Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 2045–2053
Marseille, 11–16 May 2020

A Contract Corpus for Recognizing Rights and Obligations

Ruka Funaki,[1,2] Yusuke Nagata,[2] Kohei Suenaga,[2] Shinsuke Mori[3]
[1]LegalForce Inc., Tokyo, Japan
[2]Graduate School of Informatics, Kyoto University, Kyoto, Japan
[3]Academic Center for Computing and Media Studies, Kyoto University, Kyoto, Japan
ruka.funaki@legalforce.co.jp, nagata.yusuke.88x@st.kyoto-u.ac.jp,
ksuenaga@fos.kuis.kyoto-u.ac.jp, forest@i.kyoto-u.ac.jp

Abstract

A *contract* is a legal document executed by two or more parties. It is important for these parties to precisely understand their rights and obligations that are described in the contract. However, understanding the content of a contract is sometimes difficult and costly, particularly if the contract is long and complicated. Therefore, a language-processing system that can present information concerning rights and obligations found within a given contract document would help a contracting party to make better decisions. As a step toward the development of such a language-processing system, in this paper, we describe the annotated corpus of contract documents that we built. Our corpus is annotated so that a language-processing system can recognize a party's rights and obligations. The annotated information includes the parties involved in the contract, the rights and obligations of the parties, the conditions and the exceptions under which these rights and obligations to take effect. The corpus was built based on 46 English contracts and 25 Japanese contracts drafted by lawyers. We explain how we annotated the corpus and the statistics of the corpus. We also report the results of the experiments for recognizing rights and obligations.

Keywords: contract, legal document, structuring text, information extraction, document understanding

1. Introduction

A contract is a legal document that outlines the agreements between two or more parties. It states the rights and the obligations of each party. These statements legally bind the parties. Therefore, a contract that contains imprecise statements may result in a lawsuit that costs a great deal of time and money. To prevent such trouble, many companies hire professionals, such as in-house lawyers, who are responsible for drafting and reviewing contracts. When a legal worker reviews a contract, he or she often pays attention to the following issues: (1) whether the contract endows a desirable right to his/her party and (2) whether the contract incurs unduly heavy obligations on his/her party. Precisely understanding these issues is, however, often a time-consuming task. The interest in computer-assisted *contract-review assistants* is growing in the area of legal tech to mitigate the cost of reviewing a contract.

A contract-review assistant applies a natural language processing (NLP) methodology to help a legal worker to understand the semantics of a contract. However, there has been little investigation into NLP specialized for legal documents such as contracts. One of the main challenges is understanding the endowed *rights* and incurred *obligations* in a contract, which is paramount in the contract review process, as we mentioned above.

As a step toward an NLP-based method for recognizing the rights and the obligations described in a legal document, in this paper, we present our attempt at building an annotated corpus of contracts. Building a contract corpus is difficult unless the creators are familiar with legal affairs. Our corpus consists of contracts drafted by lawyers with annotations on the legal semantics of the contracts.

Our corpus has annotations in the contract text to indicate the spans of the following expressions: (1) parties involved in the contract, (2) rights endowed to a party, (3) obligations endowed to a party, (4) conditions for these rights and obligations to take effect, and (5) exceptions of conditions for these rights and obligations to take effect. We defined an annotation standard and asked two annotators to annotate contracts in English and Japanese. To evaluate the effectiveness of our corpus, we conducted a preliminary experiment in which we trained a well-known BiLSTM-CRF model for sequence labeling problems that automatically recognizes the spans of word sequences for rights and obligations in a contract. We devised another module based on the machine learning technique to connect each right or obligation to a party.

The remainder of this paper is organized as follows: In Section 2., we review work related to the present paper; In Section 3., we briefly explain the general structure of a typical contract document; In Section 4., we describe the annotation language and the guidelines that we used during the building of our corpus; In Section 5., we present the detailed statistics of the corpus; In Section 6., we report the results of the experiment that we conducted using our corpus; and after presenting an envisaged application of the corpus in Section 7., we conclude the paper in Section 8..

2. Related Work

The legal domain is a recent target for NLP. However, there is a limited number of studies on the application of NLP to contracts. In this section, we introduce existing work on NLP for legal documents including contracts.

2.1. Recognition of Rights and Obligations

There have been several attempts at recognizing rights and obligations (Glaser et al., 2018; O' Neill et al., 2017; Chalkidis et al., 2018). However, there are several differences between our research and these studies. First of all, we specify an annotation standard to build a corpus. Second, the existing approaches are based on sentence classification, whereas our approach is based on the extraction of spans that consist of word sequences. Third, we also build a corpus so that we can associate relationships among spans, such as that between parties and rights.

