

法令文中において括弧書きで定義されている 法令用語とその語釈文の抽出

中村 誠¹ 小川 泰弘² 外山 勝彦³

名古屋大学大学院 {¹ 法学, ³ 情報科学} 研究科

² 名古屋大学情報基盤センター

mnakamur@law.nagoya-u.ac.jp

1 はじめに

2009年に政府の主導によって日本法令の外国語訳データベースがリリースされた [1, 2, 3]。2013年現在においても翻訳作業は継続しており、複数のプロの翻訳者による翻訳がウェブ上に順次公開されている¹。翻訳過程において、法令翻訳特有の問題がいくつか生じているが、とりわけ問題とされているのが訳語の不統一である。法令用語が不統一に翻訳されると、その解釈に影響を及ぼし、大きな問題となるが、これを解決するために標準対訳辞書を作成している。しかしながら、ここに登録されている語数が不十分であるため (バージョン 7.0 の時点で日本語見出し語数 4,482 語)、登録語数を増やすことが目下の課題となっている。また、翻訳を行う際、特に重要となる法令用語をあらかじめ抽出しておき、その英訳を用意しておくことにより、翻訳者の負担を減らすことが可能である。

本研究の目的は、法令文中の重要な法令用語とその語釈文を抽出することである。これらは、翻訳過程において訳語の統一に利用することができる。ここでいう重要な法令用語とは、法令文中の定義規定又は略称規定によって定義された語句をいう。定義規定とは、法令用語を定義するために書かれている事項である。それに対し、略称規定とは、法令中の用語の表現を簡潔かつ正確にするために、その法令だけに通じる簡略な呼称をつけるためのものである。定義規定は、法律の総則に置く場合と条文中に括弧書きで置く場合の二つの書き方がある [4]。総則に置く場合、第一条の目的に次いで第二条に定義規定が設けられる場合が多い。これらの条文から定義語句とその語釈文の抽出に関する研究成果がこれまでに報告されており [5]、構文情報タグ付き法律文コーパス [6, 7] を用いた実験により、

241 法令 109,380 文中、1,027 用語とその語釈文が得られている。

本研究においては、もう一方の定義規定、すなわち条文中に出現する括弧書きによる定義規定と略称規定を取り扱う。なお、構文情報タグ付き法律文コーパスに格納されている法令数は限られているため、この情報を利用した手法は実用的ではない。本研究では、コーパスの分析を行うことにより、構文情報を用いない方法を提案する。これにより、多くの法令用語とその語釈文の抽出が期待される。

本稿の構成は以下の通りである。2 節で法令文中の定義規定の特徴の記述と、コーパスを用いた分析を示す。これを基に 3 節では法令用語及びその語釈文を抽出するための手法を提案し、4 節においてその評価実験を行う。最後にまとめと今後の課題を述べる。

2 定義規定・略称規定の特徴

本節においては、定義規定と略称規定の特徴と、法令文コーパスを用いたその分析を行う。

2.1 条文中に現れる定義規定・略称規定

本研究で扱う定義規定及び略称規定の例をそれぞれ図 1、図 2 に示す。図中のアンダーラインで示した部分が定義規定又は略称規定である。

条文中に括弧書きで置く定義規定の場合には、次の二通りの表現方法がある。本稿では、これらをそれぞれ、トイウ形、ライウ形と呼ぶ。

トイウ形：A、B 及び C (以下「 」という。)

ライウ形： (A、B 及び C をいう。以下同じ。)

¹日本法令外国語訳データベースシステムのウェブサイト：
<http://www.japaneselawtranslation.go.jp/>

定義規定

資源の有効な利用の促進に関する法律

第四条 工場若しくは事業場において事業を行う者及び物品の販売の事業を行う者（以下「事業者」という。）又は建設工事の発注者は、その事業又はその建設工事の発注を行うに際して原材料等の使用の合理化を行うとともに、再生資源及び再生部品を利用するよう努めなければならない。

(注: 定義規定以外の括弧書きは省略)

図 1: 本研究で取り扱う定義規定

略称規定

中央省庁等改革関係法施行法

第千三百一条 中央省庁等改革関係法及びこの法律（以下「改革関係法等」と総称する。）の施行前に法令の規定により従前の国の機関がした免許、許可、認可、承認、指定その他の処分又は通知その他の行為は、法令に別段の定めがあるもののほか、改革関係法等の施行後は、改革関係法等の施行後の法令の相当規定に基づいて、相当の国の機関がした免許、許可、認可、承認、指定その他の処分又は通知その他の行為とみなす。

(注: 略称規定以外の括弧書きは省略)

図 2: 本研究で取り扱う略称規定

トイウ形の場合、「A、B及びC」という一連の用語をその直後にとりまとめて定義するものであるが、ライウ形の場合は先に定義する用語において、その後でその内容を説明するものである。このように、定義語句とその語釈文の位置関係は、互いに異なる。図 1 はトイウ形である。ライウ形の例を図 3 に示す。

