

Title	法情報の知識表現 : 法知識データベースとデータモデル
Author(s)	田中, 規久雄
Citation	阪大法学. 2002, 52(3,4), p. 259-284
Version Type	VoR
URL	https://doi.org/10.18910/55325
rights	
Note	

Osaka University Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

Osaka University

法情報の知識表現

——法知識データベースとデータモデル——

田 中 規久雄

はじめに

今日、ハードウェアの普及、ソフトウェア技術の成熟、データメディアの大容量化、ネットワークの拡大といった情報基盤の発展を基礎に、すでに多量の電子情報が生データとして、あるいはコンテンツ化されて蓄積されている。こうした事情を背景に、情報学 (informatics) の分野ではすでに蓄積された電子情報を共有、構造化し、そこから知識を発見し、さらには新たな知識を創造するためのアプローチが課題となっている⁽¹⁾。

法情報分野においても、すでに条文や判例は、商用のものも含めるなら、主要なものほとんど電子化されているといえるであろう。審議録なども徐々に電子化されつつある。しかし、他の分野もそれほど進んでいるとはいえないのではあろうが、法情報においてもその電子化形態は伝統的な紙メディアを電子媒体に置き換えることが中心であった。LEXIS や Westlaw、わが国の商用判例データベースに代表される法情報データベースもあくまでデータにはほとんど構造をもたせない「フルテキスト・データベース」として発展してきた⁽²⁾。

もちろん、それだけでもリテラルな検索に関してはその精度とスピードが格段に向上し、法律の実務、教育、研究にすばらしい成果をあげてきた。⁽³⁾しかし、法情報の存在パラダイムは紙媒体の時代からシフトしたとはいえないだろう。あくまで、高度な専門的知識を有した者が判断材料としてみる原資料という位置づけである。

しかし、最初に述べたような情報学の今日的課題に即してみれば、法情報を単なるデータとしてではなく知識として情報化し、人間の側の知識負荷を低減させ、その生産性をさらに向上させて法知識の創造に寄与することが求められる。⁽⁴⁾それはある意味では、単なるリストイテメントではない、体系書や研究書がめざしたものであり、そうした学問の営みとの連続性を有する事柄でもある。それを法情報学の課題としていいかえれば、法情報データベース (legal information-database) から法知識データベース (legal knowledge-database) への展開ということになる。そしてその具体化のためには法の情報構造、ライブラリ化、統合化などを考案する必要がある。

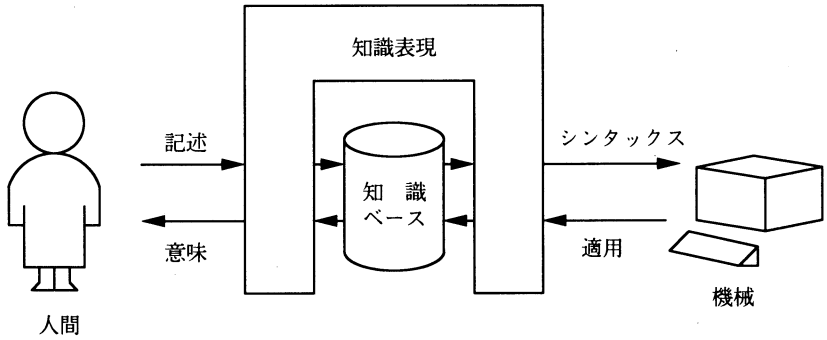
そこで本稿ではそうした課題の根底となる、法情報の「知識表現 (knowledge representation)」について、その中心概念のひとつである「データモデル」を中心に考察してみたい。

一 法知識ベースと計算機表現

「知識ベース (knowledgebase)」という言葉は歴史的に見れば、データベース (以下DBとする場合がある) をさす言葉ではなかった。元来は人工知能 (artificial intelligence) 研究、さらにはそれを中心とした知識工学、またそのひとつの応用である「エキスパートシステム (expert system)」構築の部分として想定された。

次の概念図(図一)はその典型的なものである。⁽⁵⁾

法律人工知能研究も盛んになされ格段に進化した。⁽⁶⁾そこで本稿では、その人工知能研究が開拓した知識ベースの



図一 知識表現の役割〔小山照夫, 1985〕

発想を法律データベースに導入する際の基本的な視座について考察することにする。

現代の主流計算機アーキテクチャにおいては、ソフトウェアは通常、何らかの「データ」に対して「計算」を行うことが前提とされる。これは、ソフトウェア工学においては「データ構造」と「アルゴリズム」として問題化され、知識工学においては、「知識」と「推論」という形で扱われる。前者はスタティックな、後者はダイナミックな印象を与えるが、スタティックなデータをどう扱うのかということも、知識としてスタティックに記述される点が特徴的である。

人工知能研究では当然この二つが融合することが望ましい（人間は融合している）ので、かつては、Prolog⁽⁸⁾やOPS⁽⁹⁾といった健全な推論が保証される言語にあわせた知識表現することが主流であった。⁽¹⁰⁾

たとえば民法七〇九条の不法行為を表現すると、

法律効果(損害を賠償する責めに任ず(X, Y)) : —
 ((故意(X) ; 過失(X)), 権利侵害(Y), 損害発生(Y), 因果関係(X, Y))⁽¹¹⁾