2.2. Information Extraction from Contracts

Information extraction from contracts is important because reviewers of a contract have to understand a great deal of information, such as the execution date, jurisdiction, and governing law. There are several studies concerning information extraction from contracts.

In (Chalkidis et al., 2017), they defined 11 contract element types and proposed information extraction based on a hybrid approach that combines rule-based one and classification-based one; their approach used a sliding-window method with word embedding, SVM, and logistic regression. In (Chalkidis and Androutsopoulos, 2017), they proposed an approach based on deep learning; they applied BiLSTM to the same dataset used in the former research (Chalkidis et al., 2017) and showed effectiveness of this approach. In (Chalkidis et al., 2019), they compared several neural networks such as BiLSTM, dilated-CNNs, Transformers, and BERT for the same tasks.

2.3. Building a Corpus of Legal Documents

The main purpose of our study is to build a corpus. Therefore, the studies concerning annotation for legal documents, which we discuss in this section, are related to ours.

There have been several studies on annotating legal text. In (Nazarenko et al., 2018), legal documents were annotated as XML compliant documents using LegalRuleML for the purpose of semantic search. This study is related to our research because its annotation included obligations, permissions, prohibitions, and rights, and the annotation target was legal documents.

In (Kríž et al., 2016; Kríž and Hladká, 2018), the Czech Legal Text Treebank was built, which included annotations of morphologically and syntactically annotated sentences for documents from the Collection of Laws of the Czech Republic. In the later paper, the layer of semantic relation was introduced and the relation was represented by three types of links: definitions, rights, and obligations.

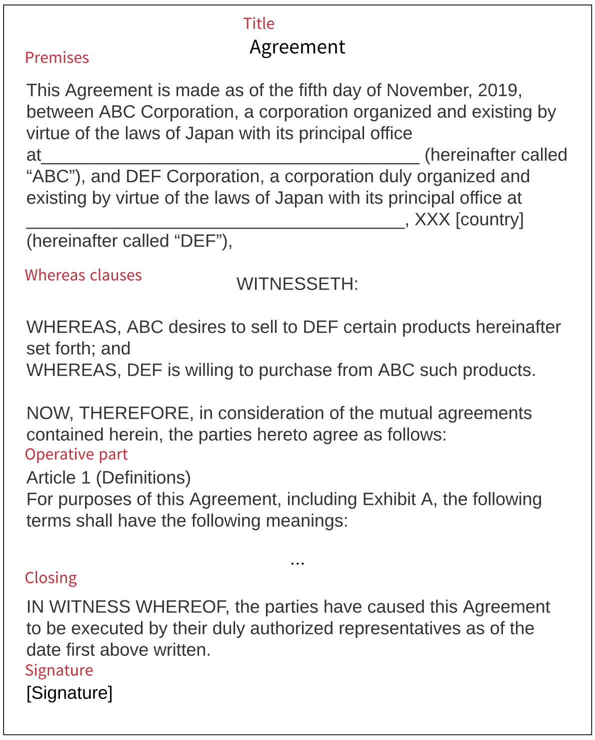
3. Contract

In this section, we briefly review the typical structure of a contract and the content written in the contract.

3.1. Structure of a Contract

The vast majority of contracts do not have a pre-determined format. More specifically, according to the principle of the *freedom of contract*, the format of a contract can be freely determined by the parties. Despite this, as a matter of practice, many contracts tend to follow a consistent format.

In our study, we use two languages: English and Japanese. There are some differences between the structure of an English contract and that of the a Japanese contract. Figure 1 shows the typical structure of an English contract, which is structured as follows. An English contract often starts with a *title* followed by *premises*, *whereas clause*, *operative part*, *closing*, *signature*, and *appendix*. We explain each component below.


Title

Agreement

Premises

This Agreement is made as of the fifth day of November, 2019, between ABC Corporation, a corporation organized and existing by virtue of the laws of Japan with its principal office at_______________________________________ (hereinafter called "ABC"), and DEF Corporation, a corporation duly organized and existing by virtue of the laws of Japan with its principal office at ______________________________________, XXX [country] (hereinafter called "DEF"),

WITNESSETH:

WHEREAS, ABC desires to sell to DEF certain products hereinafter set forth; and

WHEREAS, DEF is willing to purchase from ABC such products.

NOW, THEREFORE, in consideration of the mutual agreements contained herein, the parties hereto agree as follows:

Article 1 (Definitions)

For purposes of this Agreement, including Exhibit A, the following terms shall have the following meanings:

...

IN WITNESS WHEREOF, the parties have caused this Agreement to be executed by their duly authorized representatives as of the date first above written.

