

# Diachronic and Synchronic Analyses of Japanese Statutory Terminology

## Case Study of the Gas Business Act and Electricity Business Act

Makoto Nakamura  
Japan Legal Information Institute,  
Graduate School of Law,  
Nagoya University  
Furo-Cho, Chikusa-ku, Nagoya  
Aichi, 464-8601, Japan  
mnakamur@nagoya-u.jp

Katsuhiko Toyama  
Information Technology Center /  
Graduate School of Information Science,  
Nagoya University  
Furo-Cho, Chikusa-ku, Nagoya  
Aichi, 464-8601, Japan  
toyama@is.nagoya-u.ac.jp

### ABSTRACT

In this paper, we report on our ongoing research on the development of a diachronic legal terminology, which deals with temporal changes in legal terms. Focusing on provisions for definitions in acts, we successfully extracted legal terms, their explanations, and their relations. Thus far, we have collected 27,737 terms and 36,698 relations. Picking up two acts, we analyzed their legal terms from diachronic and synchronic points of view. Our analysis showed that similarities between the two acts can be measured by their legal terms.

### Keywords

legal text processing; legal terms; Japanese statutes; terminology

## 1. INTRODUCTION

In this paper, we report on our ongoing research on the development of a Japanese legal terminology. In general, important terms in statutes are explicitly defined prior to use. As such, we focus on the legal terms defined in a provision, each of which consists of a tuple of a legal term and its explanation. The provision for the definition of terms is typically placed in Article 2 in an act, following the act's purpose. In other words, the legal terms in a provision are regarded as governing the whole act.

Legal statutes are not only established, but also often amended by changes in social situations. In some cases, legal terms are also revised, added, or deleted depending on the scale of the amendment. Therefore, an amendment to provisions for legal terms implies a drastic change of the entire act. The terminology for legal terms must deal with temporal changes that are dependent on amendment acts.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*LTDCA2016 '16 San Diego, CA USA*

© 2016 ACM. ISBN 978-1-4503-2138-9.

DOI: 10.1145/1235

The study of the diachronic evolution of terminology can be regarded as a set. Such study deals with the narrative of the historical origins of a domain's terminology and its actual historical evolution to the present state [2]. Applied to the legal domain, this is achievable with an exhaustive number of legal terms from all of the acts in the statutory corpora.

Thus far, we have constructed a diachronic legal terminology that consists of legal terms, their explanations, and their relations. Our approach to automatic extraction is based on a rule-based method [5]. Our purpose in this paper is to analyze our terminology from diachronic and synchronic viewpoints. This thorough study makes it possible to deal with synchronic similarities in legal terms to uncover the hidden relations among acts. For example, the submission of bills is often motivated by such changes in social situations as economic and political issues, wars, and natural disasters, during which multiple acts are simultaneously enacted. Although similar terms are often defined in different acts with explanations that resemble each other, identifying such relations is difficult without knowledge. The natural language processing technique makes it possible to calculate the similarity between terms and explanations. Therefore, a diachronic legal terminology provides legal scholars with a method of analyzing the dynamics of legal changes.

We organized our paper as follows: In Section 2, we introduce the diachronic legal terminology. In Section 3, we explain the target issues. In Section 4, we explain how to extract legal terms. In Section 5, we propose a method for analyzing acts with legal terms and, in Section 6, we evaluate the terminology. We present our conclusion in Section 7.

## 2. DIACHRONIC CHANGES IN LEGAL TERMS

In this section, we introduce diachronic changes in legal terms and, in Section 2.1, we explain these changes with examples. Section 2.2 shows actual changes in definitions.

### 2.1 Examples

We cite the Gas Business Act (Act No. 51 of 1954) as an example to explain diachronic changes in legal terms. As of September 2015, this act has been amended 36 times, six of which include the revision of terms and definitions in Article

2<sup>1</sup>. Figures 1(a) to 1(c) show the diachronic changes in the terms at three time points:

- (1) At the new enactment, only two terms, “Gas Business” and “Gas Facilities,” were defined in the Gas Business Act (Act No. 51 of 1954), which came into effect as of April 1, 1954 (Figure 1(a)).
- (2) The term “Gas Business” was changed to “General Gas Utility Business,” which became a hyponym of the newly defined term “Gas Business” with the newly added term “Community Gas Utility Business,” by the Act on the Partial Amendment of the Gas Business Act (Act No. 18 of 1970), which came into effect as of October 12, 1970. Note that, unlike language changes as a natural phenomenon, the sense of legal terms was forced to change on the enforcement date (Figure 1(b)).
- (3) Although no enforcement date has been determined yet for the Act on the Partial Amendment, etc., of the Electricity Business Act, etc. (Act No. 47 of 2015)<sup>2</sup>, the number of terms defined in the Gas Business Act will be increased to 16 (Figure 1(c)). In the period between (2) and (3), the terms “General Gas Utility Business,” “Community Gas Utility Business,” “Gas Pipeline Service Business,” “Large-Volume Gas Business,” “Large-Volume Supply,” “General Gas Utility,” “Community Gas Utility,” “Gas Pipeline Service Provider,” “Large-Volume Gas Supplier,” and “Wholesale Supply” were defined, but deleted later. In addition, the term “Intra-Area Wheeling Service” was replaced with “Wheeling Service.” These were basically eliminated by social selection.