などとなる。これで直接DB記述ができるわけではないが、設計の前提にはなる。ただ、推論の健全性そのものはDBに記述する必要はな

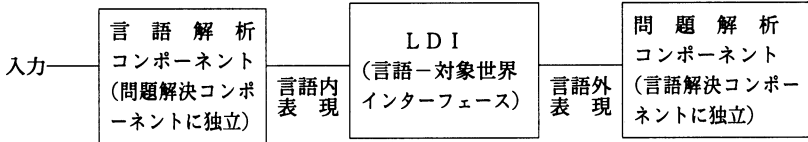
い。そもそもデータベースにおいてはアルゴリズムに対する「データの独立性」と「非手続き性」の保証が使いやすさに関わってくるのであるが、法知識データベースにおいても、人間の行う推論は数理的には不健全 (unsound) なことも多く、推論機構 (inference engine) をデータモデルに持ち込むことなく、まず推論は人間が行えることを保証するシステムの方が有用であると思われるからである。「法的思考の本領は、……事実の法的分析・構成と法的基準の意味内容の確定・形成が相互的に行われる過程が決定的に重要なのである。」⁽¹²⁾といわれる法的思考のためのデータベースは、極論すれば人間の不健全な推論をも許容するだけのシステムでなければならぬと思われる。⁽¹³⁾また、推論エンジンを組み込むエキスパートシステムの場合は、たとえば条文から複数の結論が導出されるのは問題であるが、⁽¹⁴⁾データベースでは、推論は人間が行うことを前提とすればよいのでシステム負担も軽くなる。

さて、西田豊明は図二のような言語理解のモデルを立て、入力される言語表現が問題解決コンポーネントの能力を超える場合と、その逆の場合を考察している。⁽¹⁵⁾簡単にいえば、前者がデータベースの場合、後者がエキスパートシステムの場合となるが、本稿の射程からして、前者の場合についてみる。

データベースでは、「データベース管理システム (DBMS) は限られた情報への効率的ではあるが限定された範囲内のアクセスを提供するに過ぎないが、自然言語インターフェイスをとおしてDBMSをみたとき、ユーザは提供された機能以上のことを期待しがちであり、」⁽¹⁶⁾それゆえデータベースといえども、「求める情報がデータベース中に暗黙的に存在している場合は、LDIにおいてそれをつなぎ合わせて、求める情報が得られるようにする必要⁽¹⁷⁾がある。」との認識の下、西田は、後述のリレーショナルDBMSが以下の諸点について処理可能であることを示す。

(一) データベースに暗黙的に含まれている情報を利用する。

法情報の知識表現



図二 自然言語理解システムの構成 (3 コンポーネントモデル) (西田豊明, 1989)

- (一) 部分的な回答を生成する。
 - (二) 内包的な質問に外延的情報から推定する。
 - (三) 前提についての推論。
 - (四) ただし、人工知能のアプローチでは、後述のリレーションアルスキーマやフレームといった知識表現をグラフ表現として把握し、「グラフによる階層構造がいろいろな意味を持っているのは、グラフ表現の表現力が高いことを示しているのではなくて、いわば、グラフ表現に確立した意味論がないという欠点なのである。」⁽¹⁸⁾とされ、論理の重要性が指摘されているが、エキスパートシステムの知識ベースではなく、あくまで意味論的には不健全な人間の推論判断を前提とする法知識データベースにおいてはそのことは欠点ではなく長所である。マルチパラダイムな人間存在には確立した意味論などないからである。そこで、知識ベースは推論エンジンに規定されることなく可能な限り、西田の指摘したレベルの法知識表現を行えば良いのではないだろうか。
- さて、この知識は、今日の計算機で扱いやすいように「宣言的知識 (declarative knowledge)」と「手続きの知識 (procedural knowledge)」に分類されている。
- 宣言的知識は is-a (a-kind-of) とした継承 (inheritance) や has-a (a-part-of) といった関連 (association) のような概念定義 (表記は UML; unified modeling language に準拠) 事実表現、ならびに when のようなルール記述を、手続きの知識は、ある目的にいたるにはどうすればよいか、といった表現記述を行う技法である。それゆえ、前者は what を

記述し、後者は「*raw*」を記述するといわれる。しかしこれらには互換性があり、どちらか一方だけで知識記述が可能だとされている。たとえば、「脇見運転という事実があれば、過失と評価する」は前者で、「過失と評価するために、とりあえず脇見運転の事実を探してみよ」というのは後者になる。

この二つの表現方法は計算機リソースやプログラミング効率に応じて書き分けられる。「データ・ベースは宣言型の事実としての知識⁽¹⁹⁾」といわれるように、データベース利用を用途とする法知識ベースは、宣言的知識記述を行い、人間にしろ、計算機プログラムにしろ、なにか別の推論機構が推論するためのインデックスを与えておくというのが堅実なように思える。

ただこの際留意せねばならないのは、宣言的記述を行うにしても、*Prolog*形式などのルールの記述は、エキスパートシステムにはともかく、データベースにはあまり適用的ではないかと思われることである。

たとえば、民法七〇九条の判決事例を述語的に表現した場合、知識処理を前提とした発想では、⁽²⁰⁾

賠償責任A → 事件A (故意A または 過失A, 権利侵害A, 発生損害A)

と手続的な表記を行い (A は具体的事実記述であることを示す)、これにメタデータとして識別子 (identifier) をつけることが考えられるが、これは事実表現としては推論形式をとってしまったために、データベースとしての表現の幅が狭まっている。

そこで、法知識データベースの知識記述としては、あくまで

事件A (故意A または 過失A, 権利侵害A, 発生損害A, 賠償責任A)

といった事実表現 (Prolog でいうファクトにあたる) が考えられるべきである。⁽²¹⁾

二 古典的データモデル

知識表現、すなわち、あるデータを知識として計算機が扱える形で情報化するためには、まず知識を記号化した何らかのデータモデルを想定しなければならないわけであるが、これはいずれにしろ人間の認識できる形式で行われざるを得ない。そのゆえこうしたデータモデルは、カントの先験的形式(シェーマ)論に発し、二〇世紀中葉の心理学が発達させた概念に由来して、「スキーマ(schema)」と呼ばれる²²⁾。

データの管理という広い意味でのデータベースは、ファイルによるデータの管理、ハイラーキーデータモデルといった原始的なモデルを経て、今日、代表的なデータモデルとされる、一九七〇年の「リレーショナルスキーマ(relational schema)」、一九七五年の「フレーム表現(frame representation)」、一九八四年の「オブジェクト指向スキーマ(object oriented schema)に至っている²³⁾。ただしこれらは相互排斥するものではなく、情報化する対象に応じて使い分けられている²⁴⁾。