[Signature]

Figure 1: Structure of a contract.

Title: The *title* is written as a noun phrase (e.g., non-disclosure agreement) that briefly describes the contract.

Premises: The *premises* determine the effective date and define the parties involved in the contract. Their addresses and the governing law are also included. In the corpus, the parties are annotated.

Whereas clauses: *Whereas clauses*, which are mainly observed in English contracts, explain the purpose, motivation, and background of the contract. They are sometimes called *recitals*. At the bottom of this component of the contract, consideration, which is a concept of English common law, is often written.

Operative part: The *operative part* describes the main content of the contract. Typically, a *section*, *article*, and *clause* are located at the head of the line. This component also includes definitions and general provisions. In this part, the rights and obligations of each party are defined; therefore, this part is our main target for annotation.

Closing: The closing phrase is written here.

Signature: The parties place their signatures here.

3.2. Features of a Contract as a Language Resource

A contract is a peculiar document and different from other text resources in the following aspects.

- The content is written precisely. Ambiguous expressions tend to be avoided.

Label	Description
P	Party
R	Right
O	Obligation
C	Condition
E	Exception

Table 1: Label list.

- There are expressions of dynamic term definition.
  - There is a declaration part for the parties, at the top of the contract.
  - Some keywords that are often used throughout the document are defined in the operative part.
- Coordination expressions (e.g., definition of the rights and obligations of each party) are frequently used.
- The scope of rights and obligations are limited by a condition expression or exception expression.

As described above, some peculiar expressions are often used in a contract. These expressions are primitive compared to those in the other language resources.

Although a contract is written precisely for human beings as described above, the scope of a condition or exception expression is still ambiguous for a computer. That is, many candidates of the spans are modified by such expression. This is challenging for language processing. Therefore, using our corpus, we test methods for contract understanding.

4. Annotations

4.1. Tags

We annotate a contract document with XML-like *tags* using the labels shown in Table 4.1.. The grammar of the tags is as follows:

$$
\begin{array}{rcll}
i, j, k & \in & \mathbb{N} & \\
t & ::= & \langle tn \rangle \mid \langle / tn \rangle & \text{(tags)} \\
tn & ::= & \mathrm{P}i & \text{(parties)} \\
 & \mid & \mathrm{R}j\text{-}p & \text{(rights)} \\
 & \mid & \mathrm{O}k\text{-}p & \text{(obligations)} \\
 & \mid & \mathrm{C}\text{-}rop & \text{(conditions)} \\
 & \mid & \mathrm{E}\text{-}rop & \text{(exceptions)} \\
p & ::= & \mathrm{P}j \mid \mathrm{P}j\text{-}p & \\
rop & ::= & \mathrm{R}j \mid \mathrm{O}i \mid \mathrm{R}j\text{-}rop \mid \mathrm{O}j\text{-}rop &
\end{array}
$$

A tag is either an *open tag* $\langle tn \rangle$ or a *close tag* $\langle / tn \rangle$, where tn represents the *tag name*. A tag name indicates the type of information carried in the text enclosed by the pair of open–close tags; we call the text enclosed by tags *content*. A nested structure and range duplication are not allowed. Each tag name corresponds to the *parties* involved in the annotated contract; *rights* endowed to parties; *obligations* incurred to parties; *conditions* for rights to be exercised or obligations to be incurred; or *exceptions* for rights and obligations. We explain the meaning of each tag name in detail below.

> This Agreement is made as of the fifth day of November, 2019, between <P1>ABC Corporation</P1>, a corporation organized and existing by virtue of the laws of Japan with its principal office at ____ (hereinafter called " ABC "), and <P2>DEF Corporation</P2>, a corporation duly organized and existing by virtue of the laws of Japan with its principal office at ____, XXX [country] (hereinafter called " DEF "),

Figure 2: Example of the annotation of parties.

> The Administrator may <R6-P2>participate in and assume the defense and settlement of a proceeding at its expense</R6-P2>.

Figure 3: Example of the annotation of rights.

4.1.1. Parties

A contract is signed by multiple stakeholders; we call a stakeholder who is involved in a contract a *party*. For example, if a non-disclosure agreement is signed between ABC corporation and DEF corporation, then ABC corporation and DEF corporation are parties.

It is often the case that a contract designates a denoting term for a party (e.g., “seller”, “buyer”, “provider”, and “receiver”). Although these terms denote a party in the contract, we do not treat them as parties when annotating a contract document.