## 2.2 Amendment of Legal Terms

Statutes are written in two types of languages: an object language for new enactments and metalanguage, which rewrites the description in object language, for revisions, rearrangements of provisions, and to repeal acts. While the former describes a law itself, the latter shows how to rewrite it with patch expressions. The amendment of a statute is realized by applying the latter to the former. This amendment method is called *consolidation*.

Figure 2 shows an excerpt from the acts dealing with the changes of the term “Gas Business” from Figure 1(a) to Figure 1(b). The revised act is shown in Figure 3. Note that we referred to the Japan Law Translation Database System<sup>3</sup> for the English translation of these acts. When there is no translation for the acts or act titles in the website, we manually translated them using the database. The original Japanese versions of the acts are shown in Appendix A.

## 3. TARGET ISSUES

<sup>1</sup>The Gas Business Act (Act No. 51 of 1954) includes another provision for the definitions in Article 39-2 that took effect on enforcement date of the Act on the Partial Amendment of the Gas Business Act (Act No. 18 of 1970) on October 12, 1970.

<sup>2</sup>This act provides in Supplementary Provisions that Article 2 of this Act shall come into effect as of the date specified by a Cabinet Order that has not been promulgated yet. Article 39-2 will be deleted at the same time by this amendment.

<sup>3</sup><http://www.japaneselawtranslation.go.jp/>

Related statutes are defined as those that directly or indirectly refer to each other. Finding statutes that are directly related to a target statute is not difficult; simply list up all of the statutes that refer to and are referred to by the target statute.

Some related statutes share the same or similar terms with explanations that differ from each other. They are often amended simultaneously. In this case, the amendment statute would play the key role as a bibliographic coupling. In other words, it would be difficult to find such a relation between statutes without explicit reference. By focusing on legal terms, we may be able to estimate the strength of connections between statutes. The Electricity Business Act (Act No. 170 of 1964) is related to the Gas Business Act (Act No. 51 of 1954).

We see a synchronic similarity in legal terms among related acts. In the same way, we may be able to find *hidden* related acts that are indirectly related, but not explicitly described. Therefore, we picked up these known two acts to illustrate linguistic analysis of legal terms.

## 4. LEGAL TERMINOLOGY

### 4.1 Extraction of Provisions for Definitions from Corpora

Some private companies provide advanced legal databases. The database of Dai-Ichi Hoki Co., Ltd.<sup>4</sup> serves all of the acts and ordinances that are currently in effect, precedents, literature, and so on. It stores not only all 2,501 currently effective acts, but also their old versions at any time point. In addition, the history is listed with each article, from which we can designate any two time points to create a comparison table.

The problem with this database is that old acts are only available for inspection back about 15 years, which is too short to appreciate the dynamics of legal terms. Although we employ this database due to the lack of need for a consolidation process, this is a pilot version.

We searched all the provisions for the definitions of legal terms. Since each article in an act typically has an article caption, we extracted articles whose captions include a particular string denoting definitions with a regular expression (Figure 8(a)). Restricting our target articles to those having a particular article caption, we can extract them with high precision (100%), and detect deletion of the articles by an amendment statute. The number of acts that include one or more articles for the definition of legal terms is 1,033 out of 2,501. If an act includes a number of articles for the definition of legal terms, we consider them separately. Therefore, we deal with 1,081 articles, 540 of which were revised during the last 15 years.

### 4.2 Extraction of Legal Terms

What are recognized as legal terms to be collected depends on the purpose [1, 4, 6]. In this paper, we define legal terms as those explicitly defined prior to use in a statute, each of which consists of a tuple of a legal term in the quotations and its explanation. They typically take the following forms:

