A Controlled Linguistic Model and A Representative Framework for Legal Phrase Analysis

Hirosato Nomura

Kyushu Institute of Technology

Iizuka, 820, Japan

nomura@ai.kyutech.ac.jp

Abstract

This paper describes the outline of a controlled linguistic model for the legal phrase analysis and a representative framework for the internal representation of analyzed information.

The idea for designing the controlled linguistic model comes from a fact that a legal sentence is written under the some linguistic constraints based on specific usage of legal terms and specific forms of legal sentences/texts. Such investigations are incorporated into the controlled linguistic model.

The representative framework provides a device for describing both linguistic and logical information analyzed from legal phrases. The linguistic information presents linguistic features such as lexical and syntactic/semantic characteristics/constructions. The logical information gives the logical construction of the legal meanings represented by the corresponding linguistic expressions.

1 Introduction

This paper presents the outline of a controlled linguistic model for legal phrase analysis. It also describes a representative framework for the internal representation of analyzed information from each legal phrase. This paper concerns the Japanese version of the United Nations Convention on Contracts for the International Sale of Goods /citelaw(hereafter The Contracts) as the set of example legal phrases.

The controlled linguistic model and the representative framework are intended to apply for providing information when both developing a legal knowledge-base and reasoning legal matters using the legal knowledge-base. Basically the legal knowledge-base is to store legal knowledge for legal reasoning. However, when developing it, it needs to contain linguistic information so that it can in-

corporate new information into it. The legal reasoning is conducted by applying legal knowledge while the management system including man-machine interface needs to know the correspondences between legal knowledge and its linguistic representation. This is the reason why we try to represent both linguistic and legal information by means of the representative framework.

The idea for designing the controlled linguistic model comes from a fact that a legal sentence is written under the some linguistic constraints based on specific usage of legal terms and specific forms of legal sentences/texts. A legal sentence is long and thus has a complex linguistic structure which gives us some difficulties on the computer analysis. These investigations elaborated in this paper make us possible to design an actually usable and efficient grammatical model for morphological and syntactic/semantic analyses of legal phrases.

The representative framework named here as the feature and logical structure representative framework provides a device for describing both linguistic and logical information of legal phrases simultaneously in a framework. The linguistic information presents linguistic features such as lexical and syntactic/semantic characteristics/constructions. The model adopts the case structure as the basis of the syntactic/semantic construction. The logical information gives the logical construction of the legal meaning of the sentence. The basic form of the logical construction is a pair of the legal requirement and effect.

We have ever published several technical papers on our investigations concerning the controlled linguistic model of legal phrases [2]-[6] thus we do not repeat them in this paper. This paper is intended to show the outline of our recent progress on the matter thus the details must be referred to the previous papers listed at the end part of this paper. We have examined a part of the controlled linguistic model on our computer while it is beyond the purpose of this paper to discuss about the details.

2 Lexical Unit

Proper treatment of lexical units makes us easy to analyze legal phrases. This is very important especially for the morphological analysis of Japanese sentences since Japanese is a language which uses no delimiter between words. Based on this observation, we suppose a lexical unit as longer as possible if they can provide some clear linguistic function. Such a lexical unit is called as an extended lexical unit. A compound term is also to be defined as an extended lexical unit while it is not a functional word.

A legal phrase obviously involves both legal terms and non-legal terms as its components. Thus such linguistic function can be classified into three categories. The first is for specific legal function limited to apply to the material concerned such as The Contracts. The second is for general legal function to be widely applied for writing variety of legal phrases. And the third is for general linguistic function for writing both legal and non-legal phrases.

The specific legal function appears when referring the legal contents described in other parts of the law. This is also used when defining the legal matters to be specified in the law by referring other legal matters already mentioned in the law.

The general legal function is used for specifying the legal function of a sentence. Japanese language is a verb-ending language thus the ending part used to give such function. The correspondences between legal function and its linguistic representation will be listed in somewhere in this paper. The general linguistic function is not dedicated for writing legal phrases but widely applied for writing usual sentences. An example is a function effected by Japanese single and compound particles. Another example is one indicated by expressions for announcing the text structures such as itemized statements. Sentence endings are also examples of the linguistic representations which produce such function. Other examples are functional words such as for constructing coordination.

