Towards Construction of Legal Ontology for Korean Legislation

Thi Thuy Phan[1,*], Ho-Pun Lam[2], Mustafa Hashmi[2,3] and Yongsun Choi[4]
[1]*CMC Ciber, Vietnam*
[2]*Data61, CSIRO, Sydney, Australia*
[3]*La Trobe Law School, La Trobe University, Melbourne, Australia*
[4]*Inje University, Gimhae, Republic of Korea*

Keywords: Korean Legislation, Legal Taxonomy, Legal Norms Classification, Semantic Types.

Abstract: Automating information extraction from legal documents and formalising them into a machine understandable format has long been an integral challenge to legal reasoning. Most approaches in the past consist of highly complex solutions that use annotated syntactic structures and grammar to distil rules. The current research trend is to utilise state-of-the-art natural language processing (NLP) approaches to automate these tasks, with minimum human interference. In this paper, based on its functional aspects, we propose a legal taxonomy of semantic types in Korean legislation, such as definitional provision, deeming provision, penalty, obligation, permission, prohibition, etc. In addition to this, a NLP classifier has been developed to facilitate the automated legal norms classification process and an overall F_1 score of 0.97 has been achieved.

1 INTRODUCTION

The legislation that we have nowadays is not simply a corpus of legal documents. It contains lots of information that needs to be interpreted, explained, and processed in order to determine whether an organisation complies with legislative requirements. However, working with legal documents can be both costly, time-consuming and error-prone, as it requires domain experts to understand what to be expected from the legislations with respect to its interpretation and intents.

Over the years, much research has been focused on representing information captured inside legal documents into machine understandable formalisms so that we can reason on and make sense of it using a computer, and various promising results have been obtained (Ceci et al., 2016; Lam and Hashmi, 2019).

Recently, the research focus has been shifted to the task of applying natural language processing (NLP) techniques to generate legal norms from legal documents with some success (van Engers et al., 2004; Wyner and Peters, 2011; Dragoni et al., 2015; Sleimi et al., 2018). However, most of these approaches consist of highly complex solutions that utilise annotated syntactic structures and grammar to automatically distil rules. Recently, Ferraro et al. (2019) have evaluated several state-of-the-art NLP approaches to automate the normative mining process and have identified several issues such as different types of lexical ambiguities, inconsistent use of terminologies, sentential complexities, cross-referencing between different provisions, etc., that hinder the developments in this area.

Nevertheless, at the core of these technologies is an ontology that defines the underlying *principles*, *concepts*, *assumptions*, and *legal effects* of terms, i.e., the legal taxonomy that are commonly used in a legal domain. It classifies the terms into different categories and defines their interrelations, such as whether a term is *subsumed*, *equivalent*, or in *conflict* with another term. It is a foundation stone that can facilitate the development of automated *legal analysis* and automatic *machine translation*.

The Language for Legal Discourse (LLD) (McCarty, 1989) is a first attempt to define legal knowledge in the context of legal reasoning. Since then, many different legal ontologies have been developed for a range of purposes. For instance, Legal Knowledge Interchange Format (LKIF) Core ontology (Hoekstra et al., 2007) provides the basic set of concepts of law, such as the meaning of *norm*, *liability* and *legal fact*, etc., as the basis for knowledge

*This work was done during the time when the first author was a Master student at Inje University, Republic of Korea.

Phan, T., Lam, H., Hashmi, M. and Choi, Y.
Towards Construction of Legal Ontology for Korean Legislation.
DOI: 10.5220/0010122400860097
In *Proceedings of the 12th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2020) - Volume 2: KEOD*, pages 86-97
ISBN: 978-989-758-474-9

acquisition and modelling in the legal domain. It aims to limit the set of terminologies used in LKIF applications.

Several European projects such as LYNX,[1] SPIRIT,[2] and MARCELL[3] have dedicated efforts for creating legal knowledge graphs and multilingual legal ontologies for automatically linking and translating heterogeneous legal sources such as laws, decrees, regulations, facilitating thus enterprises to remove their legal and language barriers in trade and to localise their products and services.

As can be seen above, the research efforts in this area target mostly Indo-European languages. In South Korea, the work related in this area is still in its incubation stage. Regulation technologies (RegTech) and their related products have started gaining attention from the government only until 2018.[4] Several word characters and knowledge representation (Botha and Blunsom, 2014; Cotterell and Schütze, 2015; Wieting et al., 2016), and NLP approaches (Bojanowski et al. (2016); Junho et al. (2010); Stratos (2017)) in Korean language exist, but not much useful and efficient systems have been reported.

Extracting normative information from legal documents is a process that is far from being trivial and intuitive. Legal documents are typically so complex that even human lawyers are having difficulties in understanding and applying them (Wieringa and Meyer, 1993). Thus, works on the automated transformation of Korean legislations into a machine understandable formalism are in high demand.

Hence, the purpose of this paper is to fill the gap in this area by proposing a taxonomy of semantic types for legal norms in the Korean language that can be applied to statutory texts in Korean legislations. The primary challenge to the classification of legal norms lies in the underlying legal theory with empirical observations, which is under-represented in Korean legal sciences. It is the foundation of many legal analysis and interpretation tasks, and not much work has been reported by the Korean legal informatics community.

The rest of this paper is organised as follows. An informal introduction and problems related to the Korean language will be described in Section 2. Sections 3 and 4 present the taxonomy of legal norms that we have developed on Korean legislation and the preliminary evaluation results of the taxonomy, respectively. Section 5 presents the related works, followed by conclusions and pointers to future research.

[1]LYNX: http://lynx-project.eu/

[2]SPIRIT:https://www.spirit-tools.com/

[3]MARCELL: http://marcell-project.eu/

[4]Fintech In South Korea: Regulators Step In To Boost Innovation: https://fintechnews.hk/4823/fintechkorea/fintech-south-korea-regulators-step-boost-innovation/

2 BACKGROUND

Technically, a taxonomy typically refers to a hierarchical arrangement of terminologies that describes a particular branch of science or field of knowledge (McGregor, 2005). A legal taxonomy, in addition to this, reflects also the culture and history of a given legal system. As commented by Mattei (1997), it is the product of interactions of the legal tradition and that of the new sensibilities. It provides a means where people working in the legal sector can communicate with each other, to discuss problems and exchange ideas of mutual concern among themselves.

