Extraction of Legal Definitions and Their Explanations with Accessible Citations

Makoto Nakamura[1], Yasuhiro Ogawa[2,3], and Katsuhiko Toyama[2,3]

[1] Japan Legal Information Institute, Graduate School of Law, Nagoya University, Nagoya, Japan
mnakamur@law.nagoya-u.ac.jp
[2] Information Technology Center, Nagoya University, Nagoya, Japan
[3] Graduate School of Information Science, Nagoya University, Nagoya, Japan
{toyama,ogawa}@is.nagoya-u.ac.jp

Abstract. The aim of this paper is to produce a Japanese legal terminology consisting of legal terms and their explanations that includes accessible citations. Although we have succeeded in finding over 14,000 terms with high precision, 23.1 percent of the correct explanations included citations that were inaccessible due to context-dependent format. We propose a method for revising explanatory sentences that takes into account XML-tag annotation for context-independent format for all citations. The effectiveness of this method is confirmed by our experimental results.

Keywords: Japanese statutes, Definitions, XML, Citations.

1 Introduction

The goal of this research is to construct a legal terminology for translators in which each entry consists of a legal term and its explanation. This terminology is expected to improve not only systematic translations, but also appropriate word selection depending on context. This study is related to the Ministry of Justice's Japanese Law Translation Database System project[1], which was released in 2009 [1]. The number of laws translated into English for publication has increased slowly and, as of August 7, 2013, only 339 of the over 7,800 ($< 5\%$) acts and regulations have been translated. One of the most crucial issues remaining for translation is disunity in word selection. Since a number of human translators are involved, many Japanese legal terms have a variety of English translations. We have even found that act titles are often translated differently in citations [2]. Obviously, each expression should have a specific translation for consistency of meaning. Although the government has compiled a standard translation dictionary for legal terms, the number of entries (3,594 in the latest version) is not sufficient for unified translation.

---

[1] http://www.japaneselawtranslation.go.jp

We first focused on collecting legal terms as defined by statute. We started by compiling a Japanese act corpus consisting of all acts enacted between 1947 and 2012, from which all the legal definitions can be extracted. This completeness of processing for all the acts is significant in our study. We developed an automatic method for extracting the tuples of a legal definition and its explanation from the Japanese act corpus. As a result, we succeeded in finding over 14,000 terms with high precision [3]. However, we found that 23.1 percent of the correct explanations included inaccessible citations. The notation of citations in statutes varies for provisions or acts, some of which are written with an abbreviation or a relative address from the explanation location. In other words, citations found in the explanation written in particular notations can no longer access the specified provision or act. If written with an accessible notation, the explanations would be readable and refer to the specified provision or act.

Our aim is to provide a legal terminology containing the tuples of a legal term and its explanation that also includes an accessible citation. This is achieved by replacing the citation with particular expressions to an independent notation located apart from the original location of the explanation. Our main idea is to XMLize Japanese acts and annotate citation tags in an accessible format. This task is an application of natural language processing to Japanese legal texts. Some studies on reference resolution in legal texts have been written in several European languages [4–7]. In this paper, we deal with Japanese acts by considering the characteristics of the Japanese language and Japanese statutes.

Our paper is organized as follows: In Section 2, we provide the background of this study. Section 3 discusses the types of citations included in the explanatory sentences. In Section 4, we propose a method for annotating in a context-independent format. We describe how this method works with experiments in Section 5, and we conclude and discuss our future work in Section 6.

2   Japanese Legal Text Processing

In this section, we briefly explain Japanese legal text processing. We introduce the basic structure of Japanese laws and the Japanese legislative system in the first two subsections, and conclude with a discussion of our previous study using our Japanese act corpus.

2.1   Basic Explanation of Japanese Laws

In general, laws are roughly divided into written and unwritten categories. Although unwritten laws include local customs and judicial precedents, we do not deal with these in this paper. Written laws are also called statutes, which are further divided into acts and bylaws. While acts are enacted by the National Diet (Parliament), bylaws consist of orders enacted by the cabinet and ordinances and regulations enacted by various ministries.

In this paper, we focus on statutory texts. A statute consists of a number of articles, each of which may be further subdivided into a number of paragraphs or items. Articles, paragraphs, and items have sequential numbers with different typefaces. A provision denotes an independent article or a paragraph.