Sometimes an extended lexical unit involves another word as a fragment of the string. Such a problem is not only for the legal phrases but for usual linguistic expressions. Thus in order to avoid the mistakes in the morphological analysis, we introduce a procedural function called as the compounder demon (相当でいいもん (soto-de-iimon.) It will work for recognizing candidates of extended lexical units from a sequence of usual lexical units.

3 Legal Terms

In order to write each well-organized legal phrase and then in order to construct a well-formed structure of a sequence of the legal phrases from the view point of the law expert, it is obvious that the proper use of the legal terms is crucial. A legal phrase must be a linguistic expression elaborated for representing the intended meaning correctly, consistently, clearly, naturally, and efficiently. However, the usage of some legal terms is different from the usage of them when they are used as the general words. Such differences must be encoded into the controlled linguistic model otherwise we will have no idea for understanding the correct meanings of the legal phrases.

Parts of speech of law terms are noun, verb, adverb, conjunction, auxiliary verb, and so on, thus covering almost of the parts of speech of general words. It does mean that we must concern everything which is common with usual words.

Examples of proper use of legal phrases are shown below (referred from [2][3][6][7] [8][9][10][11].) They demonstrate the difference of the usage from that in general sentences. All of the following three words are adverbs and represent the meanings concerning the immediacy of time. These words have almost the same meaning when they are used as the general words thus people do not mind which word must be used in a situation. However, they are carefully used in legal phrases since they result the different legal effects.

ただちに (**tadachini**)–*immediately* :
> 'Immediacy of time' is the strongest. If it was violated, it results in illegality frequently.

遅延なく (**chiennaku**)–*without delay* :
> 'Immediacy of time' is strong, but some delay with a good rational reason is allowed . However, if the reason is not acceptable or they were violated, it results in illegality frequently.

速やかに (**sumiyakani**)–*swiftly* :
> Even if it were violated, in terms of neglecting obligation, it never result in illegality instantly.

In order to conduct the correct analysis of such legal terms, we encode those function and meanings into the lexicon. Thus such lexical terms as well as the extended lexical units mentioned above contribute to design a part of the controlled linguistic model.

4 Sentence Function

A sentence ending of a Japanese sentence generally consists of a verb and the following several auxiliaries. Some of such sentence endings determine the sentence function as a legal sentence.

Table 1. Function of sentences

function of the sentence	the law terms which appears at the end part of sentences
Declaration of effect	ものとする (*monotosuru*) (e.g. It is that …) 推定する (*suiteisuru*) (e.g. It is presumed) みなす (*minasu*) (e.g. It is regarded) 例とする (*reitosuru*) (e.g. by way of example)
Description of rights	ことができる (*kotogadekiru*) (e.g. can) ことができない (*kotogadekinai*) (e.g. cannot)
Description of obligation	なければならない (*nakerebanaranai*) (e.g. must)
Description of prohibition	してはならない (*sitehanaranai*) (e.g. must not)
Exception of application	ただし…この限りではない (*tadashi … konokagiridehani*) (e.g. But, … it is not limited in this case) 適用しない (*tekiyousinai*) (e.g. It is not applied) 妨げない (*samatagenai*) (e.g. they do not prevent)
Application	ただし…場合に限る (*tadashi … baainikagiru*) (e.g. However, … it is limited in this case) 適用する (*tekiyousuru*) (e.g. It is applied) 準用する (*junyousuru*) (e.g. It is applied correspondingly) 例による (*reiniyoru*) (e.g. take example by)

We can see many examples of such linguistic expressions even in The Contracts. Thus it is very natural to adopt them to infer the sentence function from such a surface linguistic expression. From the fact investigated we design a set of sentence patters which can be provided by the combinations of each sentence ending and its corresponding sentence function. Such sentence patters are encoded into the lexical item of the main predicate in the sentence ending thus are designed as a part of the controlled linguistic model. It is easy to understand that they will play an important role as the clue for grasping the rough meaning of each legal phrase.