However, creating a legal taxonomy that accurately reflects the legislations and to avoid misattribution errors is not an easy task. It requires the terms selected and arranged to be mutually exclusive, thus a unique ordered structure for different terms can be created (McGregor, 2005).

2.1 Problems with Korean Language

There are a few phenomena that make NLP in Korean language a challenging task to accomplish.

Firstly, Korean has traditionally posed challenges for word segmentation and morphological analysis (Matteson et al., 2018). This is because Korean is a phonetic language with a subject-object-verb (SOV) syntax while permitting a high degree of freedom in word order (Jeong et al., 2007). In fact, Korean is a left-branching language such that the head that determines the correct phrasal category comes at the end of a phrase (Müller-Gotama, 1994). For a noun phrase that is compatible with a higher phrase type, it could be the left-branching daughter of a higher phrase, noun phrase, or verb phrase, which imposes substantial demand for the model being developed (Müller-Gotama, 1994).

Secondly, it also allows multiple concepts to be synthesised into a single *eojeol*, i.e., a Korean spacing unit similar to a word in English. As a result, depending on the context, the same *eojeol* can be analysed into different morpheme which yields different part of speech (POS) tags of morpheme combinations (Song and Park, 2019).

Thirdly, Korean is an agglutinative language such that words may contain different morphemes to determine their meanings. For example, the word "*mountain*" in English can only be derived from itself; whereas in Korean, "산을" (san-eul (mount)), "산은"

(san-eun (saneun)), "산도" (san-do (acidity)), "산이" (san-yi (sanyi)), "산이나" (san-ina (sanna)), etc., can all be derived from the root "산" (san (mountain)) (Lee, 2018). To make things even more complex, Korean also has some special rules that can apply across character boundaries, implying that morphological transformation may also occur among adjacent graphemes.

Over the years, several lexical databases have been developed. For instance, KorLex (Yoon et al., 2009) was developed by translating and mapping the English terms into Korean. Following the idea of lexical concept network (LCN) — a lexical database that provides various information of a word in terms of its relation to other lexical units, Choi et al. (2004) developed ETRI LCN for the Korean language, but for verbs and nouns only. Later, the same group of researchers also established (and maintained) another LCN called UWordMap (Ock, 2013), which consists of 514,314 words, including nouns, adjectives, and adverbs, and is the largest lexical database of its kind. However, all these works are for general purposes only, they do not cater to the needs of the legal domain that requires a more rigid understanding of the legal text.

2.2 Semantic Types

Generally, legal rules governing the behavior of citizens prescribe code of actions that citizens must follow. These codes provide applicability conditions capturing various intuitions in different situations and prescribe on how to act. Several research efforts have been reported providing classifications of legal rules defining the semantic meanings to facilitate properly interpreting of and reasoning about the legal rules. von Wright (1963) classifies legal rules as (i) *determinative rules* (a.k.a. constitutive rules) which define the concepts, or activities that cannot exist without legal rules; (ii) *technical rule* prescribing what needs to be done in order to attain some legal effects, and (iii) *prescriptive* rules that prescribe the actions and making obligatory, prohibited, or permitted regulating thus the behavior of the subject.

Gordon et al. (2009) give an extended catalog of requirements for a formal language necessary for reasoning on legal rules which includes jurisdiction (Mills, 2014), authority, rules validity (Marín and Sartor, 1999), deonticity and defeasibility of rules, normative effects (Rubino et al., 2006), contraposition (Prakken and Sartor, 1996), conflicting (Sartor, 1992) and exclusionary rules (Sartor, 1992; Prakken and Sartor, 1996), and temporal properties.

Hilty et al. (2005) provides a charaterisation of legal norms along temporal bounds and invariants properties capturing the application of norms in the time space. Whereas Hashmi et al. (2013) classify deontic effects of legal norms on the temporal validity aspects (Palmirani et al., 2011). The former provides the mapping of the norms from requirements to the enforcement while latter studies when a norm enters into force, terminated after a deadline what constitutes the violations of a legal norm, and whether a violated norm can be compensated for. Besides, they study the persistent effects of legal norms such that even after being violated a norm may still remain valid until it is performed or terminated.

More recently, Hashmi et al. (2018) discuss a taxonomy of legal terms and concepts aiming at creating a legal ontology and a socio-legal graph for sharing the Australian legal knowledge on the web. Their taxonomy is based on the legal quadrant for rule of law (Casanovas, 2019) comprising the notions of *application* and *implementation* of the rule of law which includes themes such as binding power, social dialog, privacy, trust, security, sanctions, etc., and *sources* for the legal validity (Sartor, 2008) of the legal norms emerging from regulatory dimensions such as hard law, soft law, policies, and ethics covering a range of requirements from various social, political and legal aspects.

3 THE LEGAL TAXONOMY

Failure to properly understand the real meanings of a legal term may result in the misunderstanding of legislation provisions. In practice, the way that we interpret legislations may also affect the outcome of a case.

Technically, legislation can be characterised as a combination of a set of (normative) provisions and the totality of norms that follow from executing that finite set of provisions. In the past, legislations were interpreted mainly based on the plain meanings of the text as derived from the ordinary definitions of an individual word and the overall structure of the statement (Karkkainen, 1994). Contemporary approaches to legislation interpretations focus on determining the original intent of the legislation, i.e., the goals that the legislation intended to achieve. Interpretations will be made in the context of the legislation as a whole where the interpretation of a specific provision should

[5]The parse tree in Figure 1 was generated using the syntax tree generator, Komoran3, available at: http://andrewmatteson.name/psg_tree.htm (last accessed: 21 Jun 2020)

