правительство российской федерации
Федеральное государственное автономное образовательное учреждение
высшего профессионального образования
Национальный исследовательский университет
«Высшая школа экономики»

Факультет гуманитарных наук
Образовательная программа
«Фундаментальная и компьютерная лингвистика»

Матяш Дарья Сергеевна

Деканцеляризация официальных документов

Decancelarization of official documents

выпускная квалификационная работа
студента 4 курса бакалавриата группы 181

Академический руководитель
образовательной программы
канд. филологических наук, доц.
Ю. А. Ландер

_____

« » _____ 20__ г.

Научный руководитель
кандидат технических наук, доц.
Э. С. Клышинский

_____

Научный консультант
Приглашенный преподаватель
О. С. Сериков

_____

Москва 2022

Abstract

Несмотря на большое количество исследований в области переноса стиля, не существует исследований, насколько нам известно, связанных с “переводом” официальных документов в тексты в нейтральном стиле. Во многих предшествующих работах модели переноса стиля были способны изменять стиль текста, но были склонны искажать первоначальный смысл текста или, наоборот, сушественно модифицировали смысл исходных текстов, в то время как стиль оставался неизменным. Чтение официальных документов, таких как юридические тексты, может быть сложной задачей для человека без специального образования из-за их синтаксической и морфологической сложности. Поэтому существует потребность в инструменте, который был бы способен преобразовать юридический текст в предложения, имеющие тот же смысл, но более доступные и понятные для обычного читателя. В этой статье формулируется новая задача переноса стиля текста и предлагается несколько возможных решений. В статье дается подробный обзор контролируемого подхода к деканцелизации текста. Был собран новый параллельный корпус предложений, первоначально написанных на формальном языке, и один или несколько его вариантов, написанных в нейтральном.

Ключевые слова: *перенос стиля для текстов, детекция канцеляризмов, деканцеляризация текстов, предобученные модели.*

Abstract

Despite the abundance of studies in style transfer, to our knowledge, there is no existing research connected with “translation” of official documents into neurally styled texts. In many previous works style transfer models were able to change the style of a text, but they tended to distort the original meaning, or, vise versa, the meaning of the source texts’, while the style remained unchanged. Reading official papers such as legal texts can be challenging for a person without special education due to their syntactic and morphological complexity. Therefore, there is a distinct need for a tool which would be able to convert a legal text into sentences which have the same meaning but are more accessible and comprehensible for an average reader. This paper formulates a new style transfer task and suggests several possible solutions. The paper gives a detailed overview of supervised approach to text decancelization. A new parallel corpus of sentences originally written in formal language and one or more variants of it written in neutral style was collected. We fine-tuned different style transfer models on our dataset. We developed a style transfer model which is able to translate legal texts into utterances in neutral style, while preserving the meaning of the transformed words.

Keywords: *text style transfer, officialese detection, text decancelarization, pre-trained models.*

Contents

1. Introduction

Reading official papers such as legal texts is a serious challenge for a person without special knowledge and education, owing to the linguistic features of these documents. Due to the linguistic features of these documents, reading official papers such as legal texts is a significant challenge for a person lacking special expertise and education. Most people living in the digital world draw up official documents for a broad variety of applications. Because of the length of texts, which can range from one page to a few hundreds, it might be difficult for a reader to effectively and completely comprehend the information contained in an official text. Furthermore, official documents are commonly regarded as texts syntactically and lexically complicated texts that are difficult for the average person to read. Understanding the content of official papers demands special competences and education, making it difficult for workers and people without professional qualifications to comprehend the content of documents accurately.

Style transfer is a technique that involves rewriting a text, while changing its style and preserving the content. Decancelarization is considered as a special case of style transfer. It is a specific task, whose goal is to convert the style of a text into neutral language without any semantic loss or major transformations.

However, decancelarization is a somewhat more complicated task than style transfer which can often change the meanings of words. It means that there is a risk of replacing a word or phrase in official style not with non-oficial synonyms, but with a word or words with another sense. Moreover, the balance between content preservation and style change should be found, which is a very complicated task as it requires extra resources to evaluate the models' work correctly and to make appropriate conclusions for conducting new experiments that would be able to improve results.

To our knowledge, there was no research dedicated to decancelarization itself until now . However, many researchers conducted experiments in formality transfer Toshevska and Gievska (2021). State-of-the-art solutions can help recipients read academic articles in more accessible language (Jin et al., 2022). However, we could not find any previous work specifically on legal texts. Previous works have also focused on other specific cases of style transfer, such as text detoxification (Dale et al., 2021) where the problem of finding balance between content preserving and style change was highlighted.

We can regard our work as the first approach to solving the problem of style transfer for official documents with rich legal terminology in Russian. Additionally, the solution of this problem can be compared to the task of detoxification since detoxification and decancelirazion change toxic and official words accordingly into a more neutral style. In Dale et al. (2021) two approaches were suggested. The first one included the usage of a generator model which is able to paraphrase a toxic sentence into an utterance in neutral

style. The second approach was to use one classification model which detects a toxic word, and then a synonym in a more neutral style is found.

In order to solve the problem of text decancelarization, we developed a model which can transform a stylistically marked official text into a neutral text. We also collected a parallel dataset containing 5200 utterances written in official language and their translations in neutral style. Our research was conducted as in the following way: we used style transfer techniques on unique parallel corpora composed of sentences with real legal texts written in large industrial company, along with neutral styled utterances and fine-tuned the existing generator models on our data.

2. Literature Review

Style transfer was originally proposed and widely developed for images (Gatys et al., 2016). As for text style transfer, many works were dedicated to this task, too. Individual, social, and situational changes influence the nuance and subtlety of linguistic variants. Recent studies on automatic style transfer of written text is convergent on the idea that style is an integral part of a sentence, as demonstrated by the word choices a person uses. To be more precise, a brief description of several linguistic styles and approaches to TST tasks are given below. Table 1 represents the examples of text style transfer in different applications ans spheres of life. The majority of the examples were taken from the Review on Text Style Transfer by (Toshevska and Gievska, 2021) and the example of text detoxification was received from the research of (Dale et al., 2021)

2.1. Text Style Transfer Tasks

2.1.1. Personal Style Transfer

People's words tell a lot about "themselves, such as their personality, gender, or age" (Pennebaker et al., 2003) . A few works proved that some language variations can be found across different gender and age groups not only in formal texts (Moshe Koppel et al., 2002) , but also in social media (Peersman et al., 2016), (Bamman et al., 2014) and even blog posts (Moshe Koppel et al., 2002), (Schler et al., 2006). According to the research, differences in language usage between various demographic groups exist and it can be detected with an 80 percent accuracy based on the use of specific words and phrases (Koppel et al., 2001). For example, females are more likely to employ emoticons (D. Rao et al., 2010) and terms with a positive emotional connotation (Preotiuc-Pietro et al., 2016). In addition, a possible age group of the author can also be predicted - the younger generation refer to themselves more often and utilize chat-specific e-language, while older people use more complex constructions and add more hashtags and links (Nguyen et al., 2013).

