

Title	疾患記述における異常状態オントロジー構築とその応用
Author(s)	山縣, 友紀
Citation	大阪大学, 2015, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/54003
rights	
Note	

Osaka University Knowledge Archive : OUKA

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

Osaka University

博士学位論文

疾患記述における
異常状態オントロジー構築とその応用

山縣 友紀

2015年7月

大阪大学大学院工学研究科

内容梗概

本論文は、筆者が大阪大学大学院工学研究科電気電子情報工学専攻在籍中に行った、疾患記述における異常状態オントロジー構築とその応用に関する研究をまとめたものであり、6章より構成される。

第1章では、本研究の目的と対象とする課題について述べ、本研究の課題解決に向けたアプローチを述べる。具体的には、疾患を理解する上で、異常状態の多様性、診療科横断的な疾患の共通性把握の困難性および関連知識の散在の3点の問題を対処すべき課題とし、それらの解決に向けて、オントロジー工学理論に基づいた手法を提案する。

第2章は、従来研究に対する本研究の位置づけを明らかにする。

第3章では、オントロジーを構築する上で、その基礎となる疾患記述における異常状態の統一表現モデルの提案とその記述方法について述べる。本モデルでは、臨床における異常状態の多様性という問題に対し、特性・属性分解の手法を用いることにより、一貫性のある統一記述を可能にするとともに、検査データ等の観測データと疾患の記述に見られる異常状態のような概念レベルの高い概念との相互互換を実現する。

第4章では、疾患記述における異常状態の階層的モデリングの提案について述べる。まず、オントロジー工学理論に基づき、異常状態の概念階層を3層で構造化する記述枠組みを提案する。そして、汎用的な異常状態を参照しながら、対象物依存の異常状態、さらに各診療科の疾患固有の異常状態まで特殊化を行うことで、異常状態の一般的な性質と各診療科の専門知識を区別しつつ、コンテキストにしたがって、一貫性を維持して体系的に異常状態を定義できることを示す。その結果、従来では困難とされていた診療科横断的な異常状態の共通性を把握できることを説明する。

第5章では、異常状態オントロジーの応用として、生命医学系の異常状態に関する各種リソース間ナビゲーション手法を提案する。まず、散在した既存のリソースの現状と問題点について述べ、異常状態オントロジーを骨格とする概念レベル毎の統合による関連知識を獲得する解決策を示す。次に、実際に統合を試み、その有用性を確認する。さらに Linked Data 技術を用いた応用システム開発の取り組みと統合基盤構築の全体像を示し、内外のリソースとの連携による基礎から臨床にわたる関連知識の統合における意義を考察する。

第6章では、本研究全体を通して得られた主な成果をまとめ、本論文を総括する。

関連発表論文

A. 学会誌掲載論文

- [A1] Yamagata Y., Kozaki K., Imai T., Ohe K. and Mizoguchi R.,
An Ontological Modeling Approach for Abnormal States and its Application in the
Medical Domain,
Journal of Biomedical Semantics, Vol.5, No.1:23, doi:10.1186/2041-1480-5-23, 14
pages, (2014).
- [A2] 山縣友紀, 古崎晃司, 今井健, 大江和彦, 溝口理一郎,
疾患における異常状態オントロジーの構築,
日本医療情報学会誌, Vol.34, No.3, pp. 101-117, (2014).
- [A3] 山縣友紀, 古崎晃司, 今井健, 大江和彦, 溝口理一郎,
疾患知識統合に向けた異常状態オントロジーの Linked Data 化,
人工知能学会論文誌 Vol.31, No.1, (2016). (採録決定, 15 ページ)

B. 国際会議発表論文 (査読付き)

- [B1] Yamagata Y., Kou H., Kozaki K., Imai T., Ohe K. and Mizoguchi R.,
Ontological Modeling of Interoperable Abnormal States,
In Proceedings of the 2nd Joint International Semantic Technology Conference, Nara,
Japan, Dec. 2-4, 2012, Lecture Notes in Computer Science, Vol.7774, pp.33-48,
Springer, (2013).
- [B2] Yamagata Y., Kou H., Kozaki K., Imai T., Ohe K. and Mizoguchi R.,
Ontological Model of Abnormal States and its Application in the Medical Domain,
In Proceedings of the 4th International Conference on Biomedical Ontology, Montreal,
Canada, July. 8-9, CEUR-WS.org, Vol.1060, pp.22-27, (2013).
- [B3] Yamagata Y., Kozaki K., Imai T., Ohe K. and Mizoguchi R.,
Towards the Integration of Abnormality in Diseases,
In Proceedings of the 5th International Conference on Biomedical Ontology, Houston,
United States, Oct. 8-9, CEUR-WS.org, Vol.1327, pp.7-12, (2014).

C. 研究会発表

- [C1] 山縣友紀, 古崎晃司, 溝口理一郎, 今井健, 大江和彦, 駒谷和範,
Linked Data 技術による異常状態オントロジーと生命医学系リソースの統合への
試み,
第 35 回セマンティックウェブとオントロジー研究会, SIG-SWO-035-06,
(2015).

D. 全国大会発表

- [D1] 山縣友紀, 国府裕子, 古崎晃司, 今井健, 大江和彦, 溝口理一郎,
疾患オントロジーにおけるバランス複合異常と汎用異常連鎖に関する考察,
第 25 回人工知能学会全国大会, 3G1-5, (2011).

- [D2] 山縣友紀, 国府裕子, 古崎晃司, 今井健, 大江和彦, 溝口理一郎,
異常状態オントロジーとその応用,
第 26 回人工知能学会全国大会, 1I2-R-4-3, (2012).
- [D3] Yamagata Y., Kou H., Kozaki K., Imai T., Ohe K. and Mizoguchi R.,
An Ontological Approach to Modeling of Interoperable Abnormalities in Medicine,
第 1 回生命医薬情報学連合大会, C-41-73, (2012).
- [D4] 山縣友紀, 古崎晃司, 今井健, 大江和彦, 溝口理一郎,
オントロジー工学に基づく異常状態の統一的記述モデルと体系化,
第 33 回医療情報学連合大会論文集, 1-B-1-4, (2013).
- [D5] 山縣友紀, 国府裕子, 古崎晃司, 今井健, 大江和彦, 溝口理一郎,
生命科学系 LOD と異常状態オントロジーの統合ー疾患知識横断的把握および相互運用に向けてー,
第 28 回人工知能学会全国大会, 1G4-OS-19a-1, (2014).

E. その他（関連発表）

- [E1] 古崎晃司, 国府裕子, 山縣友紀, 今井健, 大江和彦, 溝口理一郎,
臨床医学オントロジーにおける疾患連鎖モデルの考察,
第 24 回セマンティックウェブとオントロジー研究会, SIG-SWO-A1101-08,
(2011).
- [E2] Mizoguchi R., Kozaki K., Kou H., Yamagata Y., Imai T., Waki K. and Ohe K.,
River Flow Model of Diseases,
In Proceedings of the 2nd International Conference on Biomedical Ontology, Buffalo,
USA, July. 28-30, CEUR-WS.org, Vol.833, (2011).
- [E3] Kozaki K., Kou H., Yamagata Y., Imai T., Ohe K., and Mizoguchi, R.,
Browsing Causal Chains in a Disease Ontology,
In Proceedings of the 11th International Semantic Web Conference, Posters &
Demonstrations Track, Boston, USA, Nov. 11-15, CEUR-WS.org, Vol.914, (2012).
- [E4] 古崎晃司, 山縣友紀, 国府裕子, 今井健, 大江和彦, 溝口理一郎,
医療知識基盤の構築に向けた臨床医学オントロジーの LOD 化の検討,
第 27 回人工知能学会全国大会, 1N5-OS-10c-1, (2013).
- [E5] Kozaki K., Kou H., Yamagata Y., Imai T., Ohe K., and Mizoguchi, R.,
Publishing a Disease Ontologies as Linked Data,
In Proceedings of the 3rd Joint International Semantic Technology Conference, Seoul,
Korea, Nov. 28-30, 2013, Lecture Notes in Computer Science, Vol.8388, pp.110-128,
Springer, (2014).
- [E6] 古崎晃司, 山縣友紀, 国府裕子, 今井健, 大江和彦, 溝口理一郎,
医療知識基盤の構築に向けた疾患オントロジーの Linked Open Data 化
人工知能学会論文誌, Vol.29, No.4, pp. 396-405, (2014).

目次

第1章 序論	1
1.1 研究背景	1
1.2 研究内容	4
1.3 本論文の構成	6
第2章 関連研究	9
2.1 緒言	9
2.2 臨床医学オントロジーにおける異常状態オントロジーの位置づけ	9
2.2.1 疾患オントロジー	10
2.2.2 解剖学オントロジー	11
2.3 既存のオントロジーにおける異常状態関連知識の現状	13
2.3.1 生命医学系オントロジー	15
2.3.2 上位オントロジー	22
2.4 結言	26
第3章 疾患記述における異常状態の統一表現モデル	27
3.1 緒言	27
3.2 疾患記述における異常状態の表現の問題	27
3.2.1 既存のオントロジーにおける表現形式の課題	28
3.3 疾患記述における異常状態の概念定義の基本方針	29
3.4 異常状態の特性・属性分解	31
3.4.1 基本表現	31
3.4.2 拡張表現	32
3.5 観測データとの相互運用性	35
3.6 理論の実践的適用	37
3.6.1 データ表現のバリエーション	37
3.6.2 修飾語の扱い	38
3.6.3 異常状態の構成要素に関する指針	38
3.7 関連研究との比較を通じた統一表現モデルの有用性	43
3.7.1 関連研究との比較	43
3.7.2 実用面における本表現モデルの優位性	44
3.8 結言	44
第4章 疾患記述における異常状態の階層的モデリング	45
4.1 緒言	45
4.2 疾患記述における異常状態分類の必要性	45
4.3 ドメイン横断的な一貫性を指向した異常状態の is-a 階層構築	46
4.3.1 異常状態3層モデルの概要	46
4.3.2 第1層：対象物独立の異常状態 (汎用異常状態)	46
4.3.3 第2層：対象物に応じて共通に現れうる異常状態	48
4.3.4 第3層：疾患コンテキストにおいて定義される異常状態	50
4.4 統一表現モデルを用いた異常状態の is-a 階層の計算機表現	51
4.5 異常状態の上位概念分類例	53
4.5.1 感染症	53
4.5.2 循環器	55
4.6 考察	56
4.6.1 異常状態3層モデルの有用性	56
4.6.2 関連研究との比較考察	58
4.6.3 異常状態オントロジーの評価	63
4.6.4 異常状態オントロジーの意義	64

4.7	結言	65
第5章	異常状態オントロジーを媒介とした生命医学系リソース間ナビゲーション	67
5.1	緒言	67
5.2	生命医学系の異常状態に関する知識の統合	67
5.2.1	既存のリソースの統合における課題	68
5.2.2	異常状態オントロジーによるアプローチ	69
5.3	統合の試行	70
5.3.1	概念層別マッピング	71
5.3.2	外部リソース間マッピングとの比較評価	72
5.4	生命医学系リソース間ナビゲーションシステムの設計	77
5.4.1	異常状態オントロジーの Linked Data 化の概要	77
5.4.2	疾患コンパスと異常状態オントロジーの連携	77
5.4.3	応用システムの試作と実装	78
5.4.4	ナビゲーションシステムの機能	78
5.5	知識統合基盤の全体像	84
5.5.1	生命医学系リソース間ナビゲーションとその意義	84
5.5.2	臨床的な意義	86
5.6	結言	87
第6章	結論	89
6.1	研究の総括	89
6.2	残された課題と将来展望	91
	参考文献	95
	謝辞	100

目次

図 2.1	疾患の概念定義.....	11
図 2.2	心臓の記述例.....	13
図 2.3	要約内に「オントロジー」を含む生命医学文献数.....	16
図 2.4	文献タイトルにおける頻出用語の変化.....	16
図 2.5	PATO 圧力増加の is-a 階層例.....	18
図 2.6	HPO 動脈硬化の is-a 階層例.....	19
図 2.7	YAMATO における性質の is-a 階層.....	25
図 3.1	異常状態の表現例.....	32
図 3.2	データ表現のバリエーション.....	37
図 3.3	対象物の同定.....	39
図 4.1	異常状態の 3 層構造による分類階層.....	46
図 4.2	異常状態 is-a 階層における第 1 層.....	47
図 4.3	第 3 層における異常状態の概念定義例.....	51
図 4.4	異常状態 is-a 階層の計算機表現.....	52
図 4.5	汎用連鎖の生成例.....	57
図 4.6	関連研究における冠動脈狭窄の is-a 階層.....	60
図 5.1	既存リソースと異常状態オントロジーとのマッピング.....	74
図 5.2	対象物視点の動的 is-a 階層.....	76
図 5.3	異常状態オントロジー応用システム.....	79
図 5.4	文献統制語彙集の連携による文献情報の活用例.....	81
図 5.5	異常状態オントロジーと疾患コンパスの連携例.....	82
図 5.6	心不全の起こりうる原因, 結果の例.....	83
図 5.7	異常状態に関連する知識統合基盤の全体像.....	85

表目次

表 2.1	既存のオントロジーと本研究の目的との関係.....	14
表 3.1	比の表現.....	33
表 3.2	修飾語の表現例.....	38
表 3.3	属性の同定例.....	41
表 3.4	LDL 酸化の記述例.....	42
表 3.5	塞栓の記述例.....	43
表 4.1	マラリア感染における感染から発症までの定義例.....	54
表 4.2	異常状態に関連する主な生命医学系リソースとの比較結果のまとめ....	59
表 5.1	異常状態オントロジーと外部リソースとのマッピングに関する統計....	71

第1章 序論

1.1 研究背景

近年、技術の進展とともにデータも多種多様化の一途をたどり、ウェブ上ではクラウドコンピューティングにより、蓄積されているデータ量も爆発的に増加するにつれて、それらの有効活用が問題となっている。

医療分野では、従来、患者の診療記録は紙カルテに記載されており、患者個人のための一次利用を主として各機関で記録が管理されてきた。しかし、電子カルテシステムの導入など医療情報の電子化に伴い、各種統計解析や調査への二次利用が重要となってきた。例えば、地域での医療連携や国レベルでの解析を実現するためには、情報を整理し、多くの医療関係者が情報を共有できるような標準化技術やデータ交換の仕組みが必要となる。そこで、データ項目セットやコードの標準化について検討が進められている。

ここで、病名は重要な役割を果たし、病名の標準化への取り組みが進められている[MEDIS 03]。実際に、標準病名はカルテにおける罹患疾患を特定するための情報として記録に用いられ、診療報酬請求書での利用において、診療報酬を受ける根拠として広く利用されている。

しかしながら、病名自体にはその意味内容が明示されていない。より高度な情報解析を行うには、疾患のもつ意味情報を適切に処理し、疾患と関係する諸概念との関係性を処理できる高度な情報システムの構築が要請される。

ところで、疾患について、臨床では決まった定義は存在しない。また、診療の場では、これまで疾患は診断タスクの中で扱われ、個々の患者の状態を端的に示す概念として記録されてきた。臨床医は、検査値や患者に観察される所見や症状からその原因を探り、疾患との関係性を、自身のもつ専門的な背景知識で推論している。しかし、限られた診察時間内で記述されるカルテなどには、推論結果として最も特徴のある主病態と病名しか記述されておらず、原因についての詳細は記録には残っていない。

さらに、臨床は人体での実験の困難性から、例えば高血圧ラットや糖尿病マウスなど疾患のモデル動物による研究が進められ、新薬の開発などが行われている。しかし、実際には、基礎研究と臨床応用には死の谷が存在するとされ、疾患の解明に向けてはまだまだ困難な状況といえる。

疾患を本質的に理解するためには、こうした様々な疾患を取り巻く概念との相互の意味的な関係性を明らかにする深い考察が必要となる。そこで、著者らは知識の体系的理論としてのオントロジーを解決アプローチとして採用して、厚生労働省「医療情報システムのための医療知識基盤データベース研究開発事業」において、臨床医学オントロジープロジェクトを進めてきた [大江 10]。オントロジーとは、哲学用語に由来し、存在論を意味するが、人工知能分野では、Gruber はオントロジーの「概念化の明示的規約」という定義を提唱している [Gruber 93]。これをさらに発展させ、実社会の問題解決にむけた基盤概念構造構築のための理論と構築技術を開発して、実システムへの応用までを担うのがオントロジー工学である。

臨床医学オントロジープロジェクトでは、オントロジー工学者と臨床医を含む医療の専門家の協力の下、解剖学オントロジーと疾患オントロジーの開発が進められてきた。その後、本研究が対象とする異常状態オントロジーもこの臨床医学プロジェクトの一環として始まった。

まず、解剖学オントロジーでは、人体構造物の組織化を目的としている。臨床では診療科について、人体の各システムによって、循環器系ならば循環器内科、神経系は神経科、消化器系ならば消化器内科のように、診療区分がされている。人体は人工物と比べ、複雑な部品から構成されるサブシステムから成り立つシステムであるといえ、サブシステム毎に高度の専門知識を要するといえる。

そこで、解剖学オントロジーでは、各診療科で必要とされる人体のサブシステムとその部品を定義するとともに、診療科独立に存在する管や壁などの汎用構造物の定義も行っている。さらに、構造記述全般の問題への対処として、例えば、人体構造物における構成部品の捉え方について複数の視点に応じた **part-of** 関係の導入や、神経、脈管接続などの各種接続情報を識別可能な接続部品を提供するなど、オントロジー工学理論に基づいた問題解決への取り組みが行われている。解剖学オントロジーで定義される人体構造物は、本研究で扱う異常状態統一表現モデルにおいて、異常状態の対象であるオブジェクト（対象物）としての役割を担う。

次に、疾患オントロジーについて述べる。前述したように、疾患について、臨床では決まった定義は存在しない。しかし、疾患を理解するには、まず、疾患概念を定義し、これまで暗黙的であった各専門医の背景知識を明示し、疾患が何を原因として生まれ、どのような状態を引き起こしていくか明らかにし、その意味的内容を記述することが必要となる。そこで、疾患オントロジーでは、疾患を異常状態の原因、結果の連鎖の総体として定義することとした [Mizoguchi 11]。

このように、異常状態は疾患を適切に捉える上で欠かすことのできない重要な概念と位置づけられている。

しかしながら、異常状態の把握は特に医療においては困難とされている。その理由として、以下の三つの課題が存在する。

まず、医療における多様な異常状態の表現をどのように扱うかということが最初の課題となる。人体は複雑で、細胞から人体まで様々な粒度の部品から構成される。そのため、人体で起こる異常状態も多様であり、例えば細胞壊死もあれば、疼痛という人体レベルもあるというように、粒度も不均一である。さらに、異常状態は医療では実に様々な形で表現される。患者を観察した結果は、症状として電子カルテに記載されている。また、血液検査や心電図、画像といったものに関するデータとして異常値が現れる場合もある。治療においては、薬剤の使用説明書にもどのような異常状態が起こるかについての記述がある。さらに、教科書では疾患とは何かを説明するために異常状態による記述が必要となる他、基礎研究では実験動物を使った各種実験において異常状態が観察され、文献にその結果が記述される。このように、異常状態についての情報は多種多様であり、各々の利用目的に対応した表現となっている。特に、医療機関で蓄積されている臨床検査データと、教科書における疾患記述に用いられる異常状態や診療記録で専門医が各疾患に関する知識として理解している異常状態の表現の間には、その扱い方に大きく隔たりがある。データから疾患や異常状態というような概念レベルの高い表現への変換には、表現が指し示す意味を考慮する、データの内容面からのアプローチが必要となる。

第二の課題は、異常状態の共通性を的確に把握することである。臨床では、医療の高度化に伴って専門領域が非常に細分化されている。そのため、異常状態は、診療科毎に専門医が疾患固有の異常状態として扱っているのが現状である。また、総合診療科も存在するものの、初診患者の入り口として、診断ベースで各患者において観察される異常状態を症状としてのみ把握するため、異常状態としての全体像の把握は困難である。

さらに、第三の課題として、既存の散在したリソースからの関連知識の獲得が挙げられる。生命医学系では古くから用語集の開発がされており、近年では各種オントロジーの開発も盛んであり、ウェブ上では積極的にデータが公開されている。そこで、これら既存のリソースをもとに異常状態に関する知識を利活用できるような仕組みが望まれる。しかし、表層的な一致で統合しようとしても、概念レベルの不一致や概念混同が生じ、膨大なデータに埋もれたまま、結局、必要とする関係概念にたどりつけないという事が起こりうる。その結果、それぞれのデータが散在したまま有効活用できないという問題が生じている。

ここで、異常状態は、医療に限らず、機械、航空など様々な分野において故障を理解する上で重要な概念であるといえる。ときに情報の共有不足から重大な事故につながる場合もあることから、いかに異常状態の意味内容を正確に把握し、情報を互いに共有するかが要請される点においてはどの分野においても共通する。また、上述の課題についても分野が異なっても対処しなければならない点は同じといえる。異常状態は通常、時間とともに変化する。そのような異常状態の性質をどのように処理し、表

現すべきかについてはどの分野においても検討を行う必要がある。さらに、最初の課題である異常状態の多様性について考えてみると、故障における異常状態も疾患における異常状態と同様、例えば「摩耗」、「圧力低下」のように、通常、抽象概念的に捉えられるため、試験データと親和性に乏しいことが挙げられる。

また、現代では非常に高度な技術を専門的に扱えるように領域が細分化されている。そのため、例えばシステムのある構成部品のみの特化して製品開発を行っている場合、同じシステムであったとしても、別の構成部品を扱う部門の用語や技術が理解困難となってきた。したがって、人体における疾患と同様、故障が起こったとしても個別的であり、故障事象の全体像を把握することが困難となっている。さらに、別の問題として、異常状態は、通常ある故障が起こったときにはじめて認知される。つまり、事例ベースの故障診断というタスクの中でのみ扱われてきた。一方、故障を根本的に理解するには、診断タスクとは独立に、異常状態の構成要素に立ち返って明らかにしていく必要がある。

技術者や専門家は、自身の持つ経験と専門知識によって異常状態を理解することが多い。そのため、その分野に従事する専門家として当然持つべき知識については、教科書以外に遭遇する機会は少ないといえる。しかし、そのような知識の明示は計算機の問題解決能力の向上においても必要であると考えられる。さらに、第三の課題においても膨大なデータから必要な情報をいかに入手するかという課題は他の分野にも共通する。

1.2 研究内容

このような背景のもと、本研究の目標として、疾患を理解するために必要な異常状態について、その意味内容の表現と構造化における理論的な枠組みの確立とその具体的な組織化を目指す。本研究では上記の3つの課題解決に向け、異常状態オントロジー構築とその応用システムの開発に取り組む。まず、異常状態オントロジー構築では、オントロジー工学の観点に基づき、疾患記述における異常状態の意味内容を計算処理するための統一的な記述枠組みを構築し、その記述枠組みを基礎として、理論的な組織化を目指す。さらに、異常状態オントロジーを応用し、生命医学系の異常状態に関する知識を統合し、知識獲得および知識発見の支援を目指してシステム開発に取り組む。