以下、代表的な法情報、すなわち法源(legal authority)として第一的なものである、条文(legal provision)と判決文(legal sentence)を日本法に限定して考察する²⁵⁾。

二一 リレーショナル型法情報モデル

リレーショナルスキーマは、原則として一つの実体が、属性(attribute)の組という関係によって形成される一つの構造体(structure type)として表現される。これに基づいた法知識データベースの設計はどのようになされるべきであろうか。

従来の法律データベース的発想では、たとえば判決のモデル化に際して、この属性には、「事件番号」、「原告」、「被告」、「裁判所」、「適用条文」といった、形式的なデータを属性値として用いてきた。条文にいたっては、「条」、「項」、「号」などが基本的であったといえよう。さらに最大の問題は、条文と判例とが、せいぜいリンクされている程度で、法知識として一体化された形での情報化がなされてこなかったことである。

しかし、法知識データベースではそれを可能にしなければならない。具体例で考えてみよう。たとえば、民法七〇九条の不法行為は、「故意または過失」、「権利侵害」、「損害発生」、「賠償責任」などといった属性の組によって、一つの事件（判決）を表現できるのである（表一参照、本来は全文記載すべきであるが省略してある）⁽²⁶⁾。

すなわち、古典的な個々の「法律要件∥法律効果」という各属性とそれに対応する属性値（事実認定）が、一つの事件（の一部）を法知識表現している、ととらえなおすことができる。

リレーショナルであることから、複数の条文が適用されている事件にも別の条文のテーブルがあれば、たとえば事件番号を個体事件の識別子として、それを軸に統合することができる。個々の属性値についてその根拠を分離し、さらに別のテーブル（たとえば「過失相殺」テーブル）を作成しておくことも知識化の仕事となるだろう。

二―二 フレーム型法情報モデル

リレーショナルモデルは、たとえば商品管理や人事管理のように、それが適的な場面を多く有し現在でも幅広く用いられているが、不定形な知識を扱う場合にはモデル化が難しい。たとえば家計簿をつけるだけでも、直交した属性定義を行うのは難しいものであり、どうしても「その他」といった意味論的に不確定な項目を作らざるをえないという経験をした人は多いだろう。こうした問題を解決するものとして、一九七五年、ミンスキー (M. Minsky)

表一 民法七〇九条のリレージョナル表現モデル

<p>事件番号 大阪地裁平三(ワ)一七五号</p>	<p>故意または過失 前方左側にいた幼児らに気をとられ、前方および左右の注視を怠った過失により、進路前方に急に進出してきた原告に気づかず、本件事故を引き起こした</p>	<p>権利侵害 顔面外傷、全身打撲、口腔内外傷……の傷害を負った。後遺障害として、顔面の外貌に著しい醜状を残すものとして一二級一三号の認定を受けた。</p>	<p>損害発生 各傷害の治療のため、昭和六年五月一七日山口病院を経て、同日から同年六月二〇日までの三五日間と……一四日間、……実通院した。</p>	<p>賠償責任 「将来分を除く治療費二七・一万九三二五円」と「損害合計一四九八万三〇一九円」。右認定の損害額から、前記認定の既払い金五一五万四四二九円を控除すると、九八二万八五九〇円となる。</p>
<p>• • •</p>	<p>被告……は新規に開設された支店の業績工場のため、利回り保証約束をして有価証券取引を勧誘することが法により禁止されていることを知りながら、かつ年利一％での運用をすることが実現できない可能性が十分にあることを承知しながら、「年利一％で運用」する旨書面に記載して約束することにより、積極的に原告の判断を誤らせる勧誘をなし……</p>	<p>原告に本件金員および本件株式を交付させた</p>	<p>そして原告の損害については、原告が沖田支店長らに対し平成二年一月三〇日に一億一〇八万三二七四円を交付したときに同額の損害を被ったものであり、また、同年一二月二〇日に本件株式を交付したときに三三九四万円の損害を被ったものというべきである。</p>	<p>第三次的請求原因である不法行為による損害賠償請求権に基づく請求は金一億〇一五・一万六二二円および……民法所定年五分の割合による遅延損害金の支払いを求める限度で理由があるからこれを認容し……。</p>

がフレーム (frame) という知識構造概念を提唱し、一躍有力となった。

フレームとは、トップレベルが常に真であるような事柄で、その下位にはデータ (実例)、スロット (サブフレームを内包する) があり、フレームの集合がフレームシステムを形成し、フレームの取替は視点の転換を示す。値やスロットにはデフォルト (暗黙値あるいは典型値) が与えられている。フレームでは、「典型的知識」、「階層的知識」、「手続きの知識」、「推論の制御 (付加的手続き: Attached Procedure)」が可能である。⁽²⁷⁾ どれも宣言的に記述できるが、先に述べたように、後二者の記述方法はデータベースでは避けるべきであるように思われる。

フレームの考え方は、スロットの属性やスロット相互の関係が明白な、ある概念をフレームに割りあて、それぞれの具体的な属性値をトップダウンに埋めていき、一つの実態や事態を理解しようとする枠組みであるが、このフレーム理論に着目したわが国の法律学者に山本敬三がいる。⁽²⁸⁾

山本があげる例⁽²⁹⁾をフレーム表現すると、たとえば以下の様になる。

[賃貸借フレーム]

主体スロット ← 賃借人A

客体スロット ← 賃借物たる賃貸マンションの一室

内容スロット ← 当該マンションの使用収益からなる賃借権

(← 本来はここに、[賃貸料支払いフレーム]が入るべきであろう。)

[売買フレーム]

売主スロット ← 甲

買主スロット ← 乙

客体スロット→物件A

内容スロット→甲から乙への所有権の移転スロット

→[代金支払いフレーム]