To our knowledge, one of the first works dedicated to personal style transfer was

(Jhamtani et al., 2017). A parallel corpus was created of 18K sentence pairs of 36 Shakespearian plays and their modern translations using the educational site Sparknotes[1]. The corpus was used to train a phrase-based machine translation (PBMT) system to translate an utterance written in modern English to the English of Shakespeare. The PBMT system operated both on a language model built from Shakespeare's plays and probabilities for each translation. The dataset is currently open to the public and was also used in other studies in TST He et al. (2020), (Jhamtani et al., 2017).

Personal style transfer has a variety of industrial uses, including adapting famous novelists' writing styles into other novels and blending different authors' language styles to a single writer in a collaborative setting. It can also have a more edutainment purposes, like text style transfer learned from both Taylor Swift's song lyrics and romance novels (Zhu et al., 2015), (Kiros et al., 2015) .

However, this task has limitations. In TST studies, the parallel corpus that has texts from Shakespearean plays and their translations to the Modern English is the only known corpus of nowadays that supports author imitation. The corpus also has several obvious drawbacks; the size of dataset is small, and the approach is limited to transfer to only one author's style. An interesting future work may be devoted to the collection of texts written by various authors and transfer of the style belonging to a new "mixed" author.

2.1.2. Genre

The genre of a document is defined by some external parameters (Lee, 2001), one of which are target audience and purpose (Biber, 1991). Text information can be categorized into such genres as technical reports, news and advertisements. Identifying a document's genre has the potential to improve Information Retrieval systems by returning search results that match or are relevant to a specific user's search. For instance, when making a purchase decision, advertisements may be more relevant than scientific reports (Dewdney et al., 2001). A document created in a style that corresponds to the language style of a specific group of people is expected to be better comprehensible by the target users. Medical reports, for example, are generally difficult for non-experts to comprehend. Automatically converting a medical report into a layman's text may improve its readability by a wider audience.

2.1.3. Politeness

Politeness transfer (Dewdney et al., 2001) aims to regulate politeness in text. "Could you please send me the documents?" is a more polite phrase than "Send me the documents!". (Dewdney et al., 2001) created a collection of 1.39 million automatically labeled Enron corpus elements (Dewdney et al., 2001). As politeness varies from culture to culture, this dataset focuses mostly on politeness in North American English. The degree of

[1] www.sparknotes.com

politeness is essential for ”maintaining a positive face” in social relationships with others (Coppock, 2005), and it has a significant impact in the whole communication experience.

2.1.4. Sentiment

Emotions play a crucial role in human behavior (Plutchik, 1984) and someone's emotional state is often reflected in one's written or spoken language. While carrying an emotional meaning in a statement is not a normal stylistic variant of language, rewriting a phrase with toned down unpleasant feelings may be required in many situations. For emotions, several category and dimensional models have been suggested (Russell, 1980), (Ekman, 1992). Detecting a text's sentiment polarity, or whether the general sentiment of a specific text is positive or negative.

2.1.5. Offensiveness

The negative impacts of malicious online activity such as hate speech, trolling, and the use of offensive language remain to be a regular problem for almost any social media platform. The public, governments, and institutions have called for the development of systems and the establishment of interaction mediators that will automatically detect, remove, and/or label posts containing offensive language and hate speech. The detection of toxicity in user texts is an active area of study (Zampieri et al., 2020), (d'Sa et al., 2019). In recent works text detoxification is considered as style transfer task (Laugier et al., 2021). A solution to detoxification of texts written in Russian was first introduced by (Dale et al., 2021), (Dementieva et al., 2021) using RuToxic dataset (Dale et al., 2021) that was made up of 163,187 texts (131,780 non-toxic and 31,407 (19%) toxic) from the Russian social networks Odnoklassniki[2] and Pikabu[3] and a parallel corpus that contains 200 toxic sentences and their manually rewritten non-toxic utterances.

2.1.6. Formality

Language style is frequently linked to register, or the formality of a document. Although there is no universally accepted definition of formal language, the distinction between formal and informal language is well recognized. The language used in academic papers, for example, is regarded more formal than that used in social media. Longer texts, as well as those that use the passive voice, are regarded to be more formal (Sheikha and Inkpen, 2010). Detachment, precision, objectivity, rigidity, and a higher cognitive burden are all characteristics of formal writing. Texts using short words, contractions, and abbreviations, on the other hand, are deemed informal. The informal style is more subjective, less accurate, and less informative, as well as having a considerably looser shape. The usage of slang and grammatically incorrect words (Peterson et al., 2011), social distance, and shared knowledge between the writer and the audience are all considered formality

[2]https://ok.ru

[3]https://pikabu.ru

indicators in the research on automatic formality detection (Lahiri et al., 2011). A valuable feature included in writing helpers is the ability to automatically improve the formality of a written document (S. Rao and Tetreault, 2018). Since the formality of text can be affected by a variety of factors, formality transfer is likely to be more complicated than sentiment transfer. Modifications to sentence structure, text length, punctuation, and capitalization, for example, may have an impact on text formality .

Task	**Input sentence (style 1)**	**Output sentence (style 2)**
Personal style transfer	*What is your will?* (Shakespearean)	*What do you want?* (modern English)
Genre transfer	*Many cause dyspnea, pleuritic chest pain, or both.* (expert)	*The most common symptoms, regardless of the type of fluid in the pleural space or its cause, are shortless of breath ans chest pain* (layman)
Politeness transfer	*Send me the data.* (non-polite)	*Could you please send me the data?* (polite)
Sentiment transfer	*Great service but horrible food and very rude waiters!* (negative)	*Great service, tasty food and very personable atmosphere!*
Formality transfer	*Gotta hear both sides of the story.* (informal)	*You have to consider both sides of the story.* (formal)
Text detoxification	*There are commies who hate the USA.* (toxic)	*There are communists who disagree with the USA.* (neutral)

Table 1: Examples of style transfer

2.2. Text Style Data settings

For Text Style Transfer task there exist three main approaches based on the type of data for model training:

2.2.1. Parallel Supervised

The TST models are trained using known pairs of text with different styles in this type of data. With existing parallel corpus style transfer may be defined as a sequence-to-sequence task, similar to supervised Machine Translation, summarization, paraphrasing, etc. Such architectures of language models as `GPT` (Radford et al., 2019) and `T5` (Raffel et al., 2019) can be used for solving such problem. They can frequently handle a wide range of NLP tasks without any fine-tuning. Furthermore, their performance is improved even more when a small training dataset is given. For instance, in (Krishna et al., 2020), a GPT-based model was fine-tuned to transfer between various styles using an automatically generated parallel corpus. The `ruGPT`[4] and `ruT5`[5] that were recently released can give us

[4]https://github.com/sberbank-ai/ru-gpts
[5]https://github.com/ai-forever/ru-gpts

an opportunity to leverage large textual data for decancelarization task in Russian. As we collected a parallel corpus (Section 3.3), the mentioned architectures of models were used and the results are described in Section 5.