以上の考察に基づき、本研究では以下の3つの研究テーマを設定した。

(1) 疾患記述における異常状態の統一表現

(2) 疾患記述における異常状態の階層的モデリング

(3) 異常状態オントロジーを媒介とした生命医学系リソース間ナビゲーション

最初の2つの研究テーマについては、異常状態オントロジー構築に取り組むものであり、3つ目はその応用としてのアプローチとして位置づけられる。

まず、疾患記述における異常状態の統一表現に関する研究の概要について述べる。本研究では、多様な異常状態表現について、一貫性をもって統一的な記述枠組みを提供することを目指す。そのために、まず、異常状態を特性で捉えて定義することで、疾患記述における異常状態に対応づける。さらに、そのままでは概念レベルが高いため、一般的な物理パラメータを用いた属性とその値に分解するという特性・属性分解表現で、検査データなどの具体的な表現に対応可能な統一表現記述枠組みのモデルを提案する。本モデルは上位オントロジーである YAMATO [Mizoguchi 10] をベースとすることで、定量値と特性の両方を扱えることから、データから疾患記述における異常状態の概念知識までを相互互換可能とする。本モデルは3種類の記述枠組みを用意し、基本型の他、特殊型や複雑な比のような表現に対応可能な拡張表現を用いることで、一貫性を保ちながらも柔軟な記述枠組みを提案し、情報量は落とさず、かつ計算処理への負荷の少ない統一表現枠組みを実現する。さらに、実際に疾患で記述される異常状態に適用する際に必要となる、一貫性を保って適用するための指針を提案する。

次に、2つ目の研究テーマである、異常状態の階層的モデリングについて述べる。臨床医学において、疾患を理解するために必要な異常状態の構造化を考えた場合、どのような観点が必要かを考える。まず、各診療科における専門医が各疾患に関する知識として理解している異常状態を捉えることは必須である。これに加え、合併症の把握には、複数の疾患に共通する異常状態を扱うことも必要となる。とくに、治療においては人体の各器官の構造の特徴に基づいて異常状態を処理することも重要となる。さらに、医療情報の共有と相互運用を考慮した場合、診療科を横断した汎用的な異常状態も捉えて体系化することが必要となる。そこで、これらの理由から、本研究テーマでは異常状態を3層に大別し、構造化を目指す。

3層モデルでは、汎用的な概念、人体解剖構造物依存、疾患依存の異常状態の各概念を一般・特殊関係に基づいて関係性を構築 (is-a 階層化) し、各々を識別可能にする。また、再帰構造をとることにより、汎用的な異常状態は、下位から参照されるとともに、特殊化を行うことで専門医が必要とする粒度までをカバーする異常状態概念の提供を実現する。

このように、異常状態オントロジーは、データから疾患定義における異常状態までをカバーした統一的表現を行うとともに、厳格な概念化と本質属性の継承を充足するように一貫性をもって組織化することで、各診療科における疾患固有の異常状態から、診療科横断的に複数の疾患における異常状態の共通性を把握することができる。これにより、診療科間の異常状態比較解析や各種評価などの計算機による知的処理を可能にし、診療科間の異常状態の相互運用性を向上させるとともに、疾患を理解するために必要な異常状態の全体像の把握に貢献すると考えられる。

3つ目の研究テーマは、異常状態オントロジーの応用として、本オントロジーを媒介として生命医学系のリソースを統合することで、一般的な異常状態から疾患固有の異常状態まで、概念レベルに応じて横断的にナビゲートする枠組みを提案する。そしてその提案に基づき、統合の試みとその結果を用いた **Linked Data** 技術 [Heath 11] によるナビゲーションシステムの開発を行う。本システムでは各疾患で起こる異常状態について、既存のリソースから関連知識の獲得を可能にするとともに、基礎から臨床まで異常状態に関する知識記述の橋渡しを支援する。また、疾患オントロジーとの連携により、診療科横断的に疾患で起こりうる異常状態について、原因、進行、および結果として現れる症状を把握することにより、従来では困難とされていた疾患の臨床像の反映を可能にする。

1.3 本論文の構成

本論文は6章から構成されている。第2章では、疾患記述における異常状態のオントロジーに関わる従来研究に対する本研究の位置づけを明確にする。まず、筆者が属する研究グループが参画している臨床医学オントロジープロジェクトにおいて開発されている疾患オントロジーおよび解剖学オントロジーと、本研究で構築する異常状態オントロジーとの関係性について述べる。続いて、既存の生命医学系オントロジーおよび上位オントロジーについて、本研究の3つの目的である 1) 異常状態の統一表現、2) 異常状態の階層的モデリング、および 3) 生命医学系リソース間ナビゲーションの観点から、従来研究の問題点を指摘し、本研究の位置づけを明確にする。

第3章から第5章では本研究の成果について示す。3章および4章は異常状態オントロジーの構築について述べる。第3章では疾患記述における異常状態の統一表現モデルの提案とその記述方法について述べる。まず、統一記述を行うための課題を述べる。次に、統一的に異常状態を記述するには、異常状態の概念定義が必要となる。そこで、異常状態の概念定義の基本方針について説明する。そして、特性・属性分解の手法による統一表現モデルを提案し、検査データ等の観測データと疾患の記述に見られる異常状態のような概念レベルの高い概念との相互互換を実現する。続いて、提案

モデルを疾患記述に用いられる異常状態に適用する際、一貫性を保つための指針について述べる。さらに、関連研究との比較を通し、本提案手法の有用性について考察する。

続く第4章では、疾患記述における異常状態の階層的モデリングの提案について述べる。上述の統一的表現モデルを基礎として、疾患を理解するために要求される概念レベルについて、オントロジー工学理論に基づき考察した結果、3層構造でモデル化を提案する。本提案枠組みでは、汎用的な異常状態から、対象物依存の異常状態、さらに各診療科の疾患固有の異常状態まで特殊化を行うことで、一般的な知識から専門知識まで一貫性をもって定義可能であることを示す。そして、本モデルに基づき、疾患記述に用いられる異常状態の上位概念分類を行う。さらに、異常状態オントロジー構築を行う意義について議論する。

第5章では、異常状態オントロジーの応用として生命医学系の異常状態に関する各種リソース間ナビゲーション手法について述べる。まず、既存のリソースの統合における課題について述べ、異常状態オントロジーによる解決アプローチを提案する。次に、異常状態オントロジーの概念レベル毎に既存のリソースとの統合を試み、提案手法の有効性を確認する。さらに、Linked Data 技術を用いた応用システムの開発に取り組む。そして、異常状態オントロジーを媒介とした基礎から臨床にわたる異常状態に関する知識の統合基盤構築の全体像を示し、その意義を考察する。

第6章では研究全体を通して得られた主な成果をまとめ、本論文を統括する。

第2章 関連研究

2.1 緒言

本章では、関連研究に対する本研究の位置づけを述べる。本研究は、疾患記述のための異常状態オントロジーの構築とその応用に関するものである。そこで、まず、2.2節では、筆者が属する研究グループが参画している臨床医学オントロジープロジェクトにおいて構築されている、「疾患記述」(定義)を対象とした疾患オントロジーおよび、異常状態の発生個所となる「人体構造物」を対象とした解剖学オントロジーと、本研究の異常状態オントロジーとの関係性について説明する。

続いて、2.3節では、既存研究のオントロジーについて、本研究の対象である「異常状態」に注目し、本研究の目的とする異常状態の統一表現、異常状態の階層的モデリング、および散在したリソース間ナビゲーションの観点から問題点を指摘し、本研究の位置づけを明らかにする。2.3.1節では生命医学系オントロジーのうち異常状態を含むものについて、2.3.2節では異常状態を扱うために不可欠な概念である、属性、特性、属性値などの基本的な概念を扱う上で必要な上位オントロジーについて論じ、本研究の位置づけを述べる。

2.2 臨床医学オントロジーにおける異常状態オントロジーの位置づけ

本節では、まず、筆者が属する研究グループが参画する臨床医学オントロジープロジェクトにおいて開発している疾患オントロジーおよび解剖学オントロジーと、本研究が対象とする異常状態オントロジーの関係性を述べ、その位置づけを明確にする。

本研究の目的 1) および 2) はともに「疾患記述」における異常状態の統一表現と階層的モデリングにより、異常状態オントロジーの構築を目指しており、注目すべき対象は「異常状態」である。そして、疾患記述の対象である各診療科の「疾患」については、本研究の異常状態オントロジーとは別に、「疾患オントロジー」で概念定義されている。

次に、異常状態を把握するには、それが起っている人体の部位の同定が必要となる。しかし、「部位」自体の概念定義については本研究の範囲外であり、別途、解剖学オントロジーで行われている。

2.2.1 疾患オントロジー

本研究は疾患記述における個々の「異常状態」のみを対象としている。異常状態の定義は「疾患」とは区別されており、疾患記述における「疾患」そのものの概念定義は、疾患オントロジーで進められている。

臨床医学オントロジープロジェクトの開始時には、既に国際的な疾患分類 ICD-10 [ICD]が開発されていたが、オントロジー理論上、継承情報が適切でなく一貫性がないなどの問題が存在した [Shulz 07]。そこで、本プロジェクトにおける疾患オントロジーでは、臨床医学知識において重要となる「疾患」についてオントロジー理論的に適切な概念定義を行い、疾患概念の組織化を目指すこととされた。

疾患とはいわゆる病気の概念であるが、疾患そのものの定義については、広く受け入れられているものは、現状では存在しない。Barry Smith の研究グループでは、哲学に基づいた存在論の立場として、疾患を傾向 (Disposition)として定義している [Scheuermann 09]。しかし、臨床では、病名によって表現される患者の人体で起こっている複雑で多様な疾患をよりの確に捉えたモデルが必須となる。そこで、疾患について考察すると、疾患では異常状態が次の異常状態を引き起こす原因となり、伝播しつつ、次の連鎖が進んでいき、重篤な影響を及ぼしていくことが見出せる。したがって、疾患オントロジーでは、疾患について異常状態を基本構成要素とし、「疾患とは異常状態が形成する原因と結果の連鎖の総体である」と概念化することとした [Mizoguchi 11]。これをオントロジー工学で捉えると、疾患は「人体に現れる状態変化による因果連鎖から構成される従属的実在物」と定義される。疾患オントロジーでは、臨床専門医が各診療領域で対象とされる個々の疾患について、異常状態を表すノードとその因果関係を表すリンクから構成される有向グラフ (以下、疾患連鎖と呼ぶ)を用いて表現される。異常状態は、異常状態の原因または結果として各疾患に現れる。疾患の is-a 関係は、疾患定義に用いられる因果連鎖 (以下、疾患定義連鎖と呼ぶ)の範囲の包含関係で表される。すなわち、疾患の下位概念の定義は、疾患定義連鎖の範囲を原因側 (上流)・または結果側 (下流)の方向に広げたものとなる。例えば虚血性心疾患の疾患定義連鎖が「冠動脈狭窄→血流量低下→心筋虚血」であるとすると、心筋虚血の結果に心筋壊死が加わった状態を疾患定義連鎖とする心筋梗塞は虚血性心疾患の下位概念となる (図 2.1)。2013 年 3 月の時点において、疾患オントロジーには、13 診療科 6,302 の疾患および疾患を構成する 21,669 の異常状態が臨床医によって定義されており、2014 年 5 月の時点では、そのうち 7 診療科の 3030 疾患の全 13,047 の異常状態について洗練作業が進められている。

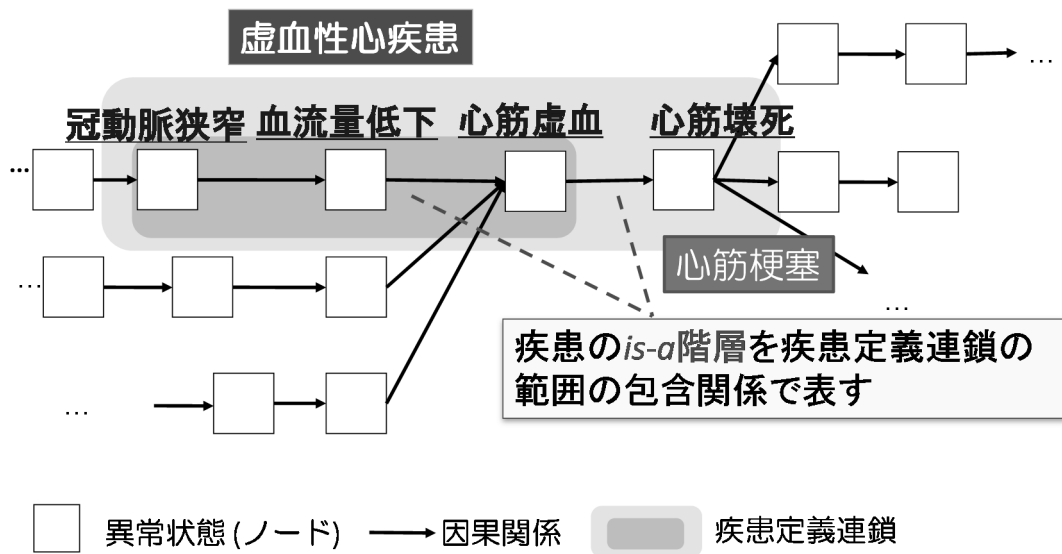


図 2.1 疾患の概念定義

このように、異常状態は、疾患の原因や結果としての役割を果たす。しかし、ある疾患では、異常状態は原因となるが、別の疾患ではその異常状態が結果として現れる場合もある。

本研究目的 1) 異常状態の統一表現においては、これら複数の疾患であらわれる異常状態の定義の統一的記述をすることが必要となる。また、本研究目的 2) 異常状態の階層的モデリングにおいては、そのコンテキストとなる疾患と独立した視点で、個々の「異常状態」について客観的に概念定義することが重要となる。

そこで、本研究では、疾患と独立して、異常状態そのものを概念定義し、組織化する理論的な枠組みの開発を目指し、異常状態オントロジーを構築することとなった。

2.2.2 解剖学オントロジー

本研究では、異常状態の多様性に対処するため、目的 1) では、異常状態を一貫した方法で表す統一表現の記述枠組みの確立を目指している。臨床では、診療記録や臨床検査、さらに投薬の際のいずれにおいても、異常状態が「人体のどこで起こっているか」という異常個所（対象物）を客観的に把握することが重要となる。それには、「人体構造物」を表す概念が一貫性のある方法で提供されなければならない。しかし、「人体構造物」の概念化そのものについては、本研究が対象とする「異常状態」とは明確に区別して行われるため、本研究の対象外としている。そこで本研究では、異常個所を表すために、臨床医学オントロジープロジェクトにおいて構築されている解剖学オントロジー [Imai 09] を用いることとした。

解剖学オントロジーでは、様々な構造物が持つ共通構造を表す汎用構造物という汎用的な概念を導入することで、それぞれの解剖学構造物が定義されている。汎用構造物には、壁状構造物、弁構造物、管状構造物などがある。例えば、胃壁も心臓壁も同じ空間を区切るという機能をもつ壁状構造物として定義されている。別の例として、食道、気道、腸、血管等については、すべて管状構造物という汎用構造物を用いて定義されている。これにより、管状構造物－血管－動脈－冠状動脈といった解剖学構造物の is-a 階層が表される。一方、人体を構造物とその部品の関係から考察すると、単なる部分-全体関係だけでは、臨床で使われている概念として不十分なことがある。例えば、胃について考えた場合、胃壁と内腔から構成されると考えることもできるし、胃を噴門部、胃底部、胃体部、幽門部のように分けて考える場合もある。さらに、胃角のような任意の領域に部分名称がついている場合も存在する。このように、複数の観点での記述に対応できるように、オントロジー工学理論に基づき、人体の部分構造については part-of 関係を4種類に細分化できるとされている [古崎 10]。

また、血管、骨、神経については構造物の部品に接続ポートという仮想の概念を導入し、血管の接続や分岐など、構造物間の接続関係をたどれるように記述がされている。

図 2.2 に心臓の記述例を示す。心臓は心房と心室から構成され、さらに各々左右の部品として、左心房/右心房、左心室/右心室をもつ。図の例のように、心室は中空性構造物として定義されている。また、心臓には心房、心室の他、僧帽弁などの弁構造物も存在し、血液の流れを制御する機能をもつ。

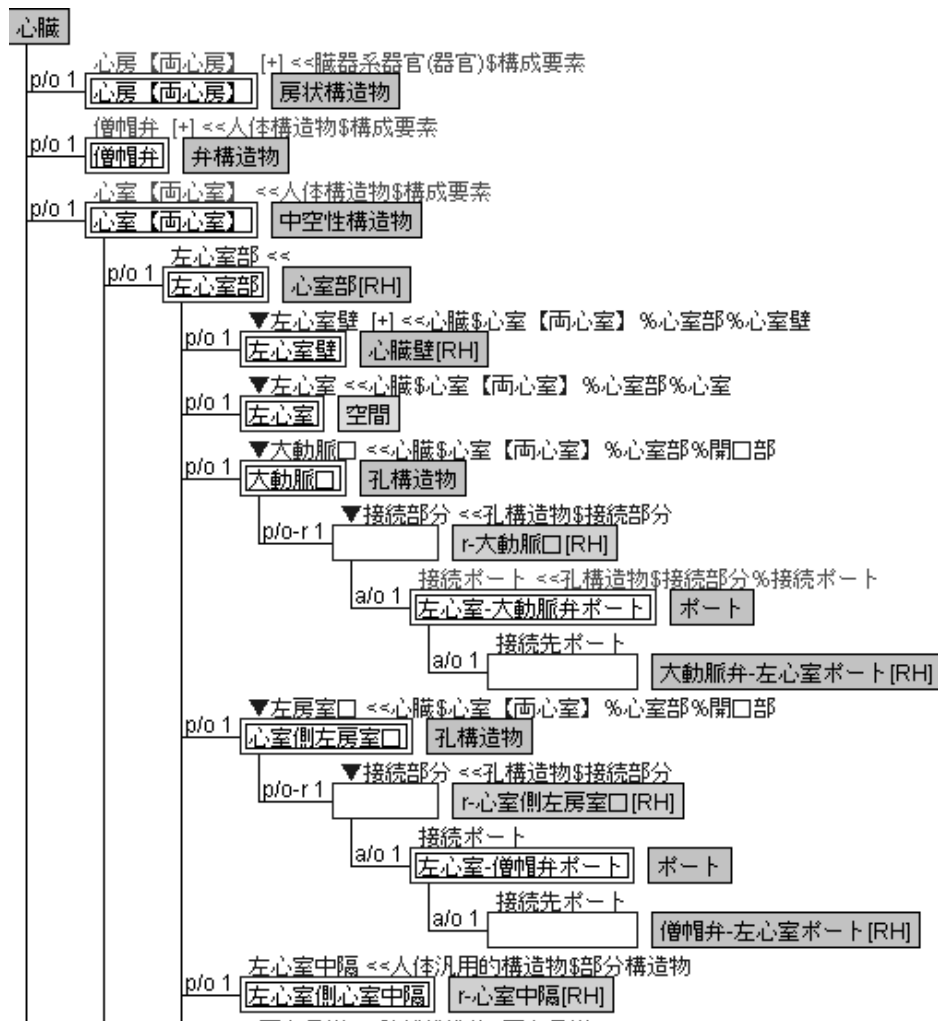


図 2.2 心臓の記述例

解剖学オントロジーで提供される人体構造物の各概念は、本研究の目的 1) 異常状態の統一表現という点において、異常状態が起こる部位についての統一表現記述に用いられる。さらに、研究目的 2) の階層モデリングの目的においては、4.3.3 節で述べる対象物に応じて共通に現れうる異常状態 (対象物依存異常状態) において、器官固有の異常状態のコンテキストとなる役割を果たす。

2.3 既存のオントロジーにおける異常状態関連知識の現状

本節では、従来研究のオントロジーについて、本研究の 3 つの目的である、1) 異常状態の統一表現、2) 異常状態の階層的モデリング、および 3) 生命医学系リソース間ナビゲーションの観点から、問題点を指摘するとともに、本研究の位置づけを考察する。

まず、本研究が対象とする異常状態知識に既存のオントロジーについて調査を行った。表 2.1 は、既存のオントロジーについて、本研究の3つの目的との関係の有無、すなわち各目的がそれぞれのオントロジーの構築目的に含まれているか否かを示したものである。

表 2.1 既存のオントロジーと本研究の目的との関係

	生命医学系オントロジー					上位オントロジー		
	PATO	HPO	SNOMED-CT	MeSH	LOINC	BFO	DOLCE	YAMATO
1) 異常状態の統一表現	○	○	—	—	○	○	○	○
2) 異常状態の階層的モデリング	○	○	○	○	—	—	—	—
3) 生命医学系リソース間ナビゲーション	○	○	○	○	—	—	—	—

○：関係あり，—：関係なし

その結果、生命医学系オントロジーのうち、一般的な表現型オントロジーとして Phenotypic Quality Ontology (PATO) [Gkoutos 05] およびヒトの表現型である Human Phenotype Ontology (HPO) [Köhler 14] は、本研究の統一表現、階層的モデリング、生命医学系リソース間ナビゲーションのいずれの目的とも関係することがわかった。しかし、これらは異常状態の観点で構築されたものではないので、2.3.1.1 節に示すように、本研究の対象とする「異常状態」のオントロジーとして利用するには問題が生じる。

次に、Systematized Nomenclature of Medicine -Clinical Terms (SNOMED-CT) [SNOMED-CT] や Medical Subject Headings (MeSH) [MeSH] は、臨床用語の標準用語集や文献のための統制語彙集として著名であり、用語の分類階層を提供している点において、本研究の2) 階層的モデリングと関連する。しかし、これらはオントロジー理論の成熟する前から開発されているため、複数の視点が混在し、異常状態に特化した概念の厳格な階層的モデリングという目的においては利用が難しい。さらに、大規模な階層構造かつ ad-hoc に散在した異常状態を既存のリソースと関連付けることは困難であり、3) リソース間ナビゲーションという目的で利用する際にも問題が発生する。

さらに、生命医学ドメインとは独立した、この世に存在する対象について高い抽象レベルで概念定義を目指している上位オントロジーとして、Basic Formal Ontology (BFO) [Grenon 04]、Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) [Guarino 98]、および Yet Another More Advanced Ontology (YAMATO) がある。これらは、持続物や生起物の定義等、非常に抽象度が高いため、2) の異常状態の階層的モデリングについては対象の範囲外であるが、1) の統一表現の記述枠組みにおいて、異常

状態の本質をとらえるために必須の概念である、属性や特性、そして属性値などの定義に貢献する。

以下、上述の生命医学系オントロジーと上位オントロジーについて概説するとともに、表 2.1 で示した本研究の目的との関係からみた問題点と、その位置づけについて詳細に説明する。

2.3.1 生命医学系オントロジー

同じ生命医学系を対象ドメインとしていても、視点や目的の違い、あるいは生物種の違いにより、様々なオントロジーが開発されている。そこで、本節では、まず、生命医学系オントロジーの開発の現状について述べる。次に、本研究の対象となる異常状態を含有する生命医学系オントロジーについて調査を行い、異常状態がどのように捉えられ、位置づけられているかを論じ、本研究の目的における問題点を述べる。