2.2   Basic Explanation of the Japanese Legislative System

The rational nature of the legislative system of Japanese law maintains the notation of expressions of statutes.[2] Although the Cabinet and Diet members can submit a bill to the National Diet, most bills are introduced by the Cabinet. In this case, the proper authority for that law basically makes a draft of the bill. Once this is accomplished, the authority negotiates with other authorities. The Cabinet Legislation Bureau then closely examines the draft in terms of inconsistency with other statutes, expressions, formats, and so on. As a result, even the usage of commas and periods is maintained. When a Diet member submits a bill, it is reviewed by the Legislation Bureau of the House of Representatives or Councilors.

Not every country's legislative system is similar to that of Japan. In the United Kingdom, the legislature's description check is not as strict, as in most cases the bill is drafted outside of the ministry. In the United States, there is no organization or system for the legislature's description check. In Asian countries other than Japan and Korea, often each ministry independently prepares a draft of a bill without coordinating with other ministries. As a result, the notation of bills differs among ministries. Moreover, in some countries bills are often modified during deliberation in the national assembly, while bills mostly pass the National Diet in Japan as drafted.

Since this political process results in consistencies in notation, this strict wording style may be an idiosyncratic feature of Japanese statutes. This suggests that simple text processing is sufficient to locate important terms or phrases with conventional expressions.

2.3   Definition of Legal Terms and Their Explanations

Although several methods for the extraction of legal terms have been proposed [8–11], the organization of terminology differs depending on their purposes. In this paper, a legal term is a term explicitly defined in an act prior to use and includes both the legal term and its explanation. These terms typically take the form of

- An independent provision or
- A statement in parentheses.

Figure 1[3] shows examples of definitions found in both provisions and parentheses, where Article 2, paragraphs (1), (2), and (3) are independent provisions that define the terms "*administrative organs,*" "*incorporated administrative agencies, etc.,*" and "*official statistics,*" respectively. A defined term is placed in quotations

---

[2] This section is based on our discussion with Prof. Matsuura of the Graduate School of Law, Nagoya University.

[3] Hereinafter, Japanese sentences are immediately followed by their English translations. We referred to the Japanese Law Translation Database System for English translations [1].

統計法（平成十九年法律第五十三号）/ Statistics Act (Act No. 53 of 2007)

第二条　この法律において「行政機関」とは、法律の規定に基づき内閣に置かれる機関若しくは内閣の所轄の下に置かれる機関、宮内庁、内閣府設置法（平成十一年法律第八十九号）第四十九条第一項若しくは第二項に規定する機関又は国家行政組織法（昭和二十三年法律第百二十号）第三条第二項に規定する機関をいう。/

Article 2　(1) The term "*administrative organs*" as used in this Act means organs established within the Cabinet or organs established under the jurisdiction of the Cabinet pursuant to the provisions of laws, the Imperial Household Agency, organs provided in Article 49, paragraph (1) or paragraph (2) of the Act for Establishment of the Cabinet Office (Act No. 89 of 1999) or organs provided in Article 3, paragraph (2) of the National Government Organization Act (Act No. 120 of 1948).

2　この法律において「独立行政法人等」とは、次に掲げる法人をいう。/

(2)　The term "*incorporated administrative agencies, etc.*" as used in this Act means juridical persons listed as follows:

一　独立行政法人（独立行政法人通則法（平成十一年法律第百三号）第二条第一項に規定する独立行政法人をいう。次号において同じ。）/

(i)　*Incorporated administrative agencies*　(meaning incorporated administrative agencies provided in Article 2, paragraph (1) of the Act on General Rules for Incorporated Administrative Agencies　(Act No. 103 of 1999; the same shall apply in the following items); (snip)

3　この法律において「公的統計」とは、行政機関、地方公共団体又は独立行政法人等（以下「行政機関等」という。）が作成する統計をいう。/

(3)　The term "*official statistics*" as used in this Act means statistics produced by administrative organs, local public entities, or incorporated administrative agencies, etc.　(hereinafter referred to as "*administrative organs, etc.*").

Fig. 1. Example of definitions

("*term*") and its explanation is underlined. Defined terms are properly extracted using pattern match [12]. In addition, definitions appear in parentheses in Article 2, paragraph (2), item (i) and paragraph (3) for the terms "*incorporated administrative agencies*" and "*administrative organs, etc.*," respectively.