Typical examples of such legal expressions appearing at the end part of legal sentences are listed in the table 1. The relationships between the linguistic expressions at the end part of a sentence and the function of the sentence are also listed in the table.

The column on the table indicating the classification of < function of the sentence > is not necessarily applicable to other legal phrases since it is based on the analysis of The Contracts. However, this does not mean that this classification is useless. There are many laws and they are concerning different matters thus it is clever to provide such classifications for each group of laws. This can be understood from another fact that they might be written in the different time or age thus the vocabulary and the sentence style might be different.

The class indicated in the Table as < description of prohibition > might be specific for The Contracts. Linguistic expressions such as 与えてはならない (*ataetehanaranai*) and 与える必要はない (*ataeruhitsuyouhanai*) appear in The Contracts. All of such representations have the judicature as their grammatical subjects. This might be an example that the classification shown in the Table 1 is depending on the material of The Contracts.

5 Case Structure

Legal sentences have not so many obligatory case elements. In The Contracts common case elements are those of seller (agent or patient), buyer (agent or patient), goods (object), amount of the goods (quantity), amount of the money (quantity), place of action (location), and time or duration of action (time). These case elements are actually represented as linguistic components which are marked by case markers. Other type of case elements are represented as subordinate clauses which give us some hard problems to design the case structure system.

There are some problems on providing a set of the case elements when they are represented as subordinates. For

example, a Japanese word 'toki' has two surface representations: one is とき which is a phonetic representation called as Hirakana and the other is 時 which is a Kanji representation. There is no problem if it is used as a usual noun but there is a problem when it is used as KEISHIKINEISHI which corresponds to a relative pronoun in English. If 'toki' is written by a Kanji such as 時, then it is sure that it represents the meaning of time thus the clause might possibly act as a case element. However, if it were described by Hirakana which produces a phonetic expression, then it may only be a complementizer thus has no meaning.

We suppose a set of obligatory case elements as small as possible. Such case frames relate to sentence patterns mentioned somewhere else in this paper. Such sentence patterns lead us to design an efficient strategy for the analysis procedure of the legal phrases. It is true that it is very hard to recognize the correct sentence structure without applying the legal expertise since a legal phrase used to have an amazing construction. Even for such ungentle constructions of linguistic expressions, such strategy can be expected to work well by means of the guidance by the sentence patterns which are not necessarily eager to invoke expert knowledge.

Japanese sentences used to omit many case elements even if they were the obligatory case elements. English, for example, puts some linguistic constraints on a sentence structure construction such as a nominal subject, a verb with a grammatical constraint on the number agreement to the subject, and null or one or two compliment/objects, and their distribution in the sentence. However, Japanese has no such grammatical constraint. Any case elements can be omitted. Even when some of case elements are not omitted, their distribution and the order are free. Thus the syntactic analysis is not easy. On the other hand, the construction of a noun phrase and a verb phrase have linguistic constraints respectively thus we can adopt the similar strategy of the syntactic analysis employed for English analysis.

The fact that Japanese used to omit its case elements makes a requirement for supplying them in order to fill the case slots in the case frame. It is needed for the proper understanding of the meaning of the legal sentence. If the agent of an event specified in a legal phrase, for example, could not be identified since it was omitted in the linguistic expression, then the legal reasoning will fail for producing the correct result. Thus, such omitted case elements must be supplied by inferring from contextual information. However, again, unfortunately it is almost impossible to identify or guess such omitted case elements from the contextual information because

of the state of the art of the current naive technology of natural language processing. Instead, such sentence patterns mentioned can suggest the clever way of finding such omitted components without applying unclear semantic and contextual information to be analyzed. Our examination on the sample legal phrases from The Contracts worked well by applying such sentence patterns provided in the controlled linguistic model.

6 Sentence Types

Information on the semantic governing of the legal meaning represented by a legal phrase can be applied for defining the sentence types. We suppose four types as the sentence types[3].

Type 1 is for the semantic governing to the Contract Object. It describes properties of a contract object and the manner how to deal with it when the provision applying.

Type 2 is for the semantic governing to the Patient. It specifies the right or the duty of the Patient.

Type 3 is for the semantic governing to the Provision Unit. It presents the relationships between provision units.