生命科学分野では、近年、ゲノム配列解読を目指して機関を超えた国際協力が必要であったことから、ゲノム解読完了後も、その流れで積極的に生物種を問わず様々なデータが公開されている。さらに、遺伝子の機能解析などに関する用語の統制語彙の構築をめざし、3種のモデル動物データベース(ショウジョウバエ, 酵母, マウス)の共同プロジェクトとして設立されたコンソーシアムによって Gene Ontology が開発された [Ashburner 00]。Gene Ontology は、遺伝子に関する概念を3つの観点：①細胞内の遺伝子が働く場所(部分構造)、②分子機能、③生物学的プロセスで分類し、生物種共通の用語として遺伝子に機能を関連づけることができるため、マイクロアレイなどの実験における遺伝子のアノテーションに幅広く用いられるようになった。Gene Ontology の開発以後生物医学分野では、オントロジーの開発が盛んになるとともに、オントロジーの利用自体も増加している。National Center for Biotechnology Information (NCBI) の開発した文献検索サービスシステムの PubMed [PubMed] を用いて、「オントロジー」という用語が要約内に付された生命医学文献を調査したところ、例えば、2000年には「オントロジー」を用いた文献は100件にも満たなかったが、2013年には約1600件となっている(図 2.3)

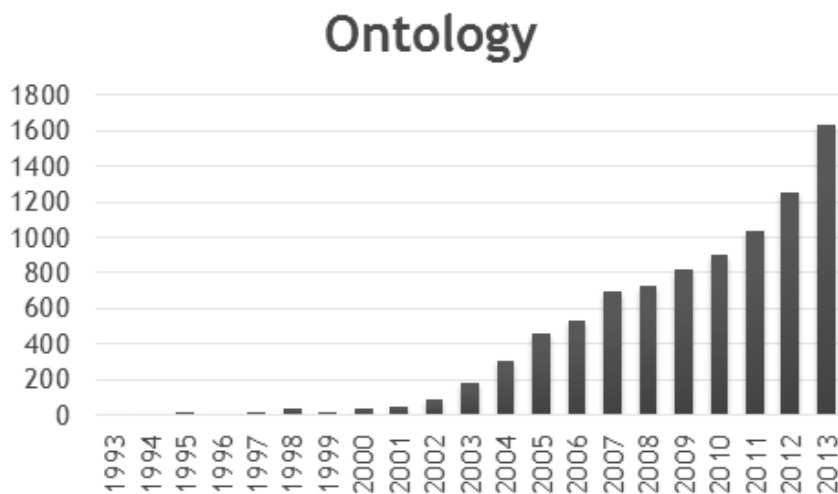


図 2.3 要約内に「オントロジー」を含む生命医学文献数

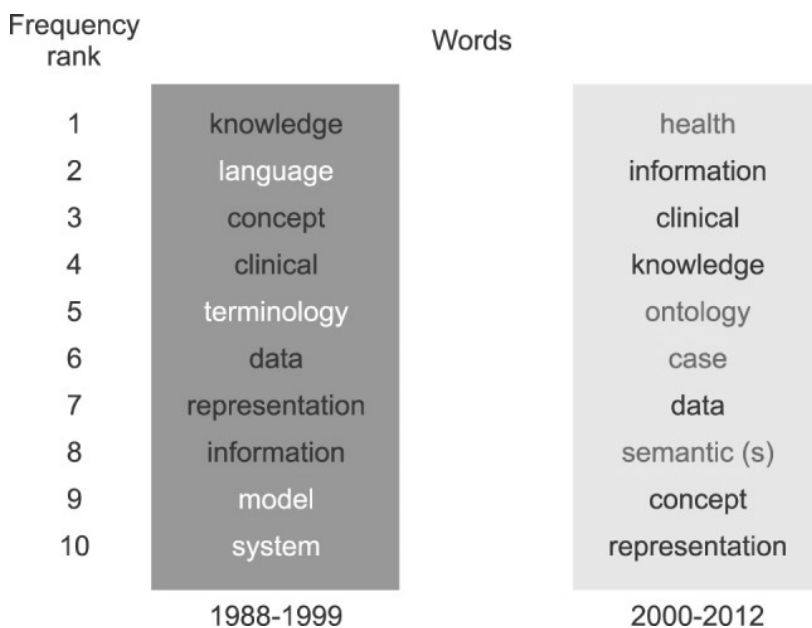


図 2.4 Schulz (2013)より引用 文献タイトルにおける頻出用語の変化

また、医療情報分野において、タイトルにオントロジーが含まれるものは 1990 年代にはなかったが、2000 年以降、頻出上位に入っていることが報告されている (図 2.4) [Schulz 13].

さらに、Web 上のレポジトリサイトでオントロジーを公開する動きも世界的に活発化している。そこで、既存の関連研究について、各種データレポジトリサイトを対象に、異常状態が含まれる用語集やオントロジーにどのようなものがあるかを調査することとした。

まず、生命医学系のデータレポジトリサイトとして著名なサイトである、National Centers for Biomedical Computing (NCBO)の Bioportal [Bioportal], Open Biological and Biomedical Ontology Foundry (OBO Foundry) 提供の Open Biological and Biomedical Ontologies (OBO) [Smith 07, OBO], European Bioinformatics Institute (EBI) 提供の Ontology Lookup Service (OLS) [Côté 10, OLS], および文献検索サービスシステム Pubmed 等を用いて、異常状態を含む既存の用語集やオントロジーの調査を行った。その結果、既存の用語集やオントロジーにおいて、異常状態そのものに特化したものは存在しないことが分かった。

そこで次に、異常状態に関する用語を包含するものについて調査したところ、オントロジーとしては、表現型を扱うオントロジーである Phenotypic Quality Ontology (PATO) [Gkoutos 05] およびヒトの表現型については Human Phenotype Ontology (HPO) [Köhler 14] があり、医学関連の用語辞書や標準語彙体系として、国際医療用語集 Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT) [SNOMED-CT] や米国国立医学図書館が提供する医学文献の統制語彙として Medical Subject Headings (MeSH) [MeSH] があることがわかった。そこで、本研究の3つの目的である異常状態の 1) 統一表現, 2) 階層的モデリング, および 3) 生命医学系リソース間ナビゲーションの観点からこれら既存のオントロジーについて説明する。

2.3.1.1 表現型オントロジー: PATO, HPO

生命科学分野全般においては、実験による観測や臨床所見を通じて異常状態知識を把握するのが一般的であり、表現型を強く意識した形で異常状態に関する知識が存在する。表現型とは、ある生物のもつ遺伝子型が観察される性質（形質）をいい、表現型を扱うオントロジーとしては、Phenotypic Quality Ontology (PATO) [Gkoutos 05]が著名であり、ヒトの表現型については Human Phenotype Ontology (HPO) [Köhler 14]が存在する。本節では、表現型オントロジーについて概説するとともに、本研究の目的である異常状態の 1) 統一表現, 2) 階層的モデリング, および 3) 生命医学系リソース間ナビゲーションの観点から問題点を述べる。

PATO はショウジョウバエ、ゼブラフィッシュ、マウスなどいくつかのモデル動物を対象とした表現型オントロジーの研究プロジェクトがメンバーとなり、2002年から開発されている。その目的は、生物種を限定せず、複数の生物種での一般的な表現型のアノテーションをするためである。PATO では、サイズ、色、重さなど表現型を表すために必要な性質の種類（属性）が分類されて階層化されている。また、その下位には生物で観察される性質が定義されており、例えば、圧力 (pressure) の下位に圧力増加 (increased pressure) などがあり、これらの概念と本研究の対象となる異常状態が対応するといえる (図 2.5)。2015年5月現在、2457クラス存在する。

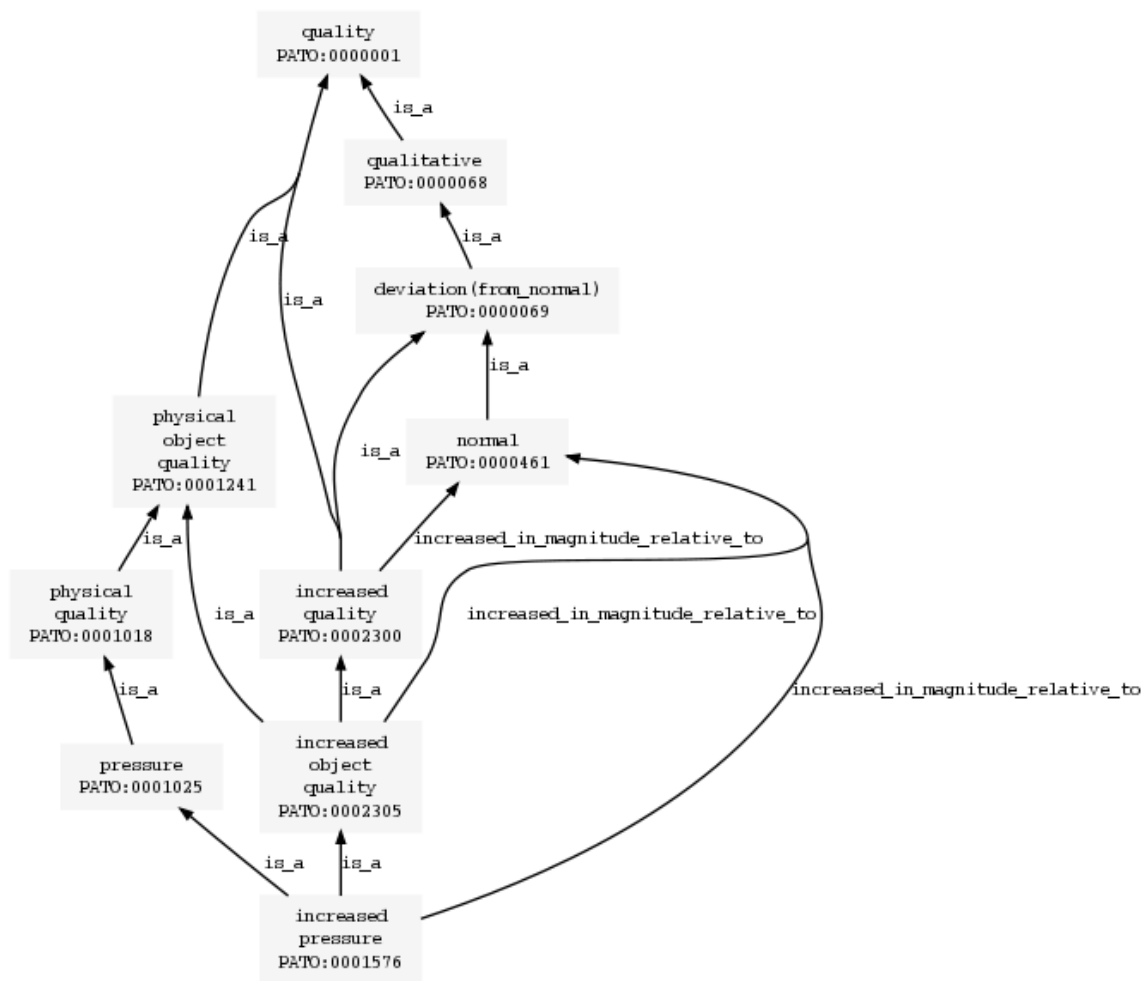


図 2.5 PATO 圧力増加の is-a 階層例 (<http://www.ebi.ac.uk/ontology-lookup/>)

HPO は、ヒトに特化した表現型オントロジーである。HPO は、ヒトの遺伝病 DB である OMIM [Hamosh 05] で記述される表現型に対する標準語彙の整備を目的として、2007 年に開発が開始された。2015 年 5 月現在、約 10000 クラスを有し、本階層は大きく、1) 遺伝形式 (Mode of Inheritance), 2) 臨床経過 (Onset and Clinical Course), および 3) 表現型異常 (Phenotypic Abnormalities) の 3 つに分かれる。異常状態は主に 3) の下位概念に含まれる。例えば、動脈硬化 (Arteriosclerosis) は 3) に存在し、血管異常 (Abnormality of the Vasculature) を上位概念にもつ。このように、HPO は人体の各器官のどこで観察されるかに従って階層が構築されている (図 2.6) [OLS].

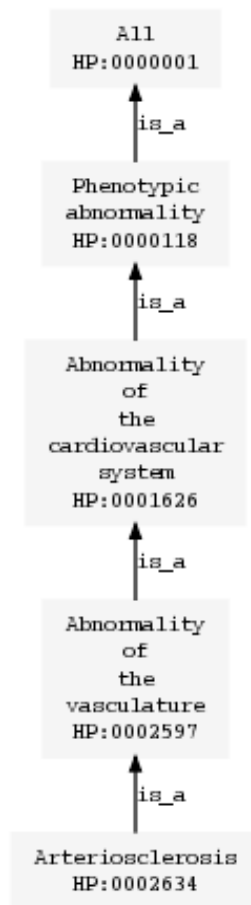


図 2.6 HPO 動脈硬化の is-a 階層例 (<http://www.ebi.ac.uk/ontology-lookup/>)

以上から、前述の表 2.1 で示したように、HPO、PATO は、本研究の 3 つの目的 1) 統一表現、2) 階層的モデリング、3) 生命医学系リソース間ナビゲーションの目的に対応しているとも考えられる。しかし、実際、異常状態という観点で利用しようとした場合には問題が生じる。

まず、本研究の目的 1) 異常状態の統一表現における問題点を述べる。PATO および HPO は測定データなどの定量値を扱っていない。临床上では検査データの値まで扱える統一表現を必要とするため、これらのオントロジーは問題があるといえる。そこで、本研究は検査データから異常状態まで統一的に扱える理論的な枠組みの確立を目指すこととし、この課題については 3 章で述べる。

次に、これら表現型オントロジーは、本研究の第二の目的である異常状態の共通性把握のための階層的モデリングにおいても問題がある。PATO および HPO は、いずれも遺伝子を原因として観察される性質（表現型）を扱っている。しかし、臨床では遺伝子は同定されない異常状態が多く存在する。例えば、心窩部痛などは遺伝子レベルではまだ解明されていない。さらに、虚血のような概念もカバーされていない。そこで、本研究では、汎用的な異常状態から、临床上起こりうる異常状態として、人体の

各構造物で観察される性質の他、心窩部痛や虚血のような状態もカバーし、異常状態の階層的モデリングを行い、複数疾患の共通性を把握できる構造化を目指すこととした。本内容は4章で説明する。

さらに、本研究の第三の目的において、既存の散在したリソースから関連知識を統合する際、カバーされていない概念を関連付けて、必要な情報を取得することは困難といえる。そこで、本研究では、異常状態オントロジーを媒介とし、概念レベルに応じて統合し、関連知識を獲得する応用研究を行い、それについては第5章で述べる。

2.3.1.2 臨床検査データ標準コード規格 LOINC

本節では、異常状態の統一表現に関連するものとして、検査データの標準規格について著名な Logical Observation Identifiers Names and Codes (LOINC) について説明する。LOINC は、1995 年に臨床検査、診察情報の項目とそのコード用の規格として Regenstrief Institute によって開発された [McDonald 03]。既存の ASTM [ASTM], HL7 [Dolin 11]などの臨床検査データの交換を目的としている。見出し語は SNOMED-CT で統一され、データは CSV 形式で提供し、LOINC TABLE とよばれている。LOINC では、検査項目を 6 項目 (成分 (Component), 測定性質 (Property), 検査のタイミング (Timing), システム (System), スケール (Scale), 検査方法 (Method))によって分類している。is-a や part-of などの関係表現はもたず、各項目を略語で表記し、「:」で結合して表す。例えば、グルコース (GLUCOSE) 負荷試験における経口 (PO)グルコース 100g (100 G) 投与後 2 時間 (2H POST) 血清/血漿 (SER/PLAS) グルコース (GLUCOSE) 濃度 (MCNC) は、"GLUCOSE^2H POST 100 G GLUCOSE PO:MCNC:PT:SER/PLAS:QN"のように記述される。

このように、LOINC は、本研究の1)の異常状態の統一表現の目的においては、検査データの統一表現としては適した形式となっている。しかし、本例に示すように、濃度のもつ値など、値自体は定義されていない。さらに、異常状態は定義されていないため、「高血糖」というような異常状態表現との互換性がないため、多様な異常状態を統一表現するには適切でない。

2.3.1.3 SNOMED-CT

Systematized Nomenclature of Medicine -Clinical Terms (SNOMED-CT) は、臨床用語集 (Clinical terminology) であり、現在 27 カ国が加盟する International Health Terminology Standards Development Organisation (IHTSDO) が現在管理、維持を行っている[IHTSDO]。SNOMED-CT は、米国病理学会によって開発された病理学用語 SNOMED から始まり、その後、英国医療用語集 CTV3 (Read code) [O'Neil 95]を統合して今に至っている [Stearns 01]。その特徴として、臨床情報に必要な用語を幅広くカバーし、2015 年 1 月現在、Bioportal で公開されている SNOMED-CT のクラス数は約 30 万となっている。最近では、国際的な標準用語として、病院間での電子医療情報システム等に使用され

ている。SNOMED-CT は、一概念 (Concept) につき 1 つの固有の ID として 6-18 桁の数字が付され、複数の表記 (Description) が関係づけられている。Description は、いわゆる人間が可読の用語が記述され、その数は、約 800,000 となっている。SNOMED-CT の階層は、19 カテゴリに大きく分類されており、トップレベルのカテゴリには、例えば、所見 (Clinical finding/disorder)、身体構造 (Body structure)、生物種 (Organism)、物質 (Substance) などがある。SNOMED-CT では、一つの概念が複数の上位概念を持つことが認められ、すなわち多重継承で階層化されているといえる。異常状態は、通常、臨床所見/異常 (Clinical finding/ Disorder) の下位概念に含まれており、例えば、臨床所見/異常の下位に「心筋壊死 (Myocardial necrosis)」 (finding) がある。概念間を結ぶ関係は、上位/下位 (is-a 関係) の他、finding site, method, など 65 種類用意されている。動脈閉塞 (Occlusion of artery) の場合、is-a 関係では、動脈異常 (Disorder of artery)、軟組織病変 (Soft tissue lesion) の 2 つが上位概念となっており、finding site 関係として動脈構造 (Arterial structure) が記述され、その結果として動脈閉塞は動脈構造で観察されることが表現されている。

ここで、本研究の目的における SNOMED-CT との関係性について述べる。本研究の目的 2) 異常状態の階層的モデリングという点において、SNOMED-CT は上述のように、大規模な臨床用語をカバーしているため、異常状態の階層化に貢献し、複数の異常状態の共通性把握も容易ではないかとも考えられる。しかし、SNOMED-CT の臨床所見/異常の階層をみると、下位概念にも、疾患など異常状態ではない他の概念が混在している。また、多数の視点が混ざった多重継承による階層化によって、30 万の用語から異常状態だけを抽出し、適切に組織化することが困難となっている。2) の目的においては、一定の視点に基づき、異常状態のもつ性質の継承関係について一貫性のある方法で異常状態に特化して厳格に構造化することが求められる。

さらに、本研究の目的 3) において、生命医学系リソース間ナビゲーションという観点においても、疾患と異常状態との識別が不明確なまま関連付けると、そのままナビゲートした場合、臨床上、診断での疾患同定などにおいて問題を生じる可能性がある。

2.3.1.4 MeSH

Medical Subject Headings (MeSH) は、米国医学図書館が提供する生命医学文献のインデックスおよびアノテーションのための医学用語統制語彙である [MeSH]。2014 年版では、27149 の用語を有し、同義語を含むと 219000 の概念数をもつ。生命医学文献情報データベース MEDLINE は、1 文献につき、その内容を表す MeSH の用語 (MeSH ターム) が 10 個程度付与されている。それらの文献は PubMed という公開文献検索サービスサイトで検索できるようになっており、収録件数は約 2400 万件を越えている [PubMed]。階層は解剖構造 (Anatomy)、生物種 (Organisms)、疾患 (Diseases)、化合物と薬物 (Chemicals and Drugs) などの 16 に分類されており、異常状態関連用語は、疾患

(Diseases) の下位に存在するものが多い。例えば、代謝性疾患 (Metabolic Diseases) の下位には、糖代謝異常 (Glucose Metabolism Disorders) があり、さらに下位には疾患である「糖尿病 (Diabetes Mellitus)」とその異常状態「高血糖(Hyperglycemia)」が兄弟概念として存在する。

このように、MeSH についても、SNOMED-CT と同様、疾患と異常状態の混在や複数の視点による多重継承から 2) の目的には不適切といえ、3) についても SNOMED-CT と同様の問題を抱える。

以上から、本研究目的における既存の生命医学系オントロジーの問題を解決するため、本研究では、異常状態について、1) 統一表現の記述枠組みの確立を目指すとともに、疾患と独立に、2) オントロジー工学理論に基づいて、ゼロから異常状態を厳格に組織化し、汎用レベルから疾患固有の異常状態までをカバーする階層的モデリングを目指すこととした。その結果、異常状態オントロジー構築の理論枠組みを確立した。さらに、3) 異常状態オントロジーを骨格とすることで、これら生命医学系リソースの異常状態に関する知識を統合し、適切な情報へナビゲートするという第三の目的も達成することを可能とした。

2.3.2 上位オントロジー

前節で述べたように、本研究の目的 1) 異常状態の統一表現において、既存の生命医学系オントロジーではその表現枠組みの確立に不十分であることがわかった。疾患記述における異常状態について統一表現を行うためには、各状態のもつ本質的な性質そのものを客観的に理解することが重要となる。そこで、本節では、このような問題意識から、異常状態の本質的な性質を考察するため、ドメイン独立で性質に関する概念を提供している上位オントロジーを関連研究として考察する。

上位オントロジーとは、ドメイン独立に、対象世界を説明するために抽象度の高いレベルで必要な概念の定義を目指して開発されたオントロジーである。序文で述べたようにオントロジーとは元来「存在論」を意味し、哲学に由来していることから、上位オントロジーでは、哲学的な考え方を尊重している [加地 05]。著名なものとして BFO [Grenon 04]、DOLCE [Guarino 98]が存在する他、日本では YAMATO [Mizoguchi 10] が開発されている。上位オントロジーの提供する概念には、どのドメインでも使用可能な一般的な性質が含まれる。以下、これら 3 つの上位オントロジーについて、基本的な概念の思想を概説するとともに、本研究の目的 1) 異常状態の統一表現記述に必要な性質、量の概念の捉え方について述べる。続いて、本研究で採用している YAMATO を取り上げ、BFO、DOLCE と比較を行い、異常状態表現において重要

とされる値の変化の扱いについて、YAMATO と他の二つのオントロジーとの違いを述べる。

2.3.2.1 BFO

Basic Formal Ontology (BFO) は、Buffalo 大学の Barry Smith が中心となり、哲学の考え方に基づいて開発されたオントロジーである。BFO は、以下の3つの軸にそって実体 (Entity) について考察し 34 の概念で分類しようというものである。

- (1) Continuant (持続物) vs. Occurrent (生起物)
- (2) Dependent (従属的) vs. Independent (独立的)
- (3) Type (種) vs. Instance (インスタンス)

持続物は、3次元空間の存在物として、例えば物などが代表される。性質は BFO では Quality とよばれており、持続物の下位概念とされている。また、持続物については、従属的に存在するか独立的に存在するかで識別される。性質は、何かのものに内在しており、他への依存なしには存在できない、従属的持続物 (Dependent Continuant) として分類される。なお、生起物とは、プロセスやイベントなど時空間上で同定される存在物を指す。BFO では、存在そのものに注目するため、量は実体としての扱いではなく、量を表現する場合、性質のインスタンスとして定義される。

米国の生命科学系のオントロジー構築を目的としたオープンコミュニティ OBO Foundry [Smith 07] では BFO を規範とする上位オントロジーとして採用している。