2.2.2. Non-Parallel Supervised

The Delete, Retrieve, Generate (Li et al., 2018) method is one of the approaches that only uses non-parallel data. It is based on the idea that words in a phrase may be split into those that contain style information and those that carry the semantics of sentence. For the Delete method, a vocabulary with unwanted words in source style or a model that can detect unneeded words in used for point-wise removal. It has two most problematic limitations. Firstly, the vocabulary cannot contain all of the words in source style. Moreover, some new words generated with new prefixes and suffixes can be missed as they were not included in the vocabulary. Secondly, though the grammatically correct output is expected, a deleted word in source style can contain the meaning crucial for understanding the whole sentence, which means that the original meaning of the sentence can be lost. As for the Retrieve and Generate components, a word in target style is retrieved with most similar word from a corpus with words in source style. However, these techniques match style transfer tasks when one or a few words should be deleted or changed into one or several words with the same meaning but used in the target style. For instance, this approach was successfully used in (Dale et al., 2021) for text detoxification task when toxic words were removed from the text or changed to neutrally styled word. Another example of pointwise editing approach that do not require parallel corpus is Masked Language Modelling (MLM). An MLM that was trained on a dataset with style labels finds a replacement word depending on both the context and the style label. Mask & Infill (X. Wu et al., 2019) is an example of such a model. The `CondBERT` is similar to this architecture, too, and was used in many successful cases of style transfer with no parallel data (Dale et al., 2021), (X. Wu et al., 2019). As toxic words were mostly one-word long nouns and adjectives, the semantic structure of source sentences was not changed and the techniques described above were effectively used in (Dale et al., 2021) and (Dementieva et al., 2021). Vice versa, officialese is characterized by preference of passive voice over active voice, inversed word order, usage of complex words, etc. It means that officialese cannot be simply changed or excluded from the official document, which makes this technique and other pointwise editing models unsuitable for our data.

2.2.3. Unsupervised

The style labels are available for supervised training of the TST models in both parallel and non-parallel supervised data scenarios. A more difficult situation would be totally unsupervised, in which only an unlabeled text corpus is given and TST models are trained unsupervised to conduct text style transfer without knowing the style label.

This approach was considered the most unreliable and unpredictable due to quite difficult strategy of the evaluation of models' outputs. It can be done automatically but with the methods that can not perfectly correlate with human assessment. Moreover, this approach was found as one of the most computationally expensive and as for most reliable evaluation of models human markup is needed.

Due to the facts described above, we decided to collect a parallel corpus that will contain sentences from official documents and their translations manually rewritten in neutral style and conduct experiments using `ruGPT` and `ruT5` models.

3. Data

We collected a parallel corpus which contains pairs of sentences in official and neutral styles. First of all, we chose the best type of official documents which was both rich with officialese and did not contain too much factual data. Secondly, we asked two assessors to translate sentences in formal style to neutrally styled utterances in one or more possible ways. As a result, we got more than 5200 pairs of sentences in formal style and their equivalents in neutral style from almost 300 official documents written in last 3 years.

3.1. Preliminary analysis

Firstly, we conducted research in official documents of bank. We analyzed different types of documents used on everyday basis by bank employees during the last 3 years. For instance, internal notes, orders, contracts and supplementary agreements, requirements and demands are used on many purposes in the bank. In order to collect a parallel corpus which can help train and fine-tune generator models most efficiently, we established the following criteria:

1. **Length**
   Short text documents mostly contain key information and the markers of official style are underrepresented while long documents in most cases consist of redundant information or too much different named entities and irrelevant words in neutral style. In order to make a preliminary statistical analysis of the texts of each type of documents, we counted the following parameters:

   - average number of tokens in documents;
   - average number of tokens per sentence;
   - average number of sentences in documents.

   As shown in the Table 2, internal notes and protocols contain short sentences, which mostly contain only factual information and officialese words might rarely occur. On the other hand, wordy contracts cannot be collected for future translation into the

neutral style due to the following reason. As we decided to collect a parallel corpus, we had to consider the length of one document for manual translation, too. If the length of the document reaches 35-40 pages on average, it would be substantially more complicated to process the file, so we decided not to take lengthy documents into consideration while choosing the best type of official documents.

Type of document	**Avg. # of tokens in documents**	**Avg. # of tokens per sentence**	**Avg. # of sentences in documents**
internal note	170	21	8
protocol	195	35	5
assignment	200	18	12
order	670	28	23
contract	8000	43	220
inner regulatory document	8300	51	350

Table 2: Preliminary statistical analysis of texts in different official documents

2. **Percentage of named entities**
   In some of the types of documents such as internal notes and assignments officialese are underrepresented as these texts generally contain key information and rarely have some linking words or expressions in formal style. Such documents are rich with named entities and cannot make up a parallel corpus with various words and phrases in formal style. After we found out that not all the types of documents can perfectly match our goal, we used transformer model `ruBERT`[6] in order to automatically calculate precisely the average percent of named entities in texts (see Table 4). It was preliminary fine-tuned on company's internal textual data to detect different bank named entities, such as names of posts and positions, names of companies, departments, names of special documents, locations, dates, etc. with F1-score of 0.93. If such named entities as employee's full names and positions, different locations, addresses, names of companies and organizations, occur in most sentences of the currently observed text, we excluded from consideration such documents. According to the Table 4, most words in protocols refer to special named entities, so this type of official documents cannot be used for collection of representative textual data for text decancelarization task. Internal notes and assignments also contain quite a lot of factual information and it it better not to use such types of official documents. Finally, orders were the most "NER-officialese" - balanced

[6]https://huggingface.co/DeepPavlov/rubert-base-cased

type of documents, which contained both information and, what is more interesting for the current research, wordy sentences which often contained complicated word phrases and syntactical structure.

Type of document	**Avg. % of named entities in document**
internal note	39
protocol	64
assignment	35
order	13
contract	15
inner regulatory document	8

Table 3: Percentage of named entities occured in different official documents in the company

3. **Structure**
   As for this aspect, we calculated the average number of paragraphs, nested lists, lists and tables in order to find such a type of official documents that would consist of less tables and nested lists, be less convenient for comprehension (more solid texts with less divisions into paragraphs and lists). We searched on such type of documents as it was crucial to collect parallel corpus with sentences which are most complicated for understanding.

Type of document	**Avg. # of lists**	**Avg. # of nested lists**	**Avg. # of tables**
internal note	2	1	0
protocol	2	0	2
assignment	3	0	1
order	2	1	2
contract	6	5	6
inner regulatory document	4	2	3

Table 4: Avarage numbers of lists, nested lists and tables occured in different official documents in the company

According to the Table 4, internal notes, protocols and orders contain the smallest amount of structurized information, while in contracts and inner regulatory documents a bunch of textual data is organized in lists and tables, which, of course, can facilitate the comprehension of lengthy documents.