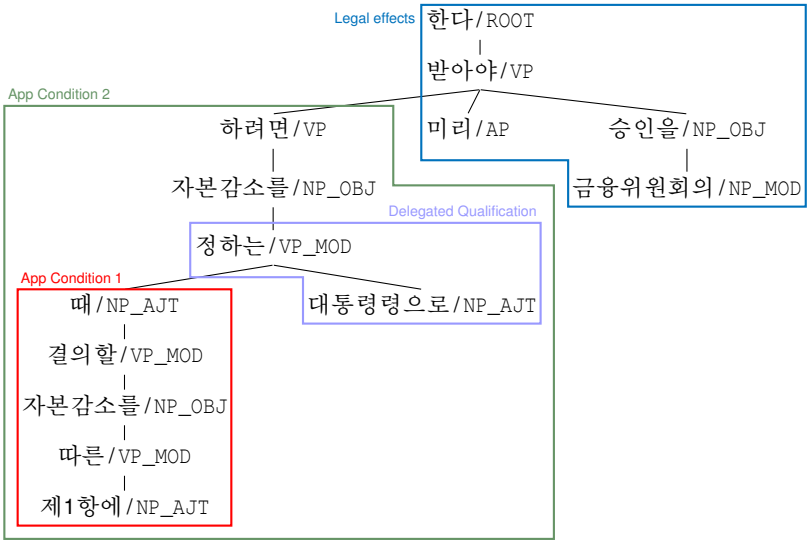

Figure 1: Parse tree of the statement[5]: "제1항에 따른 자본감소를 결의할 때 대통령령으로 정하는 자본감소를 하려면 미리 금융위원회의 승인을 받아야 한다." (statement extracted from the 2nd paragraph of Article 18 in Insurance Business Act).

be determined consistently with respect to other provisions. Hence, to meet the needs of both the legal practitioners and design requirements of the automated machine translation and interpretations process, having a clear definition of terminologies is crucial to avoid any ambiguity and eliminate potential misinterpretations in the legislations.

3.1 The Basic Concepts

In this subsection, we present a basic set of definitions that will be used throughout the entire discussion of the taxonomy of the Korean legislation. These concepts could optionally appear in some semantic types but mandatory to the others.

3.1.1 Applicability Condition

The applicability conditions, a.k.a. preconditions, specify when and under what circumstances a norm becomes applicable (or activated). In Korean language, this can be detected as a phrase or clause that ends with the following words.

- 면 (where), except the case (없으면) (unless, if not): <situation>
- 경우, 경우에는, 경우에도 (where/in case): <situation>
- 할 때 (in dealing with/when) except the case of 때부터 (from the time) and 때까지 (by the time): <situation>
- 자는, 자가 (a person who...): <a qualification for individual, or legal entities>.

3.1.2 Legal Effects

Legal effects are the normative effects that follow from applying a norm, such as *obligation*, *permission*, *prohibition*, and also other articulated effects introduced for the law (see Sartor, 2005; Rubino et al., 2006, for details).

While applicability conditions can be optional, legal effects are a mandatory component of every legal norm.

Figure 1 shows the parse tree for the statement[6]: "제1항에 따른 자본감소를 결의할 때 대통령령으로 정하는 자본감소를 하려면 미리 금융위원회의 승인을 받아야 한다", meaning that "Where a stock company intends to reduce its capital (정하는 자본감소를), as prescribed by Presidential Decree (대통령령으로), in resolving the reduction of its capital under paragraph (1) (제1항에 따른 자본감소를), it shall obtain approval (승인을 받아야 한다) from the Financial Services Commission (금융위원회의) in advance (미리)." It illustrates how the applicability conditions and the legal effects that it inferred (in this case, an *obligation* to obtain an approval beforehand) are written in Korean language and how applicability conditions can be nested together.

From the parse tree, we can also notice that, as Korean is a left-branching language, the legal effects always appear as the *rightmost* component of the tree,

[6]This statement is extracted from the 2nd paragraph of Article 18 in Insurance Business Act. The English translation available from: https://elaw.klri.re.kr/kor_service/lawView.do?hseq=43318&lang=ENG

제86조 (등록의 취소 등)
Article 86 (Revocation of Registration)

① 금융위원회는 보험설계사가 다음 각 호의 어느 하나에 해당하는 경우에는 그 등록을 취소하여야 한다.

1. 제84조제2항 각 호의 어느 하나에 해당하게 된 경우
2. 등록 당시 제84조제2항 각 호의 어느 하나에 해당하는 자이었음이 밝혀진 경우
3. 거짓이나 그 밖의 부정한 방법으로 제84조에 따른 등록을 한 경우
4. 이 법에 따라 업무정지 처분을 2회 이상 받은 경우

Where an insurance solicitor falls under any of the following subparagraphs, the Financial Services Commission shall revoke his or her registration:

1. Where he or she falls under any of the subparagraphs of Article 84 (2);
2. Where he or she is found to fall under any of the subparagraphs of Article 84 (2) as at the time of his or her registration;
3. Where he or she makes a registration under Article 84 by false or other unlawful means;
4. Where he or she is subject to a disposition of business suspension under this Act on at least two occasions.

② 금융위원회는 보험설계사가 다음 각 호의 어느 하나에 해당하는 경우에는 6개월 이내의 기간을 정하여 그 업무의 정지를 명하거나 그 등록을 취소할 수 있다. 〈개정 2014.1.14〉.

1. 모집에 관한 이 법의 규정을 위반한 경우
2. 보험계약자, 피보험자 또는 보험금을 취득할 자로서 제102조의2를 위반한 경우
3. 제102조의3을 위반한 경우

Where an insurance solicitor falls under any of the following subparagraphs, the Financial Services Commission may order him or her to suspend his or her work for the specified period of up to six months, or revoke his or her registration: 〈Amended by Act No. 12262, Jan. 14, 2014〉

1. Where he or she violates the provisions of this Act governing insurance solicitation;
2. Where he or she, as an insurance policyholder, an insured person or a person that is to receive insurance money, violates Article 102-2;
3. Where he or she violates Article 102-3;

...

[전문개정 2010.7.23]
[This Article Wholly Amended by Act No. 10394, Jul. 23, 2010]

Figure 2: Insurance Business Act: Article 86 (Revocation of Registration) (adopted from: http://www.law.go.kr/ 법령/보험업법, English translation available at: https://elaw.klri.re.kr/kor_service/lawView.do?hseq=43318&lang=ENG).

which is the feature that we use in the NLP classifier that we developed, and will be discussed in Section 4.

3.1.3 Cross-referencing

Similar to other jurisdictions, legislation in Korean is divided into parts that promotes clarity for presentation, structure and expression. As described in (Xanthaki, 2014), drafting legislation as this allows legal drafters to demonstrate the intuition behind the legislation, maintain the coherence of the legislative text, and can stress the interrelation between different provisions.