2.3.1.1 節で述べた上述の PATO も複数の生物種の一般的な性質を捉えるために BFO の性質の表現を用いて定義を行っている。

2.3.2.2 DOLCE

Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) は、イタリアのトレントにある Laboratory for Applied Ontology の Nicola Guarino が中心となり、認知科学の立場で開発された上位オントロジーである。DOLCE では、実体を表す Particular について、持続物 (Endurant) と生起物 (Perdurant)、性質 (Quality)、抽象物 (Abstract) に分類し、BFO とは異なり、性質 (Quality) を持続物とは独立して概念化している。性質については個物に内在するという考え方であり、DOLCE では、さらにその性質を、どの個物にも内在していない性質の種類 (属性) と値で記述できるように、Quality Type と値に対応する Quale を各々定義している。その結果、個別のモノが持つ性質 Quality について、例えば、ある葉の色が緑から茶に変化するとする。この場合、Quale を緑から茶にすることで、色について、値の変化が扱える点が特徴的である。なお、Quale は、個物のインスタンスとは独立した Quality Space に存在する

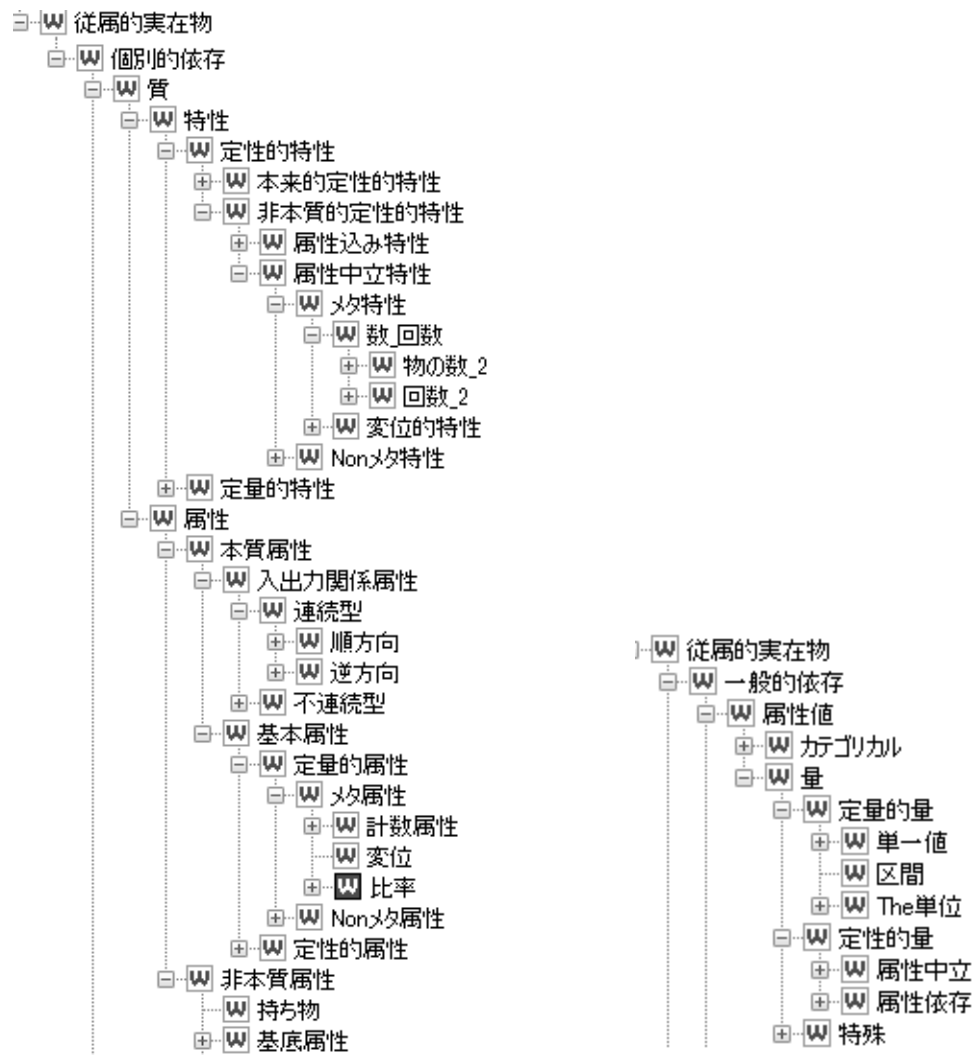
とされており、赤は緑の補色であるというような個物と独立した値の比較が可能となる。

2.3.2.3 YAMATO

Yet Another More Advanced Ontology (YAMATO) は、溝口が開発した工学的上位オントロジーで、工学的観点から構築されている。生起物、持続物の扱いは BFO, DOLCE に共通する。YAMATO では、性質に関し、以下のような特徴をもつ。

- 特性関連表現 (従属的持続物)を特性 (Property), 一般性質タイプ (属性) (Generic Quality Type), 属性値 (Quality Value) に大別している。
- 属性値の下位としてカテゴリ値と量を定義し、量を定量値, 定性値に分類している。
- 性質のロール概念を導入することで、コンテキスト依存の性質の表現を可能としている。
- 存在論としての表現と実在するデータ表現を識別することで、両者を統一的に扱う枠組みを提供している。

YAMATO では、性質の変化を扱えることを特徴とする。DOLCE と同様に性質の種類とその値を扱える他、DOLCE をさらに拡張し、値については、量を扱うために詳細な定義が用意されている (図 2.7)。例えば、大きい、小さいというような定性値、および定量値の他、胎生、卵生などのカテゴリ値などが記述可能な点が特徴的であり、また変位量、相対比較値などにも対応している。さらに、性質については、一般的な物理パラメータである面積、体積、速さなどの一般性質タイプ、コンテキストに依存した断面積などの性質ロールタイプその他、性質インスタンス、性質インスタンスタイプなども提供されているので、BFO, DOLCE との相互互換を可能にする。なお、BFO の Quality は YAMATO の Property と対応する概念であり、DOLCE の Quality Type が YAMATO の Generic Quality Type に対応する。さらに、ロール概念とは、例えば「長さ」という属性がある対象物の「高さ」や「深さ」を担うなど、コンテキストに依存した概念のことをいう [Kozaki 07]。YAMATO では、BFO, DOLCE の存在論としての捉え方に加え、測定データについても表現可能な工学的配慮がなされている上、両者を識別しつつ、統一的に扱うことができるよう定義がされている (図 2.7)。



(a) 性質の階層分類 (b) 性質の値の is-a 階層
 図 2.7 YAMATO における性質の is-a 階層

以上、上位オントロジーでは、異常状態の概念定義において、異常状態のもつ性質と、その値を扱う一般的な枠組みを提供していることを確認し、YAMATO が定量値も処理するための量の記述に貢献できることを考察した。

しかし、上位オントロジーは、一般的な性質の概念提供にとどまり、臨床の異常状態概念は全く定義されていない。そのため、臨床で起こる疾患記述のための多様な異常状態を記述するには、それが人体構造物の中のどの部位で起こっているかの同定が必要となる。加えて、どのような観点で異常値が変化するかについて異常状態毎に同定する必要がある。そこで、本研究では、上位オントロジーの本質的な性質の捉え方を基礎として、本研究の目的 1) 疾患記述における異常状態の統一表現について具体的に検討することとした。

2.4 結言

本章では、関連研究について本研究の目的を参照し、それらの位置づけを明確にするとともに、本研究の目的での利用における問題点について説明した。

次章では、異常状態の多様性に関する課題解決に向けたアプローチとして、異常状態統一表現モデルを提案する。

第3章 疾患記述における異常状態の統一表現モデル

3.1 緒言

本章では、本研究の目的 1) 疾患記述における異常状態の統一表現を目指し、疾患記述における異常状態オントロジーのための統一表現モデルを提案する。

まず、統一表現モデルを提案する際に、考察の基盤とする疾患記述における異常状態の表現の問題について説明する。次に、異常状態の統一表現を記述するために必要な異常状態の概念定義について、2.3 節で述べた既存のオントロジーの問題点について、臨床における異常状態に実際に適用しながら詳細を説明する。さらに、既存研究の問題を解決し、本研究における異常状態を統一的に記述するための異常状態の概念定義の基本方針を述べる。そして、異常状態の特性・属性分解に基づく3つの型の統一表現モデルを提案する。続いて、観測データとの相互運用性の向上について考察する。その後、実際の疾患記述に用いられる異常状態への適用を行い、一貫性を保つための指針について述べる。そして、最後に関連研究との比較を通した統一表現モデルの有用性について検討する。

3.2 疾患記述における異常状態の表現の問題

知識を体系化するためには、まず関係する対象世界を支配する概念を明確化することと知識を記述するための共通語彙を定めることが重要となる [溝口 05]。近年、知識処理の分野においては、様々な知識の意味を計算機上で適切に扱うための基盤技術として、またウェブ分野においては、ウェブ上のドキュメントやデータに計算機の意味づけを与えるセマンティックウェブの基盤として、オントロジー技術の研究や標準化が進められている。Web Ontology Language (OWL) は記述論理に意味論の基礎を置いたオントロジー記述言語として開発されたものであり [OWL]、その後、OWL2 [OWL 2] への改訂も進められている。

ここで、知識表現については、大きく2つの観点がある。形式指向に基づくものと内容指向によるものである [溝口 97]。現在、セマンティックウェブを支えるのは主に前者であり、不均一なデータを、どのように論理的に記述すれば形式的な統一を可能にするかという観点に重きを置いているといえる。一方、後者は何を明示的に記述すべきかという意味内容そのものについての表現に興味がある。その内容がどのような性質をもって、どのような種類があるのか、その性質を表すにはどのような表現が

必要か、コンテキストの変動とそれらによって変化する内容などを一つずつ明らかにしていくことで、どのような内容から構成され、具体的に形式化すべきかの解を与える。形式指向と内容指向は両者対立するものではない。内容指向アプローチはセマンティクス (意味) を明示的に扱えるような基礎づくりをサポートし、データから知識へ変換させることで、適切な意味内容をもつ知識媒体となり、さらに形式指向アプローチによって支えられた一定の枠組みの中でこれらを記述し交換することにより、知識の流通、共有、相互運用の実現が実社会においても可能になると考えられる。

以下、本研究では後者の観点に基づいて、内容指向アプローチによる疾患記述における異常状態の表現について、既存研究の問題点を考察する。

3.2.1 既存のオントロジーにおける表現形式の課題

異常状態を統一的に記述するには、まず一つの視点で異常状態を概念定義することが必要となる。2.3.2 節で示したように、上位オントロジーは、客観的に異常状態の本質を捉えるために有用であることがわかった。現在、性質を表すための枠組みとして、上位オントロジーでは BFO, DOLCE, および YAMATO が知られており、BFO では性質そのものの存在を重視した定義、DOLCE では、値の変化の扱いが可能な定義、さらに YAMATO では値について量を詳細化した定義を各々構築している。これらはドメイン独立の一般的な枠組みであるため、医療における異常状態は定義されていない。そのため、各々の性質の概念定義をそのまま用いることはできない。そこで、上位オントロジーの表現形式について、研究目的 1) 統一的表現を達成するための問題点と解決策を具体的に検討する。

BFO は例えば、臨床における「高血圧」のような表現ができる。しかし、量を扱う枠組みを用意していないため、「圧力」とその値が「高い」というように値を分離した記述ができない。臨床検査データでは、値の変化を扱えるのが必須といえるが、BFO では「圧力 180 mmHg」というように、値を分離できないので、BFO の表現形式は検査データの各値への対応が困難といえる。

一方、DOLCE は属性と値の具体的な記述枠組みを備えているため、血液の「圧力」が「高い」というように値の記述が可能である。しかし、「高血圧」のように値も包含した意味内容を示す記述ができない。また、臨床では、例えば「吐き気」のように属性への分解が困難な異常状態も存在するのが特徴的であるが、DOLCE では、これらを扱えないということになる。さらに、捉え方の違いから BFO と DOLCE の上位オントロジー同士では互換性が考慮されていない。

ここで、2.3.2.3 節で述べたように上位オントロジー YAMATO では、前述の表現全てを扱い、相互変換もサポート可能な枠組みを提供している [Mizoguchi 10]。さらに、量については、定量値と定性値も識別可能となっている。そこで、本研究では、

YAMATO をベースに、これを医療分野に適用し、疾患記述のための異常状態表現枠組みについて検討を行うこととした。

しかし、医療ドメインでは、これら一般的な性質の枠組みでは記述が困難な表現が存在する。例えば、高血糖などの比の表現や、ウイルス感染や血栓といった複数のもので関わる異常状態など複雑な表現がある。高血糖ならば糖を対象とすべきか、血液を対象とすべきか、また、ウイルス感染や血栓形成の場合、ウイルスや血栓をどのように記述すべきか、その性質をもつ対象は何かについて構造物の同定などさらに進んだ考察が求められるとともに、どのように表現の統一性を確保するかが課題となる。

3.3 疾患記述における異常状態の概念定義の基本方針

前述の課題を解決するため、本節では、疾患を理解し、疾患記述における異常状態を捉えるための異常状態の概念定義を行う。

医療ドメインでの異常状態の概念定義は、他のドメインと比べて特に困難とされている。その理由の一つとして、人体は細胞、組織から器官まで様々な粒度の要素から構成されていることがある。例えば、細胞壊死や、組織炎症というような粒度から、器官レベルでは、胃拡張などや心房中隔欠損、皮膚発赤など各々の粒度において非常に多様な状態が存在する。さらに、人体レベルでは「吐き気」というような一般的な物理パラメータでは表すことのできない表現も存在する。したがって、これら異常状態に関する知識を体系化するには、対象とする異常状態概念の本質を捉え、統一した観点から一貫性をもって概念化することが必要となる。

状態とは、「時間的に変動するのが当然とみなせる広義の属性が、あるときある属性値をとっていることを抽象化した概念」として捉える事ができ [溝口 05]、通常は、特性 (Property, (以下, P)) に対応する概念であるといえる。例えば、虚血状態の場合、ある時点において「虚血あり」として概念化することができ、状態は、時間変化可能な特性であるともいえる。

特性とは「属性 (Attribute, (以下, A)) が特定の属性値 (Attribute Value, (以下, V)) であること」を表す性質をいい、オントロジー的には、物 (対象物) やプロセスなどが bearer となり、それらに内在する性質を指す。例えば、特性(P)「高温」は、 \langle 温度(A), 高(V) \rangle の組を概念化したものである。2.3.2.3 節で述べたように、特性と属性値とは別の概念であり、例えば特性(P)「高血圧」は、「血圧が高い」の属性値(V)「高」と区別される [Mizoguchi 10]。属性値「高」は他に温度、圧力、密度など複数の属性に用いられるが、「高血圧」は、温度には用いることができず、常に \langle 圧力 (A), 高 (V) \rangle という組で捉えられるべき概念である。

そこで、本研究では、異常状態を特性 (P) と対応付けて概念定義することとし、 \langle 特性 (P), 特性値 (Property Value, (以下, Vp)) \rangle (例えば、 \langle 狭窄 (P), true (Vp) \rangle) の

ように表現することとした。特性値については、ありなしを「true/false」の真理値を用いて表現することとした。false 値については、ある患者が不顕性糖尿病の場合、今は「高血糖では無い」という状態のように、「無い」ことを明示的に記述する際に用いられる。また、狭窄の程度が重度である場合など、その状態の程度を表す必要がある場合は、〈狭窄，重度〉のように、特性値に程度表現を用いてもよいこととした。

ここで、本研究が対象とする医療ドメインでは、異常状態は様々な種類の表現を用いて記述されている。例えば、臨床医が各患者を観測した結果の症状として診療記録には記録される。また、検査データでは患者から採取した生体物質の血中濃度が数値で示される。さらに、薬剤添付文書には、使用上の注意として、例えば「腎不全（無尿等）を起こすおそれがある」というように異常状態が記述される。加えて教科書では、各疾患を理解する上で必要とする異常状態が記述されている。このように、情報媒体の利用目的に応じて異常状態の表現も多様化していることがわかる。そこで、それらを、上位オントロジー YAMATO に従って概念化すると、大きく、

(1) 定量的表現（例：血圧 180 mmHg）

(2) 定性的表現（例：血圧が高い）

(3) 特性表現（例：高血圧）

の3つに分類できることがわかった。

検査データの場合、通常、血糖値 140 mg/dl のような定量データで示される事が多く、診断においては、個々のインスタンスレベルにおける患者の状態の具体的な値を知ることがタスクとなるため、定量的表現が必須といえる。

一方、多くの教科書や辞書の場合、通常、疾患定義は異常状態によって記述されている。例えば、急性心筋梗塞症は「何らかの原因により、心筋虚血を生じ、不可逆的な心筋壊死（梗塞）に陥った状態である」[小川 01] という記述があり、糖尿病の分類と診断基準による委員会報告 [清野 12] では、「糖尿病はインスリン不足による慢性高血糖を主徴とし…」とされている。このように、疾患は、ある特定の異常状態があるかどうかで定義されるといえる。したがって、医療において疾患記述には、定量的表現を必要とせず、「心筋壊死あり」や「高血糖あり」のように特定の異常状態があるかどうかを重要視され、つまり特性として捉えることが本質的であるといえる。これは、他のドメインでも同様に、例えば、機械の故障を考えた場合、数値ではなく、「管のつまりあり」のように、特定の異常状態があるかどうかということに注目される。

しかしながら、臨床で起こっている異常状態を概念化することを考えた場合、特性のままでは概念レベルが高いため、具体的な定量的情報との関連が暗黙的になってしまう恐れがあり、また、検査データとの相互運用も困難であると考えられる。そこで、

特性を、対応する属性とその値を組み合わせたタプル<属性 (A), 属性値 (V)>に分解することとした。属性値は、定量値 (Quantitative Value, (以下, Vqt)), 定性値 (Qualitative Value, (以下, Vql))またはカテゴリ値 (Categorical Value, (以下, Vcat))のいずれかをとることが可能である。例えば狭窄 (P)は、定量値であれば<断面積 (A), 5mm^2 (Vqt)>, 定性値であれば<断面積 (A), 小 (Vql)>のように分解される。カテゴリ値とは、胎生、卵生のように順序のない本質的にカテゴリカルな値をいい、例えば妄想 (P)については、「妄想があること」のみを表すときは<妄想 (P), true (Vp)>と表現されるが、さらに嫉妬妄想 (P)のように「どのような種類の妄想があるのか」を明示的に表したい場合、<妄想性 (A), 嫉妬妄想 (Vcat)>のように分解可能である。

また、臨床医学の特徴として、対象が人体であるという性質上、悪心や変形のように具体値の把握が困難な場合が見受けられる。そのような場合については、特性を無理に属性分解せず、特性表現をそのままを用いることとした。

ところで、体内で起こっている状態は細胞、組織、器官レベルまで粒度も多様でかつ膨大であり、すべての正常状態を記述することは不可能といえる。そのため、異常状態は正常状態が分かって初めて理解できるものであると考え、異常状態を定義できないのではないかと疑念が生じる。しかし、本研究では疾患定義における異常状態に着目し、疾患を定義するために必要な状態をモデル化するという方針においては、正常状態の定義はなくとも、臨床医が判断し定義される疾患において現れる状態を「異常状態」として定義することが可能であると考え。また、人体に起こったある状態が人体に害を与えるという意味で「異常」であるかどうかを判断することについてであるが、この問いはオントロジー工学研究の範囲外であると考え。例えば、いわゆる善玉として知られている HDL コレステロールのレベルが高い状態について異常かどうか判断することは医療の専門家が専門知識に基づいて決定すべきタスクであると考えている。

3.4 異常状態の特性・属性分解

3.4.1 基本表現

属性は“あるモノが持つ性質”を表すので、その性質を持つ対象となるオブジェクトすなわち対象物 ((Object, (以下, O))を特定する必要がある。人体で起こった異常状態の場合、その異常が人体のどの部位を対象としているのかという対象物を特定する必要がある。例えば「動脈硬化」の場合は、対象が動脈であることを同定しなければならない。そこで、特性<属性, 属性値>に加え、対象物(O)を導入し、異常状態を OAV トリプル<対象物 (O), 属性 (A), 属性値 (V)>に分解することとした。これを、本異常状態表現モデルでは、基本型と呼ぶ (図 3.1 左上)。また、悪心のように本来的

に特性で表現すべきものや、現代の医学では分解困難なものは、対象物のみ同定することとし、これを特殊型と呼ぶこととした (図 3.1 右上).