- (1) An independent provision
- (2) An inserted statement in parentheses

<sup>4</sup><https://www.d1-law.com/>

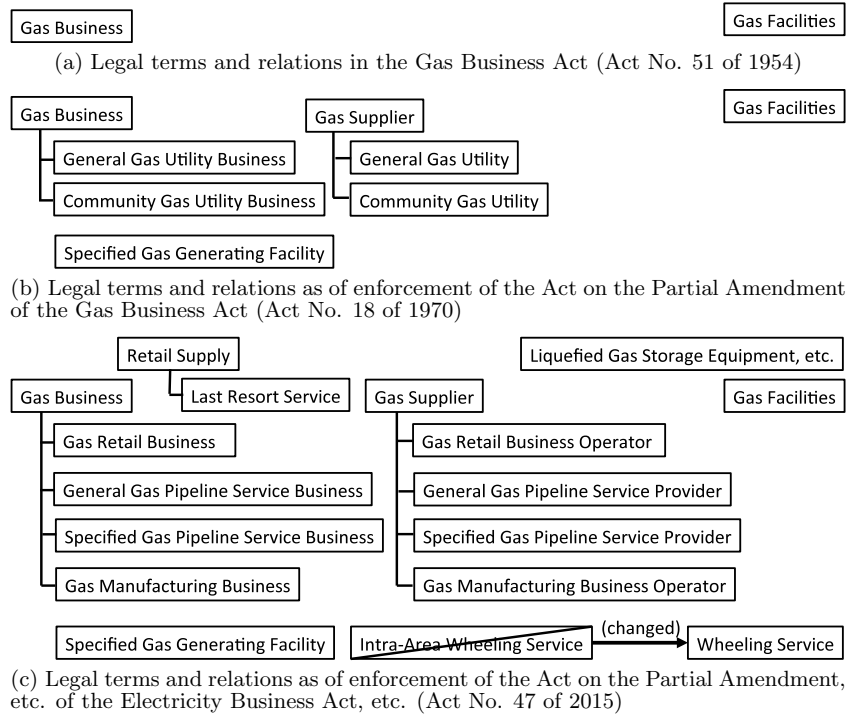


Figure 1: Dynamics of definitions in the Gas Business Act

For (1), an article often consists of a number of paragraphs, each of which defines a legal term. They are described with boilerplate expressions including a legal term and its explanatory sentence, which can be extracted with a set of regular expression rules (Figure 9(f)). The underlined phrases<sup>5</sup> in the upper box in Figure 2 match one of the rules. Some paragraphs include items for a list of either conditions for the term defined in the paragraph’s body or legal terms. A set of regular expression rules distinguishes them (Figure 8(b)). In the former case, the explanatory sentence includes all of the items. For the latter, a legal term and its definition can be extracted with a simpler set of regular expression rules (Figure 8(c)).

For (2), a defined term appears in parentheses following a phrase as its explanation in the main text. Abbreviations of terms are often defined in parentheses. An example is shown in Figure 3, where the term “Community Gas Utility Business” and the term in parentheses, “Specified Gas Generating Facility,” are defined as follows:

**Term:** Community Gas Utility Business

**Definition:** The business of generating gas at a simplified gas generating facility specified by a Cabinet Order (hereinafter referred to as a “Specified Gas Generating Facility”) and supplying such gas via pipelines to meet general demand at not less than 70 gas service points within one housing complex.

**Term:** Specified Gas Generating Facility

<sup>5</sup>Note that the original statute does not include the underlines, which were added by the author.

**Definition:** The business of generating gas at a simplified gas generating facility specified by a Cabinet Order.

We extracted the explanation of the latter, the underlined part<sup>5</sup> in Figure 3, from the beginning of the definition to just before the beginning of the parentheses.

Although the definitions in the parentheses often appear in the main text regardless of the article’s content, we deal with those in the article for the definition of legal terms. This is because legal terms can have a relation that shares the term defined in the parentheses in their explanatory sentences.

We successfully extracted legal terms, their explanations, and their relations. We found 27,737 terms and 36,698 relations. The precision of the relations was 88.0%, which might be improved with additional regular expression rules [5]. In addition, taking repealed statutes into account, they are revised as 27,737 terms and 36,698 relations.

## 5. CLUSTERING ACTS

We analyzed a set of acts in effect as of a designated date. The following paragraphs explain how to (1) define features, (2) make a word vector, and (3) cluster acts. We repeat this procedure for each year (2001-2015).

Since Japanese words in a sentence are not separated by a space, we need morphological analysis, which separates words and attaches a part of speech tag to each morpheme. We used Mecab [3]. In addition, we registered all of the act titles and legal terms in the dictionary in the morphological analyzer in advance. Therefore, each is dealt with as a single word.

Gas Business Act (Act No. 51 of 1954)

(Definitions)

**Article 2** (1) The term “Gas Business” as used in this Act shall mean the business of supplying gas via pipelines to meet general demand.

---

Act on Partial Amendment of the Gas Business Act

(Act No. 18 of 1970)

“Gas Business” in Article 2, Paragraph (1) of Gas Business Act shall be deemed to be replaced with “General Gas Utility Business” and “(excluding, however, businesses generating gas at a gas generating facility prescribed in paragraph (3) and supplying such gas via pipelines)” shall be added after the term “general demand”, (\*snip\*) and the following five paragraphs are added after Article 2, Paragraph (1):

(\*snip\*)

(3) The term “Community Gas Utility Business” as used in this Act shall mean the business of generating gas at a simplified gas generating facility specified by a Cabinet Order (hereinafter referred to as a “Specified Gas Generating Facility”) and supplying such gas via pipelines to meet general demand at not less than 70 gas service points within one housing complex.