As can be seen from Figure 2, the structure of Korean legislation is comparatively less complex than legislations of other jurisdictions. In general, a legislation may consist of different chapters, which can then be further divided into different articles, (sub)paragraphs, and items, as shown in Table 1.[7] Table 2 shows some commonly used cross-referencing patterns in Korean legislation.[8]

Besides, when referring to other legislations, the name of the referred legislations should be enclosed in square brackets. For instance, "「상법」" will be used when referring to the the "Commercial Act"[9] and "「상법」 제255조제2항" will be used when referring to "Article 255 (2) of the Commercial Act".

3.2 Semantic Types in Korean Legislations

When determining the semantic type of a legal statement, it is the legal effects part that plays an important role. It specifies the normative effects and the order of validity a legal statement has. Typically, such provisions have been made transparent by the use of

[7]Note that the word "제" can have multiple meanings in the Korean language. When used in cross-referencing, it means to "in the current legislation".

[8]Instead of writing "*paragraph #*", in some cases, for simplicity, the paragraph number will be put inside curly brackets next to the article number in cross-referencing.

[9]The word "상법" means "Commercial Act" in Korean language.

Table 1: Section name and useful terms in Korean language.

	Korean	English
Section label	편	Part
	장	Chapter
	조	Article
	항	Paragraph
	목	Item / Point
Useful term	제	in the current legislation or section
	부터	starts from
	까지	ends with
	과, 및	and

Table 2: Cross-referencing Examples.

Example	Example (in English)
이 장	This Chapter
제1조	Article 1
제32조제2항	Article 32 (2)
제2조제8호나목	Article 2 (8) point 2
제1항과 제2항	Paragraph 1 and Paragraph 2
제410조부터 제412조까지	From Article 420 to Article 412
제193조、제252조 및 제531조제2항	Article 193, Article 252 and Article 532 (2)

modal verbs (Höfler, 2019) that appear at the end of the statements. Its usage is similar to the words *shall* and *must* in English legislations which show the natural dispose of a connection to the normative value contained in the provisions, as well as the normative functions of these provisions.

Following a functional classification approach, we have analysed the modal verbs that have been used in the legal statements and identified *eleven* different categories, i.e., semantic types, that appear in Korean legislation: *Definitional provision*, *Application provision*, *Deeming provision*, *Continuation clause*, *Delegation provision*, *Penalty provision*, and different types of *Deontic provisions*, such as *Obligation*, *Liability*, *Rights*, *Permission*, and *Prohibition*. Table 3 show examples of different types of statements extracted from the Insurance Business Act (IBA). In what follows, we are going to elicit on each of these categories.

3.2.1 Definitional Provision

Definitional provisions define commonly used concepts or relevant terms that appear in (and in some cases, specific to) the legislation.

In Korean legislation, it uses "*X* 란/이 란...*Y* 을/를 말한다" to denote the pattern "*X* means *Y*". where *X* is called a *definiendum* which can be a word, a phrase, or a symbol, and is normally enclosed inside double quotes, *Y* is called a *definien* and is used to describe/define the definiendum *X*.

Notice that definitional provisions, in general, do not contain any applicability condition as the terms was defined in a general sense (within the context of the legislation) and, unless otherwise specified, it should be used without any restriction.

3.2.2 Application Provision

Application provisions set out situations or timeframes in which the law, or section(s) of law, applies (적용한다 (shall apply), or 에 따른다 (shall be governed by)), applies with some changes (준용한다 (shall apply *mutatis muntandis*)), or does not apply (적용하지/그러하지 아니한다).

In some cases, an application provision may also be used to specify the statuses (and/or timeframes) of other legislations.

3.2.3 Deeming Provision

A deeming provision indicates something to be deemed or construed ((으)로 본다) as if something else (through cross-referencing) if the two can be construed as the same thing, or the later inherits some qualities that the former does not have. In the same vein, a deeming provision can also be used to indicates something *cannot* be deemed or construed ((으)로 보지 아니한다) as something else.

However, latest research found that a deeming provision may deem things to be what they are not (Bracher, 2018). To resolve this issue, it is mandatory that a deeming provision should always be construed on its own terms under the context concerned and purposes of the legislation.

3.2.4 Continuation Clause

A continuation clause is a provision that is used to extend or limit the scope of application of a precedent legal statement. It is expected that, unless otherwise specified, the legal effects inferred by the continuation clause will be the same as (또한 같다), or applies to the same objects (과 같다) as the original statement.

3.2.5 Delegation Provision

Under normal situations, a person who is vested with a particular statutory power, duty, or function may exercise it himself/herself. However, for the sake of convenience in practice, a power, duty or function may be delegated pursuant to an instrument of delegation through a delegation provision and exercise the

Table 3: Example of semantic types of norms from the Insurance Business Act (the full Act is available at: http://www.law.go.kr/법령/보험업법, English translation available at: https://elaw.klri.re.kr/eng_service/lawView.do?hseq=43318&lang=ENG).