売主スロット→甲

買主スロット→乙

客体スロット→物件Aの代金

内容スロット→甲から乙への代金支払いの事実

先に述べたフレームの交換による事態の認知は、山本の例でいえば、「売買Ⅱ図式」、「賃貸借Ⅱ図式」のいずれを適用するかの問題となる。⁽³⁰⁾ 敷衍すれば、たとえば「債務不履行フレーム」と「不法行為フレーム」のどちらで事態を理解するかといったことも同様となる。⁽³¹⁾

フレーム理論として明示的に指摘されるまでもなく、いわゆるパンデクテン法典は条文記述にこうした概念構成をとっている。山本は法律家の認知の立場からフレーム理論を分析しており、それゆえ上記の様な記述は、条文データベースには親和的である。しかし、判例データベースにおいては、逆に個々の判例毎にこのスロットを想定していくべきであろう。⁽³²⁾ 判例文は実在しており、そこから存在するスロットを作成し、その属性値を求め、その値が得られないスロットは不存在（あるいは不知）とすることが、実在を知識表現する法知識データベースにとってはかえって好都合だと思われるからである。

そしてこの発想は、次のオブジェクト指向型法情報モデルに接合することになる。

二一三 オブジェクト型法情報モデル

フレームにおける付加手続きをもメソッドとして統合⁽³³⁾、フレームがどちらかという概念指向でプラトンのであったのに対して、実体指向のアリストテレス型なのがオブジェクト指向型データモデルである。⁽³⁴⁾

オブジェクト指向スキーマでは、一般に知識表現すべき実体に識別子をつけ、具体的な個々の実体ごとに属性を設ける。それゆえ、フレームと同様、属性はリレーションアルスキーマのような固定的な構造体を前提としない。そうしたモデリング能力の高さ、データと手続きの融合（カプセル化）といった点から、リレーションアルモデルでは表せない表現ができる。また、フレーム表現の違いとして、「フレーム形式でのオブジェクトの関係は、現実世界の意味的關係を使っていたのに対し、オブジェクト指向表現では、世界の抽象化を行うことが知識表現の第一のステップとなる。」⁽³⁵⁾といわれる。オブジェクト指向でのモデル化は次の三種類になるといわれるが、⁽³⁶⁾法情報データベースは用途一にあたり、徹頭徹尾、真性（*genuine*）⁽³⁷⁾実世界モデルが必要となる。

用途一…実世界の問題を表現し、問題の理解を容易化する。特にそのクラス図をもとにソフトウェアを開発することは考えない。

用途二…実世界を模擬実行（シミュレート）するソフトウェアを開発する。

用途三…実世界の業務を自動化するソフトウェアを開発する。

実際に、オブジェクト指向データベース（ODB）で非数値データが大半を占める人文系データベースを構築した事例では、「人文系データベースといっても多様であるが、その中で情報間の複雑な相互関係や込み入ったデータ構造を扱わねばならない場合には、ODBの持つ融通性は非常に便利である。」⁽³⁸⁾として、リレーションアル型データベース（RDB）に対する以下の様な利点をあげている。⁽³⁹⁾

- (一) ODBでは可変長配列や複合オブジェクトを用いることができるので、RDBでは複雑になる操作が簡単になる。
- (二) ODBでは資料間のリンクがオブジェクト間リンクとして直観的に表現できるが、RDBでは複雑な識別子キーが必要である。
- (三) 上記のリンクはODBでは目的リンクを選択するだけだが、RDBでは複雑な計算式となる。
- (四) 各オブジェクトクラスに特有な変換や操作をメソッドとして用意することで高度な機能が使いやすく提供できる。

法情報関連での例として、データモデルにオブジェクトと論理式を採用した工業所有権法表現システムKRIPが、特許法三三三条一項の一部である「特許を受ける権利を持つものは特許出願の前にそれを移転することができる」⁽⁹⁾をどう内部表現しているかをあげる。

Section (特 33, 1, 1).

[event (ID, 発明行為, [,],):

event (ID2, 出願, [請求の範囲 : (ID # 内容),])

→may 出願前の移転

〔主体 : (ID3 # 主体), 対象 : ID3〕

::object (ID3, 特許を受ける権利,

[客体 : (ID # 内容), self # 手続き時].

上記はデータベースでなく推論システムの知識ベースなので、上記の表現はそのままデータベースとして利用で

きるものではないが、枠組みは流用できる。

このシステムの利点としては、以下の諸点があげられている。

- (一) 手続法をシミュレーションとして扱ったので、時間とともに変化する状況を扱うのに都合がよい。
- (二) 区間論理式により、期限に関する規定を明確に記述できる。
- (三) オブジェクトの内部状態の履歴を保存するので、手続きの効力が過去にさかのぼって発生する現象が実現できる。
- (四) 条文に表れる名詞をクラス記述として階層的にまとめることができる。
- (五) 条文参照の順番付けを行って条文の体系化を行うことにより、法令文が大局的に矛盾を生じる問題を解決できる。

データの実際の情報化を、推論ではなく検索を念頭に行なう限りにおいて、(四)を代表格として、これらは法知識データベース構築にとっても有用なオブジェクト型法情報モデルの利点であるだろう。

リレーショナルスキーマに対する、オブジェクト指向スキーマの利点は、事件ごとに条文や判例といった第一的な法情報だけでなく、世論や政策動向といった、付加的な情報も随時追加できる点である。そのことはもちろん情報のそうした意味的な種類に依存しない。ニュートラルな表現でいえば、「複合オブジェクトの表現が容易」⁽⁴¹⁾、すなわち、オブジェクトの属性がどのようなものであるかを完全にしてからでないと情報化できないというわけではないことである。リレーショナルスキーマでこれを前提とするには、原理的にはアドホックなテーブルを次々追加し、後でそれを正規化(normalize)することになる。

