

## 英文契約書における内容の抽出

— 一条文における権利・義務記載部分の内容を表す句構成セットの作成 —

相良 かおる      渡邊 勝正

奈良先端科学技術大学院大学 情報科学研究科  
〒 630-0101 奈良県生駒市高山町 8916-5  
TEL: 0743-72-5306      FAX: 0743-72-5309  
E-mail: {kaoru-s,watanabe}@is.aist-nara.ac.jp

本稿では、英文契約書の条文に含まれる権利・義務に関する記載内容の構造を表現するために、「句構成セット」を導入し、その定義と作成方法について述べている。句構成セットは、句または特定の語に対応する「句構成要素」の順序付けられた集まりである。句構成要素は、分類タグと句構成パターンにより構成される。句構成パターンによって、条文は句に分割される。すでに、34の条項の内、秘密保持条項に関する95の条文から、人手により、57の動詞を中心にした143種類の句構成セットを求めている。さらに、これらの句構成セットに含まれる421種類の句構成要素と語の条項内での出現頻度を基にして、662の句構成パターンを求めた。残りの33の条項について、できるだけ機械的に句構成セットを作成する方法を検討している。

キーワード： 英文契約書、内容抽出、抽出要素、句分割、句構成要素、分類タグ

## Information Extraction From an English Contract

A definition and a making method of phrase constituent sets that represent contents of the right and obligation in contract sentences

Kaoru SAGARA and Katsumasa WATANABE

Graduate School of Information Science,  
Nara Institute of Science and Technology  
8916-5 Takayama, Ikoma, Nara, 630-0101 JAPAN  
TEL: +81-743-72-5306      FAX: +81-743-72-5309  
E-mail: {kaoru-s,watanabe}@is.aist-nara.ac.jp

This paper describes a definition and a making method of phrase constituent sets that represent contents of the right and obligation included in a contract sentence. A phrase constituent set is an ordered set of phrase constituents, each of which consists of a phrase constituent pattern with a classification tag. A contract sentence is divided into some groups of words with a meaning role by the phrase constituent pattern. We made 143 kind of phrase constituent sets from 95 sentences in confidentiality articles manually. The confidentiality articles are one of 34 articles for our work. In addition, we made 662 kinds of phrase constituent patterns from the 421 phrase constituents included in the sets and article frequency of words in the constituents. We are examining a method to make phrase constituent sets as mechanically as possible from 33 articles on the remainder.

keywords : contracts in English, information extraction, extraction element, phrase segmentation, phrase constituent, classification tag

## 1 はじめに

近年、一部の大企業に限られていた国際取引契約の締結が、中小企業やベンチャービジネスの場においても行なわれるようになってきた。それに伴い、多くの英文契約文書（以下、契約書という）が作成され、保管されている。

一方、SGML(Standard Generalized Markup Language)、XML(eXtensible Markup Language)などの出現で、従来紙面で保管されてきた文書を構造化された電子化文書の形で保管することが可能となってきた。

その結果、作成された契約書から重要な情報を見つけ出し、抽出することができれば、SGMLおよびXMLを利用することで、契約書を単なる記録としてではなく、新たな取引契約の締結に向けての重要な情報源として、積極的に利用することができる。

そこで我々は、技術取引に関する英文契約書の内容を抽出するため、その抽出要素の作成について研究を行っている。

具体的には、技術取引契約に関する34種類の条項についての例文集(全2315条文)を対象に、条項単位の抽出項目を作成する。

現在、34種類の条項の内、秘密保持に関する95の条文(以下、95条文という)を対象に、人手により権利・義務の記載部分を表す143種類の抽出要素を求めている。

この作業に要した時間は、条文の分割、分類タグの決定およびタグ付けなど、全てを含めて約80時間であった。それらの大半は、辞書や文法書などで意味や語法を調べる、または分類タグを決定するなど、コンピュータで処理することのできない作業に費やされた。そこで、残り33条項2220条文から抽出要素を作成していく過程においては、既に求めてある抽出要素に含まれる情報を利用し、人手による作業の低減を図る必要がある。

そのひとつとして、求めた抽出要素と語の条項内での出現頻度データから、句を分割するための句構成要素パターンを作成し、句分割に掛かる作業量の低減を図る手法を検討中である。