The second item is further divided into two types:

– A defined term appears in parentheses following its explanation in the main text, as shown in Article 2, paragraph (3). The term "*administrative organs, etc.*" is the defined term in this example. Abbreviations are often defined in this way.

– A sentence in parentheses explains the legal term just before the parentheses, as shown in Article 2, paragraph (2), item (i) "*Incorporated administrative agencies.*"

簡易生命保険法の一部を改正する法律（平成三年法律第三十号）/

Act for Partial Amendment of the Postal Life Insurance Act (Act No. 30 of 1991)

簡易生命保険法（昭和二十四年法律第六十八号）の一部を次のように改正する。/

The Postal Life Insurance Act (Act No. 68 of 1946) shall be partially amended as follows:

第二十四条第二項中「七十二万円」を「九十万円」に改める。/

The term "720,000 yen" in Article 24, Paragraph (2) shall be replaced with the term "900,000 yen."

Fig. 2. Example of an amendment act

Legal terms or explanations in parentheses are easily extracted by analysis of the character string, but analysis of the content outside of the parentheses is not so simple [13]. The difficulty comes from wording that is peculiar to statutory sentences. Despite the presence of high-quality dependency parsers for Japanese, we cannot count on their performance with legal texts. Since legal sentences are designed to avoid ambiguity of expression, they are likely to be long and syntactically complicated, which often leads to a parsing failure. Therefore, we employ a simple method based on pattern match and thus do not rely on a syntactic parser.

2.4   Japanese Act Corpus

We compiled a corpus of all of the Japanese acts, consisting of 9,915 acts enacted up to 2012 since enforcement of the new constitution of Japan in 1947. The size of the corpus is 252 MByte. This Japanese act corpus is based on articles of legislation in official gazettes. Since most of these acts, especially the older ones, are digitally scanned, there are many typographical errors that are not included in the published versions. We developed a preprocessor to address these typographical errors.

Since amendment acts describe how to revise pre-existing acts using amendment language [14], it is difficult to properly extract legal terms unless accurately consolidated. Figure 2 shows an example of an amendment act. Despite the terms in quotes, there is no term definition in this provision. Since the terms in quotes are supposed to be consolidated into the main clause of the original act, a tuple of a legal term and its explanation is unlikely to be extracted from the amendment act. Therefore, we eliminate in advance all acts concerning the amendment or repeal of pre-existing acts, which can be inferred from their titles, as well as supplementary provisions in other acts, which may include amendment.

2.5   Evaluation of Our Previous Study

Table 1 shows the total number of definitions and their explanations collected in our previous experiment [3]. Since some terms are defined in multiple acts, the

Table 1. Analysis of collected definitions and their explanations

Definitions	#Tokens	#Types	Correct	Incorrect	Precision	Recall
in provisions	5,250	3,799	98 (15)	2	0.980	0.980
in parentheses	9,624	6,030	84 (27)	16	0.840	0.392
Total	14,874	9,368	182 (37)	18	0.910	0.511

number of types differs from that of tokens. The precision scores were calculated from 100 samples chosen at random. The figures in parentheses under "Correct" denote the number of correct explanations, including citations in which their specified provisions or acts are inaccessible, as will be described in detail in Section 3. The recall scores were calculated based on the assumption that all legal terms in quotations are perfectly obtained with our method. Since 11,004 terms in parentheses are still unprocessed by the current method despite detection of only the terms, the recall score for the terms in parentheses was estimated. According to our experimental results, 14,874 tokens and 9,368 types of terms were extracted with high precision.

3 Citations Included in Explanatory Sentences

A number of the correctly extracted explanations include citations to other items, paragraphs, articles, or acts. Our analysis revealed that $23.1 (= (15 + 27)/(98 + 84) \times 100)$ percent of the correct explanations include citations for which the specified provisions or acts are inaccessible. Further investigation revealed that 23.9 percent of the whole includes the inaccessible citations.

The citation format is categorized as follows:

1. The absolute addressing method (**TYPE1**), which is expressed as the full notation of the location for the reference consisting of the title of the act, the article number, and so on, as shown in Article 2, paragraph (1) in Fig. 1.
2. The absolute addressing method with abbreviated expression (**TYPE2**), which shows only the notational difference between the reference address and the current address or the address previously referred to, as shown in Article 2, paragraph (4) in Fig. 3.
3. The relative addressing method (**TYPE3**), which shows the relative distance from the current address, such as a previous article, following items, and so on, as shown in Article 5, paragraph (2) in Fig. 3.