しかし、(一)オブジェクト指向言語(推論機構)では、オブジェクト同士の関連づけは、インスタンス(フレ

ムでいうスロット)にポインタを埋め込み、関連の意味をメソッドとして手続き的に表現せねばならないが、データベースのためのデータモデルとしては、それが宣言的に可能になることが求められること、(二)オブジェクト指向プログラミングでは、オブジェクトがもつデータや内部手続きはできるだけユーザや他のオブジェクトに見せないこと(情報隠蔽)が良いことだとされるが、データベースの場合は必ずしもそうではなく、ユーザに明らかにしたほうが良いように思われることが、オブジェクト指向データモデルを知識ベース(プログラミング)として設計するか、データベースとして設計するかの大きな違いとして問題になるように思われる。⁽⁴³⁾

オブジェクト指向データモデルは、その発生がプログラミング(計算、推論)の側からなされたということから、どうしてもオブジェクト指向プログラミング言語の立場、すなわち計算過程を自然にモデリングしようとする能動的オブジェクトの追求に偏りがちのように思えるが、何度もいうように法情報データベースにおいては、むしろ計算のありかたに依存しないで、実体を自然にモデリングしていくという受動的オブジェクトの構築をめざしたほうが良い。⁽⁴⁴⁾ リテラルには情報化されていないデータを推論して出力するといった事も法知識データベースとしてはあってもよいが、その際には、その推論過程を明示する方が、法律関係者ユーザを前提とした場合望ましいと考えるからである。

先の西田の指摘の、「(四)前提についての推論」を取り上げてみよう。単純な例として、事件発生日と、被告甲の生年月日がデータ化されているが、被告甲の事件当時の年齢はリテラルにはデータ化されていなかったという。この際、「被告甲の事件当時の年齢」という検索語を入力すると、システムの方で事件発生日と、被告甲の生年月日から、その結果を推論出力するのはいいが、データベースのユーザインタフェイスとしては、その根拠と過程を明示するべきであろう。

また、典型的なオブジェクト指向データモデルでは、オブジェクトの多面的な性質、複数の表現の問題は、オブジェクトの異なる側面の可能なすべての組み合わせに対するクラスを定義するという、多重継承 (multiple inheritance) を用いて多面性の表現を行ってきた。しかし、(一)クラス数が指数関数的に増加し、クラス階層の見通しが悪い、(二)継承される属性は単一文脈なので、属性名が衝突し、この解消のための規則が必要である、(三)オブジェクトが属するクラスが変更できないので、属性の動的な生成、変更、消滅を記述できない、といった点、すなわち法学理論にひきつけていえば請求権の競合問題のようなスキーマのコンフリクトという点が原理的問題となり、様々な改善が試みられている。⁽⁴⁵⁾

三 法律オントロジーと自然言語処理

三―一 オントロジー工学の導入

上述してきた知識表現はそうした点については無関心で、たとえば、ある人間Aが、かつては学生であり、今は会社員であるということを表示するのも区間論理式を用いたり、メソッドで定義したりとシステム設計者は様々な方法を開発してきた。また運用任せの副作用として、多重継承のようなオブジェクト指向身中の虫をも生み出している。

そのような問題に、オブジェクト指向の立場に立ちながら根底的な反省をせまるものが、オントロジー工学である。その問題意識は、知識の表現方法ではなく、「表現されるべきは何か」、にある。誤解を恐れずにいえばオントロジー工学は「知識哲学」の側面を有しているのである。

先に、「宣言的知識は is-a (a-kind-of) といった継承や has-a (a-part-of) といった関連のような概念定義」と述べ

たが、実はこれすら簡単には定義できないというのが、オントロジー工学の問題提起の一つである。

たとえば、法情報でいっても通常の場合、事件事実から出発する判例データベース設計の立場から見た場合は「債務不履行 *non-performance*、債権発生原因」になるが、パンデクテン法典である民法の条文データベース設計の立場から見れば「債務不履行 *non-performance*、債権発生原因」 ∇ となり、実装においてはそのシステムの目的に応じて、恣意的に定義せざるを得ない。

溝口理一郎らは、この概念化に際する諸問題を集合論に基づいて、(一)概念間の関係のみを概念化した「関係概念」と、インスタンス自体を必ず含む「全体概念」とを定義し、「関係」にはその双方が必ず含まれていること、(上記債務不履行の例)、(二)ある条件下でのみ成立する「ロール (role) 概念」(たとえば「妻」と本質属性定義(たとえば「人間」)とを定義し、例外として後者だけにドメイン依存の恣意的な決定を許すこと、などを導入し解決しようとしている。⁽⁴⁶⁾

全文データベースにおいても、今日、このメタ知識表現ともいうべきオントロジーを核に、ネットワーク上の分散情報源を統合して扱うシステムの研究がなされている。⁽⁴⁷⁾

三二 自然言語処理との統合

このドメインとして法律条文ないしは判例データベースを想定すると、それら法源の中での閉世界仮説が成り立つわけであるが、そこで重要なオントロジーとして、当然頭に浮かぶのは法概念オントロジーであろう。⁽⁴⁸⁾ 確かに検索技術として概念検索を導入する立場は多い。⁽⁴⁹⁾ しかし、エキスパートシステムではともかく、データベースでは詳細な法概念オントロジーは不要ではないだろうか。私見ではそれよりも日本語の統語法 (syntax)、つまり日本語

文法を、法知識データベースのオントロジーとして採用する方が頑健 (robust) なのではないかと考える。なぜなら統語法はまさに、語あるいは語と語、文あるいは文と文の「関係」に「全体概念」(たとえば品詞)、と「関係概念」(たとえば文脈自由文法の格構造)とを有する、日本語ドメインに限定される、弱構造ではあるがかなり整備されたオントロジー体系だからである。⁽⁵⁰⁾