Type 4 is for the semantic governing to the Judge. It describes the judgment on the contracts.

We can relate each sentence type with the sentence function listed in the Table 1. As a result, we can define the assertion of each sentence type. The sentence type 1 asserts the effectuation declaration, the sentence type 2 asserts right, duty, and prohibition declarations, the sentence type 3 asserts the application exception and the application compliment, and the sentence type 4 asserts the judgment provision.

By incorporating these investigations mentioned above together, we can define triangular relationships among sentence types, sentence functions, and sentence assertions which will be the most important component in the controlled linguistic model.

7 Coordination Structure

Coordinate structures appear very often in legal phrases[5]. This is not specific in The Contracts. Many coordination structures have extra embedded coordination structures. In The Contract more than half of the total coordination structures have some kind of embedded coordination structures. Sometimes it is impossible to identify each scope of a coordination without applying legal expert knowledge. Thus they used to make an amazing

linguistic structures which give us serious difficulties for the analysis of the legal phrases.

There are several types of coordination structures. The first type is one constructed by coordination markers such as conjunction words. Such conjunction words are 「又は」 (mataha - 'or'), 「及び」 (oyobi - 'and'), 「若しくは」 (moshikuha - 'or'), 「かつ」 (katsu - 'and'), 「また」 (mata - 'or'), 「並びに」 (narabini - 'and'), and 「あるいは」 (aruiha - 'or'). Their parts of speech are conjunctions. Others are 「と」 (to - 'and'), 「や」 (ya - 'or'), and 「か」 (ka - 'or'). They are conjunctive particles. Sometimes comma is used as the coordination marker. Other type of markers are 「その他」 (sonota - 'other') and 「その他の」 (sonotano - 'other') which are not conjunctions while they act as conjunctions in some cases. These conjunctive markers coordinate not only nouns or noun phrases but clauses and sentences.

The second type of the coordination has no explicit marker. Some kind of verb phrases called as Renyouchushi whose literal meaning is 'stopping with the inflection of Renyoukei without representing the following functional words' makes such coordination structures. The third type of the coordination is constructed by, for example, itemized statements such as itemized noun phrases.

Fortunately there is some rules for making proper use of conjunctive markers which might indicate the level of the coordination embedding. For example, if three disjunctive conjunctions made a three leveled noun coordination structures, then the top level has 「又は (mataha)」, the second level is indicated by 「若しくは (moshikuha)」, and the third level is designated by 「あるいは (aruiha)」. However, there are some varieties of the usage of conjunction words depending on the type of the coordination. Many of them can be ruled out while some exceptions remain.

Such investigations concerning the constraints of the usage of coordination markers are incorporated into the controlled linguistic model. By applying the controlled linguistic model, the analysis procedure can reduce the number of the possible constructions concerning the co-ordination structure drastically. In other word, if we do not adopt such a constraint, we must count up the astronomical figures on the candidates of the coordination structure.

8   Contextual Structure

Contextual structures are constructed by co-references, conjunctions, itemized statements, and text structures.

The Contracts is divided into four parts and each part contains articles consists of one or a couple of sentences possibly with attaching itemized descriptions. This aspect is very different form the usual texts.

Legal phrases refer to other sentences or their fragments frequently. Such a reference produces a complication of the co-reference structure. However, we can recognize that there are some style of co-references indicated by lexical terms such as 「この条約 (konojouyaku - 'this law')」, 「この条約の第一部 (konojouyaku-no-daiichibu - 'the first part of this law')」, 「前条 (zenjou - 'the previous article')], 「前項に規定する要件 (zenkou-ni-kiteisuru-youken - 'the requirement specified in the previous item')」, et al without making any serious effort.

The problems are on the contextual diversity when conducting a legal reasoning. Even for a small legal event, the judgment must be derived from the legal specifications indicated over the several or usually lots of articles in the law. Sometimes it needs to refer to other laws and even to the daily life customs for making the insightful and gentle reasoning which is the problem beyond the current state of the art of natural language processing.

It seems that there are a lot of extra contextual complexities which must be elaborated further.