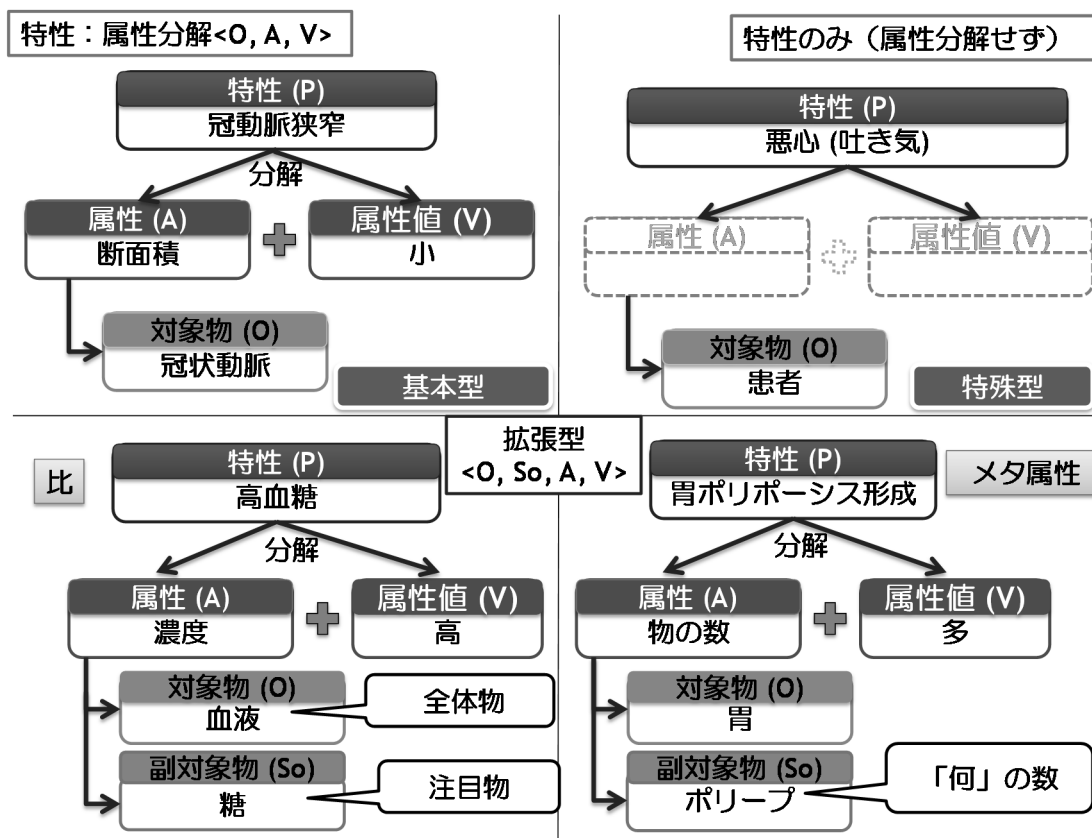


図 3.1 異常状態の表現例

3.4.2 拡張表現

一方、特性の中には、基本の OAV トリプル分解表現では記述が困難な場合がある。例えば、高血糖の場合、単純な対象物 (O) だけでは、血液と対象物とすべきか、糖を対象物と扱うべきか判断に迷うことが生じる。そこで、対象物 (O) 以外に、注目物を表す「副対象物 (Sub-Object, (以下, So))」を導入し、特性を<対象物 (O), 副対象物 (So), 属性 (A), 属性値 (V)>に分解して表現することとした。これを拡張型という (図 3.1 下段)。

3.4.2.1 比

比を表現する場合、その比をもつ対象物を特定した上で、さらに何に注目して比較を行っているかを表現しなければならない。例えば、「高血糖」の場合は、濃度は「血

液」のもつ性質であるため、血圧と同様に対象物 (O) は血液といえる。しかし、濃度とは、一般に、全体物に対する注目物の割合を示すため、血液に加え、注目物である「糖 (グルコース)」も表現する必要がある。

そこで、副対象物(So)を用いて注目物を定義することで、比を表現することとした。

比については、何を注目するかによっていくつかの種類に分けられ、それに応じて、何を副対象物とするかも変わる(表 3.1)。以下、その詳細を述べる。

表 3.1 比の表現

No.	比の分類	異常状態 (P)	対象物 (O)	副対象物 (So)	属性 (A)	属性値 (V)	比
1	m/n (注目物/全体物) 例	high m ratio 高血糖	全体物 血液	注目物 糖	比率 濃度	高 高	m/n 糖/血液
2-1	m/n (mに注目) 例	high m ratio 高UACR	対象物 尿	m アルブミン	比率 濃度	高 高	m/n アルブミン/クレアチニン
2-2	m/n (比自体に注目) 例	high m/n ratio A/G 比増加	全体物 血液	m/n A/G	比率 比率	高 高	m/n アルブミン/グロブリン

1. m/n 比 (無単位) : 全体物 (混合物) (n) に対する部分 (m) の割合による表現

全体物 (混合物) を n , 全体物に含まれる注目物を m とした場合、全体物に対する注目物の割合の表現について、比をもつ全体物 (n) を対象物 (O) とし、「何の比」であるかを示す注目物 (m) を 副対象物 (So) とする。例えば、血糖濃度は、血液中に占める糖 (グルコース) の割合を言うので、全体物は血液、注目物は糖といえる。したがって、高血糖は、 \langle 血液 (O) , 糖 (So), 濃度 (A), 高 (V) \rangle の 4 つ組で表現され、対象物は血液、副対象物は注目物である糖となる (表 3.1, 1)。

2. m/n 比 (同一対象物内) : 同一対象物のもつある属性 (m) に対する別の属性 (n) の割合による表現

2-1 m/n 比 (m に注目) : 同一対象物のもつある属性 (n) に対する注目属性 (m) の割合による表現

同一対象物内の属性について、ある属性 (n) に対し、属性 (m) に注目した割合を表現する場合、m を副対象物として定義する。例えば、アルブミンは、健常時では尿中に認められず、腎疾患患者でのみ尿中に存在することが知られている。ところが、その量は微量のため、全体物である尿と比較が困難である。したがって、尿中のクレアチニン濃度に対するアルブミンの濃度の割合 (尿中アルブミン/クレアチニン比 (UACR)) を測定することがある。この場合、アルブミン濃度およびクレアチニン濃度のいずれの「濃度」も尿がもつ属性であるので、尿が対象物であり、注目すべきアルブミンを副対象物に定義すればよい。したがって、UACR が高い場合、本モデルで

は<尿(O), アルブミン (So), 濃度 (A), 高 (V)>で表現される (表 3.1 2-1). なお, クレアチニンについては, 尿中のアルブミンが微量であるため, 補正するために用いられるものであり, 3.6.2 節で示す修飾語として表現される.

2-2 m/n 比 (比そのものに注目)

次に, 同一対象物内における属性 m と n の二つの比について, 一方に注目するのではなく, 比そのものに注目する場合がある. 例えば, アルブミン/グロブリン比 (A/G 比) は, 血中のグロブリン濃度に対するアルブミン濃度の比をいうので, 対象物は全体物である血液といえる. また, 注目すべきはアルブミンでもグロブリンでもなく, それらの比そのものであるため, A/G 比減少の場合, <血液 (O), A/G (So), 比率 (A), 低 (V)>と表現される (表 3.1 2-2). A/G については, 理論上は割合なので, 属性(A)とも考えられるが, 本研究では, 属性はなるべく一般的な物理パラメータを用いることとし, 副対象物に注目物としての比をいれて一貫性を保持することとしている.

3.4.2.2 メタ属性

「胃ポリープ形成」では, ポリープの色や大きさは, 本来ポリープ自体が持つ性質であるが, ポリープの数の多さについては, ポリープ自体が本来持つ性質ではない. ここで, YAMATO では, メタ属性という概念を導入しており, 例えば, 「カーブが多い道路 (The road is curvy)」において, カーブの数の観点から表現し, 道路についてカーブの数が多いこと (curvy) を道路の性質 (メタ属性)とみなす記述が可能となる. これと同様に, 臨床的には, 胃に生成したポリープの数の多さに対し, メタ属性を用いることで胃の異常状態として表現することができる. ただし, 何の有無, あるいは「何の」個数であるか, (この例の場合であれば, 「ポリープの」ということ)を示す必要が生じる. そこで, 個数や回数の場合も, 「副対象物 (So)」を用いて, 「何の」個数や回数であるかを概念化することとした. その結果, 胃に多数のポリープが形成された状態である「胃ポリポーシス形成」は, <胃 (O), ポリープ (So), 物の数 (A), 多 (V)>と分解して表現することが可能である.

このように, 拡張型表現を導入することで, 比やメタ属性などの複雑な表現についてもその対象物 (O) を明確に分離しつつ, 副対象物 (So) を用いて必要な情報を明示することで, 適切に概念化することが可能となる. これにより, 異常状態表現モデルに柔軟性が生まれ, かつ一貫性を保持したまま様々な異常状態を表現できることがわかる.

これまで, 3 診療科, 12 疾患を構成する異常状態とその上位概念を含めた計 107 個の異常状態について本異常状態表現モデルを適用可能か予備的に検討した結果, 統一的一貫性のある表現ですべて記述可能であることを確認した.

3.5 観測データとの相互運用性

前節までは、疾患記述における異常状態表現を中心に議論した。一方、各医療機関には個々の患者の膨大な検査データが蓄積されており、実用性の面においてはこれらとの相互運用性を確保する必要がある。

血液検査等、各種臨床検査において観測データは通常、動脈断面積 20 mm^2 のような定量データで示される。これは、3.3 節で述べた定量値(Vqt)を用いて、動脈を対象物 (O) とし、断面積(A)が 20 mm^2 (Vqt) という定量値をもつという OAVqt 形式で表現される。定量値は、各医療機関で用いられる一般的な閾値をもとに変換すれば、定性値に変換できる。

一方、異常状態「動脈狭窄」について、特性「狭窄」も「断面積という属性が小さい」という属性とその定性値に分解しておけば、〈動脈 (O) , 断面積 (A) , 小 (Vql)〉という〈OAVql〉形式を介して、定量データと異常状態間の相互運用が可能となる。

別の例として、「血糖値 260 mg/dl 」というデータについても同様に、血液を対象 (O) とし、糖 (So) の濃度 (A) が 260 mg/dl (Vqt) という定量値をもつという〈OSoAV〉形式で表現し、さらに「血液 (O) の糖 (So) 濃度 (A) が高い (Vql)」という定性変換を介して、データと異常状態「高血糖」と相互運用が可能となる。

ここで、各検査で設けている標準的な閾値から大幅に超えた定量値がデータ観測され、データから異常状態に変換する際、「重度の高血糖」のように、異常状態の程度を記録しておくことが疾患の同定において必要な場合がある。その場合は、特性値(Vp)を設定しておくことにより、severe のような程度値を記述できる。

次に、特性値について、特性は、本来は〈高血糖, あり〉あるいは〈高血糖, 真〉という意味になり、あり/なしや true/false は高血糖の程度とは独立した概念である。しかし、別の記法を導入することは複雑さを増すため、高血糖ありの場合の詳細化として mild/moderate/severe といった程度値があると捉えることによって、あり/なしや true/false を程度値と同様に扱うこととした。また、特性値 (Vp) について、3.3 節で述べたように、例えば、ある患者が不顕性糖尿病の場合、今は「高血糖では無い」という状態を表す必要もあるため、〈高血糖, false〉も用意している。なお、程度値については、あまりに多様な表現を許すと、散逸し統一性にも欠けるため、mild/moderate/severe のような最小限の表現にとどめておくことが望ましい。どこまで閾値を超えると severe にするかなどは各学会で用いられるガイドライン等の標準的なものを使うと相互運用性が高まると考えられるが、時代によって具体的な閾値は変化する場合もあるので、本オントロジーでは具体値を設けていない。

以上をまとめると、本モデルでは、臨床上あり得る表現 3 種類を全てサポートしている。まず、計測データを扱うためには、定量値 (Vqt) を用いた〈O, A, Vqt〉で表現が必要となる。次に、定量値について、閾値をもとに定性値 (Vql) に変換すること

で、定性的表現に変換される〈O, A, Vql〉. そして、その属性 (A) と定性値 (Vql) を特性 (P) に取り込み、特性化することで、異常状態〈O, P, Vp〉の表現に互換可能となる結果、疾患記述における異常状態としての「冠動脈狭窄」あり/なしで議論することができることになる. したがって、本モデルでは、データからそのデータのもつ「内容」としての異常状態知識への変換処理を可能にする. なお、比などの拡張型を用いた記述形式についても同様に定量表現から定性表現を介した特性への変換が可能である.

このように、本研究の提案する統一表現モデルでは、

- 1) 計測データとしての 〈OAV〉
- 2) 異常状態表現としての 〈OPVp〉
- 3) さらに、拡張形式として計測データ〈OSoAV〉とその異常状態表現としての 〈OSoPVp〉

という相互運用可能な記述形式を得た.

OPVp 表現は、結果的に OAV と同型になっているので、統一形式による情報の処理がしやすく、「重度高血糖」のような程度表現を含む異常状態を OAV に変換することも可能としている.

なお、本研究では、定性化の際、閾値を自由に設定することが可能であり、また、そうすることが本質的であると考えている. 定性化に用いられる閾値は、時代とともに変化する場合や、人種、性別、あるいは医療機関によって異なる場合がある. 例えば、血糖値の場合、WHO では、高血糖の判断基準となる空腹時血糖値の閾値は、140 mg/dl(1980) から 126mg/dl(1999) に改訂されている [WHO 99]. しかし、どのような閾値を与えられようと、その閾値より高い (もしくは、低い) という意味は、特性を定義する上で変わらない. 高血糖は、閾値が変化しても、血液中の糖の濃度が「高い」という意味において不変である. さらに、その「高い」という値を具体的に同定するのは、インスタンスレベルでの診断タスクであるといえる.

本研究では、異常状態は単一の異常状態を対象としている. しかし、異常状態が複数同時に起こる場合も想定され、それにより閾値は変化する可能性がある. 例として、高血圧と高血糖が同時に起こり、各々の閾値が変化したとする. その場合、閾値として新しい値を設定することが可能である. しかし、たとえ閾値は変化したとしても、「高い」という意味は、特性の定義では、両者において不変といえる. なお、1つの異常状態について、複数属性をもつ場合の対処については、後述の 3.6.3.3 節で述べる.

3.6 理論の実践的適用

本節では、これまで述べたオントロジー工学的手法に基づいた異常状態表現の記述枠組みに基づいて、臨床で用いられている実際の各種表現へ適用する。

3.6.1 データ表現のバリエーション

臨床用語には血液と圧力が組み合わさった「血圧」や「白血球数」のような複合概念が数多く存在することを確認している。これらは、オントロジー工学的には、正しくは、「圧力」、「物の数」を使った表現が適切と考えられる。しかし、これらはオントロジーとは別に、データ表現のバリエーションとして、臨床データとの相互運用性の向上を目指すためにも、今後検討すべき課題と考えている。なお、現在までに、対象物+属性複合型の OA 型 (例：血圧)、副対象物+属性付加型の SoA 型 (例：白血球数)、および対象物+副対象物+属性付加型の OSoA 型 (例：血糖値)の3種類の複合表現 (複合属性)があることを見出している (図 3.2)。

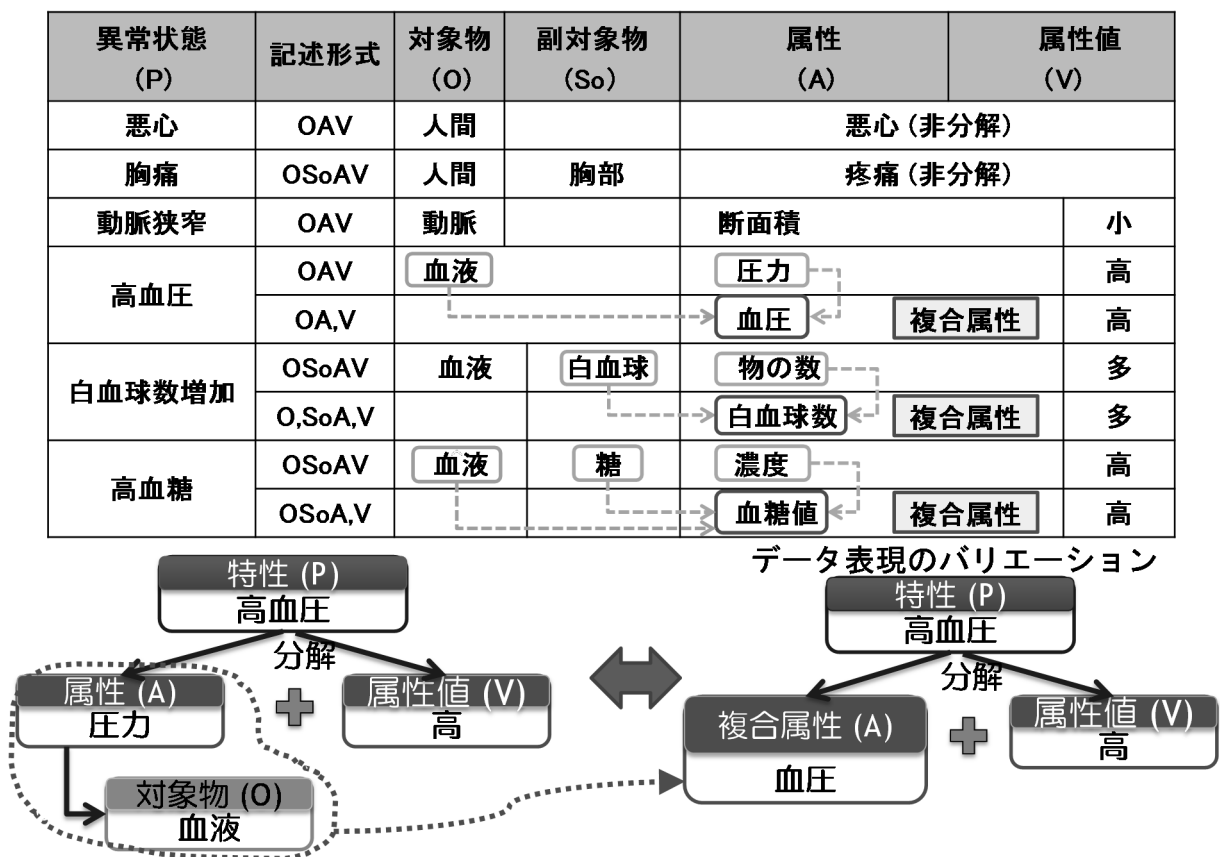


図 3.2 データ表現のバリエーション

3.6.2 修飾語の扱い

臨床の場では、診療記録等において、より患者の病態を詳しく説明できるように、「一過性高血圧」、「妊娠期高血圧」といったような修飾語が付加された異常状態が存在する。オントロジー理論で考えると、これらはいずれも「高血圧」という概念で定義されるべきといえる。しかし、実用上の観点からは、修飾語付きの用語にも柔軟に対応できることが望ましい。そこで、これら修飾語についてはオントロジーにおける「高血圧」とは別に扱い、付加情報として修飾語に関する情報を属性として別途保持しておき、必要に応じて修飾語と異常状態から自動生成して修飾語付きの異常状態概念「一過性高血圧」等を生成できるよう準備を進めている。これらの異常状態については、記述方針を示したガイドラインを作成しており、それに従って、全疾患における異常状態表現を整理した(表3.2)。

表 3.2 修飾語の表現例

異常状態 (P)	対象物 (O)	副対象物 (So)	属性 (A)	属性値 (V)	修飾属性	修飾値	被修飾物
高血圧 【一過性高血圧】	血液		圧力	高	状態持続時間	一過性	高血圧
食道粘膜黒色化領域 【食道を首座とし、胃食道接合部に明瞭な境界を有する、全周性の黒色調粘膜】	食道	粘膜	色調	黒	病変範囲(領域)	全周性	粘膜
					境界性	明瞭	粘膜
					隣接区域	胃食道接合部	粘膜

【】: 修飾語付きの異常状態を示す

今後、洗練作業中の7診療科の各疾患の異常状態について表現モデルに基づいた記述を進めていく。

3.6.3 異常状態の構成要素に関する指針

本節では、提案モデルを臨床医が記述している各診療科の疾患の異常状態に適用し、さらに統一的に記述する過程を支援するために以下の指針を提案する。

3.6.3.1 対象物(O)の同定

人体構造は様々な粒度の部品から構成される。そのため、何に注目するかにおいて異常状態の対象の同定が困難なことがある。そこで、対象物の選択基準について、「何

を対象物とするかについては、その属性を持つと見なし得る複数の候補の中で、解剖学オントロジーで定義されているもののうち、粒度の一番小さいものを使う」という方針とする。例えば、「冠動脈狭窄」の場合、心臓の疾患を定義しているコンテキストで、かつ、そこに現れる異常状態を定義したいと考える。そこで表現したい対象は冠動脈であって、心臓でもなければ、患者でもない。一方、心不全の場合、機能を持つ対象としては、心臓全体を考えるので、この場合の対象は「心臓」となる (図 3.3)。

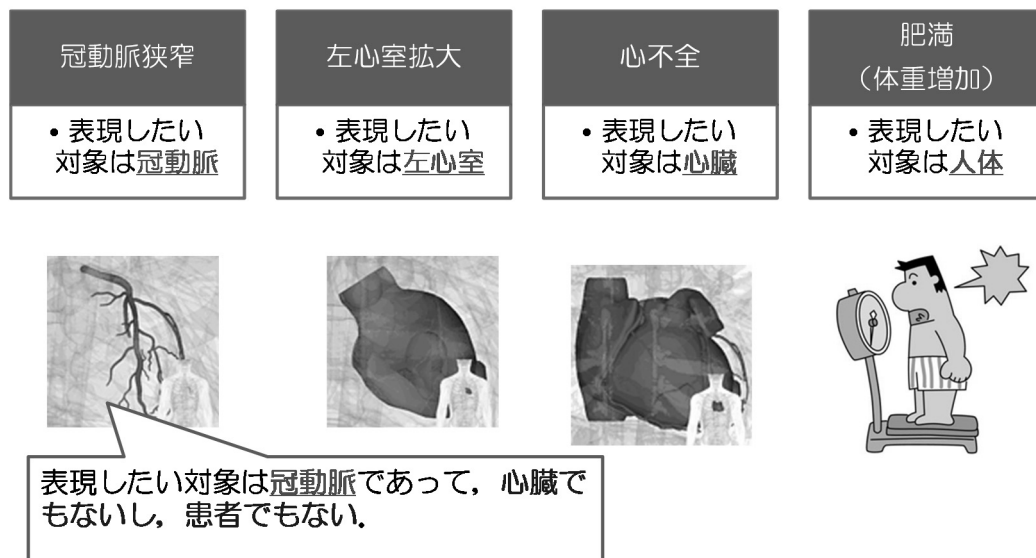


図 3.3 対象物の同定

次に、3.4.2.1 節で述べた高血糖の例を考える。粒度の一番小さいものを対象とする方針ならば、血糖値は、「糖分が多い」のだから、糖が対象物ではないかともいえる。しかし、この場合は、血液を対象物とする。その理由として、確かに糖だけに注目すれば、糖 (O) の量 (A) が多い (V) といえるかもしれない。しかし、道路にあるカーブの数が多いことは、カーブの性質ではないことと同様、「糖の量が多い」ことは糖の性質ではない。それは「糖の塊」の性質である。更に例を挙げると、机の上にリンゴがたくさんある場合、「リンゴの量が多い」はリンゴの性質ではない。そして、血糖値の場合、「糖分の多さ」について考えるときには、「血液」という全体物があつた上で、注目した糖の割合の多さを考えねばならないことを考えると、血液が対象物であることが分かる。つまり、血液を、「糖濃度 (糖の割合)」の観点について考え、そしてその観点で属性を見直すと、糖 (So) と濃度 (A) に分離される。したがって、高血糖は、 \langle 血液 (O), 糖 (So), 濃度 (A), 高 (V) \rangle と表現される。別の例として、表現したい対象「胃壁ヘリコバクターピロリ菌付着」についても、粒度の小さいものであれば、ヘリコバクターピロリ菌が一番小さくみえるので、ヘリコバクターピロリ菌が対象物ではとも考えられる。しかし、胃壁ヘリコバクターピロリ菌付着状態の場合、菌の数

の多さについて考えるとき、「胃壁」があった上で、はじめて菌の数え上げができるといえる。したがって、胃壁が対象物となる。次に、「菌の数」の観点について考え、そして属性を見直すと、ヘリコバクターピロリ菌 (So) と物の数 (A) に分離される。したがって、「胃壁ヘリコバクターピロリ菌付着」は、<胃壁 (O), ヘリコバクターピロリ菌 (So), 物の数 (A), 多 (V)>と記述される。

3.6.3.2 副対象物(O)の同定

3.4.2 節で述べたように、本モデルでは、拡張表現の際、副対象物(So)を導入することとしている。そこで、本節ではどのように副対象物を使えばよいかについてその考え方を述べる。異常状態については、まず、何 (O) が、何の観点 (A) で、どのようにおかしいか (V) を記述が必要となる。例えば、低身長という場合、患者(O)の身長という観点 (A) で、値が低い (V) という記述を行う。しかし、以下に述べるように対象物 (O) と副対象物 (So) の導入については、慎重に対応すべきである。例えば、健常人と比べて脚の長さという観点で記述したいとする。その場合、まず、副対象物と属性の両方を一つにして考え、患者の「脚長」のように、両方を1つにまとめたものを属性(A)とみなす。そして、属性(A)を見直して、副対象物 (So) と属性 (A) に改めて分離する。すると、この場合、<人体 (O), 脚 (So), 長さ (A), 短 (V)>に分離される。一方、患者の異常と考えず、脚短縮を脚自体の異常として考えたい場合、脚が対象物となり、<脚(O), 長さ(A), 短(V)>となる。特別に前者であることが要請されていない場合は、3.6.3.1 節で述べた決定法に従って、後者が推奨される。

3.6.3.3 観点(A)の同定

1) 複数の属性がある場合の扱い

ある対象物の異常状態を表現する場合、観点に対応するのが属性(A)といえる。例えば、ある皮膚の異常 (皮膚発赤)を考えた場合、何の観点 (A) かについては、色の観点だといえるので、属性は色調 (A) とし、その値が赤 (V) を持つといえる <色調 (A), 赤 (V)>。ここで、異常状態の中では、複数の属性をもつものが存在する。例えば、網状皮斑という異常状態の場合、皮膚の形状 (A) という観点では斑 (V) という値をもつが、これに加え、外見 (A) では網目状 (V) という値を持つといえる。そこで、本異常状態のように、複数の観点を持つ場合、それらの OAV の組み合わせで表現記述を行うこととした (表 3.3)。

表 3.3 属性の同定例

異常状態 (P)	対象物 (O)	副対象物 (So)	属性 (A)	属性値 (V)
皮膚発赤	皮膚		色調	赤
網状皮斑_1	皮膚		形状	斑
網状皮斑_2	皮膚		外見	網目状

2) 心拍数についての留意事項

回数について考えた場合、①一回完結型 (例：海外出張する) ②繰り返しをもって1つと数えるもの (例：振動数)の2種類がある。

① 一回完結型

頻尿の場合は、一回単位で数えるので、OAV 分解：＜患者 (O), 排尿 (So), 回数 (A), 多 (V)＞となり、複合属性表現では、＜患者 (O), 排尿回数 (SoA), 多 (V)＞となる。

② 繰り返し型

光や音では、振動が本質といえ、「振動数」で1つと考えるべきものといえる。また、波長も周期性、繰り返しをもって1つと考えるので、上述のような一回完結型とは区別すべきと考える。これを統一表現モデルの記述枠組みで表すと、＜対象物(O), 振動数 (A), 多 (V)＞のほうが、＜対象物 (O), 振動する (So), 回数 (A), 多 (V)＞より自然であり、物理で通常使われている概念とも対応しているといえる。そこで、心拍数について考えた場合、心拍数は、1回単位で数えるというより、拍動する繰り返し数をまとめて1つの属性と考えるのが自然であり、周波数や振動数に近い存在といえる。したがって、心拍数増加の通常の一表現では、拍動数 \div 周波数/振動数ととらえて、＜心臓 (O), 拍動数 (A), 多 (V)＞とすべきであると考えられる。

脈拍数も同様に、1回単位で考えるより、心臓の拍動由来の拡張の繰り返し数をまとめて1つの属性として考える。なお、脈拍数増加の場合は、心臓の拍動に基づいて流入血液量が増えた際に動脈が拡張している現象であって、自ら運動はしていないといえる。そこで、属性は拍動数ではなく、＜動脈 (O), 拡張数 (A), 多 (V)＞とする。

3) 行為について

行為についても、表現が難しい。原則として、行為やプロセスを対象物 (O)とすることはなく、異常は行為 (プロセス)そのものに帰することはなく、行為者に帰すべきものであるからである。患者自体の異常として表現する場合は、対象物 (O) は患者となる。次に、何の観点かを考えた場合、複数の観点が考えられる。

- a) 行為回数の観点で記述したい場合、まず、行為回数を属性とみなし、さらに、行為 (So) と回数 (A) とに分離される。

例) 徘徊行為異常<患者 (O), 徘徊行為 (So), 回数 (A), 多 (V)>

- b) 物理的パラメータの観点で記述したい場合, 一般的な物理パラメータを属性 (A)とする.