牧之内顕文らは、早くよりこの点を実質的に理解し、データベースへの問い合わせに際して、スピーチアクトを⁽⁵¹⁾データ構造として導入し、統語知識を用いてそれをマッピングしたものをデータベースに渡すことを提案している。またデータベースにおいては、「データを常に意味的に正しい状態に保つために、データが満たさねばならない規約を表現した」⁽⁵²⁾完全性制約 (integrity constraints) を記述することがめざされるが、統語法の意味制約から概念階層の記憶検索の効率を高めるアプローチもあり、統語法がオブジェクトの制約規則記述となる可能性もある。それゆえ、条文知識データベースに関しては、格構造に基づいた意味解析による「法律要件」法律効果」といった弱い法概念オントロジー抽出だけでよいと考えている。⁽⁵³⁾

またユーザインタフェイスの視点からは、自然言語処理を導入することによって統語による法概念の客観的な指示なども可能になるのではないだろうか。

代わりに——タグドコーパスの必要性

データベースは、知りたいことを平易に出力できることが目標である。具体的にいえば、前掲山本の例で、「数匹の猫を飼っており、悪臭などのために他の住民Bに迷惑を及ぼしている」⁽⁵⁵⁾Aを賃貸マンションから追い出す」と検索すれば、関連判例群が要件ごとに明示されて出力されるなどである。

その実現のためには研究分野のプリンシプルを越えたマルチパラダイムの技術、すなわちここでは、法律学＋知識工学（人工知能、自然言語処理）＋ソフトウェア工学のあわせ技が必要とされる。

これらの統合の核になる一つの有力な提案に、検索対象のデータにタグを埋め込んでおくという、タグドコーパス (tagged corpus) の考え方があつた。

「たとえば、法律知識ベースでは、その時点までの判例に基づいて情報がエンコードされる。ところが、それまでと異なる新しい観点からの判例がでてきたときには、その新しい観点も加えたより広い観点からの法律知識ベースの再構築が必要となる。このような場合に、タグドテキストコーパスの形式による知識ベースは威力を発揮する。すなわち、タグドテキストコーパスの形式での情報格納は、情報アクセスだけでなく、格納情報に対する新しい情報の融合および古い情報の更改などのために有効である。このような情報の表現の仕方は、『意味の表現』や『理解の表現』の1つの方法を与えることとみなすことができる。」とされている。⁵⁶⁾

このタグは、「形態素 (morpheme)」「格 (case)」「意味 (meaning)」「文脈 (context)」の四レベルから構成され、先に述べた自然言語オントロジーによる法知識データベース構築に必須である。⁵⁷⁾

タグづけの技術としては、今日、XML (extended markup language) の有用性が主張されている。⁵⁸⁾

さらに、条文や、判例のタギングについてのデータ標準構造規格策定をも視野に入れていいのではないだろうか。条文については、すでに法制局関係者などが、かなり緻密に書き方を示している。⁵⁹⁾ 推論規則としてだが、アメリカでは州法を立法段階からプログラムの形で書いておく実験がなされている。⁶⁰⁾ おそらく、判例についても、いわゆる要件事実教育として定式化された部分がある程度規格化することは不可能ではないと思われる。今後の研究課題としたい。

なお最後になったが、大阪大学名誉教授の田中茂樹先生には、専攻分野は異なるとはいえ、フロンティア研究の厳しさ、難しさを、折に触れご教示いただいた。心より感謝を捧げる。

(1) たとえば、日本学術会議や情報処理学会などが共催する、「情報学シンポジウム」の二〇〇三年の課題は『データの共有と知識の発見・創造』であり、これが焦眉の課題であることを示している。<<http://www.jpss.or.jp/katsudou/sig/signp/f/#symposium>>。

(2) 根岸正光「フルテキスト・データベースの応用動向」、情報処理学会『情報処理』三三巻四号、一九九二年、四一四頁、参照。

(3) ただし、誤植はおくとしても、テキスト化や漢字コードの問題などから、資料的価値は下がり、もはや第三次資料的位づけである。そこで、今度は逆に第一次資料である「判決原本」を、美術工芸品のデータベース化と同様の手法で電子化し、原典性を保存することが別の課題となってくる。特集「失われゆく情報の復元・保存技術、情報処理学会『情報処理』四三巻九号、二〇〇二年、九三九頁」、参照。なお、学説などの法情報についての電子化はあまりなされていない。

(4) なお、「法データ」、「法情報」、「法知識」、「法の存在」、「法の認識」といった基礎概念の考察については本稿の射程にないことをご容赦頂きたい。

(5) 小山照夫「知識表現言語」、情報処理学会『情報処理』二六巻二二号、一九八五年、一五二―一九頁。

(6) 吉野一他『法律人工知能』創生社、二〇〇〇年は、その集大成である。

(7) 辻三郎「知識表現と推論制御」、情報処理学会『情報処理』二六巻二二号、一九八五年、一四七―一五五頁。

(8) 加賀山茂『法律家のためのコンピュータ利用法』有斐閣、一九九〇年、二七―四六頁、参照。

(9) 田中博、下井優一『エキスパートシステム構築の方法』パーソナルメディア、一九八七年など。

(10) 現段階でのその方向での最高到達点は、「複合述語論理式 (CEP: compound predicate formula)」による表現ではないかと思われる。前掲吉野、一四九頁、和田悟「法律知識ベースの構築」、『明治大学情報科学センター年報』九号、一九九六年、三七頁、参照。