We annotate a party that appears in a contract using a pair of open–close tags whose tag names are Pi, where i, which is called an *ID*, is a natural number. We use the natural number i to distinguish different parties. IDs are assigned in the order of appearance in the contract; the first party is assigned ID 1, the second is assigned ID 2, and so on. IDs are used in the remainder of the contract to refer to a party. In the example of the above non-disclosure agreement, the first party that appears in the agreement (say, ABC corporation) is annotated as $\langle$P1$\rangle$ABC corporation$\langle/$P1$\rangle$. If the second party is DEF corporation, then it is annotated as $\langle$P2$\rangle$DEF corporation$\langle/$P2$\rangle$. Figure 2 shows an actual example of a contract document annotated with Pi tags.

4.1.2. Rights

In a contract, *rights* are designated typically following keywords represented by, for example, *may* or *is entitled to*. We annotate the part of a contract in which a right is endowed to parties using the tag Rj-p, where p is a hyphen-connected list of Pi that denotes a set of parties. Specifically, the text enclosed by a pair of open–close tags with the name Rj-p endows some rights to the parties denoted by p. The ID j is added to this tag to distinguish the different rights given to the parties p; this ID j may be referred to when we annotate conditions and exceptions for this right to be exercised (see Sections 4.1.4. and 4.1.5.). Figure 3 is an example of an actual annotation.

4.1.3. Obligations

In a contract, *obligations* are typically designated following keywords represented by, for example, *shall*, *will*, or *must*. Our corpus also annotates the text in which obliga-

> The Consultant shall <O2-P1>perform the Services in a timely and professional manner consistent with industry standards</O2-P1>.

Figure 4: Example of the annotation of an obligation.

> Target and Acquirer will <O7-P1-P2>use their best efforts to maintain and preserve its business organization, employee relationships, and goodwill intact</O7-P1-P2>, and will <O8-P1-P2>not enter into any material commitment</O8-P1-P2> <E-O8>except in the ordinary course of business</E-O8>.

Figure 5: Example of the annotation of an obligation that depends on multiple obligations.

tions are incurred to parties. The text enclosed by a pair of open–close tags with the name Ok-p incurs an obligation to the parties p. The ID k is used to distinguish different obligations, which may be referred to from the annotations for conditions and exceptions. Figure 4 shows an example of the actual annotation. Additionally, Figure 5 is another example in which the obligation depends on multiple parties.

4.1.4. Conditions

Some of the rights and the obligations specified in a contract are often subject to certain *conditions* under which they are effective. These are described using keywords represented by, for example, *if*, *when* or *in the event that*. For example, in a European call option contract, the right to buy some assets is endowed *at a certain time in the future*. Annotating the part of a text that specifies these conditions is crucial for understanding a contract.

We use a tag whose name is C-*rop* for annotating conditions, where *rop* is a hyphen-connected list of Rj and Ok. It denotes the set of rights and obligations specified earlier; we define condition tags (and the exception tags explained in Section 4.1.5.) so that it can refer to a *set* of rights and obligations rather than a single right or obligation. This design is used because a single part often specifies a condition that is related to multiple rights and obligations in a contract. Figure 6 shows an example of an actual annotation of conditions.

Figure 7 shows an additional example of the actual annotation of conditions, which has multiple conditions for a single obligation.

4.1.5. Exceptions

A contract often uses *exceptions* for rights and obligations. Typically, exceptions are described using keywords such as *except for* or *unless*. To annotate exceptions specified in a contract, we designate a tag E-*rop* where *rop* denotes a set of IDs for rights and obligations. The text enclosed by a pair of open–close tags with the name E-*rop* mentions an exception to the definitions of the rights and obligations denoted by *rop*. Figure 8 shows an actual example of annotating exceptions.

> <C-O45-O46>In the event that the Service Provider infringes or is likely to infringe the intellectual property rights of other third parties</C-O45-O46>, the Service Provider shall <O45-P2>immediately notify the Company thereof</O45-P2> and <O46-P2>resolve such matter at its own risks and expenses</O46-P2>.

Figure 6: Example of the annotation of a condition.

> The obligations of the Issuer to <O8-P1>consummate the transactions contemplated by this Agreement shall be subject to fulfillment of the following conditions on or prior to the date of Closing</O8-P1>:
> (a) <C-O8>The representations and warranties of the Investor set forth in Article 3 shall be true and correct on and as of the date of Closing</C-O8>.
> (b) <C-O8>All proceedings, corporate or otherwise, required to be taken by the Investor on or prior to the date of Closing in connection with this Agreement, and the Debt Exchange contemplated hereby, shall have been duly and validly taken, and all necessary consents, approvals or authorizations required to be obtained by the Investor on or prior to the Closing shall have been obtained</C-O8>.
> (c) <C-O8>The Investor shall have delivered the Notes and evidence of the Advances to the Issuer for cancellation</C-O8>.
> (d) <C-O8>The Investor shall have delivered to the Issuer such other documents, certificates or other information as the Issuer or its counsel may reasonably request</C-O8>.

