>