TYPE2 and TYPE3 must coincide with the current address.

From the provisions shown in Fig. 1 and Fig. 3, the above-mentioned tuples are enumerated with their citations underlined as follows:

1. *Administrative organs*: Organs established within the Cabinet or organs established under the jurisdiction of the Cabinet pursuant to the provisions of laws, the Imperial Household Agency, organs provided in Article 49, paragraph (1) or paragraph (2) of the Act for Establishment of the Cabinet Office (Act No. 89 of 1999), or organs provided in Article 3, paragraph (2) of the National Government Organization Act (Act No. 120 of 1948).

統計法（平成十九年法律第五十三号）/ Statistics Act (Act No. 53 of 2007)

第二条 / **Article 2**

4　この法律において「基幹統計」とは、次の各号のいずれかに該当する統計をいう。/

(4)　The term "*fundamental statistics*" as used in this Act means statistics falling under any of the following items:

　　一　第五条第一項に規定する国勢統計/

　　(i)　Population census statistics provided in Article 5, paragraph (1);

　　二　第六条第一項に規定する国民経済計算/

　　(ii)　National accounts provided in Article 6, paragraph (1); (snip)

第五条　総務大臣は、本邦に居住している者として政令で定める者について、人及び世帯に関する全数調査を行い、これに基づく統計（以下この条において「国勢統計」という。）を作成しなければならない。 /

Article 5　(1) With regard to persons specified by a Cabinet Order as those residing in Japan, the Minister of Internal Affairs and Communications shall conduct a complete census concerning individuals and households and produce statistics based on such a census (hereinafter referred to as "*population census statistics*" in this Article).

2　総務大臣は、前項に規定する全数調査（以下「国勢調査」という。）を十年ごとに行い、国勢統計を作成しなければならない。（以下略）/

(2)　The Minister of Internal Affairs and Communications shall conduct a complete census as specified in the preceding paragraph (hereinafter referred to as the "*population census*") every ten years and produce population census statistics. (snip)

Fig. 3. Example of explanations including specified citations

2-1. *Fundamental statistics*: Population census statistics provided in Article 5, paragraph (1).

2-2. *Fundamental statistics*: National accounts provided in Article 6, paragraph (1).

 3. *Population census*: A complete census as specified in the preceding paragraph.

We intend to construct a terminology from these tuples in which each explanation is written in an independent phrase apart from the original act. In other words, citations should not be written in the notation of TYPE2 or TYPE3. From this standpoint, we can review explanations as follows:

− Item 1. requires citing to specified provisions for which the full notation of TYPE1 is accessible.
− Both items 2-1. and 2-2. refer to provisions in other articles belonging to the same act as these provisions. In this case, addresses are expressed in TYPE2, that is, only the notational difference between the reference address and that of the explanation location or the address previously referred to.

– Item 3. employs TYPE3, referring to the preceding paragraph. It is impossible to reach the designated provision unless the location of this explanation is clear.

In this paper, we propose a method to replace the TYPE2 and TYPE3 notations to that of TYPE1.

4 Approach to Revision of Explanatory Sentences

We propose a method for revising the explanatory sentences that takes into account annotation of the absolute address for all the citations, as follows: (1) XMLizing Japanese acts; (2) extending the XML format to absolute addressing; and (3) running the extraction method with revision of the addresses. We address these issues in the following subsections.

4.1 XML Tagging to the Japanese Act Corpus

The Japanese Law Translation Database System project provides law data in XML format, as well as document type definition (DTD), for Japanese statutory laws including definitions for 103 elements and 75 attributes. According to the *One Source Multi Use* policy, users can easily reuse and reformat this law data for their own purposes [1].[4]

We applied the XML format to Japanese acts in our corpus introduced in Section 2.4. Figure 4 shows an example of an act in the XML format. Note that this act is translated into English for readability.

The project also developed a tool for automatic annotation using pattern match based on the strict wording style mentioned in Section 2.2. Although actual statutory data in the XML format are released after manual modification, only five percent of all the statutes has been completed so far due to the progress of translation. We divert this tool to our Japanese act corpus.