(\*snip\*)

(5) The term “Gas Business” as used in this Act shall mean a General Gas Utility Business or Community Gas Utility Business.

**Figure 2: Excerpt from the Gas Business Act (Act No. 51 of 1954) and the Act on Partial Amendment of the Gas Business Act (Act No. 18 of 1970)**

We deal with a set of acts that are effective at a particular time point. We chose the date of April 1st in each year, since acts typically come into effect on that date, which is the start of the fiscal year in Japan.

We employed a vector space model with *tf-idf* weighting. Each act is represented by a word vector, which reflects words in the context of the target act. All content words appearing in the context are used as features. The region of the context is limited to the articles of definitions of legal terms in the act. At this point, a similarity between two acts can be calculated by cosine similarity. We defined a context matrix as follows, each row of which stands for a vector of the act.

$$\begin{matrix}
 \text{act}_1 \\
 \text{act}_2 \\
 \vdots \\
 \text{act}_1
 \end{matrix}
 \begin{pmatrix}
 \text{an}_1 & \cdots & \text{an}_j & \text{lt}_1 & \cdots & \text{lt}_m & \text{cw}_1 & \cdots & \text{cw}_n & \\
 0 & \cdots & 0 & 0.305 & \cdots & 0 & 0 & \cdots & 0.02 & \\
 0.01 & \cdots & 0 & 0 & \cdots & \cdots & 0.01 & 0.07 & \cdots & 0.01 \\
 \vdots & & & & & & & & & \vdots \\
 0 & \cdots & 0.02 & 0 & \cdots & 0 & 0.01 & \cdots & 0 & 
 \end{pmatrix},$$

where  $\text{act}_i$  denotes an act, and  $\text{an}_j$ ,  $\text{lt}_k$ , and  $\text{cw}_l$  denote the name of the act, a legal term, and a content word, respectively. Each numeral in the matrix corresponds to the value of *tf-idf* for the column term.

We used bayon<sup>6</sup>, which is a clustering tool for large-scale

<sup>6</sup><https://github.com/fujimizu/bayon/>

Gas Business Act (Act No. 51 of 1954)

[As of October 12, 1970]

(Definitions)

**Article 2** (1) The term “General Gas Utility Business” as used in this Act shall mean the business of supplying gas via pipelines to meet general demand (excluding, however, businesses generating gas at a gas generating facility prescribed in paragraph (3) and supplying such gas via pipelines). (\*snip\*)

(3) The term “Community Gas Utility Business” as used in this Act shall mean the business of generating gas at a simplified gas generating facility specified by a Cabinet Order (hereinafter referred to as a “Specified Gas Generating Facility”) and supplying such gas via pipelines to meet general demand at not less than 70 gas service points within one housing complex.

(\*snip\*)

(5) The term “Gas Business” as used in this Act shall mean a General Gas Utility Business or Community Gas Utility Business.

**Figure 3: Excerpt from the Gas Business Act (Act No. 51 of 1954) as of October 12, 1970**

data sets. We employed repeated bisection clustering. We set an option to ‘-1 1’, which specifies the limit value of cluster partitions.

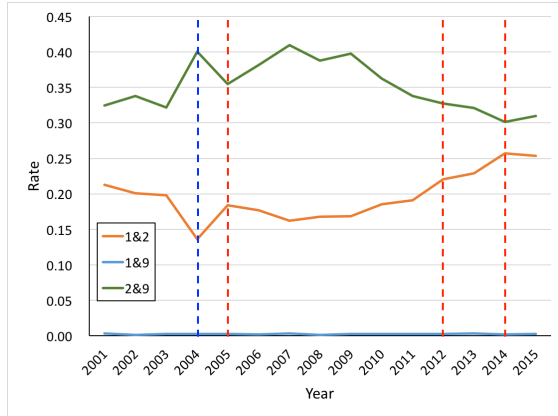
## 6. EVALUATION

Table 1 shows the chronological table of clusters for the Electricity Business Act and Gas Business Act. The column ‘#Acts’ denotes the numbers of the effective acts that have articles for definitions in that year. The number of clusters denoted by ‘#Clst’ changes due to the clustering option. The column ‘Clst ID’ lists the cluster ID to identify the cluster. The numerals in the column ‘Act IDs in the clusters of 1 & 2’ correspond to the acts shown in Table 2. They are in descending order of similarity to the center of the cluster. The IDs 1 and 2 denote the Electricity Business Act and Gas Business Act, respectively. The ‘Sim’ column shows the cosine similarity between the two acts.

Note that the clustering result is not fixed, but contains a random factor. In fact, the clustering ‘Clst ID’ might have no meaning, but we could see the change in the relation between the two acts. They belong to the same cluster for the years 2005, 2006, and 2014, at which the cosine similarity shows high values.