Figure 7: Example of the annotation of a condition that has multiple conditions for a single obligation.

Remark 1 (Comments) *Certain parts (e.g., titles and headers) of a contract document are not relevant to the rights and obligations of the parties. To allow an annotator to comment out such a part, our annotation language also provides a syntax to comment out text. The comment symbol is denoted by # and it represents as ignorance to the end of the line.*

4.2. Guidelines for Annotation

To prevent an annotation from fluctuating depending on the annotator, we define the following guidelines.

1. The content of a right and an obligation must not include the subject of a phrase.
2. The content of a right and an obligation must include all the information, for the text to be understandable, but must be as minimal as possible.
3. The content of a right and an obligation must include at most one verbal phrase; if several verbal phrases are used in conjunction, then each phrase must be annotated by a single tag.
4. If a negative phrase is annotated, then the negative expression (e.g., "not") must be included in the annotated text.
5. The content must not include multiple sentences in principle. Such an annotation that includes multiple

> The Grantee may <O3-P2>not transfer Units or any rights hereunder to any third party other than by will or the laws of descent and distribution</O3-P2>, <E-O3>except for transfers to a beneficiary or as otherwise permitted and subject to the conditions under Section 5.4 of the Plan</E-O3>.

Figure 8: Example of the annotation of exceptions.

> The Provider shall <O14-P2>not, except for the use of the Company, its Subsidiaries or Affiliates, use, copy or duplicate any of the aforementioned documents or objects</O14-P2>, <O15-P2>nor remove them from the facilities of the Company or such Subsidiaries or Affiliates</O15-P2>, <O16-P2>nor use any information concerning them except for the benefit of the Company, its Subsidiaries and or Affiliates, either during the Engagement Term of the Agreement or thereafter</O16-P2>.

Figure 9: Example of an annotation that includes *nor*.

sentences is allowed only if the annotated text is not understandable within a single sentence. When verbs connected with *and* share the same subject, the tag is not split.

6. For the case in which multiple rights and obligations are connected with "nor", later content includes *nor* (Figure 9).

7. The content must not include a period at the end of a sentence.

These guidelines are for English annotation, but we also created guidelines for Japanese annotation. Although, Japanese grammar differs from English grammar in many ways, the guidelines were created in a similar manner.

5. Annotated Corpus

5.1. Statistics for the Corpus

We annotated 46 English contracts and 25 Japanese contracts following the manner described above under the direction of lawyers. Table 2 shows the statistical information of the corpus.

5.2. Interesting Examples

While building the corpus, some notable annotations were found. We present some of them here.

Figure 10 shows an interesting example in which a condition defined by a C tag is augmented later. In this example, the condition annotation marked by C-O11-O12 defines a condition on the obligation marked by O12-P2, whose condition is augmented by the annotation C-O12.

Figure 11 is a difficult case. The *indemnifying party* who has the obligation marked by O15 is not determined at the time that the contract is executed. Therefore, all possible parties are listed as O15-P1-P2-P3.

Figure 12 shows the case of a passive sentence. The subject is often omitted in a passive sentence. This is the same case as that above. Everyone can clearly understand from the sentence that the company should pay a bonus to the employee. This is why the subject is omitted.

> <C-O11-O12>If the Employee is prevented from work due to sickness or accident</C-O11-O12>, he shall <O11-P2>inform the Employer without delay</O11-P2> and shall <C-O12>at the Employer's request</C-O12> <O12-P2>submit a medical certificate within three work days</O12-P2>.

Figure 10: Example of condition augmentation.

> <C-O15>As long as the indemnifying party is defending any such claim in good faith</C-O15>, the indemnified party will <O15-P1-P2-P3>not settle such claim</O15-P1-P2-P3>.

Figure 11: Example of undetermined parties.

6. Experiment

We explain the process of rights and obligations recognition using our annotated corpus. In this section, we only use the English corpus for the experiment.

We separate the prediction steps into three parts. First, we extract the spans that correspond to parties, rights, obligations, conditions, and exceptions (see Figure 13). Next, we associate the rights and obligations of reference relations with parties (see Figure 14). Then, we associate the rights and obligations of reference relations with conditions and exceptions (see Figure 15). Through these steps, we achieve the recognition of which parties have what rights and obligations are given in what cases is achieved. We explain the three steps below.