Semantic type	Example (Korean)	Example (translation in English)
Definitional provision	"생명보험업"이란 생명보험상품의 취급과 관련하여 발생하는 보험의 인수, 보험료 수수 및 보험금 지급 등을 영업으로 하는 것을 말한다.	The term "life insurance business" means the business of underwriting insurance, receiving premiums, paying insurance proceeds, etc. which arise in selling life insurance products.
Application provision	상호회사의 입사청약서나 사원에 대한 통지 및 최고(催告)에 관하여는 「상법」 제353조를 준용한다.	Article 353 of the Act on Corporate Governance of Financial Companies shall apply *mutatis mutandis* to mutual companies.
Deeming provision	이 경우 "보험회사"는 "자회사"로 본다.	In such cases, "insurance company" shall be construed as "subsidiary".
Continuation clause	보험계약을 이전하지 아니하게 된 경우에도 또한 같다.	This shall also apply where it decides not to transfer its insurance contracts.
Delegation provision	제1항과 제2항에 따른 출연금의 납부방법 및 절차에 관하여 필요한 사항은 대통령령으로 정한다.	Necessary matters concerning procedures for and methods of paying contributions under paragraphs (1) and (2) shall be prescribed by Presidential Decree.
Penalty provision	제91조제1항에 따른 금융기관보험대리점등 또는 금융기관보험대리점등이 되려는 자가 제83조제2항 또는 제100조를 위반한 경우에는 1억원 이하의 과태료를 부과한다.	Where an insurance agency, etc. of a financial institution prescribed in Article 91 (1) or a person that intends to become an insurance agency, etc. of a financial institution violates Article 83 (2) or 100, he/she or it shall be punished by an administrative fine not exceeding 100 million won.
Deontic provisions		
Obligation	보험협회는 정관으로 정하는 바에 따라 다음 각 호의 업무를 한다. . . .	The insurance association shall perform any of the following affairs, as prescribed by the articles of association: . . .
Liabilities	상호회사의 채무에 관한 사원의 책임은 보험료를 한도로 한다.	The liability of the members of every mutual company for the debts of their company shall be limited to their insurance premiums.
Rights	보험계약자나 보험금을 취득할 자는 피보험자를 위하여 적립한 금액을 주식회사가 이 법에 따른 금융위원회의 명령에 따라 예탁한 자산에서 다른 채권자보다 우선하여 변제를 받을 권리를 가진다.	A policyholder or a person who is to receive insurance proceeds is entitled to be paid the amount accumulated for the insured in preference to any other creditors from assets deposited by the relevant stock company pursuant to orders issued by the Financial Services Commission under this Act.
Permission	주식회사는 그 조직을 변경하여 상호회사로 할 수 있다.	A stock company may convert its organization into a mutual company.
Prohibition	보험대리점 또는 보험중개사는 자기 또는 자기를 고용하고 있는 자를 보험계약자 또는 피보험자로 하는 보험을 모집하는 것을 주된 목적으로 하지 못한다.	An insurance agency or insurance broker shall not be mainly engaged in soliciting any insurance contract which is to make himself/herself or a person who employs himself/herself as the policyholder or the insured.

power in the name of the delegated (Victorian Government Solicitor's Office, 2008).

A delegation provision should state clearly the nature of powers, duties, or functions being delegated, as well as the entitlements, conditions, and restrictions that it may have on the delegate. In Korean legislation, it can be distinguished by the phrases: (으)로/가 정한다 (shall be prescribed by), or 로/가 정하여 고시한다 (shall be determined and announced).

3.2.6 Penalty Provision

The primary function of a penalty provision is to stipulate potential consequences (legal effects) when a breach of legislation, i.e., a violation of a prescribed requirement, or non-performance of an obligation has occurred, and it can be identified with the phrases: 벌금에 처한다, 처벌한다, 부과한다, or 부과할 수 있다 (shall be punished).

Besides, a penalty provision may also stipulate the conditions under which a government agency may/shall revoke, suspend, or cancel the penalties stipulated by the legislation (취소할 수 있다 (may revoke/suspend/cancel) or 취소하여야 한다 (shall revoke/suspend/cancel)).

3.2.7 Deontic Provision

Deontic concepts of *obligations*, *dispensations* (exception from obligations), *liabilities*, *rights*, *permissions*, and *prohibitions* are important concepts in legalisms and legal reasonings, and is used to *manipulate* or *restrict* the behaviour of an entity.

For one, the use of 야 한다 (shall do) in Korean legislation makes it clear that an entity has an obligations (i.e., a duty to comply), or committing herself to such action. Whilst the provision indicates a specific state a legal entity should be into, the phrase 법인으로 한다 (shall be a juristic person) is used. If the provision, however, requires an entity to carry out some specific actions, then the phrase 업무를 한다 (shall perform tasks) will be used instead.

In addition to this, the following phrases are used to determine the liabilities of an entity: 책임을 진다 (shall take responsibility), 책임을…못한다 (may not be released from responsibility), 책임은…한도로 한다 (the liability shall be limited to), and 의무를 지지 아니한다 (may not take any responsibility).

Rights, on the other hand, dictate the principle of entitlement that one may have under some specific conditions and *cannot* be infringed by other person, government, or authorities, and are expressed with the phrases: 권리를/을 가진다 (is entitled/have the right to), 권리와 의무는…승계한다 (rights and duties shall be succeeded).

Similarly, *permissions* refer to a licence to do something, or in some cases, an entity is authorised to do an act which, in principle, without such authorities, such actions would have been unlawful.[10] In Korean language, this can be identified by the terms ㄹ or 을 수 있다 (may do).

From a legal reasoning perspective, both permissions and rights are similar in nature as they can be considered as a dual of obligations i.e., if an entity has the obligation to perform a task, then she should have the permission (or right) to carry out such task (note that the reverse might not be true). The main difference between the two is that the entitlement enjoyed by an entity under rights cannot be infringed or retract; while the case for permissions may still be subject to other conditions as prescribed in the legislation.

Lastly, *prohibitions* prescribe the states or actions that should *not* be undertaken by a legal entity or a violation will appear. It can be identified by the phrases: 아니 된다 (no…shall do), 못한다 (shall be prohibited), and 수 없다 (not permitted).

Table 4 shows a summary of semantic types and their corresponding terminologies in Korean legislation.

4 EXPERIMENTAL ANALYSIS

4.1 Dataset

To evaluate the taxonomy discussed in the previous section, an empirical analysis has been undertaken. The used dataset comprises 1,237 sentences which constitute the statements from three different Korean legislations, namely: Insurance Business Act (IBA), Banking Act (BA), and Financial Holding Companies Act (FHCA).

In the preprocessing phase, the raw text of these legislations was segmented into sentences. As sentences in the Korean language are ended with a *period*, punctuation marks e.g. comma, colons, semicolons, etc., will be ignored. In the case of enumerations or lists, the same rule is applied. That is, all

[10]Notice however that, in the literature, there are some discussions in the legal reasoning domain that explicitly permitting an action makes little sense when such action has not (generally) been prohibited before. Besides, such permission may limit the effects of an obligation (or a prohibition). As the discussion of this topic is outside the scope of this paper, we refer the interested reader to (Hansen, 2014) for details.

Table 4: Semantic types and their corresponding terminology in Korean Legislations.