After the described above preliminary analysis of all the existing official documents in the company, we found the most appropriate type of official texts - orders. Inner regulatory documents take second place in our ”competition” for the most appropriate type of official

document. Consequently, our developed models could be used for decancelarizarion of lengthy inner regulatory documents and be helpful for bank employees who have to work with such type of textual data.

3.2. Instruction for manual decancelarization

Then, we investigated which linguistic features systematically occur in official documents. We studied internal notes, orders and other official documents used within the bank. As a result, we developed an instruction[7] how to primarily detect officialese words or word phrases in document and then translate them into neutrally styled utterances, paying attention to both lexical (e.g. professional vocabulary, words, rarely used in everyday life) and grammatical aspects (e.g. passive forms) of text. The linguistic features used in official documents frequently are shown in Table 5.

Feature	**Example**
Passive Voice	*Договор прологируется Заказчиком* 'The Contract is prolonged by the Internal Client'
Verbal nouns	*осуществление приказа* 'implementation of the order' *выполнение поручений* 'execution of orders'
Professional vocabulary	*пролонгирование* 'prolongation'
Action verbs + verbal nouns	*осуществлять тиражирование ПО* 'to carry replication' *производить проверку ПО* 'perform verification of software'
Derived prepositions + nouns	*в течение выделенного срока* 'during the allotted period' *согласно приказу* 'according to the order'
Indirect word order	*Принимающей стороне передает Заказчик информацию о...* 'The Customer transmits to the receiving party information about…'

Table 5: Frequently used linguistic features in official documents

3.3. Data collection

The data was collected in two steps. Firstly, we asked assessors to write all possible neutrally styled variants of each sentence written in formal language. We had three assessors: one of them had legal education and others are graduates in technical fields. On the first iteration, 1300 pairs of sentences from official documents and their possible "translations" to neutral style were collected. It took us 3 working days (24 hours of work) of three assessors. We discussed the gathered texts and found the following recurring mistakes in "translation" of assessors:

[7]https://github.com/MatyashDare/text_decancelarization/blob/materials/Instruction.pdf

1. the more tokens in one sentence, the more the probability of missing one or a few officialese (fortunately, it proved our preliminary assumption not to take lengthy sentences from inner regulatory documents for parallel corpus collection as it was presented in Section 3.1);

2. the deeper a word in official style is situated in syntactic tree of clause, the, the higher the probability that it will be missed (in most cases, there were mostly archaic pronouns (*иная информация* 'other information') and dependent nouns (*в части оптимизации процесса тиражирования материалов* 'in terms of optimizing the process of replication of materials')

3. in phrases consisted of verb with "to do/ to make" semantics and verbal noun the less the noun looked like a loan-word, the higher was a probability that assessors missed the officialese in text (*провести проверку* 'to do an inspection' was rarely changed to *проверить* 'to check' than textitосуществлять тиражирование 'to carry out replication' was "translated" into *распространить* 'replicate', 'spread')

4. derivative prepositions (*в целях* (in order to), *согласно (документу)* (according to (the document))) were the most frequent morphological category that was not translated to simplier constructions

5. assessor with legal education missed significally more officialese than people who did not work with official documents on daily basis, which proved our assumption that people without special legal education cannot easily comprehend official texts and they are able to detect the most complicated for understanding parts of texts.

After all of the manual translations were discussed with assessors and all the mistakes were analyzed, only 1000 pairs of official and non-official utterances left. These pairs were then used for training of a binary classifier that could be able to distinguish officialese and non-official texts (see Section 4.1).
On the second step, we received 5200 pairs of official and non-official sentences, which were than used for style transfer experiments (Section 4.4). The collection of this amount of data took as 5 days of work (almost 40 working hours) of three assessors, which is quite fast for manual data collection. After we applied our binary classifier to all the sentences of corpus, and we had to manually rewrite 100 a few "informal" sentences that still contained offisialese. These final 5200 pairs made up our parallel corpus.

3.4. Parallel corpus

The parallel corpus contains 5200 pairs of sentences in official and neutral style from almost 300 official documents. Each sentence in official style has 3 different translations

to neutral style on average, which can be quite informative for style transfer models during the training. Each formally styled sentence is detected as officialese by our binary classifier (Section 4.1) and each neutrally styled translation is predicted as non-officialese by the classifier.

We calculated corpus statistics. According to the Table 6, sentences from original official documents contain longer words and are written with more tokens. Official texts have also greater amount of non-repeating tokens and lemmas. Therefore, our manually translated sentences are shorter and should be simpler for comprehension and understanding.

Parameter	**Officialese**	**Non-officialese**
Avg. # of symbols	200	172
Avg. # of tokens	25	21
Avg. length of tokens	10	6
Avg. # of unique tokens	20	17
Avg. # of unique lemmas	19	15

Table 6: Statistics on parallel corpus

The most popular and representative uni- and bi-grams in official and non-official texts are shown in Figures 1 and 2 respectively. According to the Figure 1, such constructions as ”derivative pronoun” + ”noun” (*согласно Распоряжению* ‘according to the Order’, *на основании обратной связи* ‘based on the feedback’) and ”adjective” + ”noun” (*календарный день* ‘calendar day’, *конфиденциальная информация* ‘confidential information’) are in the number of the most frequent bigrams in official texts. All of the shown in Figure 1 uni- and bigrams are typical for officialese. Consequently, our theoretical knowledge about officialese and the formulated instruction for manual translation to neutral style highly correlate with original official textual data.
In accordance with Figure 2, the most frequent uni- and bigrams were prepositions with nouns that denote the name of document or some types of company’s initiatives (*приложение к (Документу/Распоряжению)* ’appendix to (the Document/ the Order)’), or paraphrasing constructions for changing officialese into more neutrally styled texts(*как написано в (Распоряжении)* ’as it is written in (the Order)’ is one of possible constructions for decancelarization of the phrase *согласно (Распоряжению)* ’according to (the Order)’). However, a few bigrams can depict the type of text - frequently occured bigrams *пилотный проект* ’pilot project’ or *рабочий день* ’business day’ relate to business-oriented themes, but they still do not belong to official style.

4. Methods

4.1. Officialese detection

We trained different Machine and Deep learning models on binary classification task (0 - official text, 1 - neutrally styled text). In Table 8 in Section 6 top-7 best results were shown. The best model was the binary classifier based on `Naive Bayes classifier` and `TfidfVectorizer` that predicts probabilities whether texts written in formal or neutral style. We trained the model on 300 randomly selected sentences from neutrally styled news by "Lenta.ru"[8] mass media in Russian and 1000 sentences in official style and manually their translated to neutral style pairs obtained on the first iteration (Section 3.3). The model achieved F1-score of 0.87 on test dataset. As a result, our binary classifier provides an adequate result for detecting toxic texts and can be used for evaluation of the strength of style transfer. Since we want our models to perform decancelarization task, we expected the outputs of style transfer models to be non-formally styled. Based on this assumption, we compute the accuracy (see 4.2.1).