4.2 Method for Annotating Absolute Addresses

We extend the XML format to the annotation of absolute addressing. Although we dealt with all the strings in a plain text file in the previous version, the new method is restricted to strings in the 'Sentence' tag or other content tags. The string is annotated with an 'a' tag if it matches an act title in the list of act titles or the notation of TYPE2 or TYPE3. If the newly annotated 'a' tag is nested, the inside annotation is eliminated. Finally, all the 'a' tags are reviewed and an attribute of 'href' is added. Obviously, both 'a' and 'href' are loanwords from HTML.

Figure 5 shows XML tags for citations where the provisions correspond to the ones shown in Fig. 3 and part of Fig. 1. The new attribute 'id' is added

---

[4] The DTD is downloadable at http://www.japaneselawtranslation.go.jp/

Statistics Act (Act No. 53 of 2007)

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE Law PUBLIC "-//JaLII//DTD J-STATUTE 1.0//EN" "jstatute.dtd">
<Law OriginalPromulgateDate="May 23, 2007" LawType="Act" Lang="en" Year="19"
                                              Era="Heisei" Num="053">
  <LawNum>Act No. 53 of May 23, 2007</LawNum>
  <LawBody>
    <LawTitle>Statistics Act</LawTitle>
    <EnactStatement>
      All provisions of the Statistics Act (Act No. 18 of 1947) shall be revised.
    </EnactStatement>
    <TOC>
      <TOCLabel>Table of Contents</TOCLabel>
                          *** snip ***
    </TOC>
    <MainProvision>
      <Chapter Num="1" >
        <ChapterTitle>Chapter I General Provisions</ChapterTitle>
                          *** snip ***
      <Article Num="2" >
        <ArticleCaption>(Definitions)</ArticleCaption>
        <ArticleTitle>Article 2</ArticleTitle>
        <Paragraph Num="1" >
          <ParagraphNum>(1)</ParagraphNum>
          <ParagraphSentence>
            <Sentence>The term "administrative organs" as used in this Act means
                          *** snip ***
            Act (Act No. 120 of 1948).</Sentence>
          </ParagraphSentence>
        </Paragraph>
        <Paragraph Num="2" >
          <ParagraphNum>(2)</ParagraphNum>
          <ParagraphSentence>
            <Sentence>The term "incorporated administrative agencies, etc." as used in
            this Act means juridical persons listed as follows:</Sentence>
          </ParagraphSentence>
          <Item Num="1">
            <ItemTitle>(i)</ItemTitle>
            <ItemSentence>
              <Sentence>Incorporated administrative agencies (meaning incorporated
                          *** snip ***
              1999; the same shall apply in the following items);</Sentence>
            </ItemSentence>
          </Item>
          <Item Num="2">
            <ItemTitle>(ii)</ItemTitle>
            <ItemSentence>
              <Sentence>Juridical persons specified by a Cabinet Order among those
                          *** snip ***
              administrative agencies is required for their incorporation.</Sentence>
            </ItemSentence>
          </Item>
        </Paragraph>
      </Article>
                          *** snip ***
      </Chapter>
    </MainProvision>
    <SupplProvision >
      <SupplProvisionLabel>Supplementary Provisions</SupplProvisionLabel>
                          *** snip ***
    </SupplProvision>
  </LawBody>
</Law>
```

Fig. 4. Example of an act in XML format (English version)

```
Statistics Act (Act No. 53 of 2007)