例) 歩行速度低下異常<患者 (O), 歩行 (So), 速度 (A), 遅 (V)>

- c) 行為の output の異常をいたい場合, output を対象物とする.

例) 便の色の観点なら, 便白色異常<排出便 (O), 色 (A), 白 (V)>

例) 分泌量が少ないなら, インスリン分泌量低下<分泌物 (インスリン) (O), 量(A), 小 (V)>

3.6.3.4 状態変化の扱い

本研究では, 異常状態は, 基本は結果(変化後の状態)として記述する. 疾患を理解する上で, 例えば高血圧の場合, 血圧が連続的に上昇している状態 (動的な状態)をいっているわけではなく, その結果として, 「血液の圧力が高い」状態が起きているということが重要である. そのため, 変化後の「起こった状態」として異常状態を記述することとした.

例1) LDL 酸化

例えば, 低比重リポ蛋白(Low-Density Lipoprotein (LDL)) の酸化を考えた場合, 本研究では, 変化後の「酸化された状態」として記述する. また, その結果, 血中に酸化型 LDL コレステロールの濃度が増加した場合, 別途, 物理的パラメータ「濃度」(A) を用いて「血中酸化型 LDL コレステロール濃度増加」を記述する (表 3.4).

表 3.4 LDL 酸化の記述例

異常状態 上位	異常状態 (P)	対象物 (O)	副対象物 (So)	属性 (A)	属性値 (V)
酸化	LDL酸化	酸化型LDL		量	大
高濃度/濃度増加	血中酸化型 LDLコレステロール 濃度増加	血液	酸化型LDL コレステロール	濃度	高

例2) 塞栓

塞栓では, 栓子によって (既に)閉塞した状態のことを表現する. 栓子の移動 (流入)を記述したい場合, 別に栓子移動 (流入)を記述する. さらに, 栓子が血栓から剥離したことを記述したい場合, 「血栓から栓子剥離」について記述する (表 3.5). なお, 栓子については, コレステロールなど具体的に同定できる場合, その物質を記述する.

表 3.5 塞栓の記述例

異常状態 上位	異常状態 (P)	対象物 (O)	副対象物 (So)	属性 (A)	属性値 (V)	修飾属性	修飾値	被修飾物
閉塞	冠動脈塞栓	冠動脈				閉塞物	栓子	異常状態
異物付加 【物質】	冠動脈内栓子 流入	冠動脈	栓子	流入量	大			
剥離	血栓から栓子 剥離	血栓				剥離物	栓子	

3.7 関連研究との比較を通じた統一表現モデルの有用性

3.7.1 関連研究との比較

まず、PATO、HPO はともに上位オントロジー BFO を採用形式として採用しており、特性形式で表現される。また、SNOMED-CT も一つの異常状態について、finding site という我々の対象物に対する概念をもっているのので、我々の特性表現と同様の形式であるという点において、類似している。

ここで、これらと本研究の表現形式の相違について、PATO、HPO、SNOMED-CT は属性表現をもたないが、本オントロジーは特性(P)を属性(A)と属性値(V)に分解した形式をもつ点で異なる。また、PATO、HPO では、対象物を同定していないが、本研究では3.4.1節で述べたように、対象物(O)を記述する枠組みを基本表現として備えている。例えば、本研究では冠動脈狭窄を<冠動脈(O)、断面積(A)、小(V)>のように分解可能であるが、PATO、HPO は分解がなされず、「狭窄」、「冠動脈狭窄」として表現される。

次に、2章で述べたように、LOINC については、臨床所見データを表現するため、我々の O(So)A に対応する形式をもつ。ただし、LOINC は属性値 V をもたない(例: 高い)。

ここで、本研究の表現モデルは、関連研究から単独の表現形式の内の一つを採用しようとしているのではなく、存在する各々の表現形式をオントロジー理論に基づいて定式化・組織化しつつ、相互の運用可能性を実現している。それは、PATO、HPO、LOINC などの既存のリソースを単に採用すれば済む問題では無い。各表現形式間の相互運用性を担保する統合的な理論的考察があつて、初めて異常状態のコンピュータ処理可能なモデルが確立できたと考える。

3.7.2 実用面における本表現モデルの優位性

本節では、実用的観点からみた本表現モデルの優位性について考察する。まず、HPO, PATO, SNOMED-CT は、OAV 分解表現をもたないので、データとの相互運用性に欠ける。臨床の検査データとの表現と相互運用性をかなえるには、3.5 節で示したように、特性形式だけではなく、定量値 (Vqt) を表現しうる属性表現としての OAV 形式との互換を必要とする。次に、LOINC はデータ間の相互運用は、OSoA 形式で統一されている点において優れている。しかし、臨床の検査データとの表現と異常状態との相互運用性をかなえるには、定量値 (Vqt) まで持つ必要がある。本オントロジーでは定量データも OAV 形式で扱えるので、異常状態の特性表現との互換が可能であり、臨床において、電子カルテシステムなどで臨床検査から疾患における異常状態まで記述できる表現能力をもつ。これらは単なる理論的な成果にとどまる物では無い。既存のリソースの再利用では不可能な異常状態の統一的な取り扱いが初めて可能になるという実用的な貢献をする。

以上より、本異常状態表現モデルは、疾患記述における異常状態を、まず、特性で捉えるという特徴をもつ。特性の利点として、細かいパラメータ変動に影響をうけないことがあげられる。また、抽象度を上げることで、診断タスクと明確に分離ができるため、疾患の原因をとらえるような状況下では、特性のままの表現が適切であると考えられる。これまで述べてきたように、本研究では特性を属性分解することで、臨床データから疾患記述における異常状態まで一貫した統一的表現で記述することが可能となり、相互運用性が確保できる点において有用であり、臨床での診療記録の統合管理に貢献すると期待される。

3.8 結言

本章では、異常状態オントロジーを構築する際に基盤とする異常状態の統一表現モデルについて述べた。特に、多様な人体の異常状態を統一的な視点で捉えるとともに、計測データ等との相互運用可能な特性属性分解モデルについて述べた。

次章では、異常状態統一表現モデルをもとに、疾患記述に必要な異常状態の組織化を目指した異常状態異常状態オントロジー3層構造による階層的モデリングを提案する。

第4章 疾患記述における異常状態の階層的モデリング

4.1 緒言

本章では、本研究の目的 2) 疾患記述における異常状態の階層的モデリングについて、3 章で提案した異常状態の統一表現モデルをもとに記述された異常状態について、分類階層を構築する。まず、疾患を記述するための異常状態の分類の必要性について検討を行う。次に、3 層構造による階層的モデリングを提案し、具体的な適用例を説明する。本モデルは診療科横断的に共通性を把握できることを特徴とし、その有用性について考察するとともに、関連研究の比較を通して異常状態オントロジー構築の意義について述べる。

4.2 疾患記述における異常状態分類の必要性

知識を体系化するには、その対象世界がどのような概念に依存しているか、概念間の関係性を明示するとともに、必要な概念知識を収集し、組織化することが必要となる。臨床医学における異常状態概念の場合、各診療科において専門医が特定の疾患を定義する上で必要な状態、すなわち疾患コンテキストに依存した異常状態を適切に概念化することが求められる。一方、臨床医学知識全般を考えれば、各疾患コンテキストに特化した概念だけではなく、診療科を超えた複数の疾患に共通する汎用的な異常状態を用意する必要がある。

また、知識を組織化する際、今後電子カルテシステム等の各種システムとの情報の共有化を考慮すると、領域間の知識の相互運用性の面においても、各専門医の固有の見方から中立な立場で診療科横断的に客観的に捉える見方で体系化していくことが重要であるといえる。

さらに、疾患依存、疾患非依存の 2 種類の知識に加え、臨床における異常状態は、通常、人体構造に依存して現れるので、人体構造物への依存関係も明示しなければならないといえる。

そこで、本研究では、異常状態オントロジーのモデリングとして、3層による構造化を行うこととした (図 4.1).

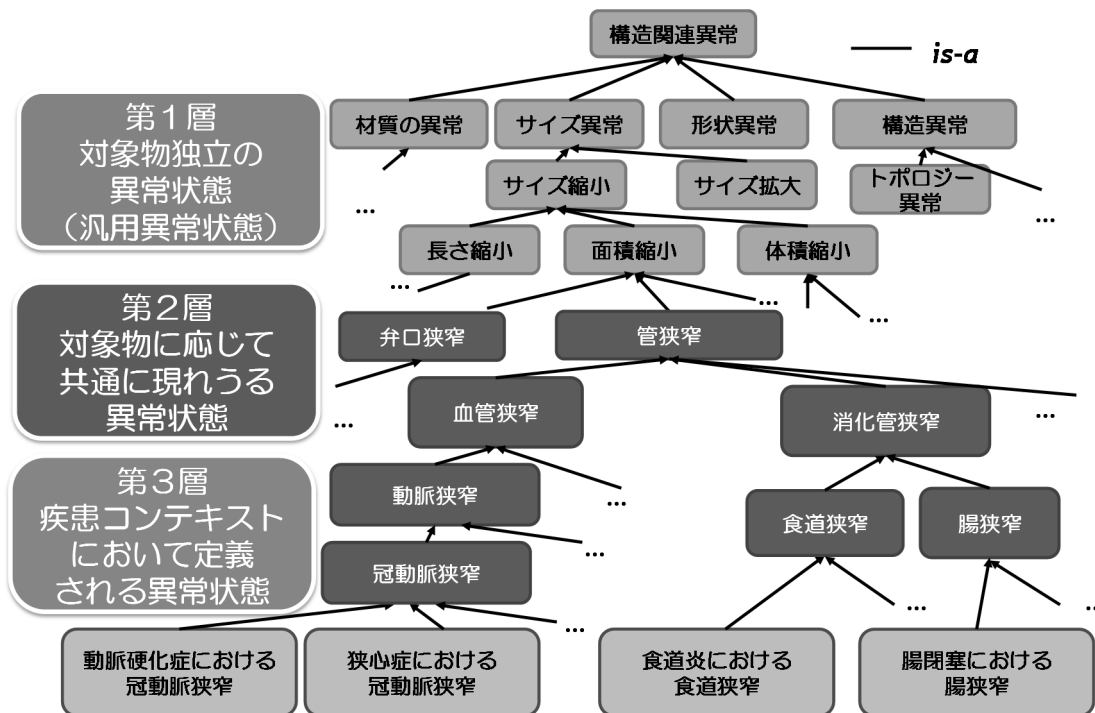


図 4.1 異常状態の3層構造による分類階層

4.3 ドメイン横断的な一貫性を指向した異常状態の is-a 階層構築

4.3.1 異常状態3層モデルの概要

本研究の異常状態オントロジーは、概念レベル別に3層からなる構造を特徴とし (図 4.1), オントロジー工学理論に基づいたオントロジーとしての整合性を備えつつも、臨床で使う立場も意識した構造となっている。以下、各概念階層の詳細を述べる。

4.3.2 第1層：対象物独立の異常状態 (汎用異常状態)

本節では、まず、異常状態オントロジーの最上位の構造について説明する。最上位 (第1層)の異常状態は、最も汎用的かつ一般的な異常状態 (汎用異常状態)から構成されており、対象物に依存せずに定義される (対象物非依存)。

一般に、異常状態は、構造に関するものと、機能に関するもの(機能障害)に分けることができる。臨床医学においても同様に、異常状態は、臓器や器官などの人体構造物に関する何らかの異常(器質的異常)と、機能性に関する異常の2種類に関するものが多いといえる。そこで、本異常状態オントロジーでは、異常状態概念を構造関連異常、機能関連異常、その他の異常状態に大別することとした。(図4.2)。



図 4.2 異常状態 is-a 階層における第 1 層

構造関連異常とは、構造に関する異常状態をいい、材質の異常、形状異常、サイズ異常、構造異常が下位概念として定義される。材質の異常とは、構造体の材料としての性質に何らかの異常が生じた状態として定義され、熱変性等の物理学的変性と酸化、腐食などの化学的変性、硬化/軟化、析出、凝固などが下位概念に存在する。形状異常とは、部分を限局しない全体形状としての変化に関する異常をいい、変形が例として挙げられる。サイズ異常は、構造体の大きさに関する異常状態が定義されており、長さ、面積、体積の拡大/縮小がある。構造異常とは、そのもの自体のアイデンティティ

を保ちながらもその部分構造や部品の構造に関する何らかの異常をいう。下位には、構造的欠損、異物付加や新生成物形成などの構造的過剰、トポロジー異常がある。トポロジー異常とは、位相構造に関する異常状態をいい、逆位、トランスロケーション、亀裂、癒着などの結合異常や閉鎖/開放、孤立など空間に関する異常を含む。

次に、機能関連異常とは、機能物が発揮する機能に何らかの異常が生じた状態をいい、機能亢進、機能低下、機能停止、機能不全、機能喪失が定義される。機能関連異常についても、構造関連異常と同様、様々な対象物に用いられる極めて一般性の高い各分野から参照される概念といえる。

構造、機能以外のその他異常状態として、属性値関連異常 (パラメータ異常) と、生起物 (Occurrent) 異常が定義されている。パラメータ異常は、閾値と比較して属性値が大きい状態であるパラメータ増加 (大きい) 異常と小さい状態のパラメータ低下 (小さい) 異常に分けられ、下位には、速度、重さ、圧力、濃度などの一般性のある物理的なパラメータを用いて表現される、圧力増加/減少、速度増加/減少、重さ増加/減少などが概念化されている。

なお、サイズ異常は上位概念をパラメータ異常ではなく、構造関連異常として定義されている。その理由は、長さ、面積、体積はパラメータで表現できる属性ではあるが、異常状態の本質を考えた場合、胃、血管狭窄も胃拡張も、まず構造に注目して「構造に関する何らかの異常がサイズにある」として捉えた上で、それら属性 (面積、体積) の定量値が何らかの異常を示しているというアウトプットとして観察されるといえるからである。

次に、生起物 (Occurrent) 異常では、時間軸上で異なる部分 (Temporal Part) を持つものに関する異常が定義されており、下位には行為/動作関連異常と現象関連異常が含まれる。前者はエージェントの行為やエージェントの部分動作に関する異常 (例: (ある部分の) 過収縮) を指し、また、後者は漏れや逆流のようなエージェントがいない、いわゆる現象に関する異常状態が定義される。なお、オントロジー工学的には厳密にはプロセスとイベントに分かれるが、臨床上観察される状態によっては、プロセスとイベントの識別が困難な場合が多いので、両者を識別せず、現象とする方針とした。

このように、第1層の異常状態は具体的な対象物の種類を特定せずに汎用的に定義することができるのが特徴であり、人体に限らず人工物にも共通する対象物非依存の最も基本的な概念といえ、例えば、圧力増加のように、医療分野に限らず、機械等の分野においても共通の定義として用いることが可能である。

4.3.3 第2層：対象物に応じて共通に現れうる異常状態

第2層では、対象物に応じて共通に現れうる異常状態 (対象物依存異常状態) が定義される。本定義は、概念が特定のコンテキストのもとで担う役割 (ロール) を捉える

ことで、そのコンテキストに依存した概念の定義を行うロール理論 [Kozaki 07] に基づいて概念化されている。具体的には、異常状態の対象物をコンテキストとして、その対象物に固有の異常ロールを汎用異常状態が担うことによってできる状態として、第2層の対象物依存状態が定義される。第2層の最上位では、複数分野に共通する汎用構造物を対象とした異常状態が定義されている。汎用構造物とは、汎用的に見られる共通構造を記述したもので、管状構造物、壁状構造物、弁構造物などがあげられる。人体は多数の器官から構成されているが、血管、食道、腸などは共通して、何らかの物体を通す「管状構造物」であるといえる。別の例では、僧帽弁も三尖弁も流体制御機能をもつ弁構造物であるといえる。このように、汎用構造物は複数の器官の共通的な性質を捉えるのに適しており、異常状態についても、これら汎用構造物には各器官に共通する特徴的な異常が見られる場合が多い。例えば、管状構造物の場合、管の内腔断面積が縮小した「管狭窄」という異常状態が生じやすいことがいえる。そこで、第1層の「面積縮小」を参照し、管状構造物を対象物として特殊化することで「管狭窄」を定義し、さらに、管状構造物を血管で現れると「血管狭窄」、食道で現れると「食道狭窄」のように各器官に固有の異常状態として特殊化することで、異常状態の共通性と器官固有の特殊性を扱うことが可能である。

第2層では、下位に行くにつれて、人体構造物を解剖学オントロジーの is-a 階層に従って特殊化することで、器官固有の異常状態が定義される。「血管狭窄」の場合、「動脈狭窄」から「冠動脈狭窄」や「腎動脈狭窄」のように各専門領域で必要とされる粒度まで特殊化を行うことができる。

なお、人体解剖構造物依存の異常状態については、現在我々のプロジェクトで開発中の解剖学オントロジーの階層に依存している。そのため、複雑な解剖構造の場合、階層が深くなるというように、第2層の階層構造の深さは相対的となっている。例えば、異常状態オントロジーにおいて、「冠動脈狭窄」の場合、第2層の最上位から管狭窄、血管狭窄、動脈狭窄、冠動脈狭窄のように第2層は4階層である一方、「心室拡大」の場合、最上位から腔状構造物体積拡大、心室拡大となっており、第2層は2階層となる。また、細胞や組織など器官より粒度の小さい人体構造物は、本研究では部品というより、材料として扱っている。そのため、汎用構造物依存の異常状態を第2層の最上位とせず、例えば、「心筋細胞壊死」の場合、最上位から細胞壊死、筋細胞壊死、心筋細胞壊死というように階層化される。

ここで、オントロジー構築について、下位概念をどこまで詳細にまで構築するかが課題となる。例えば、「ポリープ形成」を場所が特定されただけの「胃ポリープ形成」状態までを「ポリープ形成」の下位概念として記述しておくのは理論上、冗長であると考えられる。しかし、各ドメインで、現実の状況に適用することを考慮した場合、疾患や故障では、異常状態は、隣接器官/構造物が互いに影響を及ぼし、その結果次の異常状態を引き起こすことが多いと考えられる。例えば、狭窄の場合、冠動脈狭窄と

気道狭窄は明らかに異なった概念として区別すべきものであり、冠動脈狭窄は、虚血性心疾患を引き起こす危険性を生じ、気道狭窄は喘息を引き起こすという点において、どの器官/構造物でおこっているかが重要となる。そこで、対象とするドメインで利用されるために必要十分な状態をモデル化するというオントロジー工学的な立場をとった場合、異常状態オントロジーの構築においては、場と取り巻く環境を詳細レベルまで考慮する必要がある。下位概念として、特定の器官/組織/細胞レベルの異常状態まで用意することが妥当であると考えられる。

ただし、第2層で定義される異常状態は対象物がどこまで詳細に特定されたとしても、どのようなコンテキストで現れたものであるかというコンテキストとは独立な点において次の第3層と概念的に分離されるべきものであり、特定の疾患に現れる異常状態とは明確に区別される。例えば、「高血圧」は「動脈硬化症」という疾患の原因となることが知られているが、他の疾患でも汎用的に現れうる異常状態である。「高血圧」が「血圧が高い状態」という性質は、いずれの疾患で現れるかというコンテキストに独立して定義できる共通な性質であるが、どのような機序で現れ、他にどのような影響を及ぼすかといった性質は、高血圧が現れる疾患のコンテキストに依存した性質であると言える。さらに、「高血圧」をより一般化した「圧力増加」を考えるならば、それは医療に限らず様々なドメインにおいて一般的におこりうる異常状態であるといえる。

4.3.4 第3層：疾患コンテキストにおいて定義される異常状態

第3層では特定の疾患コンテキスト下で現れる異常状態を概念定義する。本定義では、第2層の異常状態を参照し、疾患コンテキスト下で、疾患を構成する異常として特殊化される。また、各異常状態は、ロール理論に基づき、その異常状態がコンテキストとする疾患において、どのような異常状態の原因や結果となるかを捉えることで概念化が行われている。第3層は、2.2.1節で示した疾患オントロジーにおいて記述される7診療科（循環器内科、神経内科、消化器内科、腎臓内科、代謝内分泌内科、アレルギーリウマチ内科、整形外科）の疾患毎に、それらを構成する異常状態が定義される。

図4.3に第3層の各異常状態の定義の具体例を示す。本定義は、代謝内分泌科の代表的な疾患である糖尿病とその下位疾患の1型糖尿病およびさらにその下位疾患である自己免疫性1型糖尿病について、これらの疾患を構成する異常状態の1つである「インスリン作用不足」を異常状態オントロジー第3層で概念定義した例である。

疾患「糖尿病」では、「インスリン作用不足」という異常状態が起こることが知られている。そこで、異常状態オントロジーの第2層「インスリン作用不足」を特殊化し、第3層では、異常状態「糖尿病におけるインスリン作用不足」が概念定義されている。

また糖尿病では、インスリン作用不足の結果、「高血糖」が引き起こされることから、インスリン作用不足の結果に「高血糖」が記述される(図 4.3 最上部).