The provisions of the definitions of legal terms in the Gas Business Act were amended in 2004, while the Electricity Business Act was amended three times in 15 years; 2005, 2012, and 2014. Although the amendment of the Electricity Business Act in 2005 followed that of the Gas Business Act in 2004, they were implemented by the same amendment act. Therefore, it would be reasonable that they were amended for the same purpose and in the same way. In fact, they belonged to the same cluster after both were amended. Likewise, the Electricity Business Act’s amendment in 2014 might have been close to the Gas Business Act.

Other acts are amended every year, which affects the val-



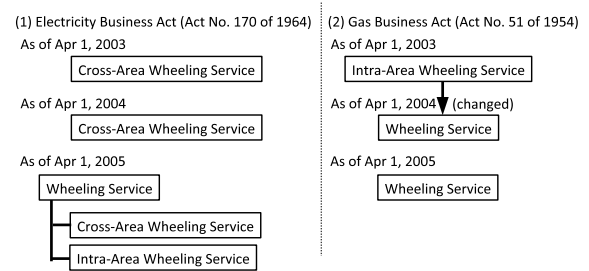
**Figure 4: Coupling rate for two acts (1: Electricity Business Act, 2: Gas Business Act, 9: Hot Spring Act)**

ues of word vectors with *tf-idf* weight. As a result, the cosine similarity between the two acts changes even when neither is amended.

Since the Act ID 9: the Hot Spring Act (Act No. 125 of 1948) sometimes appears in the same cluster as the Gas Business Act (Act No. 51 of 1954), we investigate the relation between these three acts, adding the Electricity Business Act (No. 170 of 1964). The more similar the acts from the viewpoint of the cosine similarity, the higher the probability that they are in the same cluster. Figure 4 shows the coupling rate for two out of the three acts every year by 10,000 trials, where the indices 1, 2, and 9 denote the Electricity Business Act, Gas Business Act, and Hot Spring Act, respectively.

The blue and red dashed lines denote the years that the articles for definitions in the Gas Business Act and the Electricity Business Act are amended, respectively. The orange line represents the coupling rate between the Electricity Business Act and the Gas Business Act.

Since only the Gas Business Act was drastically amended in 2004, the similarity between the two acts decreased. Instead, it returned in 2005 due to a similar amendment. Figure 5 illustrates the change of legal terms about the service provided by these two acts, which explains what happened in the two years. Although the Act on the Partial Amendment of the Electricity Business Act and the Gas Business Act (Act No. 92 of 2003) amended both acts, the Gas Business Act preceded the Electricity Business Act for one year with respect to the enforcement date. The Electricity Business Act had originally defined “Cross-Area Wheeling Service.” The amendment act added “Intra-Area Wheeling Service” and “Wheeling Service” as a broader concept of these two terms. On the other hand, the Gas Business Act had originally defined “Intra-Area Wheeling Service.” Since the amendment act assimilated the name of the service with the Electricity Business Act, the term “Wheeling Service” was defined not as a broader concept of but as a replacement with the term “Intra-Area Wheeling Service.” As a result, there is no common term between the two acts about the service as of April 1, 2004. This is one of the reasons why the coupling rate suddenly dropped down in this year. As of



**Figure 5: Change of legal terms by the Act on the Partial Amendment of the Electricity Business Act and the Gas Business Act (Act No.92 of 2003)**

April 1, 2015, the acts shared a common term again, which got the coupling rate back. Therefore, we could say these two acts are connected, and amendment acts occur close to each other except the case in 2004. As a result, these two acts retain a high degree of similarity in 2012 and, finally, the coupling rate comes to a peak in 2014. We demonstrated the change in relation by the enforcement of amendment acts.

The blue line shows the coupling rate between the Electricity Act and the Hot Spring Act, which have little relation.

The green line denotes the coupling rate between the Gas Business Act and the Hot Spring Act, which seems to complement the orange line. We compared the legal terms between these acts. Unfortunately, they seem to have little relation, but we did find the word ‘gas’ in an explanation. This is because the *tf-idf* value was blown up.

Although this does not indicate a hidden relation to the Gas Business Act, we could suggest a method for finding it.

## 7. CONCLUSION

In this paper, we extracted synchronic similarity in the legal terms among related acts. Related statutes are defined as ones that directly or indirectly refer to each other. Finding directly related statutes of a target is not difficult; simply list all of the statutes that refer to and are referred to by the target. On the other hand, finding indirect relations between statutes without explicit references is difficult. We introduced a clustering method for statutes represented by legal terms at a particular date. Although we picked up two acts, further analysis will reveal relations among legal terms.

## 8. ACKNOWLEDGMENTS

This research was partly supported by the Japan Society for the Promotion of Science KAKENHI Grant-in-Aid for Scientific Research (S) No. 23220005, (A) No. 26240050, and (C) No. 15K00201.