4.2. Evaluation

In order to conduct a thorough evaluation of a style transfer model, we have to consider the following output of the model:

- the style of the source sentence is changed;
- the content of the source sentence is preserved;
- it yields a grammatically correct sentence.

Individual metrics are used in the majority of works on style transfer to evaluate these parameters. (Pang and Gimpel, 2018) points out, however, that these three components are frequently inversely correlated, thus they must be combined to obtain the balance. As we need a compound metric to find a balance between them, our evaluation setup follows this principle.

*4.2.1. Style transfer accuracy (**STA**)*

In order to evaluate style transfer accuracy (STA), we trained a binary classifier described in 4.1 on held-out set of official orders that were not manually translated to neutrally styled phrases for parallel corpus collection and texts collected from "Lenta.ru" news written in neutral style. We count an average of all the predicted classes (1 - neutral style, 0 - official language) for each sentence and the level of officialese. The higher the level of STA, the better style was transferred.

[8] https://github.com/yutkin/Lenta.Ru-News-Dataset

*4.2.2. Content preservation (**CP**)*

We look into content preservation from two perspectives. First, we compute well-established word-based metrics in order to count the number of matching substrings in the fastest and most efficient way:

- **BLEU** score - ngram precision for n from 1 to 4. It calculates how similar a candidate sentence is to a reference sentence based on n-gram matches between sentences and correlates with human evaluation up to 0.817 (Papineni et al., 2002). The larger the metric, the closer generated texts to their references;

- **METEOR** - this metric is also based on word overlap (**WO**) calculation. METEOR is determined as the harmonic mean of unigram precision and recall, with recall weighted more heavily (Denkowski and Lavie, 2014). Though METEOR is better correlated with manual evaluation (Yamshchikov et al., 2019) than BLEU, in practice, BLEU is the most extensively used metric for measuring the semantic similarity between the source phrase and the style-transferred result (Yang et al., 2018). As the BLEU score, the larger the METEOR metric, the less differences between the generated and source sentences are;

As our source and target sentences have great differences in syntactic structure, words and word phrases, the parameters above do have much weight in final evaluation.
Secondly, we calculate cosine similarity (**CS**) between sentence-level embeddings of both source and transformed texts. There are several models specially trained to translate text into a vector so that the vectors of texts that are close in meaning are geometrically close to each other. One of them is `LaBSE`[9] which was trained to bring the embedding of texts with the same meanings in different languages closer to each other (Feng et al., 2020). As `LaBSE` works better than analogues on most tasks of coding Russian sentences without additional training, the similarity of the meaning of pairs of texts as the cosine proximity of their embeddings are measured with `LaBSE` embeddings.

*4.2.3. Fluency (**FL**)*

Natural language outputs demand fluency as a minimum criterion. As to evaluate the quality of the generated sentence automatically, we use perplexity (`PPL`). We utilize the `ruGPT2Large`[10] language model for this metric. As a result, we can state that this model can provide us with a fair score for perplexity. The lower the PPL, the more natural a sentence is generated, so we considered the 1/PPL as one of the factors in aggregated metric (see Section 4.2.4).

[9] `https://huggingface.co/cointegrated/LaBSE-en-ru`
[10] `https://github.com/vlarine/ruGPT2`

*4.2.4. Aggregated metric(**GM**)*

Following (Dementieva et al., 2021) and (Pang and Gimpel, 2018), we use a combination of STA, CS and FL. Other content preservation parameters are not included in the combination but rather are reported to help understand the model properties. Specifically, we calculate GM the geometric mean of STA, CP and 1/PPL:

$$\mathrm{GM} = (\max(\mathrm{STA}, 0) \times \max(\mathrm{CP}, 0) \times \max(1/\mathrm{PPL}, 0))^{\frac{1}{3}}$$

4.3. Baseline

As a baseline, we used a paraphrasing model[11] based on T5 architecture introduced by David Dale[12]. This model was trained on a large neutrally styled ParaNMT-Ru-Leipzig dataset[13], so we do not expect the output of the model to be generated in official style.

4.4. Models

4.4.1. Baseline + modifications

Two techniques for gaining a higher level of original content preservation and fluency of the generated text were described. The first one is the `num_beams` parameter: when the parameter is increased, the level of text fluency improves. However, it also requires more time for text generation in this case. Secondly, the `bad_word_ids` parameter was found important for better content preservation. A list of tokens that we do not expect to be generated can be added to the paraphraser model, so the style transfer accuracy metric can also improve. The list of unwanted tokens was made up from the list of officialese from the dictionary of officialese[14], top-100 `max_features` parameter from the best binary classifier (Section 4.1) and top-100 important features from the Attention matrices of the Self-Attention layer obtained from the second best binary classifier - fine-tuned transformer model `ruBERT-base-cased`.

4.4.2. ruGPT3

GPT is a transformer decoder model that uses attention instead of earlier recurrence- and convolution-based architectures (Sherstinsky, 2020), (O'Shea and Nash, 2015). The model's attention mechanisms allow it to selectively focus on parts of input text that it predicts would be the most relevant. We fine-tuned `ruGPT3-small`, `ruGPT3-medium`, `ruGPT3-large` for text generation given a new token <DECANC> that was used as special token for generation decancelarized variant of the text before the mentioned token. We used optuna[15] open-source library for searching for the best hyperparameters. It took us about 40 GPU hours to find the best parameters for all 3 models. Promt-tuning (Lester

[11] https://huggingface.co/cointegrated/rut5-base-paraphraser
[12] https://habr.com/ru/post/564916/
[13] https://storage.yandexcloud.net/nlp/paranmt_ru_leipzig.zip
[14] https://avtoram.com/slovar-kantselyarizmov/
[15] https://github.com/optuna/optuna

et al., 2021) is a technique of adding trainable embeddings to a sequence of tokens embeddings and optimizing them given a frozen network. We used this technique in order to improve the outputs of the `ruGPT3` models.

4.4.3. ruT5

The `T5` (Raffel et al., 2019) is an encoder-decoder architecture that could be applied on a number of purposes, including text style transfer. We fine-tuned `ruT5-base` and `ruT5-large` models on our parallel corpus on text style transfer task. Due to the similarity of the decancelarized text to the original one, it is likely for the model to copy the original input to get the low value of the Cross-Entropy loss. To focus on the rare dissimilarities during the optimization, we adopted the Focal loss (Lin et al., 2017) for our task.