次に糖尿病の下位疾患である 1 型糖尿病では、糖尿病と同様インスリン作用不足が起こる. そこで、「糖尿病におけるインスリン作用不足」を特殊化し、「1 型糖尿病におけるインスリン作用不足」が定義される. 異常状態「1 型糖尿病におけるインスリン作用不足」では、上位概念から結果の「高血糖」を継承しつつ、さらに原因として、「膵β細胞破壊」が起こることから、「膵β細胞破壊」を原因に新たに追加している.

さらに、1 型糖尿病の下位疾患である自己免疫性 1 型糖尿病では、「自己免疫性 1 型糖尿病におけるインスリン作用不足」が定義されており、原因が「膵β細胞破壊」に加え、「自己免疫異常」が原因でインスリン作用不足を生じることが定義されている.

このように、第 3 層の異常状態は、上位疾患を構成する異常状態からその性質を参照し、下位疾患で、その疾患固有の異常状態として、各々のもつ性質を特殊化することで概念化されている.

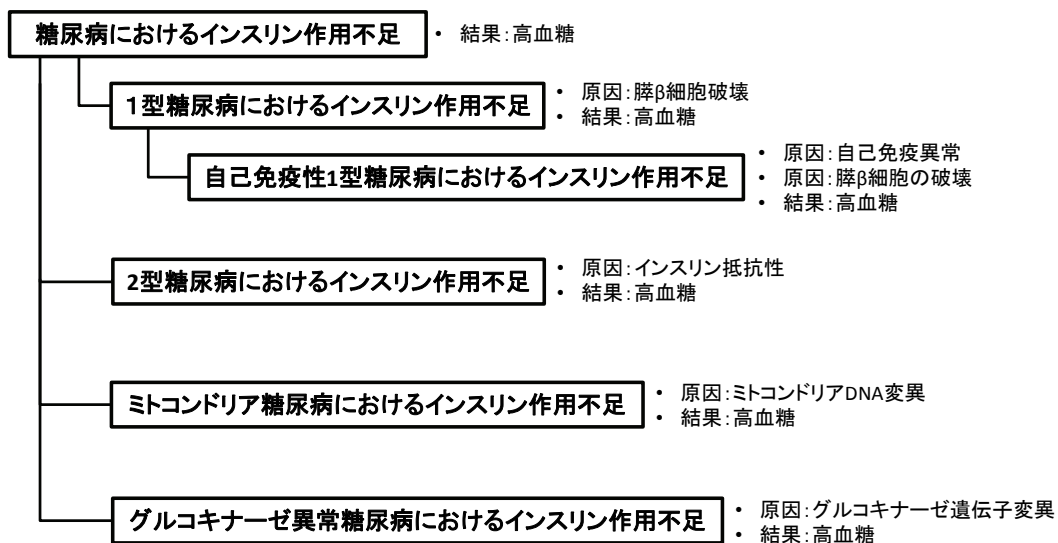


図 4.3 第 3 層における異常状態の概念定義例

4.4 統一表現モデルを用いた異常状態の is-a 階層の計算機表現

これまで述べてきた異常状態オントロジーはオントロジーエディタ「法造」 [古崎 02, 法造] を用いて構築されている. 法造は、オントロジーをグラフィカルに記述する環境を提供しており、オントロジーの意味定義, 属性の継承, ロール定義について

記述できる。オントロジー記述言語として Extensible Markup Language (XML) ベースのフレーム型言語を基本としている他、Resource Description Framework Schema (RDF(S)), Web Ontology Language (OWL) 形式でのエクスポート/インポート機能をもつ。図 4.4 に、異常状態オントロジーにおける異常状態定義の「法造」による計算機的表現を示す。対象物非依存の最も一般的な異常状態 (第1層)は、汎用異常状態の下位概念として、基本概念として定義され、is-a 関係が階層化される。図 4.4 右上のサイズ縮小の下位概念として「面積縮小」があり、属性「面積」(A) とその値が「小」(V)として定義されている。対象物に応じて共通に現れうる異常状態 (対象物依存異常状態, 第2層)は、対象物をコンテキストとした対象物固有の異常として定義される。図の左上段では、汎用構造物である「管状構造物」に固有 (管状構造物依存)の異常状態として、「管狭窄」が定義されている。「管狭窄」の右側の長方形は、汎用異常状態への参照を示し、汎用異常状態から定義内容が継承されるとともに、下位概念では定義する際、必要に応じて継承された定義内容を特殊化することができる。「管狭窄」の場合、「面積縮小」から継承された「面積」をさらに「断面積」に特殊化し、管に固有の「管の断面積が小さい」という内容が表現される。対象物依存の異常状態の is-a 関係は、対象物を特殊化することにより階層化される。図 4.4 左中段では、管をさらに特殊化し、人体構造物固有の異常状態「血管狭窄」、「動脈狭窄」、「冠動脈狭窄」が定義されている。さらに、「冠動脈狭窄」が疾患「狭心症」を構成する異常状態である場合、「狭心症」をコンテキストとして、「冠動脈狭窄」を参照する「狭心症における冠動脈狭窄」(第3層)が定義される (図 4.4 右下)。

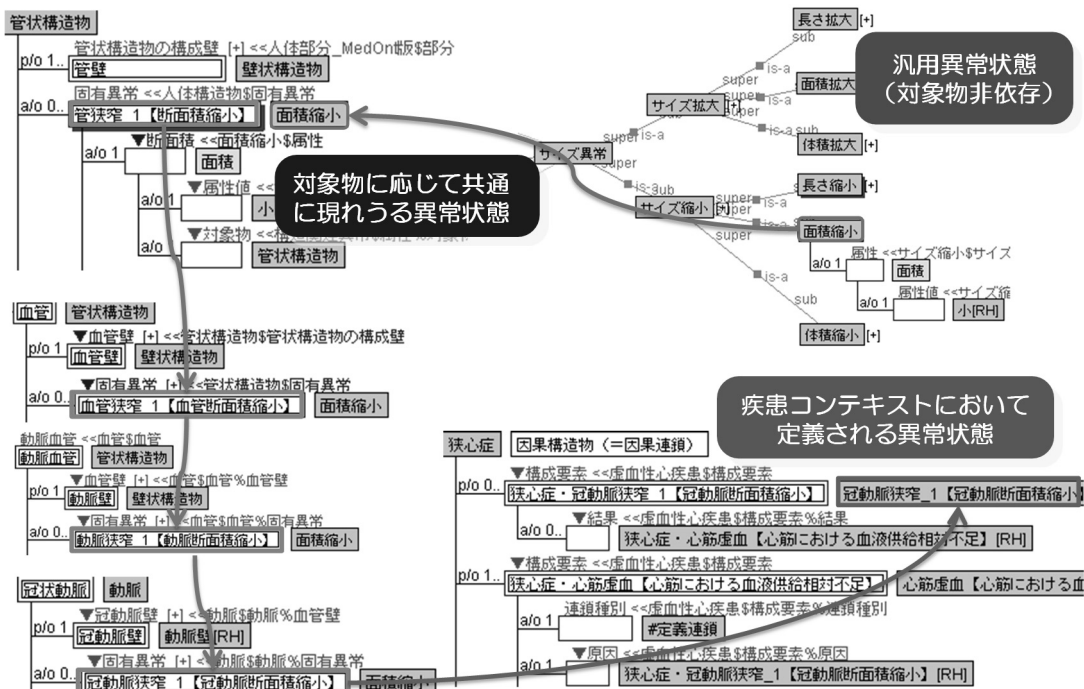


図 4.4 異常状態 is-a 階層の計算機表現

以上のように、異常状態オントロジー構築では、3章で示した特性・属性分解記述に基づく統一表現を基礎として3層に階層化することで、共通性と特殊性を明確に分離しながら、コンテキストに応じた異常状態を表現することができる記述となっている。具体的には、第1層で定義したできるだけ汎用性をもたせたドメイン非依存かつ構造物非依存の異常状態を、第2層で特殊化し、汎用的な構造物（例えば管状構造物）に共通する異常、さらに特定ドメイン内においても汎用的レベルから、対象物の is-a 階層に沿った特殊化が行われる。そして第3層では、それらの異常状態を、さらにそのドメインの特定コンテキスト構成する異常状態まで特殊化していく。

このような異常状態のオントロジー構築によって、例えば、循環器内科内の「虚血性心疾患における冠動脈狭窄」状態と、「大動脈閉塞症における大動脈狭窄」とは、同じ「動脈狭窄」という上位概念をもつことが明示できる。加えて、他の診療科である脳外科の「脳梗塞における脳血管狭窄」とは「血管狭窄」という共通性を持ち、さらに、消化器内科の「腸閉塞における腸管狭窄」とは、「管狭窄」という上位概念において共通するという共通性があることがわかる。つまり、各診療科（専門領域）で詳細に専門疾患における異常状態として、別々に扱われていたものが、共通性を捉えることにより、診療科を超えて扱うことが可能となることから、様々な応用が可能であると考えられる。

4.5 異常状態の上位概念分類例

本節ではいくつかの典型的な異常状態について、本研究モデルに基づき上位概念の分類を行い、本モデルの有効性を考察する。

4.5.1 感染症

感染症は、病原体が体内に侵入後、各臓器で定着し、その後様々な症状をもたらす。そこで、これらを3章の統一表現モデルに基づき概念定義を行うとともに、4.3節の3層モデルに基づき分類することで、各異常状態について識別可能であるか、マラリア感染症を例に検討した。結果を表4.1に示す。

まず、マラリア原虫は血液への感染を起こす。そこで、「血液マラリア原虫感染」状態として考えると、上位概念は、異物付加＜具体物 (O), 付加物 (So), 物の数 (A), 多 (V)>といえ、構造体に何らかの異物が付着した状態で捉えることができる。そこで、異物付加の属性を継承し、対象物 (O) と副対象物 (So) を特殊化することで、病原体感染定着＜器官 (O), 病原体 (So), 物の数 (A), 多 (V)>という状態が定義される。さらに、これらの対象物 (O) と副対象物 (So) を特殊化することで、血液マラリア原虫感染＜血液 (O), マラリア原虫 (So), 物の数 (A), 多 (V)>と定義される (表

4.1). 次に、血液に感染した後、マラリア原虫は血液から移動し、体内の肝細胞組織に定着することが知られている [メルク]. このとき、血液マラリア感染と同様に病原体感染定着を特殊化し、対象物を新たに定着した肝細胞組織とすることで、肝細胞組織マラリア原虫定着<肝細胞組織 (O), マラリア原虫 (So), 物の数 (A), 多 (V)>が定義できる. ここで、マラリア原虫が定着した肝組織では炎症が生じるので、肝細胞組織炎症<肝細胞組織 (O), 炎症細胞 (So), 物の数 (A), 多 (V)>と状態が表される. 炎症について考えると、肝臓という器官の材質が異常となった状態といえるので、上位概念として、第1層の汎用的異常状態は「材質の異常」となる. さらにマラリア原虫は赤血球内に移動し、赤血球で増殖する. 赤血球内の原虫増殖については、第1層の上位概念は、異物付加ではなく、パラメータ異常の「物の数が多い/増加【many】」となり、それを特殊化し、第2層では病原体増殖 (P) <器官 (O), 病原体 (So), 物の数 (A), 多 (V)>とし、これをさらに特殊化することで、赤血球内マラリア増殖<赤血球 (O), マラリア原虫 (So), 物の数 (A), 多 (V)>となる. その後、赤血球が破裂し (上位概念は空間異常の「裂ける」), 患者は貧血を起すことが知られている. このとき、貧血は、<血液 (O), ヘモグロビン (So), 濃度 (A), 低 (V)>となり、上位概念はパラメータ低下の「濃度低下」となる. また、他に発熱や、脾臓の重さが増大する脾腫という症状が起こることが知られている.

表 4.1 マラリア感染における感染から発症までの定義例

異常状態 上位	異常状態 (P)	対象物 (O)	副対象物 (So)	属性 (A)	属性値 (V)
異物付加	血液マラリア原虫感染	血液	マラリア原虫	物の数	多
異物付加	肝細胞組織マラリア原虫定着	肝細胞組織	マラリア原虫	物の数	多
材質の異常	肝細胞組織炎症	肝細胞組織	炎症細胞	物の数	多
物の数が多い/増加 【many】	赤血球内マラリア増殖	赤血球	マラリア原虫	物の数	多
濃度低下	貧血	血液	ヘモグロビン	濃度	低
高温/温度上昇	発熱	人体		温度	高
重量増加	脾腫	脾臓		重さ	重

このように、異常状態オントロジーでは、統一表現を用いることで一貫性のある記述を可能とするとともに、感染から発症までの各状態について定義できることが確認された。各状態が、人体構造物そのものの異常か、材質が異常となっているか、あるいは、それ以外のパラメータの異常であるかについては、第1階層で定義した汎用異常状態からいずれを上位概念とするかで表す。さらに、パラメータの異常の場合、各異常状態の属性 (A) として温度や濃度などの一般的物理パラメータのもつ値がどのように変化したかが表される。このように、3層構造により、汎用異常状態からどのような性質を継承し、その意味内容は人体のどの臓器 (O) を対象として、注目物 (So) について何の観点で (A) どのように値 (V) が変化するのかについて把握が可能となる。

4.5.2 循環器

次に、疾患オントロジーにおいて、循環器の専門医が記述した疾患 577 のうち、医療情報開発センターの提供する MEDIS 標準マスター [MEDIS 03] に対応のある主要 173 疾患 886 の異常状態を対象として、3章の統一表現を行い、さらにこれら異常状態について、第1層の汎用異常状態としてどのような上位概念をもつかについて検討を行った。

その結果、各異常状態の対象物 (O) は、動脈血管として大動脈や冠状動脈、血液、心臓の左/右心室、心房、心筋の他、弁 (僧帽弁、大動脈弁、三尖弁) など循環器系の人体構造物を主としていた。異常状態としては、血液を対象とした異常状態では、血圧増加 (高血圧)、血流量低下などが多くみられ、心筋については心筋壊死、心筋炎症、心筋虚血、冠状動脈血管では冠動脈狭窄、冠動脈血栓形成が複数の疾患で現れていた。また、上位概念別にみると、新生成物形成、現象関連異常、ノンパラメーター異常、圧力増加、材質の異常の他、サイズ異常 (面積縮小、長さ拡大) の他、孔あきや裂けるなど結合に関する異常が多かった。新生成物形成とは、新しく何らかの物質が生成される異常をいうが、循環器内科では血栓形成が多い一方、ポリープ形成はみられなかった。現象関連異常では、心筋酵素の逸脱という臓器からの物質の漏出や逆流、ノンパラメーター異常では、疼痛とその下位概念である胸痛などが心筋梗塞をはじめ、心膜炎、動脈硬化症、動脈閉塞症、肺高血圧症など複数の疾患で起こることが確認された。これらの結果は循環器内科で把握されている臨床像をそのまま反映できていることが確認された。このように、本モデルでは細胞から人体というレベルまで様々な粒度の異常状態について、各々を汎用的な概念から性質を継承し、分類することで一貫性した観点で概念化を行うことができることが確認された。

4.6 考察

本研究では、異常状態の統一表現をもとに3層構造によるモデリングを行うことでオントロジー構築の理論枠組みを実現した。そこで、本節ではまず、3層モデルの有用性について考察する。次に、関連研究と異常状態オントロジーとの比較を行う。さらに構築されたオントロジーを評価し、その意義について述べる。

4.6.1 異常状態3層モデルの有用性

本節では、構築された階層構造の有用性について考察する。

1) 概念の共通性の把握

異常状態オントロジーの is-a 階層の上位の層に上がることによって、疾患固有の異常状態から、複数の疾患で起こりうる異常として一般化され、疾患独立に共通性を捉えることができる。その結果、診療科を超えて異常状態を捉えられることから、相互の比較が可能となり、異なる診療科間でどの異常状態がどのくらい共通性があるか把握が可能となると考えられる。例えば、消化器内科の「腸疾患における腸間膜動脈狭窄」について考えると、第2層の「腸間膜動脈狭窄」については、消化器内科の腸疾患でしか現れないことから、消化器内科の腸疾患固有といえる。しかしながら、第2層の「腸間膜動脈狭窄」の上位概念をたどって得られる「動脈狭窄」では、複数の診療科の疾患における異常状態と共通性があることがわかる。例えば消化器内科で見られる「腸間膜動脈狭窄」も循環器内科で見られる「虚血性心疾患における冠動脈狭窄」も、共に同一の「動脈狭窄」を上位概念として持つことを示すことができる。さらに、神経内科の「脊柱管狭窄症における脊柱管狭窄」とは、第2層上位の汎用構造物依存の異常状態「管狭窄」で共通する。このように、異常状態オントロジーの概念階層を用いれば、様々な概念レベルでの共通性が明示できることから、今後、異常状態の比較解析によって、類似症例解析による治療介入などへの貢献も期待される。

さらに、これまで全く違った診療科領域で扱っていたもの同士に共通性を見出す可能性がある。場 (器官コンテキスト) に応じて高度に分化し、複雑で特殊化された異常として個別に扱われてきた概念も、器官非依存の異常状態として共通性を捉えて整理すればもっとシンプルに異常状態を扱える可能性がある。また、ラベルが全く異なっても、実際は場 (器官) が異なるだけで本来は同じものと扱ってよいような概念の異常状態も発見できる可能性がある。このように、異常状態オントロジーは異常状態の本質を捉え、異常状態の意味内容を明示することを可能にするといえる。

2) 診療科横断的な複数疾患で起こりうる異常状態の把握

異常状態 (単独)の共通性をみつけることができれば、さらに異常状態間の関係の共通性も把握可能となる。2.2 節で説明した疾患オントロジーでは疾患を異常状態の因果関係 (疾患連鎖)として捉えて定義している。各疾患定義における異常状態は、4.3.4 節の第3層の疾患を構成する異常状態として定義され、臨床医が診療科別に記述をおこなっている。そこで、各診療科で定義された全疾患から因果関係を集め、異常状態オントロジーを参照することで同一の異常状態を含む因果連鎖を結合させるよう計算機処理を行うと、7診療科の1万を超える異常状態から起こりうる全ての因果関係を含む汎用的な連鎖 (以下、汎用連鎖とよぶ)の生成が可能となる(図4.5)。

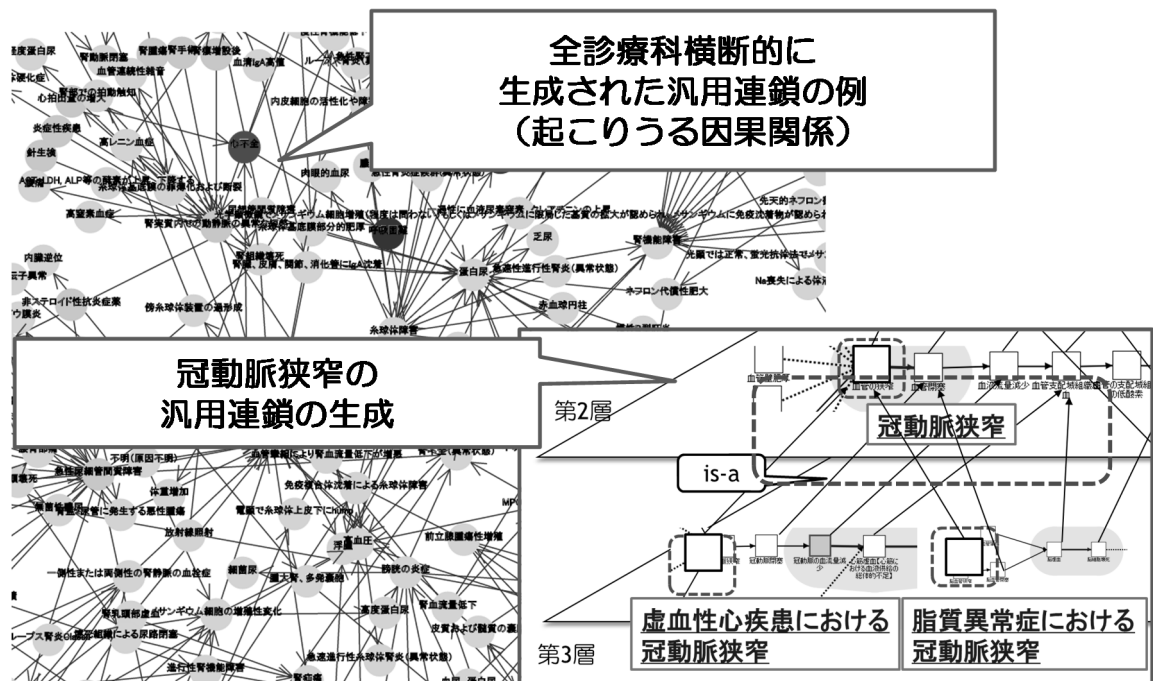


図 4.5 汎用連鎖の生成例

例えば、異常状態オントロジーでは、第3層で虚血性心疾患固有の異常状態として「虚血性心疾患における冠動脈狭窄」状態が定義され、疾患オントロジーでは、循環器内科で専門医が虚血性心疾患において、異常状態の因果関係「冠動脈狭窄→血流量低下→心筋虚血→…」を記述している。そこで、因果を構成している異常状態について、異常状態オントロジーを参照して個々の異常状態の概念階層を全て第2層に上げて一般化しておけば、汎用連鎖の生成が計算処理される。その結果、他の診療科における別の疾患 (例: 代謝内分泌科の脂質異常症)で、脂質蓄積の結果、冠動脈狭窄が起こった場合、「脂質蓄積→冠動脈狭窄」と代謝内分泌科の臨床医が記述していたのに加え、先の汎用連鎖を参照し「脂質蓄積→冠動脈狭窄→血流量低下→心筋虚血→…」

のような一般的な因果関係を導くことができる。このように、本研究で提案した異常状態オントロジーを用いて、7診療科を横断し、1万を超える異常状態を is-a 階層構造と因果関係との観点から捉えることが可能であり、その結果、断片的な知識を統合し、診療科横断的な医学的知見の獲得に貢献することが期待される。

なお、本研究では、「特定の患者に起こった事象の表現をするための情報モデル」ではなく、「疾患定義に必要な異常状態とその因果連鎖を表現するモデル」を提供することを目的としている。そのため、因果関係については、「そのような現象が実際に起こったか否かを判定した結果」を記述するのではなく、臨床医がこれまで経験した既に起こったことを抽象化し、(疾患を定義するための臨床の世界での共通の合意として得られる)原因と結果として定義している。したがって、実際に生起した二つの事象間に本当に因果関係があったかどうかを同定するための知識の記述は、本研究の範囲外としている。また、臨床医がインスタンスレベルで個々の患者で起こっている状態について、1万以上の異常状態を参照しながら判断することは、現実的に対応も困難と考えている。しかし、概念階層を上げると、汎用構造物レベルでは、共通する汎用構造物依存の異常状態(第2層上位)の因果関係がある程度のパターンに収束するので、それら実際に起こっている因果関係の比較は可能であると考えている。

本研究で得られる理論上起こりうる因果関係と、実際の知見とを比較することで因果生成プロセスの基本原理がみえてくる可能性がある。現在、予備的な調査では、第2層の汎用構造物依存の異常状態として、管状構造物の異常について、「管の流量低下→供給不足」の因果の原因が、管の狭窄(断面積縮小)と一時的な収縮に大別され、その原因が、前者は異物や新生成物の付加、壁の肥厚、外部圧迫のパターンがあることを見出している。そして、実際の疾患でも、例えば虚血性心疾患では、上記を特殊化した「冠動脈血流量低下→虚血」と、その原因として、冠動脈への血栓形成または冠動脈壁肥厚による狭窄、あるいは冠動脈攣縮(一時的収縮)が起こることを確認している。現在、本パターン処理について、応用システムの調整を行っており、今後汎用レベルの因果関係の把握についてより深い考察をしていきたい。

なお、実際に起こっている異常状態はさらに複雑な要因が組み合わさる可能性がある。将来的にはこれらにも対処し、より臨床の場の実用を考慮した枠組みを別途検討したいと考えている。例えば、実際には、異常状態が同時に起こることも存在する。そこで、これらにも対応できるように、疾患オントロジーでは、異常状態について同時に起こる(AND)、いずれかが起こる(OR)の記述枠組みも用意している。

4.6.2 関連研究との比較考察

本節では、2章で示した既存の用語集および表現型オントロジーと比較を行い、オントロジー工学的な観点における本研究の優位性について考察する。まず、用語集で

は SNOMED-CT および MeSH , オントロジーとしては, 表現型一般を扱う PATO, およびヒトの表現型を扱う HPO を対象とし, 異常状態を捉える際のオントロジー的な課題を調査するため, どのように表現され, 組織化されているかについて本オントロジーとの違いを検討する. 調査対象の概要と, 表現形式および組織化の仕方 (is-a 階層の有無, クラス継承が多重継承か単一継承か, 提供する概念レベル, 上位オントロジーの参照の有無)について比較した結果を表 4.2 に示す. 次に, 組織化については「冠動脈狭窄」を例に, その上位概念の is-a 階層についてより詳細な調査を行い, その結果を図 4.6 に示す. 以下, これらの結果をもとに考察を述べる.