- (11) 拙稿「プログラミングによる法学基礎教育」、教育システム情報学会『第二六回全国大会講演論文集』二〇〇一年、二二三頁。
- (12) 田中成明『法的思考とはどのようなものか』有斐閣、一九八九年、二九頁。
- (13) 一般に知的システムの実装においては、ユーザには最小限の情報だけを見せ、その推論形態なども頑健な部分は見せないというシステムの隠蔽性が高い程良い。たとえば、藤村茂、富田昭司、飯間昇、鈴木明「オブジェクト指向知識表現 a u k を用いた知的システム構築用シェル A U K」、『情報処理学会論文誌』三一巻一号、八六頁、参照。しかし、データベースでは必ずしもそうではない。
- (14) たとえば、兼岩憲、東条敏「法律知識の事象的／属性的読みを区別した推論システム」、『情報処理学会論文誌』四〇巻七号、一九九九年、二八九二頁、参照。
- (15) 西田豊明「自然言語理解のための知識表現と推論」、情報処理学会『情報処理』三〇巻一〇号、一九八九年、一二三二頁。
- (16) 前掲西田、一二三三頁。
- (17) 前掲西田、一二三四頁。
- (18) 松本裕治「知識表現——論理的アプローチに焦点を当てて」、情報処理学会『情報処理』二七巻八号、一九八六年、九一六頁。
- (19) 上野春樹『知識工学入門(改訂二版)』オーム社、一九八九年、一九頁。
- (20) 本稿では触れなかったが、関係モデルの限界を越えるものとして、このような演繹データモデルが提唱されたこともあった。
- (21) こうした発想は、Prolog による知識表現においてさえも必要とされていた。たとえば、中島秀之「知識表現用語としての Prolog/KR」、『情報処理学会論文誌』二五巻二号、一九八四年、一八〇頁、参照。
- (22) それゆえ本来は人間の側の認知スキーマをも同時に考察すべきであろうが、幸運なことに法律専門家が自覚して構造化したデータモデルであれば、当然その専門的認知スキーマがデータ構造に反映するものと考えられる。なお、その他の人工知能知識表現で重要なプロダクションシステム、意味ネットワークなどについては本稿の趣旨とは離れるので

省略する。

(23) 増永良文、田中克己「次世代データベースシステムの展望」、情報処理学会『情報処理』三二巻五号、一九九一年、六〇二頁―、参照。

(24) やや異なる観点から、宇田川佳久「高水準データモデル」、情報処理学会『情報処理』三三巻九号、一九九一年、一〇〇八頁、参照。

(25) ただし、内容は仮定的なもので、法律学的な意味での正確性がないことは本研究の趣旨からご容赦いただきたい。

(26) もちろん、現実の判決はこのように単純なものではなく、場合によっては適用条文すらよくわからないこともあるが、まずは単純なモデルから構造化していかざるを得ない。ただし、属性値間の包含関係（たとえば、損害発生を権利侵害の要素ととらえるなど）は、法的推論エンジンや、法律家の側で処理しうる程度の精度を出しておく必要がある。厳密な取り扱いについては後述のオントロジー工学を参照のこと。

(27) フレーム理論については、小川均「フレーム理論に基づく知識表現言語」、情報処理学会『情報処理』二六巻一二号、一九八五年、一四九七頁、参照。

(28) 山本敬三「民法における法的思考」、田中成明編『現代理論法学入門』法律文化社、一九九三年、二二四頁―。なお、認知心理学的理解については、大村敦志『典型契約と性質決定』有斐閣、一九九七年、三一〇頁―。なお双方とも人間の法認知として「プロトタイプ論」に着目している。プロトタイプ論については、see, George Lakoff, *WOMEN, FIRE, AND DANGEROUS THINGS*, UCP, 1987, PART I: Categories and Cognitive Models, p. 5. (邦訳：池上嘉彦、川上誓作他訳『認知意味論』紀伊国屋書店、一九九三年。)

(29) 前掲山本、二二八頁、二三四頁。

(30) 前掲山本、二三四頁。

(31) 人間はこれの中で並列処理しているのですばやく判断できるといわれる。生天目章、大澤洋一「並列推論ネットワーク…スキーマの表現と並列探索」、『情報処理学会論文誌』三三巻十二号、一九九一年、一五五八頁。

(32) これをいいかえれば、条文データベースにはルールによる法的思考の似姿を、判例データベースには事例からの法思考の似姿をインプリメントしたデータモデルを設計すべきだといえることができるだろう。前掲大村、三二四頁、松浦

好治「法的推論——模範例による法思考」、長尾龍一、田中成明編『現代法哲学——法理論』東大出版会、一九八三年、一六七頁、参照。

(33) 前掲小山、一五三二頁、前掲小川、一四九九頁、参照。

(34) 双方は形式的な類似性があるので、ソフトウェア工学における「オブジェクト指向プログラミング」が、知識工学の「フレーム表現」を呑み込んだようである。溝口文雄「オブジェクト指向概念による知識表現言語」、情報処理学会「情報処理」二九巻四号、一九八八年、三八二頁。

(35) なお、同時に「ユーザにとっては、その対象の知識を直接的に表現するという操作性は失われていくという見方もできる。」とも指摘されていることには、マンマシンインタフェイスの視点だけからでなく、留意する必要がある。前掲溝口、三八五頁。

(36) 磯田定宏「実世界モデル化有害論——オブジェクト指向モデル化技法の解明」、『電子情報通信学会論文誌D—I』J八三—D—I巻九号、二〇〇〇年、九四八頁。

(37) ちなみに、法律エキスパートシステムは用途二も内包するが、最終的な目的は用途三になると思われ、この場合には、擬似 (Pseudo) 実世界モデル化が必要であるものと思われる。前掲磯田、九四九頁、参照。

(38) 丸山勝巳「ODBと全文検索エンジンの連携による人文系DB構築システムと電子図書館」、情報処理学会「情報処理」四〇巻三号、一九九九年、八二〇頁。

(39) 前掲丸山、八一四頁、参照。

(40) 新田克己、長男順太郎、水島哲也「工業所有権法の知識表現システムKRIP」、『情報処理学会論文誌』二七巻一
号、一九八六年、一〇五一頁。

(41) 牧之内顕文「次世代データベースシステム」、『電子通信情報学会誌』七八巻一号、一九九五年、七一頁。なお、牧之内は他に、データと操作のこん包(カプセル化)によるデータ構造と操作アルゴリズムのユーザインタフェイスからの隠蔽化、プログラム部品の部品化、アプリケーションの管理ができる、データオブジェクトと計算中の一時的オブジェクトの区別が不要、さらにはシームレスなユーザインタフェイスへの出力、といった点をODBの長所としてあげている。