なお、定義規定と似て非なるものとして、「略称規定」というものがある(図 2)。これは、法令中の用語の表現を簡潔かつ正確にするために、しばしば現れる長い表現や複雑な表現に対して、その法令だけに通じる簡略な呼称をつけるためのものである。例を以下に示す。

- …(以下「 」という。)
- …(以下単に「 」という。)
- …(以下「 」と総称する。)

ライウ形

保険業法 第二百七十二条の三十七

2 少額短期保険持株会社（少額短期保険業者を子会社とする持株会社であって、第二百七十二条の三十五第一項各号に掲げる取引若しくは行為について保有者となる承認を受け、同項の承認を受けて設立され、又は同条第三項ただし書の承認を受けているものをいう。以下同じ。）は、外国の法令に準拠して設立されたものを除き、株式会社であって次に掲げる機関を置くものでなければならない。

図 3: ライウ形で書かれた定義規定

本研究においては、定義規定と略称規定を区別する必要は無く、すべての略称規定をトイウ形として取り扱う。

2.2 コーパスによる分析

本節における法令文の分析には、構文情報タグ付き法令文コーパス [6, 7] を用いた。このコーパスは、241 法令 109,380 文からなり、Cabocha [8] による構文解析結果を基に、法令文特有の記述に対応した修正を施したものである。

括弧書きによる定義規定の出現頻度は、正規表現を用いて得ることができる。コーパス全体から求めた結果、トイウ形は 5,297 回、ライウ形は 2,124 回出現したことが確認された。また、コーパスの構文情報を利用することにより、定義規定の範囲を同定することができる。すなわち、定義規定の括弧の直前にある語句に直接もしくは間接的に掛かる語句がその範囲であると考えることができる。

図 1 における定義規定の例では、括弧の直前にある語句である「者」にその前にある語句がすべて直接もしくは間接的に掛かっているため、定義規定は文頭から開始することがわかる。これに対し、図 4 の例では、文頭の「外国人が」が、略称規定の括弧のあとにある「行う」に掛かっているため、略称規定に含まれない。

本研究において、括弧書きによる定義規定は、その性質上、文の先頭の主語である名詞句に現れやすいという仮説を置く。本節においては、コーパスの分析によってこれを検証する。定義規定が文頭から始まるかどうかの判別は、文頭の語句が直接もしくは間接的に括弧前の語句に掛かっていればよい。

略称規定

排他的経済水域における漁業等に関する主権的権利の行使等に関する法律

第三条 外国人が我が国の排他的経済水域（以下単に「排他的経済水域」という。）において行う漁業、水産動植物の採捕及び探査（以下この条において「排他的経済水域における外国人の漁業等」という。）に関しては、この法律の定めるところによる。

図 4: 文頭から開始しない略称規定の例

表 1: 文頭から括弧始まりまでの字数頻度分布

字数区分	トイウ形		ライウ形	
	頻度	累計	頻度	累計
1 - 10	133	133	270	270
11 - 20	383	516	75	345
21 - 30	243	759	52	397
31 - 40	134	893	8	405
41 - 50	104	997	4	409
51 - 100	80	1,077	14	423
100 -	7	1,084	3	426

表 1 が表しているのは、定義規定が文頭から始まり、読点（「、」）を用いずに定義規定の括弧に至ったときの文頭から括弧までの文字数の頻度分布及びその累計である。このように、文頭から定義規定が始まっているか判別するのに、おおよそ 100 文字が目安となっていることがわかる。

次に、図 4 に示すような、文頭から定義規定の括弧前まで 100 文字以内であり、更に読点が無いにもかかわらず、コーパスの構文情報により定義規定が文頭から始まっていない場合の出現頻度を求めた。その結果、トイウ形は 353 回出現したが、そのうち 212 個 (60.1%) が実際には定義規定が文頭から始まっているにもかかわらず、構文解析の誤りのためにこれが検出できていなかったものであることが確認された。これに対し、ライウ形は 140 回出現したが、構文誤りのために検出できていなかったものは 14 個 (10.0%) であった。

以上をまとめると、構文情報付き法律コーパスを用いた分析により、条文の文頭から始まる定義規定は抽出しやすいという結論を得た。対象となる条文は、文頭から 100 文字以内に定義規定の括弧があり、かつ、それまで読点がないものである。これに該当する箇所

が見つかったら、文頭から括弧の始まりまでと、括弧内の記述を処理することで、法令用語及びその語釈文を高い精度で抽出することができる。

3 提案手法

前節の分析から、プレインテキストコーパスから法令用語及びその語釈文を抽出する手法を示す。なお、条文中にある 2 つめ以降の定義規定においては、語釈文に照応解析が必要であるケースが多いため、今回は扱わない。また、定義規定が括弧文内にあるとき（すなわち二重ないしは三重以上の括弧となる。）も扱わないこととする。