本稿では、これらの抽出要素の作成方法について述べる。

## 2 契約書の構造

本手法は、条文の構文上の特徴を利用している。そこで、本章では、条文の特徴について述べる。

契約書は、条項ごとに内容がまとめられており、大まかな構造化がなされている。条項は、(1)契約の種類に関係なく、どの契約書にも通常含まれる一般条項(不可効力、仲裁、契約終了等)と、(2)契約の内容に依存する固有条項(技術援助、品質管理、技術的情報など)に大別される[3～7]。また、契約書に記載されている内容は、以下の4つに大別される。

1. 権利または義務の記述
2. 権利または義務が生じる事由
3. 権利または義務が生じた際の措置
4. 契約内容の詳細

一般に、権利または義務を表す記述は、以下に示す語句(helping verbs)によって、識別される。但し、“shall”、“may”、および“will”の直後に続く動詞が全て権利または義務を表すとは限らない。

1. shall
2. may
3. will
4. shall be deemed (to)
5. shall have (the/a) right to
6. be entitled to
7. reserve (the/a) right to
8. be permitted to
9. be obligated to
10. be liable to
11. be responsible (to/for)
12. be under the obligation to
13. have (the/an) obligation to

加えて、合意事項として当事者間の権利または義務を記述する場合があります、この場合は以下の語句によって識別することができる。

1. agree (to/that)
2. undertake (to/that)

尚、2315 条文の中で、前述の権利・義務を示唆する語句を含む条文は 1704 条文ある。

また、権利または義務が発生する条件についての記載は、以下の語句によって識別される。

1. if
2. unless
3. in the event (of/that)
4. in case (of)
5. subject to
6. when
7. upon
8. once
9. provided, however that
10. except (to/for/that)
11. except as provided (herein/hereinafter)
12. notwithstanding
13. but (not/that)
14. save (for/that)
15. in the absent of

したがって、前述の権利または義務を示唆する語句を含み、かつ条件を示唆する語句を含まない節が、権利または義務の記載となる。そこで、本研究では、前述の語句を指標として、権利・義務の記載部分を求め、抽出要素の作成を行う。

尚、前述の示唆語、権利・義務を表すもの 13 種、合意事項を表すもの 2 種、条件を表すもの 15 種の計 30 種類の語句は、英文契約書に関する文献 [3 ~ 16] から抜き出したものを基に、2315 条文中にその存在を確認し、また確認の際に気づいたものを追加し、まとめたものである。

次章では、抽出要素の基になる句構成セットについて述べる。

### 3 句構成セット

条文の内容の構造を句構成セットとして表現する。句構成セットは、句または特定の語に対応する句構成要素の順序付けられた集まりである。句構成セットには、必ず動詞句の機能を持つ句構成要素が 1 つ含まれる。句構成要素は、句の機能を表す分類タグと、句構成パターンからなる。分類タグは、句のタイプを表す識別子(句タイプ識別子)と、句の意味的な概要を表す代表語からなる。以下に、定義をまとめる。

句構成セット = { 句構成要素 }<sup>+</sup>  
句構成要素 = 分類タグ: 句構成パターン  
分類タグ = 句タイプ識別子 - 代表語  
句構成パターン = { タグ付き語 }<sup>+</sup>  
タグ付き語 = 文字列/品詞タグ |  
                  文字列/\* | \*/品詞タグ  
品詞タグ = Penn Treebank 品詞体系  
\* = 空白以外の文字列

以下に、“disclose” および “be disclosed” を含む句構成セットの一例を示す。動詞 “disclose” を含む句構成セットは 5 つの句構成要素からなり、動詞 “be disclosed” を含むものは、6 つの句構成要素からなっている。

< disclose >

1. **np-person**: Licensee/NN
2. **shall-not**: shall/MD not/RB
3. **vp-disclose**: disclose/VB
4. **np-incorporeal+**: All/DT and/CC any/DT of/IN data/NNS ,/, drawings/NNS ,/, specification/NN and/CC other/JJ information/NN
5. **to-person**: to/TO any/DT third/JJ party/NN or/CC parties/NNS