9   Controlled Linguistic Model

By incorporating investigations mentioned above, we can design a controlled linguistic model for legal phrases. The current version is not the complete one but a preliminary model to be applied for further elaboration to the details. The very brief outline of the controlled linguistic model is listed in the Figure 1 while its full size needs a lot of space for listing.

The controlled linguistic model encodes both general grammatical information as well as specific constrains investigated from the usage of legal terms. It also defines the details of the case structures and the subordinate structures. Such constraints will play a significant role in the analysis process of legal phrases.

We have been developing a lexicon from the stand point of linguistic data. We have also been elaborating for a tagged corpus of legal phrases. The tagged corpus involves lexical and structural information as well as expected semantic and contextual information.

Lexical tags are grammatical ones such as part of speech, inflections, phonology, idioms, and the concept. Lexical information relates to function of terms which makes some kind of linguistic ontology. Sentence patterns are stored in the lexical items. Assertions men-

tioned earlier are conceptual tags attached to sentence patterns.

Syntactic tags are dependency structures containing case structures, embedding structures, subordinate structures, and coordinate structures. The case structures are related to the sentence patterns. Additional type of tags are on the linguistic constraints of the legal terms usage. they are written in the form of rules.

Semantic tags are attached to the syntactic tags. Extra tags are provided for representing semantic forms which will be used for constructing the meaning of the legal phrase.

Contextual tags concern text styles and co-reference structures while the investigations for these tags are expected to elaborate.

Current data-base which will turn to the controlled linguistic model is not well-developed while it needs more than several hundreds of pages for printing out thus is not listed here.

Figure 1. A controlled linguistic model

legal phrases

```
:: = < legal phrases  type 1:Participant > |
     < legal phrases  type 2:Contract Object > |
     < legal phrases  type 3:Provision Unit > |
     < legal phrases  type 4:Judge > |

< legal phrases  type 1:Participant >
  :: = < legal effect  participant > |
       < legal condition >
       + < legal effect  participant > |
       < sentence pattern 1.1 > |
       < sentence pattern 1.2 >
                  : (almost)
< sentence pattern 1.1 >
  :: = (SUB : participant) + (OBJ : obligation)
       + ou(e.g. be charged) |
       (SUB : seller) + (OBJ : article)
       + hikiwatasu(e.g. deliver)
       + 「nakerebanaranai」 (e.g. must) |
                  : (almost)
< sentence pattern 1.2 >
  :: = (SUB : participant) + (OBJ : contract) +
       (CON)
       + kaijosuru(e.g. cancel)
       + 「kotogadekiru」 (e.g. can) |
       (SUB : seller) + (OBJ : article) + (CON |
       TIM)
       + hikiwatasu(e.g. deliver)
       + 「nakerebanaranai」 (e.g. must) |
                  : (almost)
```

10 Feature and Logical Structure Representation

In order to supply enough information for constructing and updating the legal knowledge-base, the system must provide both linguistic and logical information analyzed from legal phrases. This is especially invoked when incorporating new information into the legal knowledge-base. Basically the legal knowledge-base is to store the logical construction of legal knowledge analyzed from linguistic representations of legal specification. However, if it does not contain any linguistic information, then it has no idea for combining newly coming information with stored information. This is the reason why we design the representative framework which can encode both linguistic and logical construction of legal knowledge.

Such kind of information is also needed when transferring query from natural language interface to the reasoning engine and vice versa. This is for the legal expert system by which people consult some legal reasoning. Such a system is also used for learning legal knowledge. In such an expert system, the communication between the user and the expert system is carried out through linguistic means. Thus the system must provide a facility for converting the linguistic representation into the logical representation.

The future and logical structure representative framework (FLS) provides a device for meeting such a requirement. It tries to represent both kind of information in a uniform manner by the representative framework. Thus it has two faces: one face is for linguistic interest and the other for logical interest. The reasoning engine may face to the logical side while the natural language interface may concern only to linguistic aspect.

The feature and logical representative framework provides a feature structure representation device which is a formalization of the so-called unification representative scheme.

The component of the representative framework is a representation of the pair of a feature label and a feature value. It allows both the iterative and recursive constructions of the pair thus is very universal.