## 9. REFERENCES

- [1] S. Höfler, A. Bünzli, and K. Sugisaki. Detecting Legal Definitions for Automated Style Checking in Draft Laws. Technical Report CL-2011.01, University of Zurich, Institute of Computational Linguistics, 2011.
- [2] K. Kageura. *The Dynamics of Terminology*. John Benjamins, 2002.
- [3] T. Kudo, K. Yamamoto, and Y. Matsumoto. Applying conditional random fields to japanese morphological

**Table 1: Chronological table of clusters for the (1) Electricity Business Act and (2) Gas Business Act**

Date (Apr 1)	# Acts	# Clst	Clst ID	Act IDs in the clusters of 1&2	Sim
2001	769	156	4 139	1, 3 2, 4, 5, 6, 7, 8	0.199
2002	789	162	27 141	1, 3, 5 2, 9, 4, 10, 11, 12	0.201
2003	823	172	4 118	1, 3 2, 5, 4, 6, 7	0.200
2004	851	176	39 95	2, 6, 7, 4 1, 3, 5, 13	0.182
2005	881	179	143	1, 3, 5, 2, 4, 13	0.208
2006	902	194	179	1, 3, 2, 5, 4, 6	0.206
2007	919	183	94 111	1, 3, 5, 13 2, 6, 7, 4, 14	0.201
2008	948	191	53 172	2, 4, 5, 6 15, 3, 16, 1, 13, 17	0.200
2009	966	189	119 187	2, 4, 5, 6, 7 3, 1, 18, 19, 20	0.198
2010	988	195	24 161	2, 9, 10 1, 3, 21, 22	0.200
2011	999	206	6 49	1, 3 2, 4, 10	0.199
2012	1,021	211	8 111	1, 3 2, 6, 7, 4, 10	0.189
2013	1,034	202	106 169	3, 1, 23, 24 2, 5, 6, 4, 7, 9, 10	0.190
2014	1,051	213	124	1, 3, 2, 25	0.213
2015	1,074	217	62 198	2, 5, 4, 7 23, 3, 1, 24, 26	0.210

analysis. In *Proceedings of EMNLP-2004*, pages 230–237, 2004.

- [4] G. Lame. Using NLP Techniques to Identify Legal Ontology Components: Concepts and Relations. In *Law and the Semantic Web*, LNAI3369, pages 169–184. Springer, 2005.
- [5] M. Nakamura, Y. Ogawa, and K. Toyama. Development of Diachronic Terminology from Japanese Statutory Corpora. *Journal of Open Access to Law*, 4(1):16 pages, 2016.
- [6] R. Winkels and R. Hoekstra. Automatic Extraction of Legal Concepts and Definitions. In *Legal Knowledge and Information Systems - JURIX 2012: The Twenty-Fifth Annual Conference*, pages 157–166, 2012.

## APPENDIX

### A. THE ORIGINAL STATUTES

Figures 6 and 7 show the original version of Figures 2 and 3, respectively. Also, Table 3 shows the original Japanese act titles of Table 2.

### B. REGULAR EXPRESSION RULES FOR EXTRACTION AND IDENTIFICATION

**Table 2: List of Acts**

ID	Act Title
1	<b>Electricity Business Act</b> (Act No. 170 of 1964)
2	<b>Gas Business Act</b> (Act No. 51 of 1954)
3	Act on Promotion of Power-Resources Development Tax (Act No. 79 of 1974)
4	Heat Supply Business Act (Act No. 88 of 1972)
5	Industrial Water Supply Business Act (Act No. 84 of 1958)
6	Act on the Supervision of Construction Project to Install Specified Gas Appliances (Act No. 33 of 1979)
7	Act on the Securing of Safety and the Optimization of Transaction of Liquefied Petroleum Gas (Act No. 149 of 1967)
8	Act on the Prevention of Disasters in Petroleum Industrial Complexes and Other Petroleum Facilities (Act No. 84 of 1975)
9	Hot Spring Act (Act No. 125 of 1948)
10	High Pressure Gas Control Act (Act No. 204 of 1951)
11	Gravel Gathering Act (Act No. 74 of 1968)
12	Act on Water Washing Coal Business (Act No. 134 of 1958)
13	Waterworks Act (Act No. 177 of 1957)
14	Act on Rational Use and Proper Management of Fluorocarbons (Act No. 64 of 2001)
15	Act on Development of Areas Around Electric Facilities for Electricity Generation (Act No. 78 of 1974)
16	Act on Special Measures Concerning the Promotion of Establishment Area for the Nuclear Power Facilities etc. (Act No. 148 of 2000)
17	Specified Multipurpose Dam Act (Act No. 35 of 1957)
18	Electrician Act (Act No. 139 of 1960)
19	Act on Regulation of Electrical Contracting Business (Act No. 96 of 1970)
20	Basic Act on the Advancement of Utilizing Geospatial Information (Act No. 63 of 2007)
21	Act on Special Provisions of the Customs Act, etc. Attendant upon the Enforcement of the Convention on International Transport of Goods Under Cover of TIR Carnets (TIR Convention) (Act No. 65 of 1971)
22	Special Measures Act on Support for Independence of Homeless People (Act No. 105 of 2002)
23	Act on Special Measures Concerning Procurement of Electricity from Renewable Energy Sources by Electricity Utilities (Act No. 108 of 2011)
24	Act on the Promotion of Use of Nonfossil Energy Sources and Effective Use of Fossil Energy Materials by Energy Suppliers (Act No. 72 of 2009)
25	Act on Special Measures for Prevention of Mining Pollution by Metal Mining etc. (Act No. 26 of 1973)
26	Basic Act on the Advancement of Utilizing Biomass (Act No. 52 of 2009)