< be disclosed >

1. **np-incorporeal+**: All/DT information/NN ,/, the/DT data/NNS ,/, specification/NN ,/, drawing/VBG or/CC any/DT of/IN them/PRP
2. **shall-not**: shall/MD not/RB
3. **vp-disclosed**: be\_disclosed/VB
4. **to-person**: to/TO any/DT other/JJ person/NN ,/, firm/NN or/CC corporation/NN
5. **by-person**: by/IN Licensee/NN
6. **without-consent**: without/IN the/DT prior/RB written/JJ consent/NN of/IN Company/NNP

これらは、前述の 95 例文から、人手で求めた権利・義務および合意事項についての 57 種類の動詞を含む句構成セット 143 個のうちの 2 例である。

本研究では、34 種類の条項(全 2315 条文)からこれらの句構成セットを求め、体系化したものを抽出要素とする。

## 4 句構成要素

句構成要素は、句の機能を表す分類タグと、句構成パターンからなる。句を分割する際、名詞句、形容詞句、副詞句、動詞句の単位に分割を行い、動詞句に関しては、更に第2章記載の権利・義務を示唆する語(通常は助動詞)と動詞に分割する。従って、動詞句に付ける分類タグは、他の句につけるものとは異なっている。

第5章では分類タグについて述べる。

## 5 分類タグ

分類タグは、句タイプ識別子と代表語をハイフン(-)で繋げたものである。句タイプ識別子には、以下のものを用いる。

### 5.1 句タイプ識別子

**np** : 名詞または名詞句

**vp** : 動詞または動詞句

**adj** : 形容詞

**adv** : 副詞

**前置詞、形容詞、副詞、名詞** : 前置詞または複合前置詞から始まる句

複合前置詞とは、前置詞と他の語とが組み合わさってできた前置詞をいう。複合前置詞については、5.3.3節で述べる。

**助動詞等** : 第2章記載の権利・義務を示唆する語句については、その語句を指定する。

### 5.2 代表語

以下に、秘密保持条項に関する句構成セットに用いた代表語の一覧を示す。今後他の33条項について句構成セットを求める毎に、代表語は随時追加更新される。

代表語は原則として名詞の語義を指定する。迷いが生じた場合は、最初にある名詞を代表語とする。尚、2315条文全てについて句構成セットが求めた時点で、代表語の見直しを行い、包含関係を明らかにする予定である。

人または法人 : person

物 : object

無体物 : incorporeal

有体物 : material

所有物 : possession

製造物 : product

規定物 : specify

条項 : article

権利 : right

義務 : obligation

状態 : state

目標 : purpose

理由 : reason

原因 : cause

行為 : action

同意 : consent

手法・手段 : method

場所 : place

時間 : time

数量・量 : quota

範囲 : scope

金銭 : mony

### 5.3 分類タグ付け規則

以下に、分類タグ付けの規則について述べる。

#### 5.3.1 一般規則

1. 同格の語が等位に列挙されている場合は、分類タグの末尾に+を指定
2. 句の後ろに修飾部が付加されている場合は、分類タグの末尾に\*を指定
3. “not”や“no”などの否定を表す語については、現れた句の分類タグに代表語として付加

#### 5.3.2 句が動詞または動詞句の場合

能動態の場合 : 代表語として動詞を指定

受動態の場合 : 代表語としてbe動詞を省略して動詞の過去分詞を指定

使役動詞の場合 : 句タイプの識別子として“causative”を指定

複数の動詞の等位接続 : 最初に記述されている動詞を代表語とし、分類タグの末尾に+を指定

### 5.3.3 前置詞から始まる句の場合

句タイプ識別子：通常は、前置詞を指定する。複合前置詞から始まる句で、同義の単純前置詞がない場合は、語義を表す形容詞、副詞、名詞を指定する。付録の句構成セット番号1の9の“pursuant\_to”は、単純前置詞“to”が形容詞“pursuant”に続いて複合前置詞を構成している例で、句タイプ識別子に形容詞“pursuant”を指定する。なお、複合前置詞については、複合語として前処理の段階で語の連結を行う。従って、新たに見つかった複合前置詞については、逐次、前処理用のデータに追加する。尚、文献[17～19]から抜き出した前置詞および複合前置詞の一覧から2315条文に含まれる115種の前置詞および複合前置詞を求めている。