6.1. Extraction of Spans

The first step is the extraction of spans, which correspond to parties, rights, obligations, conditions, and exceptions. As illustrated in Figure 13, given the contract text input, the system performs word segmentation and converts words to the word embedding. It also removes stop words and low-frequency words, and normalizes numbers. Then, we train sequences of combinations between a word and a BIO tag (Ramshaw and Marcus, 1995) concerning the labels of parties, rights, obligations, conditions, and exceptions (see Table 4.1.) using BiLSTM-CRF (Huang et al., 2015; Lample et al., 2016), where the dimension of the hidden layers is 256 and the dimension of the input embedding vector is 128. From the dataset, we used 32 contracts for training and 14 contracts for testing.

6.2. Association of Parties with Rights and Obligations

The second step is the association of parties with rights and obligations, which is illustrated in Figure 14. The system determines which parties with rights and obligations are associated from extracted spans in the prior step. Let $N_p \in \mathbb{N}$ denote the number of parties, $N_r \in \mathbb{N}$ denote the number of rights, and $N_o \in \mathbb{N}$ denote the number of obligations. The associations of parties with rights are a many-to-many

Language	#Contracts	#Sent.	#P	#R	#O	#C	#E
English	46	6754	98	311	1050	498	72
Japanese	25	1994	49	92	438	268	7

Table 2: Statistics for our corpus.

<O8-P1>The remaining 50% of any positive Quarterly Bonus amount will be held in a reserve account for the Employee</O8-P1>.

Figure 12: Example of a case in which the subject is omitted.

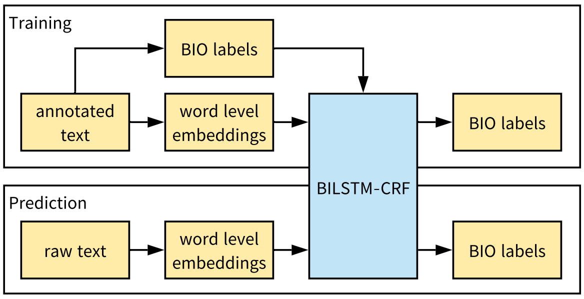


Figure 13: Flow of span extraction.

mapping. Therefore, all $N_p \times N_r$ patterns are predicted using binary classification. Likewise, $N_p \times N_o$ patterns of associations of parties with obligations are predicted in the same manner.

To achieve this, we calculate vectors as features of the parties from tagged contract text. We include the information in the surrounding text to a *party vector* because the context information represented by the word sequences before and after the text of the party is important. We also calculate *rights / obligations vectors* so that they the information of the surrounding text is included due to the similar reason. We prepare these vectors by concatenating three vectors, that is, (1) a vector generated from text before the open tag, (2) a vector generated from inner text between the open tag and close tag, and (3) a vector generated from text after the close tag. Each vector is calculated as the average of the word embeddings, which is calculated using word2vec (Mikolov et al., 2013).

Then, we combine these two vectors as a *joint vector*. From the input of the joint vector, we train binary classification using Logistic Regression.

6.3. Association of Conditions and Exceptions with Rights and Obligations

We associate conditions and exceptions with rights and obligations (see Figure 15). Let $N_c \in \mathbb{N}$ denote the number of conditions, $N_e \in \mathbb{N}$ denote the number of exceptions, $N_r \in \mathbb{N}$ denote the number of rights, and $N_o \in \mathbb{N}$ denote the number of obligations. The association of conditions and exceptions with rights and obligations is a many-to-many mapping respectively. Therefore, there are $N_c \times N_r$ patterns for associations of conditions with rights, $N_c \times N_o$ patterns for associations of conditions with obligations, $N_e \times N_r$ patterns for associations of exceptions with rights, and $N_e \times N_o$ patterns for exceptions and obligations calculated using binary classification.

Label	Precision	Recall	F1
P	0.99	0.65	0.79
R	0.87	0.61	0.72
O	0.90	0.74	0.81
C	0.92	0.80	0.85
E	0.42	0.42	0.42

Table 3: Span extraction score.

The number of these patterns is large. Therefore, the candidates are narrowed down using heuristics. Generally, associated pairs, such as a pair of conditions and rights, are located close to each other. Using this knowledge, we only use the candidate pairs that are close to each other as the text location.

The vectors of conditions and exceptions are also calculated from the surrounding text as well. From the input of the joint vector, we train binary classification with logistic regression and recognize the reference relationship.

6.4. Evaluation

To evaluate the performance of the model, we perform training in each step.

6.4.1. Results of Span Extraction

We summarize the span extraction scores for the first step in Table 3. Each score in the table is the average score for the Begin (B) label and Inside (I) label in BIO tagging.