        <Article Num="2" id="at2">
          <ArticleCaption>(Definitions)</ArticleCaption>
          <ArticleTitle>Article 2</ArticleTitle>
          <Paragraph Num="1" id="at2pr1">
            <ParagraphNum>(1)</ParagraphNum>
            <ParagraphSentence>
              <Sentence>The term "administrative organs" as used in this Act means
                        organs established within the Cabinet or organs established
                        under the jurisdiction of the Cabinet pursuant to the provisions
                        of laws, the Imperial Household Agency, organs provided in
                        <a href="H11HO089.html#at49pr1">Article 49, paragraph (1)</a> or
                        <a href="H11HO089.html#at49pr2">paragraph (2) of the Act for
                        Establishment of the Cabinet Office (Act No. 89 of 1999)</a> or
                        organs provided in <a href="S23HO120.html#at3pr2">Article 3,
                        paragraph (2) of the National Government Organization Act
                        (Act No. 120 of 1948)</a>.</Sentence>
            </ParagraphSentence>
          </Paragraph>
                              *** snip ***
          <Paragraph Num="4" id="at2pr4">
            <ParagraphNum>(4)</ParagraphNum>
            <ParagraphSentence>
              <Sentence>The term "fundamental statistics" as used in this Act means
                        statistics falling under any of the following items:</Sentence>
            </ParagraphSentence>
            <Item Num="1" id="at2pr4it1">
              <ItemTitle>(i)</ItemTitle>
              <ItemSentence>
                <Sentence>Population census statistics provided in
                          <a href="#at5pr1">Article 5, paragraph (1)</a>;</Sentence>
              </ItemSentence>
            </Item>
            <Item Num="2" id="at2pr4it2">
              <ItemTitle>(ii)</ItemTitle>
              <ItemSentence>
                <Sentence>National accounts provided in
                          <a href="#at6pr1">Article 6, paragraph (1)</a>;</Sentence>
              </ItemSentence>
            </Item>
                                *** snip ***
          </Paragraph>
        </Article>
          <Article Num="5" id="at5">
            <ArticleCaption>(Population Census Statistics)</ArticleCaption>
            <ArticleTitle>Article 5</ArticleTitle>
            <Paragraph Num="2" id="at5pr2">
              <ParagraphNum>(2)</ParagraphNum>
              <ParagraphSentence>
                <Sentence Num="1" Function="Main">The Minister of Internal Affairs and
                          Communications shall conduct a complete census as specified
                          in <a href="#at5pr1">the preceding paragraph</a> (hereinafter
                          referred to as the "population census") every ten years and
                          produce population census statistics.</Sentence>
                <Sentence Num="2" Function="Proviso"> *** snip *** </Sentence>
              </ParagraphSentence>
            </Paragraph>
          </Article>
```

Fig. 5. Example of XML tags for citation (English version)

to the tags, denoting articles, paragraphs, items, and so on for the absolute address. For example, the attribute '`id="at2pr4it1"`' means the item is located at 'Article 2, paragraph (4), item (1).' Likewise, a citation is tagged with its absolute address denoted by the attribute '`href`' in the '`a`' tag. The attribute takes on the notation of *"(Statute)#(Provision),"* where the former part denotes the cited statute expressed by a part of the uniform resource identifier (URI), which enables access to the statutory database run by the Ministry of Internal Affairs and Communications[5]. It is left blank if the source statute is cited. The latter uses the same format as the '`a`' tag. For example, the attribute '`href="H11HO089.html#at49pr1"`' shows Article 49, paragraph (1) of the Act for Establishment of the Cabinet Office (Act No. 89 of 1999). This act can be seen on the online database[6].

4.3   Extraction Method with Revision of Citation Addresses

In the process of extraction, a text may include a citation expression with an '`a`' tag, which needs to be replaced with its absolute address. Given an '`a`' tag with an absolute address, the citation is correctly decoded to a Japanese expression.

Examples are shown as follows:

1.′ `<a href="H11HO089.html#at49pr1">Article 49, paragraph (1)</a>`
   ⇒ Article 49, paragraph (1) of the Act for Establishment of the Cabinet Office (Act No. 89 of 1999)

2-1.′ `<a href="#at5pr1">Article 5, paragraph (1)</a>`
   ⇒ Article 5, paragraph (1) of Statistics Act (Act No. 53 of 2007)

2-2.′ `<a href="#at6pr1">Article 6, paragraph (1)</a>`

   ⇒ Article 6, paragraph (1) of Statistics Act (Act No. 53 of 2007)