以下に複合前置詞の例を示す。太字の見出しは句タイプ識別子である。

- about** : ~に関して  
with respect to  
in respect of  
in connection with  
with regard to  
concerning
- instead** : ~代わりに  
instead of  
in place of  
in lieu of
- at\_cost** : ~の費用負担で  
at the expense of  
at one's own expense of  
at the cost of  
at one's burden of cost  
for account of  
at the risk of and for the account of  
at one's own risk
- before** : ~より前に  
prior to

## 6 句構成セットの作成

今回、以下の手順で、95条文から57種の動詞からなる句構成セット143組を作成した。

1. サンプルとなる条文を、前述の語句および前置詞または複合前置詞の直前で分割する

2. 分割した語群から人手により、意味的にまとまりのある語群(句という)を求める
3. 品詞タグの妥当性のチェックを行う
4. 分類タグの決定および付加を行う
5. 権利・義務に関する句を抽出する
6. 省略部分の補完を行う
7. 動詞をキーとした句構成セットの分類を行う

### 6.1 句構成セットの作成規則

以下に句構成セットの作成規則を示す。

文法上の直前の句で補完 省略部分を補完する場合、文法上直前にある同じ機能を持つ句によって補完する。

例) 付録の句構成セットの番号2の1および番号3の1は、句の機能としては主語であるが、原文においては省略されている。そこで、前の句構成セット1の主語の部分で補完を行う。

解釈の曖昧な副詞句は除外 「直ちに」、「速やかに」など、解釈の曖昧なものは、機能をもつ句からは除外する。

暫定規則 例外事項および但し書きについては、存在の有無のみを記述する。  
すなわち現時点では、権利・義務に関する記載部分のみを扱う。

### 6.2 結果

表1は、上記の手順で作成した句構成セットの内、出現頻度の高い動詞の上位5個を示したものである。尚、ここでの重要度とはTF.IDFを応用した値であり、詳細は文献[20]を参照のこと。表2は、23回出現する動詞“disclose”の内、能動態の句構成セット17個中の分類タグをまとめたものである。

表3は、ある動詞と組を成す他の分類タグをまとめたものである。番号1は、動詞“take”と“use”が同じ機能を持つ句と組を成すことを表す。

今回求めた143セットの句構成セット中、動詞句以外の分類タグが同じである句構成セットは、132であった。

表 1: 権利・義務の記載に関する動詞 (上位 5 種)

動詞	# 1	重要度	# 2
disclose	23	6.81	50/97
use	17	1.93	35/640
take	10	2.14	13/122
keep	9	3.59	11/74
hold	9	3.46	11/56

# 1: この動詞を含む句構成セットの数  
 # 2: 秘密保持条項 95 条文における出現頻度 / 全書式集 4519 条文中の出現頻度

表 2: 動詞 “disclose” と組を成す句構成要素

修飾部, 主語 分類タグ	動詞	目的格, 補語, 修飾部 分類タグ
during-state*	vp-disclose	adj-any in-scope* in-method
np-person		np-incorporeal
may shall-not agree-to will		during-time while-time*
		to-person np-person
		about-method without-consent if-case*

表 3: 動詞と組を成す句構成要素

#	分類タグ	頻度
1	np-person shall	3
	vp (take & use)	
	np-action	
	to-purpose	
2	np-person shall	3
	vp (cause & have & require)	
	np-person	
	to-action	
3	np-person shall	2
	vp (hold & treat)	
	np-incorporeal	
	as-state	

表 4: 不適切な品詞タグの一例

不適切	適切
keep/NN in/IN	keep/VB in/IN
to/TO the/DT public/JJ	to/TO the/DT public/NN
may/MD sublicense/NN	may/MD sublicense/VB
resorts/NNS to/TO	resorts/VBZ to/TO
filing/NN patent/NN	filing/VBG patent/NN
turn/NN over/IN	turn/VB over/IN