5. Results

All the counted metrics are represented in Table 7.

model	**STA↑**	**CP↑**	**BLEU↑**	**METEOR↑**	**PPL↑**	**GM↑**
Baseline	0.43	0.7	0.64	0.76	0.85	0.25
Baseline + `bad_word_ids`	0.52	0.75	0.81	0.64	0.69	0.27
`ruGPT3-small`	0.66	0.8	0.7	0.7	0.61	0.33
`ruGPT3-medium`	0.7	0.88	0.74	0.77	0.7	0.43
`ruGPT3-large`	0.72	0.9	0.88	0.73	0.75	0.49
`ruGPT3-large` + promt-tuning	0.71	0.9	0.78	0.78	0.82	0.52
`ruT5-base`	0.71	0.88	0.88	0.76	0.8	0.49
`ruT5-base` + Focal loss	0.72	0.89	0.79	0.79	0.82	0.55
`ruT5-large`	0.74	0.9	0.75	0.78	0.82	0.55
`ruT5-large` + Focal loss	**0.74**	**0.93**	**0.89**	**0.8**	**0.83**	**0.57**

Table 7: The results of evaluation of proposed decancelarization approaches. **STA**: Style transfer accuracy. **CS**: Cosine similarity. **BLEU**: bilingual evaluation understudy. **METEOR**: Metric for Evaluation of Translation with Explicit ORdering). **PPL**: Perplexity. **GM**: Geometric mean of STA, CP and PPL. The larger↑, the better. GThe values in bold denote best scores.

Despite our hypothesis that the paraphrasing model would be able to decrease the level of officialese in official documents,the Baseline model in more than a half of the sentences did not change officialese, which is obvious from the STA metric of 0.43. The BLEU and METEOR metrics represent that the majority of original official texts were modified. Fortunately, the content was preserved, so the baseline model did not deformed the key information in official texts, which is true for all the tried models. Though the generated text seemed to look like a natural one, it contained many formal words and expressions. The usage of the `bad_word_ids` parameter substantially improved the outputs of the baseline model: officialese were not allowed to be generated and STA metric

was higher. Content was not significantly modified, but the text fluency dramatically decreased.

According to the Table 7, the best models' performance was obtained from fine-tuned `ruT5-large` model with Focal loss. The second best model was `ruT5-large` fine-tuned with Cross-Entropy loss. The number of parameters of the applied models had influence on the quality of the generated texts: the larger was `ruT5` or ruGPT3 model, the better results were gained, judging by all the metrcis. Promt-tuning for the best `ruGPT3` model improved the model's output.

Tables 9 and 10 depict how four of the best models (`ruGPT3-large`, `ruGPT3-large` + promt-tuning, `ruT5-large` and `ruT5-large` fine-tuned with focal loss) transferred the style of 4 randomly taken sentences from the test dataset. We provide the original sentence, its human translation to the neutral style and the outputs of the mentioned models. We manually highlighted the parts of texts that contain officialese in red, and the translated formally styled in green. The best (by all the metrics and out human evaluation) style transfer is in bold. According to the Table 9, the texts generated by the fine-tuned `ruT5-large` with focal loss were the closest texts to the neutral style, while the outputs of the `ruGPT-2-large` models did not replace a few segments of officialese. In some sentences different models, both `ruGPT-2-large` and `ruT5-large` did not change the officialese in the same parts of the texts. The mistake, mentioned before (Section 3.3) as one of the most frequent mistakes during human translation occured in the outputs of the models (Examples 2.2 - 2.5). The same and simultaneously erroneous outputs of the models could be seen in Examples 3.2 and 3.3. However, the sentence generated by `ruT5-large + focal loss` was considered as the more fluent and natural utterance than the human translated ones due to the straight word order. It means that this model can be used for improving and enlarging the existing parallel corpus. As for the sentences 4.2-4.5, all the models did insignificant mistake (they did not transferred passive participle into a construction with relative pronoun and verb in active voice that is easier for understanding) but a word *перечень* that belongs to the official style was replaced with a more neutrally styled *список* only by ruT5-large model, which is a representative example of how one word can be simpler and have the same meaning.

For `ruT5` models fine-tuned with Focal loss had better performance than `ruT5-base` and `ruT5-large` using Cross-Entropy as loss function. The `ruT5` models generated texts wcich had less common words with the manually translated text (judging by the BLEU and METEOR metrics), but the content was still preserved and the STA metric was the highest in accordance with baseline and `ruGPT3` models.

6. Conclusion

We presented the first study of text decancelarization for the Russian language. It is crucial to enlarge the spheres of application of style transfer', so we used this technique for processing legal texts that are both lexically and syntactically rich and quite difficult for human comprehension. After we analyzed different types of official documents used in large bank, orders that have been used in the company on daily basis in two recent years were chosen for the following manual translation into neutral style. We have prepared a parallel dataset for supervised experiments. It contained 5200 pairs of sentences in official style and their manual translations to the neutral style collected from almost 300 official documents. We fine-tuned `ruGPT3` and `ruT5` pre-trained models on the manually collected parallel corpus. The best aggregated metric was achieved by the ruT5-large model fine-tuned with Focal loss. The `ruT5-large` and `ruGPT3-large` models gained the highest style accuracy metrics, while the best content preservation values were reached by `ruT5-large` models fine-tune with cross-entropy and focal losses. However, the best perplexity metric reached by the baseline paraphrasing model though it did not decancelarized the given official texts well.

As for possible applications of our findings, we could mention that our parallel data can be used for future research in decancelarization tasks – for example, for improvements in this specific field, but also for the inverse problem – transforming neutrally styled texts into official ones. This possible future application of our findings can help people with insufficient proficiency or time to easily transform massive text data into official texts. Our parallel corpus can also be used for improvement of paraphrasing models, as it contains 3 different neutrally styled variants for each phrase in formal language. Finally, our decancelarization fine-tuned models can significantly simplify the process of comprehension of official documents and save people's time.

Additional materials

All the supplementary materials, including, code for text style transfer evaluation and experiments are uploaded in the Github repository:
`https://github.com/MatyashDare/text_decancelarization`.

References

Bamman, D., Eisenstein, J., & Schnoebelen, T. (2014). Gender identity and lexical variation in social media. *Journal of Sociolinguistics*, *18*(2), 135–160.

Biber, D. (1991). *Variation across speech and writing*. Cambridge University Press.

Coppock, L. (2005). Politeness strategies in conversation closings. *Unpublished manuscript: Stanford University*.

Dale, D., Voronov, A., Dementieva, D., Logacheva, V., Kozlova, O., Semenov, N., & Panchenko, A. (2021). Text detoxification using large pre-trained neural models. *arXiv preprint arXiv:2109.08914*.

Dementieva, D., Moskovskiy, D., Logacheva, V., Dale, D., Kozlova, O., Semenov, N., & Panchenko, A. (2021). Methods for detoxification of texts for the russian language. *Multimodal Technologies and Interaction*, *5*(9), 54.

Denkowski, M., & Lavie, A. (2014). Meteor universal: Language specific translation evaluation for any target language. *Proceedings of the ninth workshop on statistical machine translation*, 376–380.