	Term	Description
Definitional provision		
1.1	을/를 말한다	means
Application provision		
2.1	준용한다	shall apply *mutatis muntandis*
2.2	적용한다	shall apply
2.3	적용하지 그러하지 아니한다	shall / does not apply
2.4	에 따른다	shall be governed by
Deeming provision		
3.1	(으)로 본다	shall be deemed shall be construed
3.2	(으)로 보지 아니한다	shall not be deemed shall not be construed
Continuation clause		
4.1	또한 같다	the same as
4.2	과 같다	shall also apply
Delegation provision		
5.1	(으)로/가 정한다	shall be prescribed by
5.2	위탁한다	shall be entrusted
Penalty provision		
6.1	벌금에 처한다 처벌한다 과태료를 부과한다 벌금형을 과(科)한다	shall be punished
6.2	벌금을 병과할 수 있다 병과(併科)할 수 있다	a fine may be imposed
6.3	과징금을 부과할 수 있다	may impose a penalty surcharge
Obligation		
7.1	야 한다	shall do
7.2	책임을 진다	shall take responsibility
Liabilities		
8.1	책임은… 한도로 한다	the liability shall be limited to
8.2	의무를 지지 아니한다	may not take any responsibility for
8.3	책임을 진다	shall take responsibility
Rights		
9.1	권리를/을 가진다	is entitled, have the right
Permission		
10.1	ㄹ/을 수 있다	may do
Prohibition		
11.1	아니 된다	No . . . shall do
11.2	아니한다	shall not do
11.3	수 없다	No. . . may do, may not do
11.5	못한다	shall be prohibited

items in an enumeration or a list will be considered as a single sentence unless one of them ended with a period.

Next, all sentences were manually classified by the domain experts, according to the taxonomy discussed in Section 3. Table 5 shows the semantic types distributions in each of the legislations and their total occurrences in the dataset. As can be seen, some types appear regularly, e.g., definitional provisions, application provisions and most types of deontic provisions, whereas some have very low support, e.g., deeming provisions, continuation clauses, rights, and liabilities. However, as are common in other legislations, the three types of deontic provision, namely: obligations, permissions, and prohibitions together constitute to more than half of the statements found in the three legislations.

4.2 Evaluation Results

To evaluate the taxonomy of automated legal norms classification, a NLP classifier based on regular expressions has been developed. In each iteration, a statement from the dataset is selected and passed to the syntax tree generator, *Komoran3*, mentioned before. The resulting parse tree is then analysed and the semantic type of the statement is determined through applying the regular expression rules to the legal effect component of the tree.

The results are shown on the right hand side of Table 5. Thereby, different semantic types are differentiated, and the precision and recall are determined for every type individually. As evidenced by the results, the taxonomy presented in the previous section can help in effectively classifying the semantic types of statements in the Korean legislation with only limited issues appeared. This is due to the fact that legal statements are often written in boilerplate expressions where a fixed set of terminologies was used. In short, a total of 1,190 statements has been correctly classified with an overall precision and recall rate of 0.99 and 0.96, respectively.

For the statements that cannot be classified correctly, we found that they were mostly due to either the statements were so complex such that the rules that we defined in the regular expressions are not capable to handle, or the taxonomy terms have appeared at some places other than the main paragraph, which negatively impacted the performance of the classifier.

Table 5: Semantic types distribution of statements.

Semantic type	IBA	BA	FHCA	Total Occurrences	Precision	Recall	F_1 score
Definitional provisions	20	11	10	41	1.00	1.00	1.00
Application provisions	101	35	47	183	0.98	0.98	0.98
Deeming provisions	22	5	8	35	0.95	1.00	0.97
Continuation clauses	6	0	1	7	1.00	1.00	1.00
Delegation provisions	42	23	22	87	1.00	0.99	0.99
Penalty provisions	34	27	23	84	0.99	0.89	0.94
Deontic provisions							
Obligations	192	113	87	392	0.99	0.93	0.96
Liabilities	4	0	1	5	1.00	1.00	1.00
Rights	6	0	0	6	1.00	1.00	1.00
Permissions	114	79	71	264	0.97	0.99	0.98
Prohibitions	59	32	42	133	1.00	0.98	0.99
Total	600	325	312	1237	0.99	0.96	0.97

5 RELATED WORKS

The development of legal taxonomies and ontologies have received unprecedented attention, and various dedicated works has been proposed in the past two decades. For instance, Hachey and Grover (2004) presented an early attempt to address the legal norm classification problem. In the paper, the authors have classified statements from the judgements of the UK House of Lords according to their rhetorical role. Merchant and Pande (2018), on the other hand, proposed an approach to summarize legal judgements based on *latent semantic analysis* (Foltz, 2001), and was able to achieve an average ROGUE-1 score of 0.58.

Zeni et al. (2015), on the other hand, proposed a framework, GauiT 2.0, to semi-automate the legal concepts extraction and annotation process. Boella et al. (2013) presented a rather similar approach which automatically extract semantic knowledge from legal texts based on the syntactic dependencies between different lexical terms. However, the downside of these studies is that they still need substantial manual efforts and time to prepare, model training and development.

Likewise, the development of several ontologies notably IPROnto (Delgado et al., 2003), FOLAW (Valente et al., 1999), PRONTO (Palmirani et al., 2018), DOCLE/DOCLE$^+$ (Gangemi, 2007), etc., have been reported. The PRONTO, a legal ontology on GDPR, provides the legal knowledge modelling of the privacy agents, data types, types of processing operations, rights and obligations. Contrary to HL7 privacy ontology[11] to manage the health data for electronic health records, the goal of PRONTO ontology is to support the legal reasoning using defeasible logic theory. Rubino et al. (2006) presented an OWL-DL based ontology of basic legal concepts (Sartor, 2006) such as obligations, permissions, rights, erga-omnes rights, liabilities, legal power. While these studies are relevant to our work but their goal different from ours – in particular, the focus of PRONTO ontology is limited to privacy and data rights in the context of GDPR. Whereas, Rubino's work is limited as they only extract basic legal concepts.