今後、全 34 条項の句構成セットを求めた時点で動詞句による分類を行う予定である。

前述のように、これらの句構成セットを求めためには多くの時間を要する。そこで、既に求めた句構成要素から、句分割用の構成パターンを求めて、句分割を行い、仮の分類タグを付加した句構成要素を生成することを検討している。次章では、句分割用の構成パターンの作成について述べる。

## 7 句分割用の句構成パターンの作成

まず、句構成パターンを作成する前に、不適切な品詞タグの修正を行なう。(表 4 参照)。これらの不適切な品詞パターンを品詞タグ修正パターン一覧に追加すると共に、句構成要素および条文の品詞タグの修正を行なう。次に、以下の規則により 2 種類の句構成パターンを作成する。

**rule 1.** 修正後の句構成要素から条項頻度 20 以上の語については品詞タグを、他の語については語の文字列を “\*” で置換する。

**rule 2.** 修正後の句構成要素から、秘密保持条項のみに出現する語 (表 5) の部分を、その他の語については品詞の部分 “\*” に置換する。

これらの規則により作成された句構成パターンと対応する分類タグの関係は、規則 2 により作成されたものの方が規則 1 によるものより、適合性が高くなる。すなわち、規則 2 の句構成パターンに一致した句の分類タグは、句構成パターンに対応する分類タグと一致する確率が高いのに対して、規則 1 の句構成パターンに一致した句のそれは低くなる。そこで、これらの句構成パターンを識別するために分類タグの末尾に番号を付加する。

表 5: 秘密保持条項にのみ出現する語

語	頻度	条項頻度
afterwards	1	1
anticipation	1	1
disseminate	1	1
duplication	1	1
fundamental	1	1
innovation	1	1
leak	1	1
leakage	3	1
moreover	1	1
necessity	2	1
precaution	5	1
recipient	2	1
reservation	1	1
resort	1	1
servant	1	1
subassemble	1	1
submanufacture	1	1
ten	1	1
unjustify	1	1
unpublished	2	1

以上の手順から、143 種類の句構成セットに含まれる 421 種類の句構成要素から、662 種類の句構成パターンを作成している。

これらの句構成パターンを対応する分類タグで分類するとともに、表 3 にある句構成セットの分類タグを基に求めた各句間の関連情報を用いることで、条文の分割を行う。

以下の 2 つの句構成要素からは、3 種類のパターンが生成される。

<例>

語の出現する条項の数

all : 34

reasonable : 21

precaution : 1

care : 2

A: np-action: all/DT reasonable/JJ precautions/NNS

B: np-action : all/DT reasonable/JJ care/NN

(1): np-action1 : all/\* reasonable/\* \*/NN

(2): np-action2 : all/\* reasonable/\* \*/NNS

(3): np-action2 : all/\* reasonable/\* care/\*

A の句構成パターンからは、規則 1 から規則 2 から同じパターン (2) が生成される。

## 8 まとめと今後の課題

本稿では、人手により求めた秘密保持条項における権利・義務内容についての抽出要素の基となる句構成セットの定義とその作成方法について述べた。更に、人手で求めた句構成要素を用いて、条文を句に分割するための句構成パターンの作成方法について提案した。

当面の課題として以下の検討項目がある。

1. 句構成パターンを用いた効率のよい句分割手法についての検討と実装
2. 分類タグ規則の評価
3. 全 34 条項の句構成セットの作成
4. 分類タグの見直しと構造化
5. 句構成セットのデータ構造の検討
6. 条文中の権利・義務以外の内容抽出についての検討
7. 抽出項目の作成

### 謝辞

日頃から御討論いただき、本学、自然言語処理学講座の松本裕治教授と、木村晋二助教授、高木一義助手はじめ渡邊研究室の皆様へ感謝します。本研究は、一部文部省科学研究費特定領域研究課題番号 10139226 の援助による。

### 参考文献

- [1] 英和对訳 取引条件表現法辞典 第 2 巻技術取引, 国際事業開発株式会社, 1992.
- [2] Eric Brill: Some Advances in Transformation-Based Part of Speech Tagging, (AAAI-94). <http://www.cs.jhu.edu/~brill/acadpubs.html>
- [3] 日野修男, 出澤秀二, 竹原隆信, 杉浦幸彦, 水谷孝三: 英文契約書の知識と実務, 日本実業出版社, 1997.
- [4] 岩崎一生: 英文契約書 - 作成の理論と実務 -, 同文館, 1988.
- [5] 中村秀雄: 新版 英文契約書作成のキーポイント, 社団法人 商事法務研究会, 1996.
- [6] 宮野準治, 飯泉恵美子: 英文契約書の基礎知識, The Japan Times, 1997.
- [7] 山本孝夫: 英文契約書の書き方, 日経文庫 (473 G23), 1995
- [8] 阿部佳基, 長谷川俊明: ビジネス法律英語辞典, 日本経済新聞社, 1991.