ガス事業法 (昭和 29 年法律第 51 号)

(定義)  
 第二条 この法律において、「ガス事業」とは、一般の需用に応じ導管によりガスを供給する事業をいう。

ガス事業法の一部を改正する法律  
 (昭和 45 年法律第 18 号)

第二条第一項中「ガス事業」を「一般ガス事業」に、「需用」を「需要」に改め、「供給する事業」の下に「(第三項に規定するガス発生設備においてガスを発生させ、導管によりこれを供給するものを除く。)」を加え、(\*snip\*) 同条第一項の次に次の五項を加える。  
 (\*snip\*)  
 3 この法律において「簡易ガス事業」とは、一般の需要に応じ、政令で定める簡易なガス発生設備 (以下「特定ガス発生設備」という。) においてガスを発生させ、導管によりこれを供給する事業であつて、一の団地内におけるガスの供給地点の数が七十以上のものをいう。  
 (\*snip\*)  
 5 この法律において「ガス事業」とは、一般ガス事業及び簡易ガス事業をいう。

Figure 6: The original version of Figure 2: Excerpt from the Gas Business Act (Act No. 51 of 1954) and the Act on Partial Amendment of the Gas Business Act (Act No. 18 of 1970)

The regular expression rules used for extraction and identification are shown in Figures 8 and 9. Note that they are written in Japanese. There is no English translation, because they include fragments of Japanese words.

ガス事業法 (昭和 29 年法律第 51 号)  
 昭和 45 年 10 月 12 日時点

(定義)  
 第二条 この法律において、「一般ガス事業」とは、一般の需要に応じ導管によりガスを供給する事業 (第三項に規定するガス発生設備においてガスを発生させ、導管によりこれを供給するものを除く。) をいう。  
 (\*snip\*)  
 3 この法律において「簡易ガス事業」とは、一般の需要に応じ、政令で定める簡易なガス発生設備 (以下「特定ガス発生設備」という。) においてガスを発生させ、導管によりこれを供給する事業であつて、一の団地内におけるガスの供給地点の数が七十以上のものをいう。  
 (\*snip\*)  
 5 この法律において「ガス事業」とは、一般ガス事業及び簡易ガス事業をいう。

Figure 7: The original version of Figure 3: Excerpt from the Gas Business Act (Act No. 51 of 1954) as of October 12, 1970

- Regular Expression Rules
- (a) to extract articles for legal term definitions:  
 "~[\( ([^ ] \)]\*定義 [ ] \) 等及]"
  - (b) to identify presence of itemization for definitions:  
 "用語.+当該各号",  
 "規定の解釈.+次の定義",  
 "とは、次に掲げるものとする。\$"
  - (c) to extract definitions from an explanatory sentence:  
 "とは、(.+)をいう。\$",  
 "とは、(.+をいい、.+もの)とする。\$",  
 "とは、(.+をいう。.+)\$"
  - (d) to identify the legal term as a hypernym of terms in its explanatory sentence:  
 "( {[^ } ]+} (又は|若しくは|及び|並びに))+ {[^ } ]+} (又は|若しくは|及び|並びに) {[^ } ]+} \$", "( {[^ } ]+}, )\* {[^ } ]+} (又は|若しくは|及び|並びに) {[^ } ]+} \$", "~ {[^ } ]+} \$"
  - (e) to identify the legal term as a hyponym of the term in its explanatory sentence:  
 "{([ ^ } ]+)} で(あつて)?、", "{([ ^ } ]+)} \$"

Figure 8: Regular expression rules (a)-(e)