Whereas some scores, such as precision scores, were over 0.9, the precision and recall of exception extraction were relatively low. Moreover, the recall was equal to or lower than the precision for all labels. As the prior step of the association, the recall score is important because missing extraction leads to no input data for association in subsequent recognition flow. In particular, improving the recall of party extraction is crutial due to its importance in contract reviewing. We expect the results can be improved by adding additional annotated data.

6.4.2. Results of the Association of Parties with Rights and Obligations

The results for recognition flow for the second step, which is the association of parties with rights and obligations, are shown in Table 4. In the table, P-R in the association column denotes the association between parties and rights. P-O denotes the association between parties and obligations.

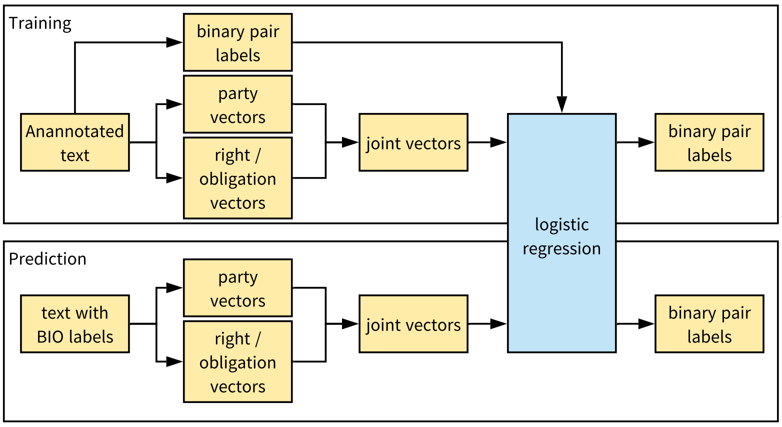


Figure 14: Flow of the association of parties with rights and obligations.

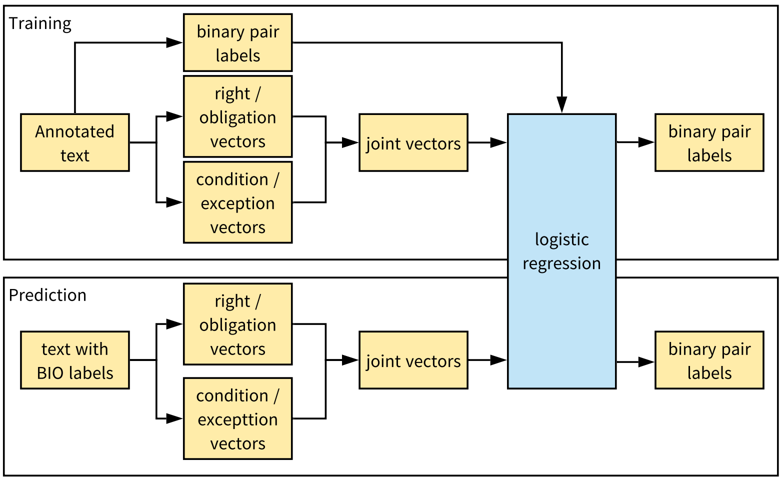


Figure 15: Flow of the association of rights and obligations with conditions and exceptions.

Association	Precision	Recall	F1
P-R	0.77	0.92	0.84
P-O	0.87	0.96	0.91

Table 4: Scores of the association of parties with rights and obligations. P-R in the association column denotes the association between parties and rights. P-O denotes the association between parties and obligations.

Association	Precision	Recall	F1
C-RO	0.93	0.96	0.94
E-RO	0.93	1.00	0.96

Table 5: Scores of the association of conditions and exceptions with rights and obligations. C-RO in the association column denotes the association of conditions with rights and obligations. E-RO denotes the association of exceptions with rights and obligations.

The results show the score exceeded the chance rate of 0.5, where the score is calculated not for total prediction, but for each prediction in binary classification in many-to-many mapping. The recall score for both P-R association and P-O association was higher than 0.9.

6.4.3. Results of the Association of Conditions and Exceptions with Rights and Obligations

The results for recognition flow for the third step, which is the association of conditions and exceptions with rights and obligations, are shown in Table 5.

In the table, C-RO in the association column denotes the association of conditions with rights and obligations. E-RO denotes the association of exceptions with rights and obligations.

Compared with the association of parties with rights and obligations, the entire score is relatively higher. This was due to the heuristics for pruning. Although the score is sufficiently high, the data size is still small. Therefore, the result may change if we increase the amount of the data.