法令用語及びその語釈文の抽出は、以下の手順に従って処理を行う。

1. 法令文を一文ずつ読み込み、トイウ形、もしくはライウ形に該当する条文を抜き出す。
2. 定義規定ではない括弧書きを除去する。これには定義規定が括弧文内にあるときも含む。
3. 文頭から数えて 100 文字以内に定義規定の括弧があり、その間に読点（「、」）がなければ、以下の処理を行う。

- トイウ形であれば、括弧内の「カギ括弧」が法令用語、文頭から括弧までが語釈文である。
- ライウ形であれば、文頭から括弧までが法令用語、括弧内の「をいう。」までが語釈文である。

この提案手法を用いると、図 1、図 2、図 3 は法令用語及びその語釈文を正しく抽出できる。それに対し、図 4 のように、文頭から定義規定又は略称規定の括弧まで読点がなく、そこにある語句が括弧の後ろの語句に掛かる場合は語釈文の抽出に失敗する。

4 実験結果と考察

実験には、前節の分析で用いた構文情報タグ付き法令文コーパス [6, 7] (241 法令 109,380 文) のテキスト部分を利用した。したがって、本実験は、クローズドテストとなる。

実験結果を表 2 に示す。Precision は、表 1 にある 100 文字目までに定義規定又は略称規定の括弧があるものの累計と、文頭から定義規定の括弧前まで 100 文字以内であり、更に読点が無いにもかかわらず、コー

表 2: 実験結果 (文頭からの定義規定・略称規定の抽出)

	トイウ形	ライウ形
Precision	$\frac{1,501}{1,652} = 90.9\%$	$\frac{440}{566} = 77.7\%$
Recall	$\frac{1,501}{5,297} = 28.3\%$	$\frac{440}{2,124} = 20.7\%$

パスの構文情報により定義規定が文頭から始まっていない場合の出現頻度から求めた。また、Recall は、トイウ形、ライウ形のパターンマッチで得られたそれぞれの出現頻度 5,297 回、2,124 回を分母にして得た。

このような単純な手法でも、Precision に関しては高い値が得られた。一方、Recall に関しては、値が低いが、これはそもそも文頭から 100 文字以内に出現するものが限られているためである。したがって、文中での定義規定と略称規定の抽出は、今後の課題である。しかしながら、従来手法が 1,027 語しか取れなかったのに対して、提案手法では、1,941 語得ることができた。

5 おわりに

本稿では、条文の文頭の主語に定義規定が現れやすいという傾向に着目して、法令用語及びその語釈文の抽出を行った。実験の結果、高い精度でこれらの抽出が行われた。この方法は、構文情報を全く必要としないため、法令文のプレーンテキストコーパスから抽出が可能である。さらに、今回の分析によってコーパス中の構文情報の修正が可能であることが示唆された。

今後の予定として、本稿では適用できなかった定義規定の抽出方法を検討する。また、本研究で得られた定義規定の分析を行う必要がある。抽出された法令用語は、それぞれの法令や条項において同じか似通った語釈文によって複数回定義されていることが多い。したがって、これらのパリエーションの分析は、翻訳作業に役立つと考えられる。

参考文献

- [1] 外山勝彦, 小川泰弘. 自然言語処理の応用に基づく法令外国語訳支援. 人工知能学会誌, 23(4):521–528, 2008.
- [2] Katsuhiko Toyama, Daichi Saito, Yasuhiro Sekine, Yasuhiro Ogawa, Tokuyasu Kakuta, Tarihiko Kimura, and Yoshiharu Matsuura. Design

and Development of Japanese Law Translation Database System. In *Proc of Law via the Internet*, 12 pages, 2011.

- [3] 外山勝彦, 齋藤大地, 関根康弘, 小川泰弘, 角田篤泰, 木村垂穂, 松浦好治. 日本法令外国語訳データベースシステムの設計と開発. 情報ネットワーク・ローレビュー, 11:33–53, 2012.
- [4] 山本庸幸. 実務立法技術. 商事法務, 東京, 2006.
- [5] Makoto Nakamura, Ryusei Kobayashi, Yasuhiro Ogawa, and Katsuhiko Toyama. A Pattern-Based Approach to Hyponymy Relation Acquisition for the Agricultural Thesaurus. In *Proceedings of AOS2012*, pages 2–9, 2012.
- [6] 山田将之, 小川泰弘, 外山勝彦. 構文情報付き法律文コーパスの設計と構築. In 言語処理学会第 14 回年次大会講演論文集, pages 604–607, 2008.
- [7] Yasuhiro Ogawa, Masayuki Yamada, and Katsuhiko Toyama. Design and Compilation of Syntactically Tagged Corpus of Japanese Statutory Sentences. In *New Frontiers in Artificial Intelligence*, LNAI6797, pages 141–152. Springer, 2011.
- [8] 工藤 拓, 松本 裕治. チャンキングの段階適用による日本語係り受け解析. 43(6):1834–1842, 2002.