Table 3: The original version of Table 2: List of Acts

ID	Act Title
1	電気事業法 (昭和 39 年法律第 170 号)
2	ガス事業法 (昭和 29 年法律第 51 号)
3	電源開発促進税法 (昭和 49 年法律第 79 号)
4	熱供給事業法 (昭和 47 年法律第 88 号)
5	工業用水道事業法 (昭和 33 年法律第 84 号)
6	特定ガス消費機器の設置工事の監督に関する法律 (昭和 54 年法律第 33 号)
7	液化石油ガスの保安の確保及び取引の適正化に関する法律 (昭和 42 年法律第 149 号)
8	石油コンビナート等災害防止法 (昭和 50 年法律第 84 号)
9	温泉法 (昭和 23 年法律第 125 号)
10	高圧ガス取締法 (昭和 26 年法律第 204 号)
11	砂利採取法 (昭和 43 年法律第 74 号)
12	水洗炭業に関する法律 (昭和 33 年法律第 134 号)
13	水道法 (昭和 32 年法律第 177 号)
14	フロン類の使用の合理化及び管理の適正化に関する法律 (平成 13 年法律第 64 号)
15	発電用施設周辺地域整備法 (昭和 49 年法律第 78 号)
16	原子力発電施設等立地地域の振興に関する特別措置法 (平成 12 年法律第 148 号)
17	特定多目的ダム法 (昭和 32 年法律第 35 号)
18	電気工事士法 (昭和 35 年法律第 139 号)
19	電気工事業の業務の適正化に関する法律 (昭和 45 年法律第 96 号)
20	地理空間情報活用推進基本法 (平成 19 年法律第 63 号)
21	コンテナに関する通関条約及び国際道路運送手帳による担保の下で行なう貨物の国際運送に関する通関条約 (T I R 条約) の実施に伴う関税法等の特例に関する法律 (昭和 46 年法律第 65 号)
22	ホームレスの自立の支援等に関する特別措置法 (平成 14 年法律第 105 号)
23	電気事業者による再生可能エネルギー電気の調達に関する特別措置法 (平成 23 年法律第 108 号)
24	エネルギー供給事業者による非化石エネルギー源の利用及び化石エネルギー原料の有効な利用の促進に関する法律 (平成 21 年法律第 72 号)
25	金属鉱業等鉱害対策特別措置法 (昭和 48 年法律第 26 号)
26	バイオマス活用推進基本法 (平成 21 年法律第 52 号)

Regular Expression Rules

(f) to extract both a legal term and definitions (1: legal term, 2: definition (noun phrase), 3: definition (verb phrase))  
 "[^「」]+)とは、?([^[「」]+)(を、|をいい、))  
 「",  
 "[^[「」]+)とは、?([^[「」]+をいう。)",  
 "[^[「」]+)とは、?([^[「」]+をいい、^[「」]+)とする。)",  
 "[^[「」]+)とは、?([^[「」]+)(をいい|とし、^[「」]+(とみなす|で定める|を含む。))",  
 "[^[「」]+)には、?([^[「」]+)とする。)",  
 "[^[「」]+)とは、?([^[「」]+(含むもの|をいうもの))とする。)",  
 "[^[「」]+)とは、?([^[「」]+を総称する。)",  
 "[^[「」]+)又は「.+」とは、?([^[「」]+)(を、|をいい、)) 「",  
 "[^[「」]+)又は「.+」とは、?([^[「」]+をいう。)",  
 "[^[「」]+)、「^[「」]+)又は「.+」とは、?([^[「」]+をいう。)",  
 "[^[「」]+)、「^[「」]+)、「^[「」]+)又は「.+」とは、?([^[「」]+)をいう。)",  
 "[^[「」]+) (、「^[「」]+)又は「.+」の意義は、?([^[「」]+)による。)",  
 "[^[「」]+)又は「.+」の意義は、?([^[「」]+)による。)",  
 "[^[「」]+)の意義は、?([^[「」]+)による。)",  
 "[^[「」]+)とは、?([^[「」]+をいい、.+を除く。)",  
 "[^[「」]+)、「^[「」]+)、「^[「」]+)、「^[「」]+)又は「.+」とは、?([^[「」]+)をいう。)",  
 "[^[「」]+)、「^[「」]+)、「^[「」]+)、「^[「」]+)、「^[「」]+)又は「.+」とは、?([^[「」]+)をいう。)",  
 "[^[「」]+)、「^[「」]+)、「^[「」]+)、「^[「」]+)又は「.+」とは、?([^[「」]+)をいう。)",  
 "[^[「」]+)、「^[「」]+)、「^[「」]+)、「^[「」]+)又は「.+」とは、?([^[「」]+)をいい、^[「」]+)とに分ける。)",  
 "[^[「」]+)とは、?([^[「」]+)をいい、以下「^[「」]+)と略称する。)",  
 "([^[「」]+)とは、^[「」]+をいい、以下「^[「」]+)と略称する。",  
 "及び「^[「」]+)は、?([^[「」]+)とする。)",  
 "[^[「」]+)及び「.+」は、?([^[「」]+)とする。)",  
 "[^[「」]+)、「^[「」]+)及び「.+」は、?([^[「」]+)とする。)",  
 "[^[「」]+)、「^[「」]+)、「^[「」]+)及び「.+」は、?([^[「」]+)とする。)",  
 "[^[「」]+)、「^[「」]+)、「^[「」]+)及び「.+」は、?([^[「」]+)とする。)",

Figure 9: Regular expression rules (f)