The feature value encodes variety of linguistic information mentioned above in this paper. In parallel it encodes logical information thus the logical structure has a parallel construction with the corresponding linguistic structure.

The linguistic construction is to represent the dependency structure of components of legal phrases. Such a dependency structure is summarized into a case structure, for example, so that we can manage a set of such

dependency structures simultaneously.

The logical construction is also to represent the dependency structure of fragments of legal meanings. The very atomic construction of such dependency is one between a legal requirement and the legal effect. However, there are not so many explicit descriptions of such words. Thus the logical construction seems to be rather simple if it is analyzed from the legal stand point since it used to suppose the huge amount of legal expertise.

If the logical construction of legal knowledge is analyzed from the linguistic point of view, there are a lot of pairs of logical dependency. Such a logical dependency appears iteratively and recursively which are parallel to the corresponding linguistic constructions. For examples, any kind of prerequisites concerning time and place for a legal event such as the individual contract or the individual sale must be recognized as the requirement thus must be a part of the logical dependency structure.

There are variety of linguistic expressions which must be analyzed as clues for making logical dependency structure. Such linguistic components must be analyzed as partners of the subordinate structure. This fact tells us a fact that those components must not be case elements. Thus we suppose a case structure as simple which encodes very semantically obligatory clues.

Any kind of dependencies is designated by a symbolic notation of 'imply' which is familiar with as that of the mathematical logic. However, we classify it into subclasses so that it can represent the details of the dependency function. Such function is determined from the sentence function already mentioned somewhere else in this paper.

11 Concluding Remarks

This paper presented a controlled linguistic model for legal phrases, which incorporates linguistic constraints coming from the usage of legal terms. It also described a representative framework called as Feature and Logical Structure Representative Framework for the internal representation of analyzed information, which provides a uniform representative framework for both linguistic construction and logical construction of legal phrases.

The controlled linguistic model has been examined on the sample phrases from The Contracts. More investigations will be elaborated and the model will be extended to cover wider variety of phenomena of linguistic constructions of legal phrases.

References

[1] K. Sono, "United Nations Convention on Contracts for the International Sale of Goods", (in Japanese)

[2] H. Nomura, and H. Iwamoto, "Linguistic Model of Law Sentences and Computer Processing", Proc. of Second Workshop of Japan - United Kingdom Bilateral Cooperative Research Programme on Computational Linguistics, 1991

[3] H. Iwamoto, and H. Nomura, "Linguistic Analysis of Law Sentences and its Application to Natural Language Processing", Proc. of Japan-Australia Joint Symposium on Natural Language Processing, 1991

[4] H. Iwamoto, K. Nagano, H. Nagai, and H. Nomura, "Analysis of Sentence Endings and its Application to Linguistic Model for Law Sentences", Proc. of International Symposium on Natural Language Understanding and AI, 1992

[5] H. Iwamoto, H. Nagai, T. Nakamura, H. Nomura, "Analysis of Coordinate Structures and its Application to Controlled Linguistic Model for Law Sentences", Proc. of Natural Language Processing Pacific Rim Symposium, 1993

[6] K. Nagano, H. Nagai, T. Nakamura, H. Nomura, "Controlled Linguistic Model Based on Constraints for Legal Phrases", Proc. of Natural Language Processing Pacific Rim Symposium, 1993

[7] Shuzou Hayashi, "Houreikaishakuno joushiki (Common Sense for Composing Legal Phrases)", Nihonhyouronsha, 1975 (in Japanese)

[8] Shuzou Hayashi, "Houreisakuseino joushiki (Common Sense for Writing Legal Phrases)", Nihonhyouronsha, 1975 (in Japanese)

[9] Nobutoshi Tajima, "Shinpan Houreiyougono kisochishiki (New-Edition Elementary Knowledge of Law Terms)", Gyousei, 1991 (in Japanese)

[10] Shuzou Hayashi, "Houreiyougono joushiki (Common Sense of Law Terms)", Nihonhyouronsha, 1975 (in Japanese)

[11] Eiichi Hoshino, "Shou Roppou (A Small Compendium of Laws)", Yuuikaku, 1990 (in Japanese)