Recently, the use of NLP techniques to automate the legal norms classifications process has been advocated. For instance, Sleimi et al. (2018) used NLP techniques to extract the legal provisions information such as modalities, actors, conditions, exceptions and violations. Hwang et al. (2018) used NLP tools and data mining techniques and extracted legal as well as domain-relative terms from the Chinese regulations and legal sources and constructed a legal ontology with legal terms and definitions from the Taiwan legislation. They considered the related attributed and relationships among the keywords and extracted 1114 legal terms and relevant definitions interpreted by the domain experts from more than 15 patterns in the Taiwan's law and regulations.

The work of Waltl et al. (2019) is closely related to us. In the paper, the authors have applied both rule-based and machine learning approach and classified German Civil Law into 9 different semantic types. such as duties, prohibitions, permission, etc. We share the common objective with these works in the context of constructing the legal knowledge ontological graph from the Korean legal sources; however, the work of Waltl et al. is limited in scope as we consider more

[11] https://wiki.hl7.org/index.php?title=Security_and_Privacy_Ontology

granular functional aspects of the Korean legislation, which results in 12 different semantic types.

6 CONCLUSIONS

In conclusion, extracting normative information captured in a legal document is a time-consuming and error-prone task. The taxonomy presented in this paper has filled a gap to legislation written in the Korean language, and to the best of our knowledge, is the first of its kind.

For future work, we plan to extend the taxonomy to cater to the wider needs of the Korean legislations analysis and investigate different NLP approaches to automate the legal norms formulation (or translation) process so that a machine understandable formalism can be inferred directly from the Korean statutory texts.

ACKNOWLEDGEMENTS

The authors would like to thank Kiet Hoang The and Wooseok Kim for helping with this project, Gabriela Ferraro for her fruitful comment and suggestions to improve the quality of the paper.

REFERENCES

Boella, G., Di Caro, L., and Robaldo, L. (2013). Semantic Relation Extraction from Legislative Text Using Generalized Syntactic Dependencies and Support Vector Machines. In *RuleML 2013*, pages 218–225.

Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. (2016). Enriching word vectors with subword information.

Botha, J. and Blunsom, P. (2014). Compositional morphology for word representations and language modelling. In *Proceedings of the 31st International Conference on Machine Learning*, page 1899–1907.

Bracher, P. (2018). Interpretation: The nature of a deeming provision. https://www.financialinstitutionslegalsnapshot.com/2018/04/interpretation-the-nature-of-a-deeming-provision/. [accessed: 29 June 2020].

Casanovas, P. (2019). *Legal Linked Data Ecosystems and the Rule of Law*, chapter Legal Linked Data Ecosystemsand the Rule of Law, pages 87–126. Springer.

Ceci, M., Khalil, F. A., and O'Brien, L. (2016). Making Sense of Regulations with SBVR. In *Proceedings of the RuleML 2016 Challenge, Doctoral Consortium and Industry Track*.

Choi, M., Hur, J., and Jang, M.-G. (2004). Constructing Korean Lexical Concept Network for Encyclopedia Question-Answering System. In *Proceedings of the 30th Annual Conference of IEEE Industrial Electronics Society*, volume 3, pages 3115–3119. IEEE.

Cotterell, R. and Schütze, H. (2015). Morphological word-embeddings. In *NAACL 2015*, page 1287–1292.

Delgado, J., Gallego, I., Llorente, S., and Garcia, R. (2003). IPROnto: An Ontology for Digital Rights Management. In *Proceedings of the 16th Annual Conference on Legal Knowledge and Information Systems*.

Dragoni, M., Villata, S., Rizzi, W., and Governatori, G. (2015). Combining Natural Language Processing Approaches for Rule Extraction from Legal Documents. In *AI Approaches to the Complexity of Legal Systems International Workshops*, pages 287–300.

Ferraro, G., Lam, H.-P., Colombo Tosatto, S., Oliveri, F., van Beest, N., and Governatori, G. (2019). Automatic Extraction of Legal Norms: Evaluation of Natural Language Processing Tools. In *JURISIN 2019*, Kanagawa, Japan. Springer.

Foltz, P. (2001). Semantic Processing: Statistical Approaches. In *International Encyclopedia of the Social and Behavioral Sciences*, pages 13873 – 13878.

Gangemi, A. (2007). Design Patterns for Legal Ontology Constructions. In *LOAIT 2007*, pages 65–85.

Gordon, T. F., Governatori, G., and Rotolo, A. (2009). Rules and Norms: Requirements for Rule Interchange Languages in the Legal Domain. In *RuleML 2009*, pages 282–296.

Hachey, B. and Grover, C. (2004). Sentence Classification Experiments for Legal Text Summarisation. In *The 17th Annual Conference on Legal Knowledgeand Information Systems*, JURIX 2004. IOS Press.

Hansen, J. (2014). Reasoning About Permission and Obligation. In *David Makinson on Classical Methods for Non-Classical Problems*, volume 3 of *Outstanding Contributions to Logic*, pages 287–333.

Hashmi, M., Casanovas Romeu, P., and de Koker, L. (2018). Legal Compliance Through Design: Preliminary Results of a Literature Survey. In *TERECOM 2018*, pages 59–72.

Hashmi, M., Governatori, G., and Wynn, M. T. (2013). Normative Requirements for Business Process Compliance. In *Proceedings of the 3rd Australasian Symposium on Service Research and Innovation*, pages 100–116.

Hilty, M., Basin, D., and Pretschner, A. (2005). On Obligations. In *Proceedings of the 10th European Symposium on Research in Computer Security*, ESORICS 2005, pages 98–117, Milan, Italy.

Hoekstra, R., Breuker, J., Bello, M. D., and Boer, A. (2007). The LKIF Core Ontology of Basic Legal Concepts. In *Proceedings of the Workshop on Legal Ontologies and Artificial Intelligence Techniques*, pages 43–63.

Höfler, S. (2019). Making the law more transparent: Text linguistics for legislative drafting. In *Legal linguistics beyond borders: Language and law in a world of media, globalisation and social conflicts*, pages 229–252.

Hwang, R.-H., Hsueh, Y.-L., and Chang, Y.-T. (2018). Building a Taiwan Law Ontology Based on Automatic

Legal Definition Extraction. *Applied System Innovation*, 1(3):22.