6.5. Summary of the Experiment

We conducted an experiment and evaluated the performance of models using the three steps in the recognition flow. We evaluated the result of each step in the flow. We

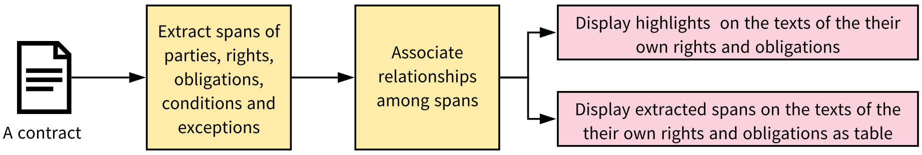

A contract
Extract spans of parties, rights, obligations, conditions and exceptions
Associate relationships among spans
Display highlights on the texts of the their own rights and obligations
Display extracted spans on the texts of the their own rights and obligations as table

Figure 16: Application image.

plan to conduct an evaluation of end-to-end recognition in future.

7. Application

In this paper, we described an automatic recognition module for rights and obligations with conditions and exceptions for each party in a contract. This module has a direct application that consists of the following functions (see Figure 16):

1. Highlighting spans that describe the rights and obligations of a specified party using some colors that depend on the type, where the colors for obligations for the party and rights for the counterpart may be similar, such as dark red and light red; and
2. Generating a table that summarizes the rights and obligations of a specified party.

These functions will help lawyers (1) understand the contract easily and quickly and (2) write rights and obligations appropriately.

The results of the module may be passed to further NLP/AI functions, such as sorting the rights and obligations in order of degree and suggesting other word sequences to adjust the degree.

8. Conclusion

In this paper, we presented our attempt at building an annotated corpus of contracts with the aim to recognize the rights and the obligations in contracts. Our corpus has annotations in the contract text to indicate the spans of the following expressions: (1) parties involved in the contract, (2) rights endowed to a party, (3) obligations endowed to a party, (4) conditions for these rights and obligations to take effect, and (5) exceptions of conditions for these rights and obligations to take effect. We defined an annotation standard and asked two annotators to annotate contracts in English and Japanese.

To evaluate the effectiveness of our corpus, we tested a well-known sequence labeling method that automatically recognizes spans of the word sequence and its label. We devised another module based on connecting each right or obligation to a party. We obtained relatively high recall and precision scores for the training data size.

Our corpus allows researchers interested in legal documents to test various NLP techniques, such as an authoring system. Researchers can also attempt to apply artificial intelligence methods after NLP, such as reasoning using the deontic logic.

Bibliographical References

Chalkidis, I. and Androutsopoulos, I. (2017). A deep learning approach to contract element extraction. In *Proceedings of the International Conference of Legal Knowledge and Information Systems (JURIX)*, pages 155–164.

Chalkidis, I., Androutsopoulos, I., and Michos, A. (2017). Extracting contract elements. In *Proceedings of the International Conference on Artificial Intelligence and Law*, pages 19–28.

Chalkidis, I., Androutsopoulos, I., and Michos, A. (2018). Obligation and prohibition extraction using hierarchical RNNs. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 254–259.

Chalkidis, I., Fergadiotis, M., and Malakasiotis, P. (2019). Neural Contract Element Extraction Revisited. In *Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS)*.

Glaser, I., Scepankova, E., and Matthes, F. (2018). Classifying semantic types of legal sentences: Portability of machine learning models. *Proceedings of the International Conference on Legal Knowledge and Information Systems (JURIX)*, pages 61–70.

Huang, Z., Xu, W., and Yu, K. (2015). Bidirectional LSTM-CRF Models for Sequence Tagging. *CoRR abs/1508.01991*.

Kríž, V. and Hladká, B. (2018). Czech legal text treebank 2.0. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC)*, pages 4501–4505.

Kríž, V., Hladka, B., and Urešová, Z. (2016). Czech legal text treebank 1.0. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC)*, pages 2387–2392.

Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., and Dyer, C. (2016). Neural architectures for named entity recognition. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT)*, pages 260–270.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. In *Proceedings of Advances in neural information processing systems (NIPS)*, pages 3111–3119.

Nazarenko, A., Levy, F., and Wyner, A. (2018). An annotation language for semantic search of legal sources. In *Proceedings of the 11th International Conference* *on Language Resources and Evaluation (LREC)*, pages 1096–1100.

O' Neill, J., Robin, C., Buitelaar, P., and O' Brien, L. (2017). Classifying sentential modality in legal language: A use case in financial regulations, acts and directives. In *Proceedings of the International Conference on Artificial Intelligence and Law*, pages 159–168.

Ramshaw, L. and Marcus, M. (1995). Text Chunking using Transformation-Based Learning 1 Introduction. *ACL third Workshop on Very Large Corpora*, pages 82–94.