- [9] 長谷川俊明：LAWDAS 法律英語辞典，東京布井出版，1996
- [10] 長谷川俊明：法律英語のカギ - 契約・文書・術語 - 法律英語・パート1，東京布井出版，1985
- [11] 長谷川俊明，法律英語のプロ - 契約と裁判，東京布井出版，1992
- [12] 早川武夫：法律英語の基礎知識，財団法人 商事法務研究会，1992
- [13] 長谷川俊明：ビジネス法律英語入門，日本経済新聞社，1994
- [14] 尾崎哲夫：ビジネスマンの基礎英語，日本経済新聞社，1994
- [15] 羽田三郎：ビジネスマンの英文法，研究社，1988
- [16] 長野格，秋山武清，豊田暁：ウェブスター ビジネス英語事典，研究社出版，1993
- [17] S. グリーンバウム，R. クワーク著，池上嘉彦 他訳：現代英語文法 大学編，紀伊国屋書店，1995.
- [18] マイケル・スワン著，金子稔，渡瀬和清，山田泰司訳：オックスフォード 実例 現代英語用法辞典 Practical English Usage，桐原書店/オックスフォード
- [19] 山口俊治：コンプリート 高校総合英語，桐原書店
- [20] 相良かおる，渡邊勝正：英文契約書における要目の抽出，電子情報通信学会技術研究報告，NLC98-19，29-36，1998.

付録 (条文と句構成セット)

<日本語訳 文献 [1] >

各当事者は、本契約に従い相手方より開示されたすべての図面及び情報を秘守するものとし、それをいかなる個人にも開示しないものとし、並びにその役員及び使用人が本条の規定を遵守し、従う事を保証するために、その取り得るあらゆる措置を講じるものとする。

<条文 52 語 文献 [1] >

Each party shall *keep* confidential all drawings and information disclosed to it by the other pursuant to this Agreement and shall not *disclose* the same to any other person and shall *take* all steps open to it to ensure that its officers and servants observe and comply with the provisions of this Article.

<前処理済条文>

Each/DT party/NN shall/MD keep/VB confidential/JJ all/DT drawings/NNS and/CC information/NN disclosed/VBN to/TO it/PRP

by/IN the/DT other/JJ pursuant\_to/IN this/DT Agreement/NNP and/CC shall/MD not/RB disclose/VB the/DT same/JJ to/TO any/DT other/JJ person/NN and/CC shall/MD take/VB all/DT steps/NNS open\_to/IN it/PRP to/TO ensure/VB that/IN its/PRP\$ officers/NNS and/CC servants/NNS observe/VBP and/CC comply/VB with/IN the/DT provisions/NNS of/IN this/DT Article/NN ./.

<句構成セット>

1.

1. np-person : Each/DT party/NN
2. shall : shall/MD
3. vp-keep : keep/VB
4. adj-state : confidential/JJ
5. np-incorporeal+\* : all/DT drawings/NNS and/CC information/NN
6. vp-disclosed : disclosed/VBN
7. to-person : to/TO it/PRP
8. by-person : by/IN the/DT other/JJ
9. pursuant-article : pursuant\_to/IN this/DT Agreement/NNP

2.

1. np-person : Each/DT party/NN ... (1-1)
2. shall-not : shall/MD not/RB
3. vp-disclose : disclose/VB
4. np-person : the/DT same/JJ to/TO any/DT other/JJ person/NN

3.

1. np-person : Each/DT party/NN ... (1-1)
2. shall : shall/MD
3. vp-take : take/VB
4. np-action\* : all/DT steps/NNS
5. to-purpose\* : to/TO ensure/VB