Jeong, H., Sugiura, M., Sassa, Y., Haji, T., Usui, N., Taira, M., Horie, K., Sato, S., and Kawashima, R. (2007). Effect of syntactic similarity on cortical activation during second language processing: A comparison of English and Japanese among native Korean trilinguals. *Human Brain Mapping*, 28(3):194–204.

Junho, J. P., Jo, Y., and Shin, H. (2010). The KOLON System: Tools for Ontological Natural Language Processing in Korean. In *JUCLIC 2010*, pages 425–432.

Karkkainen, B. C. (1994). Plain Meaning: Justice Scalia's Jurisprudence of Strict Statutory Construction. *Harvard Journal of Law & Public Policy*, 17(2):401–477.

Lam, H.-P. and Hashmi, M. (2019). Enabling Reasoning with LegalRuleML. *Theory and Practice of Logic Programming*, 19(1):1–26.

Lee, D. (2018). Morpheme-based Efficient Korean Word Embedding. MSc. Thesis, Seoul National University, Seoul, South Korea.

Marín, R. H. and Sartor, G. (1999). Time and Norms: A Formalisation in the Event-Calculus. In *ICAIL '99*, page 90–99.

Mattei, U. (1997). Three Patterns of Law: Taxonomy and Change in the World's Legal Systems. *The American Journal of Comparative Law*, 45(1):5–44.

Matteson, A., Lee, C., Kim, Y., and Lim, H. (2018). Rich Character-Level Information for Korean Morphological Analysis and Part-of-Speech Tagging. In *COLING 2018*, pages 2482–2492.

McCarty, L. T. (1989). A Language for Legal Discourse: I. Basic Features. In *ICAIL '89*, page 180–189.

McGregor, B. (2005). Constructing a concise medical taxonomy. *Journal of the Medical Library Association*, 93(1):121–123.

Merchant, K. and Pande, Y. (2018). NLP Based Latent Semantic Analysis for Legal Text Summarization. In *2018 International Conference on Advances in Computing, Communications and Informatics*, ICACCI 2018, pages 1803–1807.

Mills, A. (2014). Rethinking Jurisdiction in International Law. *British Yearbook of International Law*, 84(1):187–239.

Müller-Gotama, F. (1994). The cross-linguistic survey. In *Grammatical Relations: A Cross-Linguistic Perspective on their Syntax and Semantics*, pages 78 – 140.

Ock, C. (2013). *UWordMap*. University of Ulsan, Ulsan, South Korea.

Palmirani, M., Governatori, G., and Contissa, G. (2011). Modelling temporal legal rules. In *The 13th International Conference on Artificial Intelligence and Law, Proceedings of the Conference*, pages 131–135.

Palmirani, M., Martoni, M., Rossi, A., Bartolini, C., and Robaldo, L. (2018). PrOnto: Privacy Ontology for Legal Reasoning. In Kő, A. and Francesconi, E., editors, *Electronic Government and the Information Systems Perspective*, pages 139–152. Springer.

Prakken, H. and Sartor, G. (1996). A Dialectical Model of Assessing Conflicting Arguments in Legal Reasoning. *Artificial Intelligence and Law*, 4(3):331–368.

Rubino, R., Rotolo, A., and Sartor, G. (2006). An OWL Ontology of Fundamental Legal Concepts. In *JURIX 2006*, pages 101–110.

Sartor, G. (1992). Normative conflicts in legal reasoning. *Artificial Intelligence and Law*, 1(2-3):209–235.

Sartor, G. (2005). *Legal Reasoning: A Cognitive Approach to the Law*, volume 5 of *A Treatise of Legal Philosophy and General Jurisprudence*. Springer.

Sartor, G. (2008). Legal Validity: An Inferential Analysis. *Ratio Juris*, 21(2):212–247.

Sleimi, A., Sannier, N., Sabetzadeh, M., Briand, L. C., and Dann, J. (2018). Automated Extraction of Semantic Legal Metadata Using Natural Language Processing. In *RE 2018*, pages 124–135.

Song, H.-J. and Park, S.-B. (2019). Korean Morphological Analysis with Tied Sequence-to-Sequence Multi-Task Model. In *EMNLP-IJCNL 2019*, pages 1436–1441.

Stratos, K. (2017). A Sub-Character Architecture for Korean Language Processing. In *EMNLP 2017*, page 721–726.

Valente, A., Breuker, J., and Brouwer, B. (1999). Legal modeling and automated reasoning with ON-LINE. *International Journal of Human-Computer Studies*, 51(6):1079 – 1125.

van Engers, T. M., van Gog, R., and Sayah, K. (2004). A Case Study on Automated Norm Extraction. In *JURIX 2004*, pages 49–58.

Victorian Government Solicitor's Office (2008). Current Issues in Delegations. In *Clent Newsletter*, Administrative Law. [accessed: 30 June 2020].

von Wright, G. H. (1963). *Norm and action: a logical enquiry*. Routledge, London.

Waltl, B., Bonczek, G., Scepankova, E., and Matthes, F. (2019). Semantic types of legal norms in German laws: classification and analysis using local linear explanations. *Artificial Intelligence and Law*, 27(1):43–71.

Wieringa, R. J. and Meyer, J.-J. C. (1993). Applications of Deontic Logic in Computer Science: A Concise Overview. In *International Workshop on Deontic Logic in Computer Science*, pages 17–40.

Wieting, J., Bansal, M., Gimpel, K., and Livescu, K. (2016). Charagram: Embedding Words and Sentences via Character n-grams. In *EMNLP 2016*, pages 1504–1515.

Wyner, A. and Peters, W. (2011). On Rule Extraction from Regulations. In *JURIX 2011*, pages 113–122.

Xanthaki, H. (2014). Structure of a Bill. In *Drafting Legislation: Art and Technology of Rules for Regulation*. Hart Publishing.

Yoon, A.-S., Hwang, S.-H., Lee, E.-R., and Kwon, H.-C. (2009). Construction of Korean Wordnet "KorLex 1.5". *Journal of KIISE: Software and Applications*, 36(1):92–108.

Zeni, N., Kiyavitskaya, N., Mich, L., Cordy, J. R., and Mylopoulos, J. (2015). GaiusT: supporting the extraction of rights and obligations for regulatory compliance. *Requirements Engineering*, 20(1):1–22.