3.′ `<a href="#at5pr1">the preceding paragraph</a>`
   ⇒ Article 5, paragraph (1) of Statistics Act (Act No. 53 of 2007)

Item 1.′ shows a part of a citation that is tagged because the rest is shared with another connected with a coordinate conjunction[7]. According to the attribute `href`, the phrase "Article 49, paragraph (1)" is decoded and the act title is added by referring to the list of act titles. The remainder in the other items can also be replaced with absolute addresses corresponding to their `href` tags. Since the statute title is left blank, the title of the source statute, Statistics Act (Act No. 53 of 2007), is added. This can also be replaced with Japanese expressions according to the same procedure.

---

[5] `http://www.e-gov.go.jp/` (in Japanese)

[6] `http://law.e-gov.go.jp/htmldata/H11/H11HO089.html` (in Japanese)

[7] The regions of the citations tagged for Article 49, paragraph (1) and paragraph (2) do not correspond to that of the Japanese version due to a difference in word order. They are actually separated into "Article 49, paragraph (1) of the Act for Establishment of the Cabinet Office (Act No. 89 of 1999)" and "paragraph (2)," respectively.

5   Experimental Results

We examined the collected explanations. In the XML-tagging process, 148 out of 9,915 acts failed due to the presence of rare styles. These are still under investigation. Moreover, 15 percent of the citations were not annotated correctly according to calculation using a random sampling of 100. One reason is the failure of annotation for statutory titles.

In the previous study [3], the precision scores were calculated from 100 samples chosen at random for each of the definitions in provisions and in parentheses, of which 42 out of 182 correct explanations (23.1%) included citations with a context-dependent format. Table 2 shows an experimental result for the inaccessible citations included in the previous study [3]. A new trial with the same samples succeeded in replacing 27 out of these 42 citations with those in context-independent format except for 13 for failure of annotation and 2 for need of anaphora resolution.

Examples of revised explanations are:

2-1.″ *Fundamental statistics*: Population census statistics provided in Article 5, paragraph (1) of Statistics Act (Act No. 53 of 2007).

2-2.″ *Fundamental statistics*: National accounts provided in Article 6, paragraph (1) of Statistics Act (Act No. 53 of 2007).

3.″ *Population census*: A complete census as specified in Article 5, paragraph (1) of Statistics Act (Act No. 53 of 2007).

Figure 6 shows an act that caused a failure of annotation. The citations with underlines are expected to be replaced as follows:

– *The level of consumption*: The calculated level, prescribed in Article 1, paragraph 7 of the Protocol, of consumption prescribed in Article 1, paragraph 6 of the Protocol.

However, the tuple was extracted as follows:

– *The level of consumption*: The calculated level, prescribed in Article 1 of Act on the Protection of the Ozone Layer Through the Control of Specified Sub Substances and Other Measures, paragraph 7 of the Protocol, of consumption prescribed in paragraph 6 of Article 1 of Act on the Protection of the Ozone Layer Through the Control of Specified Substances and Other Measures.

Since the Protocol is written not conforming to the Japanese legislative rules, the notation of paragraphs differs from that of Japanese statutes[8]. In addition, the Protocol is not registered in the list of act titles. As a result, the underlined parts were replaced in the wrong way, as only articles were detected and were recognized as the ones of the current act. Updating the list of act titles and accepting an additional notation can solve this error.

---

[8] Paragraph numbers are denoted in Arabic numerals. While paragraphs in Japanese acts are cited in a text using Chinese numerals with the suffix for paragraph, ones in the Protocol are cited in Arabic numerals as they are.

Table 2. Experimental result for inaccessible citations included in the previous study

Evaluation	#Citations
Successfully replaced with an absolute address	27
Failed in annotation	13
Need anaphora resolution	2
Total	42

特定物質の規制等によるオゾン層の保護に関する法律（昭和六十三年法律第五十三号）/ Act on the Protection of the Ozone Layer Through the Control of Specified Substances and Other Measures (Act No. 53 of 1988)

第二条
5　この法律において「消費量」とは、議定書第一条６に規定する消費量の同条７に規定する算定値をいう。/

Article 2
(5) The term "the level of consumption" as used in this Act means the calculated level, prescribed in Article 1, paragraph 7 of the Protocol, of consumption prescribed in paragraph 6 of the said Article.

Fig. 6. Example of failure in annotation

6   Conclusion and Future Work