(42) このことはエキスパートシステム開発においてすら問題である。制約表現 (constraint) を導入してこれを解決しようとするものに、横山孝典、佐塚秀人「制約に基づくオブジェクト指向知識表現システム」、『情報処理学会論文誌』三一巻一号、一九九〇年、六八頁。

(43) 田中克己「オブジェクト指向データベースの基礎概念」、情報処理学会「情報処理」三三巻五号、一九九一年、五〇五頁、参照。

(44) 横田一正、西尾章治郎「演繹・オブジェクト指向データベース」、情報処理学会「情報処理」三一巻二号、一九九〇年、二三五頁、参照。

(45) 佐藤秀樹、池田峰輝、舟橋栄、林達也「多面的オブジェクト指向モデル MAORI」、『電子情報通信学会論文誌D—II』J七九—D—II巻一〇号、一九九六年、七八一頁、参照。

(46) 溝口理一郎、池田満、來村徳信「オントロジー工学基礎論」、『人工知能学会誌』一四巻六号、一九九九年、八七頁。なお、紙数の都合もあり、本稿における要約部分は理論のほんの一部の特徴を示したものである。実はすでに法律学において、集合論的自覚はないにしても、「〇〇権」という本質属性定義だけでは必然的に引き起こされるオブジェクト指向データモデルの権利概念の諸問題を解決するものとして、「関係性(＝ロール概念)」を導入しようとするオントロジー理論的試みがある。たとえば、高井裕之「関係性志向の権利論・序説(三・完)」、『民商法雑誌』有斐閣、九九巻五号、一九八九年、六三八頁、など。このように、オントロジー工学は法解釈方法論に接合する可能性もあると思われるが、それは今後の課題としたい。

(47) たとえば、岩爪道昭、白神謙吾、畑谷和右、武田英明、西田豊明「オントロジーに基づく広域ネットワークからの情報収集・分類・統合化」、『情報処理学会論文誌』三八巻三号、一九九七年、六〇六頁、伊藤史朗、上田隆也、池田裕治「分散情報源に対する情報エージェントのための事例に基づくフレームマッピング」、『電子情報通信学会論文誌D—II』J八一—D—II巻五号、一九九六年、四三三頁など。本稿の射程ではないが、広域分散対象のマルチデータベースやメディアータシステムについては、西尾章治郎、太田友一、横田一正、西田豊明、佐藤哲司「情報の共有と統合」岩波書店、一九九九年、五八頁、参照。

(48) 山口高平「汎用オントロジーを用いた法的オントロジー開発支援機構」、平成六年度科学研究費補助金重点領域研

- 究・研究成果報告書「法律エキスパートシステムの開発研究」一九九五年、一六七頁、参照。
- (49) たとえば、判例データベースに転用可能な有益な研究として、堀井千夏、今井正和、千原國宏「デジタル図書館のための概念情報を用いた科学技術論文の検索」、『電子情報通信学会論文誌D-I』J八二-D-I卷一〇号、一九九九年、参照。
- (50) 石川幹人「文書検索技術と判例データベース」、『明治大学情報科学センター年報』九号、一九九六年、二七頁、も「自然言語での原文レベルでの処理を拡張する必要がある。」とする。また本稿とはいささか趣を異にするが、形式文法の立場から、やはりオブジェクト指向データモデルに文の構造情報を利用しようとするものがある。佐藤秀樹、伊藤暢浩、林達也「オブジェクト指向データモデルGROOM-IIにおける属性文法によるメソッドの形式化」、『電子情報通信学会論文誌D-I』J七九-D-I卷五号、一九九六年、二八〇頁、参照。
- (51) 「スピーチアクトは動作の種類とそれに関係するパラメータをまとめたデータ構造である。動作がデータベースの検索であるなら、関係するパラメータとしては検索項目、検索条件、グループ化項目、順序指定項目などデータベースコマンド生成に必要なパラメータがある。」牧之内顕文、吉野利明、泉田義男「移行性のあるデータベース自然言語インターフェイス」、『情報処理学会論文誌』二九卷八号、一九八八年、七五六頁。
- (52) 前掲田中克己、五〇九頁。
- (53) 小山雅史、泓田正雄、岡田真、青江順一「格構造解析における概念階層の効率判定アルゴリズム」、『情報処理学会論文誌』三九卷三号、一九九八年、五五一頁、参照。
- (54) 拙稿「法律効果規定部の意味機能について」、『情報処理学会研究報告(自然言語処理)』二二四号、一九九八年、一頁、角田達彦、清水仁、長尾真「表層的手がかりによる六法全書法律文での要件部・効果部の抽出手法」『情報処理学会研究報告(自然言語処理)』一一七号、一九九七年、一一九頁。
- (55) 前掲山本、二二五頁。
- (56) 野村浩郷、伊佐原均、徳永健伸、中村貞吾「情報ハイウェイ時代のテキスト情報への知的アクセス」、情報処理学会『情報処理』三三卷一号、一九九六年、五一六頁。
- (57) 前掲野村他、七一八頁。

- (58) 高田伸彦、田村武志、大沢一彦「XMLによるWeb上の論文検索システムの構築」、『電子情報通信学会論文誌D—I』J八四—D—I巻六号、二〇〇一年、六五〇頁、明治大学学術フロンティア推進事業「社会・人間・情報プラットフォーム・プロジェクト」〈<http://ship.mind.meiji.ac.jp/>〉、主にファイリング目的で意味処理を考えるものではないが、Legal XML 〈<http://www.legalxml.org/>〉を参照。
- (59) たとえば、田島信威「新版 法令用語の基礎知識」きょうせい、一九九一年など。
- (60) See, Grayfed Gray, An Experiment with Normalized Statutes in an EMYCIN Expert System, CHARLES WALTER ED., COMPUTER POWER AND LEGAL LANGUAGE, QUORUM BOOKS, 1988, p. 225-.