Dewdney, N., Van Ess-Dykema, C., & MacMillan, R. (2001). The form is the substance: Classification of genres in text. *Proceedings of the ACL 2001 Workshop on Human Language Technology and Knowledge Management*.

d'Sa, A. G., Illina, I., & Fohr, D. (2019). Towards non-toxic landscapes: Automatic toxic comment detection using dnn. *arXiv preprint arXiv:1911.08395*.

Ekman, P. (1992). An argument for basic emotions. *Cognition & emotion*, *6*(3-4), 169–200.

Feng, F., Yang, Y., Cer, D., Arivazhagan, N., & Wang, W. (2020). Language-agnostic bert sentence embedding. *arXiv preprint arXiv:2007.01852*.

Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

He, J., Wang, X., Neubig, G., & Berg-Kirkpatrick, T. (2020). A probabilistic formulation of unsupervised text style transfer. *arXiv preprint arXiv:2002.03912*.

Jhamtani, H., Gangal, V., Hovy, E., & Nyberg, E. (2017). Shakespearizing modern language using copy-enriched sequence-to-sequence models. *arXiv preprint arXiv:1707.01161*.

Jin, D., Jin, Z., Hu, Z., Vechtomova, O., & Mihalcea, R. (2022). Deep learning for text style transfer: A survey. *Computational Linguistics*, *48*(1), 155–205.

Kiros, R., Zhu, Y., Salakhutdinov, R. R., Zemel, R., Urtasun, R., Torralba, A., & Fidler, S. (2015). Skip-thought vectors. *Advances in neural information processing systems*, *28*.

Koppel, M., Argamon, S., & Shimoni, A. (2001). Automatically determining the gender of a text's author. *Bar-Ilan University Technical Report BIU-TR-01-32*.

Koppel, M. [Moshe], Argamon, S., & Shimoni, A. R. (2002). Automatically categorizing written texts by author gender. *Literary and linguistic computing*, *17*(4), 401–412.

Krishna, K., Wieting, J., & Iyyer, M. (2020). Reformulating unsupervised style transfer as paraphrase generation. *arXiv preprint arXiv:2010.05700*.

Lahiri, S., Mitra, P., & Lu, X. (2011). Informality judgment at sentence level and experiments with formality score. *International Conference on Intelligent Text Processing and Computational Linguistics*, 446–457.

Laugier, L., Pavlopoulos, J., Sorensen, J., & Dixon, L. (2021). Civil rephrases of toxic texts with self-supervised transformers. *arXiv preprint arXiv:2102.05456*.

Lee, D. Y. (2001). Genres, registers, text types, domains and styles: Clarifying the concepts and nevigating a path through the bnc jungle.

Lester, B., Al-Rfou, R., & Constant, N. (2021). The power of scale for parameter-efficient prompt tuning. *arXiv preprint arXiv:2104.08691*.

Li, J., Jia, R., He, H., & Liang, P. (2018). Delete, retrieve, generate: A simple approach to sentiment and style transfer. *arXiv preprint arXiv:1804.06437*.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. *Proceedings of the IEEE international conference on computer vision*, 2980–2988.

Nguyen, D., Gravel, R., Trieschnigg, D., & Meder, T. (2013). " how old do you think i am?" a study of language and age in twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, *7*(1).

O'Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*.

Pang, R. Y., & Gimpel, K. (2018). Unsupervised evaluation metrics and learning criteria for non-parallel textual transfer. *arXiv preprint arXiv:1810.11878*.

Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). Bleu: A method for automatic evaluation of machine translation. *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, 311–318.

Peersman, C., Daelemans, W., Vandekerckhove, R., Vandekerckhove, B., & Van Vaerenbergh, L. (2016). The effects of age, gender and region on non-standard linguistic variation in online social networks. *arXiv preprint arXiv:1601.02431*.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology*, *54*(1), 547–577.

Peterson, K., Hohensee, M., & Xia, F. (2011). Email formality in the workplace: A case study on the enron corpus. *Proceedings of the Workshop on Language in Social Media (LSM 2011)*, 86–95.

Plutchik, R. (1984). Emotions: A general psychoevolutionary theory. *Approaches to emotion*, *1984*, 197–219.

Preotiuc-Pietro, D., Xu, W., & Ungar, L. (2016). Discovering user attribute stylistic differences via paraphrasing. *Proceedings of the AAAI Conference on Artificial Intelligence*, *30*(1).

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I. et al. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, *1*(8), 9.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2019). Exploring the limits of transfer learning with a unified text-to-text transformer. *arXiv preprint arXiv:1910.10683*.

Rao, D., Yarowsky, D., Shreevats, A., & Gupta, M. (2010). Classifying latent user attributes in twitter. *Proceedings of the 2nd international workshop on Search and mining user-generated contents*, 37–44.

Rao, S., & Tetreault, J. (2018). Dear sir or madam, may i introduce the gyafc dataset: Corpus, benchmarks and metrics for formality style transfer. *arXiv preprint arXiv:1803.06535*.

Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, *39*(6), 1161.

Schler, J., Koppel, M., Argamon, S., & Pennebaker, J. W. (2006). Effects of age and gender on blogging. *AAAI spring symposium: Computational approaches to analyzing weblogs*, *6*, 199–205.

Sheikha, F. A., & Inkpen, D. (2010). Automatic classification of documents by formality. *Proceedings of the 6th international conference on natural language processing and knowledge engineering (nlpke-2010)*, 1–5.

Sherstinsky, A. (2020). Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network. *Physica D: Nonlinear Phenomena*, *404*, 132306.

Toshevska, M., & Gievska, S. (2021). A review of text style transfer using deep learning. *IEEE Transactions on Artificial Intelligence*.

Wu, X., Zhang, T., Zang, L., Han, J., & Hu, S. (2019). " mask and infill": Applying masked language model to sentiment transfer. *arXiv preprint arXiv:1908.08039*.

Yamshchikov, I. P., Shibaev, V., Nagaev, A., Jost, J., & Tikhonov, A. (2019). Decomposing textual information for style transfer. *arXiv preprint arXiv:1909.12928*.

Yang, Z., Hu, Z., Dyer, C., Xing, E. P., & Berg-Kirkpatrick, T. (2018). Unsupervised text style transfer using language models as discriminators. *Advances in Neural Information Processing Systems*, *31*.

Zampieri, M., Nakov, P., Rosenthal, S., Atanasova, P., Karadzhov, G., Mubarak, H., Derczynski, L., Pitenis, Z., & Çöltekin, Ç. (2020). SemEval-2020 task 12: Multilingual offensive language identification in social media (OffensEval 2020). *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, 1425–1447. https://doi.org/10.18653/v1/2020.semeval-1.188

Zhu, Y., Kiros, R., Zemel, R., Salakhutdinov, R., Urtasun, R., Torralba, A., & Fidler, S. (2015). Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. *Proceedings of the IEEE international conference on computer vision*, 19–27.