In this paper, we focused on Japanese statutory sentences. As long as boilerplate expressions are commonly used, a simple method for surface pattern recognition is sufficient for legal text processing. Based on these characteristics, we proposed the following methods:

- Extraction of legal terms and their definitions, the number of which exceeds 14,000 tokens;
- XML tagging in terms of the document structure; and
- Annotation of absolute addresses to citations.

We were faced with the problem that the extracted explanations include citations, some of which are not accessible apart from the original statutes. This is because they are expressed in context-dependent format. We replaced these ambiguous expressions with the absolute addressing method. The effectiveness of our method was demonstrated in our experimental results.

Our goal is to provide a terminology for translation. Although explanations including citations in context-independent format are accessible for a specified provision or act, their readability is insufficient. Explanatory sentences should be independently readable. For example, the term *population census* shown in Fig. 3 should be explained as "a complete census concerning individuals and

households," taking into consideration the expression in the preceding paragraph. One method of replacing citations with referential expressions has already been proposed [15] using a machine learning method, although it dealt with only the National Pension Act. In our future work, we will integrate this method for reference resolution with our XML corpus.

References

1. Toyama, K., Saito, D., Sekine, Y., Ogawa, Y., Kakuta, T., Kimura, T., Matsuura, Y.: Design and Development of Japanese Law Translation Database System. In: Proceedings of Law via the Internet, 12 p. (2011)
2. Sekine, Y., Toyama, K., Ogawa, Y., Matsuura, Y.: The Development of Translation Memory Database System for Law Translation. In: Proceedings of 2012 Law via the Internet Conference, 21 p. (2012)
3. Nakamura, M., Ogawa, Y., Toyama, K.: Extraction of Legal Definitions from a Japanese Statutory Corpus – Toward Construction of a Legal Term Ontology. In: Proceedings of 2013 Law via the Internet Conference, 11 p. (2013)
4. Palmirani, M., Brighi, R., Massini, M.: Automated Extraction of Normative References in Legal Texts. In: Proceedings of International Conference on Artificial Intelligence and Law (ICAIL), pp. 105–106 (2003)
5. Biolioli, A., Dini, L., Mercatali, P., Romano, F.: For the Automated Mark-Up of Italian Legislative Texts in XML. In: Legal Knowledge and Information Systems - JURIX 2002: The Fifteenth Annual Conference, pp. 21–30 (2002)
6. de Maat, E., Winkels, R., van Engers, T.: Automated Detection of Reference Structures in Law. In: Legal Knowledge and Information Systems - JURIX 2006: The Nineteenth Annual Conference, pp. 41–50 (2006)
7. Martínez-González, M., de la Fuente, P., Vicente, D.-J.: Reference extraction and resolution for legal texts. In: Pal, S.K., Bandyopadhyay, S., Biswas, S. (eds.) PReMI 2005. LNCS, vol. 3776, pp. 218–221. Springer, Heidelberg (2005)
8. Lame, G.: Using NLP techniques to identify legal ontology components: Concepts and relations. In: Benjamins, V.R., Casanovas, P., Breuker, J., Gangemi, A. (eds.) Law and the Semantic Web. LNCS (LNAI), vol. 3369, pp. 169–184. Springer, Heidelberg (2005)
9. Höfler, S., Bünzli, A., Sugisaki, K.: Detecting Legal Definitions for Automated Style Checking in Draft Laws. Technical Report CL-2011.01, University of Zurich, Institute of Computational Linguistics (2011)
10. Winkels, R., Hoekstra, R.: Automatic Extraction of Legal Concepts and Definitions. In: Legal Knowledge and Information Systems - JURIX 2012: The Twenty-Fifth Annual Conference, pp. 157–166 (2012)
11. Le, T.T.N., Nguyen, M.L., Shimazu, A.: Unsupervised Keyword Extraction for Japanese Legal Documents. In: Legal Knowledge and Information Systems - JURIX 2013: The Twenty-Sixth Annual Conference on Legal Knowledge and Information Systems, pp. 97–106 (2013)

12. Nakamura, M., Kobayashi, R., Ogawa, Y., Toyama, K.: A Pattern-based Approach to Hyponymy Relation Acquisition for the Agricultural Thesaurus. In: Proceedings of International Symposium on Agricultural Ontology Service 2012 (AOS 2012), pp. 2–9 (2012)
13. Nakamura, M., Ogawa, Y., Toyama, K.: Extraction of Legal Terms and Their Explanations Defined in Parenthesis in Statutes. In: Proceedings of 19th Annual Meeting on Natural Language Processing (NLP 2013), pp. 670–673 (2013) (in Japanese)
14. Ogawa, Y., Inagaki, S., Toyama, K.: Automatic consolidation of japanese statutes based on formalization of amendment sentences. In: Satoh, K., Inokuchi, A., Nagao, K., Kawamura, T. (eds.) JSAI 2007. LNCS (LNAI), vol. 4914, pp. 363–376. Springer, Heidelberg (2008)
15. Tran, O.T., Ngo, B.X., Nguyen, M.L., Shimazu, A.: Automated Reference Resolution in Legal Texts. Artificial Intelligence and Law, 1–32 (2013)