Appendix

model	F1-score
Naive Bayes + TfidfVectorizer	**0.87**
Logistic Regression + TfidfVectorizer	0.8
Passive Aggressive Classifier + TfidfVectorizer	0.81
Random Forest Classifier + TfidfVectorizer	0.78
DeepPavlov/rubert-base-cased	**0.86**
xlm-roberta-base	0.82
bert-base-multilingual-cased	0.83

Table 8: Binary classifiers on officialese detection

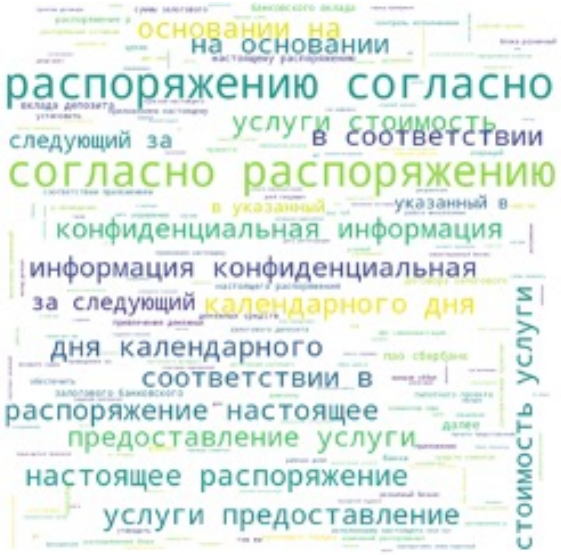

Figure 1: Word clouds of top-30 frequently occured uni- and bi-grams in official texts.

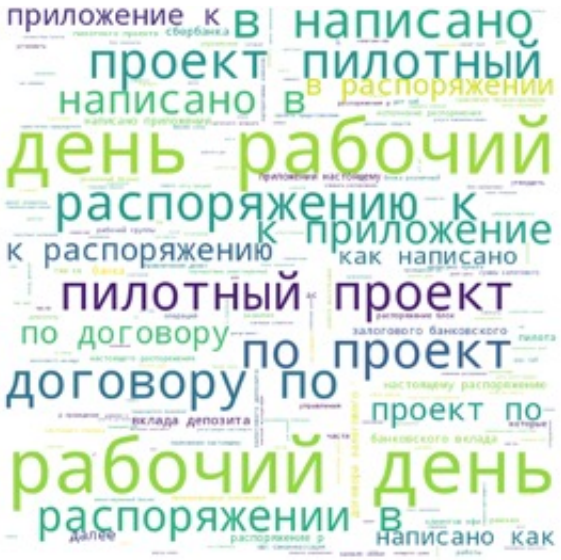

Figure 2: Word clouds of top-30 frequently occured uni- and bi-grams in non-official texts.

№	model	text
1.0	original	*Начиная с премирования по итогам работы за январь 2021 г., руководствоваться перечнем операций и коэффициентами пересчета продуктов в условные продукты (далее - УП) для менеджеров по продажам в соответствии с Приложением к настоящему Распоряжению.*
1.1	manual translation	*С января 2021 года использовать список операций и показателей того, как продукты пересчитываются в условные продукты (УП), для менеджеров по продажам (см. Приложение).*
1.2	ruGPT3-large	*Начиная с премирования по итогам работы за январь 2021 г., руководствоваться перечнем операций и коэффициентами пересчета продуктов в условные продукты (далее - УП) для менеджеров по продажам по Приложению к этому Распоряжению.*
1.3	ruGPT3-large + promt-tuning	*Начиная с премирования по итогам работы за январь 2021 г., руководствоваться перечнем операций и коэффициентами пересчета продуктов в условные продукты (далее - УП) для менеджеров по продажам по Приложению к этому Распоряжению.*
1.4	ruT5-large	*С января 2021 года использовать перечень операций и коэффициенты пересчета продуктов в условные продукты (УП) для менеджеров по продажам по Приложению к этому Распоряжению.*
1.5	ruT5-large + Focal loss	***С января 2021 года использовать список операций и коэффициенты пересчета продуктов в условные продукты (УП) для менеджеров по продажам по Приложению этого Распоряжения.***
2.0	original	*В случае подтверждения устранения Катастрофической или Серьезной ошибки Заказчик самостоятельно осуществляет тиражирование ПО в своих подразделениях и филиалах.*
2.1	manual translation	*Если подтверждается, что катастрофическая или серьезная ошибка устранена, заказчик самостоятельно распространяет ПО в своих подразделениях и филиалах.*
2.2	ruGPT3-large	*При подтверждении устранения Катастрофической или Серьезной ошибки Заказчик самостоятельно тиражирует ПО в подразделениях и филиалах.*
2.3	ruGPT3-large + promt-tuning	*При подтверждении устранения Катастрофической или Серьезной ошибки Заказчик самостоятельно тиражирует ПО в подразделениях и филиалах.*
2.4	ruT5-large	*При подтверждении устранения Катастрофической или Серьезной ошибки Заказчик самостоятельно тиражирует ПО в подразделениях и филиалах.*
2.5	ruT5-large + focal loss	***Если подтверждается, что катастрофическая или серьезная ошибка устранена, заказчик самостоятельно тиражирует ПО в своих подразделениях и филиалах.***

Table 9: Examples of models' outputs-1

№	model	text
3.0	original	*Контроль за исполнением настоящего Распоряжения оставляю за собой.*
3.1	manual translation	*Исполнение этого Распоряжения проконтролирую лично.*
3.2	ruGPT3-large	*Исполнение этого Распоряжения оставляю за собой.*
3.3	ruGPT3-large + promt-tuning	*Исполнение этого Распоряжения оставляю за собой*
3.4	ruT5-large	*Исполнение этого Распоряжения проконтролирую лично.*
3.5	ruT5-large + Focal loss	***Исполнение этого Распоряжения лично проконтролирую.***
4.0	original	*Утвердить перечень автоматизированных систем и информационных ресурсов Банка, доступных к подключению Сотрудникам (Приложение 1).*
4.1	manual translation	*Предоставить список систем и ресурсов Банка, которые доступны к подключению Сотрудникам (Приложение 1).*
4.2	ruGPT3-large	*Установить перечень автоматизированных систем и информационных ресурсов Банка, доступных к подключению Сотрудникам (Приложение 1).*
4.3	ruGPT3-large + promt-tuning	*Установить перечень автоматизированных систем и информационных ресурсов Банка, доступных к подключению Сотрудникам (Приложение 1).*
4.4	ruT5-large	*Предоставить список систем и ресурсов Банка, доступных к подключению Сотрудникам (Приложение 1).*
4.5	ruT5-large + Focal loss	*Предоставить список систем и ресурсов Банка, доступных к подключению Сотрудникам (Приложение 1).*

Table 10: Examples of models' outputs-2