表 4.2 異常状態に関連する主な生命医学系リソースとの比較結果のまとめ

内容	SNOMED-CT 臨床用語統制 語彙	MeSH 医学文献統制 語彙	PATO 表現型 オントロジー	HPO ヒト表現型 オントロジー	本オントロジー 異常状態 オントロジー
概念数	310,000	27,000	2,500	10,000	13,000
is-a 階層	有	有	有	有	有
クラス継承	多重継承	多重継承	多重継承	多重継承	単一継承
汎用異常状態	無	無	有	無	有
人体解剖構造物依 存異常状態	有	有	無	有	有
疾患依存異常状態	有	無	無	無	有
上位オントロジー	無	無	BFO	BFO	YAMATO
OP/ OAV 表現形式	無	無	OP	OP	OPVp OSoAV
データ公開	一部公開	有	有	有	公開予定 (一部公開済)

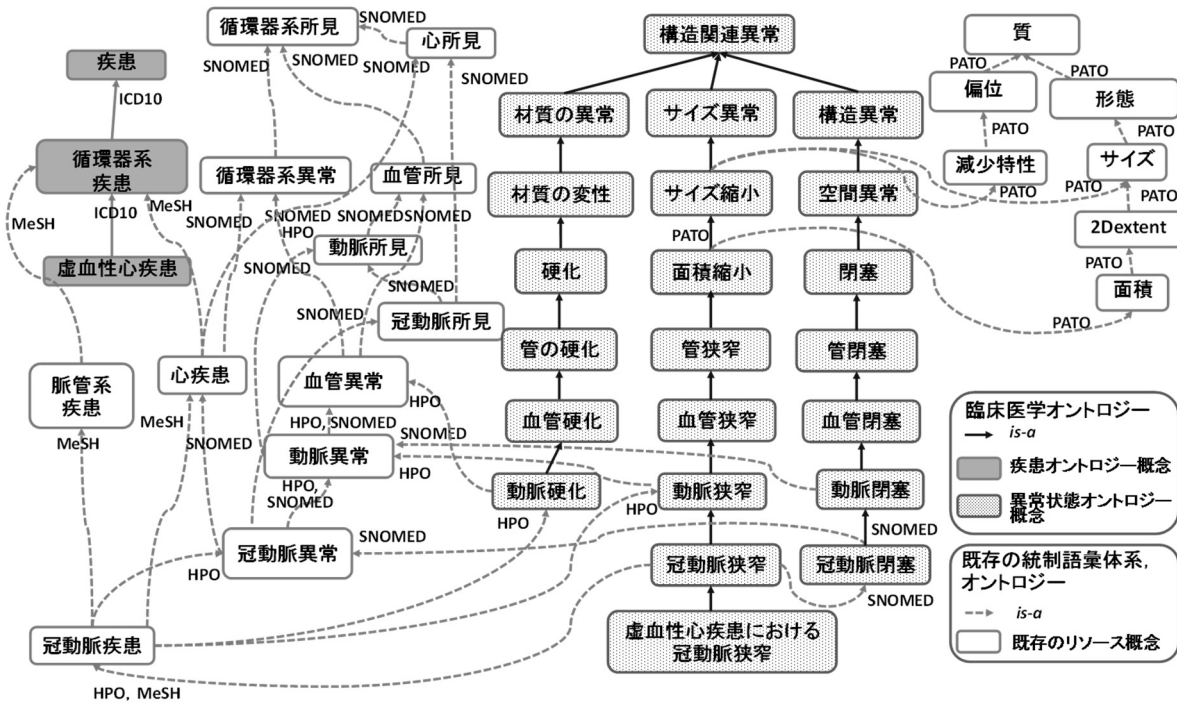


図 4.6 関連研究における冠動脈狭窄の is-a 階層

1) 表現系オントロジー (PATO, HPO)との比較

まず, PATO, HPO についてオントロジー的な観点から本研究との違いについて調査した結果を述べる. PATO, HPO は表現型を対象世界としているのに対し, 本オントロジーは異常状態全体を対象世界としている. PATO は生物種横断的な汎用的な概念, HPO は人体に特化した概念 (我々の第2層) のみ体系化しているのに対し, 本オントロジーでは, 異常状態の組織化を, オントロジー理論に基づいて汎用レベルから人体構造物依存, さらに疾患依存レベルまで, 各概念を識別しつつ, 一般的な概念を参照しながら疾患依存レベルまで一貫性をもって行っているため, 異常状態相互の相違や類似性を明示的に扱うことができる.

また, PATO の概念体系では 3.7 節で述べたように, 特性 (P)と属性 (A)がオントロジー工学的な観点において識別されていない. その結果, PATO の階層は多重継承となっている. 例えば, PATO では, 「面積縮小」は, 「面積」と「サイズ縮小」という2つの上位概念をもつ (図 4.6 中央から右部). このような多重継承は, PATO と他の標準語彙体系や用語辞書 (例: MeSH) において, 面積縮小の下位概念に相当する「冠動脈狭窄」とを統合する場合, 自動的に計算機処理で結合すると, 「冠動脈狭窄 is-a 面積」のような不適切な is-a 関係が推論されてしまう可能性がある. 本オントロジーでは, 圧力, 面積のようなパラメータは属性として扱っているため, そのような特性と属性を混同することはない.

異常状態の本質を考えた場合, 状態は変化し, 異常状態がどのような観点 (A)で, どのような値 (V)をもつかという点において, 「面積縮小 is-a 面積」ではなく, 「面積縮小 is-a サイズ縮小」を本質と捉えるべきであり, 本異常状態オントロジーは, 後者のみを本質として捉え, 適切な is-a 関係が設定されている.

HPO は, 2.3.1.1 節で述べたように, ヒトの表現型のオントロジーとして表現型の概念を構築するという構築方針において, 「その状態が人体のどこで観察されるか」という人体の部位という観点に重きをおいて上位概念が階層化されている. そのため, 「動脈狭窄 is-a 動脈異常」というように異常状態の is-a 関係については, 狭窄という特性に基づく階層構造をたどることが難しい. 一部に「腎動脈狭窄 is-a 動脈狭窄」のような関係も見受けられるが, 本モデルのような第1層の汎用的な異常状態をもたない上, 属性に関する情報がないため, それがどのような観点を, どういう値をとる異常なのかという状態としての本質は表現されていない.

また, HPO では疾患と異常状態というオントロジー的には異なる概念を識別しておらず, 疾患の上位が異常状態で有る場合や, その逆の異常状態の上位が疾患である例が存在する. 例えば, HPO では「冠動脈狭窄」の上位概念は異常状態ではなく, 「冠動脈疾患」となっており, さらにその「冠動脈疾患」は3つの上位概念: ①動脈狭窄, ②冠動脈異常, ③動脈硬化をもつ (図 4.6 左下). また, 上下の階層構造を見る限り, 動脈硬化は果たして異常状態なのか, 疾患 (動脈硬化症) なのか不明である.

本オントロジーでは、4.3節で述べたように異常状態オントロジーでは、第1層は「対象物非依存の異常状態(汎用異常状態)」(例：面積縮小)、第2層では「対象物に応じて共通に現れうる異常状態」として、下位では人体構造依存の異常状態(例：冠動脈狭窄)を定義し、第3層では「疾患コンテキストの下で定義される異常状態」(例：虚血性心疾患における冠動脈狭窄)というように一貫性をもって異常状態のみからなる階層の組織化を行っており(図4.6網掛け部分)、疾患とは明確に識別される。

2) 用語集 (SNOMED-CT, MeSH)との比較

病理学用語の記述法とコードを意味的に体系化するために開発された SNOMED-CT、および米国国立医学図書館が提供してきた医学統制用語シソーラス MeSH は、オントロジー理論の成熟する前から開発されてきた。そのため、オントロジーを意識して構築されたものではなく、上位オントロジーももたず、is-a 関係の間違った使い方や、属性の継承が正しく行われていない、多重継承などの問題を抱えていることが知られている [Schulz 07]。複数の上位概念をもつ場合、概念間の関係が複雑化しすぎて臨床上の知見と整合性のあるモデルが維持されているか問題となる可能性がある。

そこで、異常状態について「冠動脈狭窄」の is-a 階層において上位概念としてどのようなものがあるかを調べたところ、MeSH および SNOMED-CT は前述の HPO と同様、疾患と異常状態が混在した階層化が行われていることがわかった。例えば、MeSH の「冠動脈狭窄」の上位概念をたどると、「心疾患」、さらに「循環器系疾患」のように、上位概念は疾患となっている。また、SNOMED-CT では、「冠動脈狭窄」の直上は、「冠動脈閉塞」であり、さらにその階層関係を追跡すると、上位は心疾患、血管所見、循環器系異常のように、複数の視点に基づいて上位概念をもつように概念化されており、図4.6に示すように多重継承が多用されている。SNOMED-CT は、臨床用語集として著名であるが、疾患と異常状態が混在していた場合、それらの用語を利用して診療録を記述すると、各用語の意味内容が不明確となり、診療情報において信頼性が確保できない事態が生じる恐れがある。例えば、臨床医が血管造影の検査を行い、異常状態の記録として SNOMED-CT の用語「冠動脈狭窄」を用いて記述したとする。しかし、その上位概念の一つである疾患を継承して冠動脈狭窄も疾患概念であると計算機が判断して処理した場合、血管造影は疾患の同定、診断に用いられたと推論され、正確さを欠いた情報処理につながる恐れがある。

一般に、複数の観点で分類される場合、階層構築において多重継承の問題が生じる。大規模な医療用語を包含する SNOMED-CT や、MeSH は、異常状態そのものに特化したものではないため、多数の観点に基づいて分類されていることがわかる。また、表現系オントロジーの場合においても、属性あるいは発生部位など複数の観点に基づいた分類が必要となるため、多重継承にならざるを得なかったのではないかと考えられる。しかし、オントロジーの理論的な考察をせず安易に多重継承を導入することは、概念の本質的な性質の隠蔽につながる恐れがあるなど、様々な弊害を生む可能

性がある。そのため、多重継承に相当する意味をオントロジー的に適切に扱う手法が提案されている。例えば、我々がオントロジー構築に利用しているツール「法造」においては、ロール理論に基づく上位概念およびクラス制約からの多重継承の表現や is-a と IS-A の書き分けによる多重継承表現 [太田 11]、1 つのオントロジーから必要に応じて異なる観点の is-a 階層が動的に生成する技術 (動的 is-a 生成) [古崎 12] を提供しており、これらの手法を用いることで多様な観点に対応できる。また、本質的な観点から見た is-a 階層のみを構築し、異なる観点による分類は推論することで導出するという手法も提案されている [Adams 09]。

本研究ではオントロジー工学理論に基づいて、異常状態オントロジーは疾患オントロジーと独立して構築しており、異常状態と疾患を明確に分離された概念として扱うことで本課題を解決している。具体的には、異常状態の本質的な性質を、第 1 層では「対象物独立の異常状態(汎用異常状態)」、第 2 層では「対象物に応じて共通に現れる異常状態」、第 3 層では「疾患コンテキストにおいて定義される異常状態」と、レベル毎にそれぞれ統一した観点で捉えオントロジーを構築している。そのため、各レベルの is-a 階層においては、多重継承を用いずに一貫した分類観点でのオントロジーが構築されている。また、疾患固有の異常状態を第 3 層で概念化しており、異常状態と疾患との対応付けについても適切に行っていることから、診療情報の計算機処理において優位であると考えられる。更に、異常状態について、本研究で提案した分類観点以外に、異常対象となる器官・臓器、原因や結果の症状など、異なる分類観点を導入したい場合に、必要に応じて上述の動的 is-a 生成による手法を導入することも可能である。

4.6.3 異常状態オントロジーの評価

今回、本研究では、世界で初めて異常状態に特化したオントロジーを開発した。前節の関連研究との比較で述べたように、既存のリソースでは異常状態に特化したオントロジーは存在しなかった。また、異常状態概念を含むオントロジーは存在するが、多数の視点が混在し、疾患との識別もされていなかった。

異常状態オントロジーは 3 章で提案した統一表現モデルによって異常状態の定義を行い、これを基礎として 3 層構造モデリングにより構築されている。本研究では、一般的な概念から人体解剖構造物依存、疾患依存異常状態まで、各概念を一般・特殊関係に基づいて組織化(is-a 階層化)し、各々を識別可能にした。また、再帰構造をとることにより、一般的な異常状態は、下位から参照されるとともに、特殊化を行うことで専門医が必要とする粒度までをカバーする異常状態知識を提供することができる。そして、疾患と異常状態を明確に区別した上で、さらに、複数疾患における異常状態の共通性を把握可能とした。

次に、統一表現モデルについて述べる。人体での異常状態は変化し、粒度も多様である。そのため、従来、統一的記述は一般的に困難とされてきた。本研究における異常状態の統一表現モデルでは、上位オントロジー YAMATO を採用し、特性・属性分解の手法により、人体の「悪心」や「妄想」というような物理パラメータで表現が困難な概念レベルの状態記述にも対応しつつ、細胞レベルの「細胞壊死」等も表現可能とし、さらに値の変化の詳細にも対応し、検査データ等の定量値を記述可能とした。その結果、臨床での各種異常状態に応じて適切に対処し、相互互換性をもたせ、統一的に記述できる枠組みを実現した。異常状態表現については、哲学的立場を重視し、BFO のような存在論ベースで構築するという手法も考えられるが、科学という観点で臨床の知識を概念化するには、本研究のように量を扱える表現が必須となる。さらに、DOLCE のような属性形式のみでは、人体の「悪心」というような粒度の異常状態を扱えない。本研究は、BFO や DOLCE では達成できない人体で起こりうる異常状態を、様々な粒度ですべて記述できる枠組みを達成した。

さらに、本研究の理論枠組みを用いて、3 診療科の代表的な 107 の異常状態とその上位概念をあわせた総計 386 の異常状態ををすべて統一的に記述できることを確認し、各々の異常状態についても 3 層で概念定義を行った。また、臨床医が、特に難しいと考える異常状態については、オントロジー工学者と臨床医と協議の上、3.6.3 に示す記述指針を作成し、本枠組みを用いて適切に記述できることが確認できている。さらに、循環器の主な疾患 173 疾患の 886 の異常状態についても統一記述と上位概念の分類を完了し、循環器内科の各疾患の臨床像を反映していることを確認している。

4.6.4 異常状態オントロジーの意義

近年、専門領域が細分化され、各診療科で高度な専門知識に基づいて治療が行われる一方、診療科を横断して総合的に異常状態を捉えるという機会が減少していると考えられる。このことに対処する一つの機能として、本異常状態オントロジーと疾患オントロジーにおける異常状態の因果関係 (疾患連鎖) との組み合わせにより、1 万を超える異常状態の因果関係を捉えることが可能となり、診療科個別の疾患固有の異常状態から、診療科横断的な異常状態の共通性を把握することが可能となることを 4.6.1 で示した。その結果、診療科間の異常状態に関する比較が可能となり、診療科横断的な知識獲得や知識発見など計算機処理において有用であると考えている。

医療分野においては古くから、用語集や標準語彙体系の構築がすすめられ、それをもとに各種オントロジーが公開されている。しかし、これらはオントロジー理論上、理論的整合性に欠け、また、大規模な用語を多数の視点で分類しているため、一貫性をもった体系化がされていない。近年では、オントロジー的観点で捉えなおした拡張、バージョンアップなどの検討も行われているが、大規模な概念の中から、これまで述

べたような不適切な is-a 関係の使用や、間違っただ階層分類をみつけだし、正しい階層構造に抜本的に修正するのは困難な作業である。さらに、既存オントロジーや用語集については多様な観点をもつため、これらを統合しようとする、ますます関係性が複雑化し、意味の理解を困難なものとする恐れがある。また、統合した後も概念の比較解析や妥当性検証の作業も引き続き要請されると考えられる。

臨床の高次知識の意味を統合するには、内容指向の技術が必要とされる。我々の開発研究は、オントロジー工学の専門家と専門医を含む医療の専門家の協力体制の下、一貫性のある理論的にエラーのない深い考察に基づいたオントロジー構築を目指し、ゼロから概念体系の構築を進めてきた。疾患、解剖構造、異常状態をオントロジー理論に基づいて定義し、さらにこれら3つのオントロジーを特性・属性分解表現、疾患連鎖などの洗練された手法により関係性をもたせている。したがって、我々のオントロジーは既存の用語集や標準語彙体系に対し、骨格としての知識を与えることができると考えている。

臨床の現場では、専門医は検査結果や患者を観察することで、重要な異常状態を選別し、関連疾患の因果関係を高度な背景知識に基づいて推論している。しかし、その観点は臨床医にとっては当たり前のこととして診療記録には反映されず、計算機が理解して処理することはできない。疾患連鎖と異常状態オントロジーを組み合わせることで、罹患状態について、何が原因で、どのような異常状態が起こり、その結果、何がどのような観点で、どのような値が観察されて疾患が成り立っているのか、そしてその異常状態は他の疾患や他の部位ではどのような状態として現れているか、というような関係性が明示できれば、他の臨床医、研究者、さらには計算機とこれら知識の意味内容を共有することが可能になり、各種疾患解析や、医学知識基盤構築支援にも発展させることができると期待される。

4.7 結言

本章では、疾患を記述するための異常状態について、概念層を3層とした異常状態の階層的モデリングについて述べた。本研究は疾患と異常状態を識別可能な点において臨床の診療記録などにおいて有用であるといえる。また、本階層モデルでは、本研究の目的 2) 異常状態の階層的モデリングを達成し、汎用の異常状態から疾患固有の概念まで一般的な知識から医療の臨床の専門知識まで包括的に計算機処理できることを示した。従来、異常状態は各診療科別に個別の疾患にあらわれる異常状態として把握されていたが、本研究では診療科横断的に異常状態の共通性を見出すことが可能となった。関連研究の検討によって、既存のオントロジーでは複数の観点が混ざった階層は一貫性に欠け、意味の解釈に曖昧さを生じ、疾患と異常状態の混同が生じる恐れがあることが明らかになった。これに対し、本異常状態オントロジーはオントロジ

一理論に基づいてその関係性と個々の意味内容を適切に処理し、一貫性を保ちつつ構造化を行うことで、異常状態の全体像を把握することができる。したがって、既に存在する生命医学系の異常状態関連知識の骨格としての役割に貢献することが可能となり、さらなる疾患の理解につながると期待される。

次章では、本研究で提案した異常状態オントロジーを媒介として、生命医学リソースの異常状態に関する様々な知識へのナビゲーション実現のための取り組みについて述べる。

第5章 異常状態オントロジーを媒介とした生命医学系リソース間ナビゲーション

5.1 緒言

本章では、本研究の目的 3) 異常状態オントロジーを媒介とした生命医学系リソース間ナビゲーションについて、応用システムを提案する。生命医学系では、膨大なデータが日々蓄積されており、散在されたデータからいかに必要な情報を獲得するかが喫緊の課題となっている。しかし、表層上の一致だけでは、意味内容を処理できず、実際には情報の発見が困難となっている。したがって、既存のリソースの互いの位置づけやデータの意味内容が理解できるように関連付け、関連知識を獲得できるようなシステムが必要とされる。そこで、本章では、3章の統一表現および4章の階層的モデリングにより構築される異常状態オントロジーの応用研究について述べる。具体的には、異常状態オントロジーを骨格とし、既存のリソースに含まれる異常状態関連知識を概念レベルに応じて統合し、疾患の理解に必要な各種情報を獲得するための取り組みについて説明する。5.2節では、生命医学系での既存リソースに含まれる異常状態に関する知識の統合の課題について述べ、オントロジー工学理論に基づく解決手法を提案する。次に5.3節で、典型的な疾患を事例とした異常状態関連知識の統合を試み、提案手法の有効性を確認する。そして5.4節で、応用システムの試作について述べる。本システムでは、Linked Data 化した異常状態オントロジーと内外の既存のリソースを概念レベルに応じて連携することで、適切な概念にナビゲートし、関連情報を取得する機能を提供する。続く5.5節で本研究に基づく統合基盤構築の全体像を示し、その意義を考察する。

5.2 生命医学系の異常状態に関する知識の統合

近年、Web上のデータを相互に連携(リンク)させる Linked (Open) Data 技術を用いて公開されるデータ数はますます増加しており、生命科学ドメインでは、特にその動きは活発である [Whetzel 11]。また、臨床ドメインにおいても電子カルテシステムの導入など、情報のシステム化が急速に進められるとともに、組織を横断した情報共有と相互運用性の確保が不可欠となってきている。そこで、既存のオントロジーや用

語集の有用なデータを積極的に活用し、異常状態知識を統合することは多種多様な疾患のより深い理解に向けた知識基盤の構築において重要となる。

そこで本節では、まず、オントロジー工学的な観点から統合上の課題について議論するとともに、異常状態オントロジーを用いたアプローチについて提案する。

5.2.1 既存のリソースの統合における課題

現在、生命医学系では既に様々なオントロジーが開発されている。しかし、表層的な一致での統合では、必要とする関連知識の獲得が困難であり、結局、各リソース内でデータが散在化されたままで有用な知識が埋もれているという課題がある。そこで、本節ではオントロジー工学的な観点から、既存のリソースの統合における課題を考察する。

1) 概念レベルの不一致

専門家は、通常、専門領域に特化した用語集やオントロジーを構築する。そのため、自身の構築した概念の粒度や抽象レベルが外部リソースのものとは一致するとは限らず、そのままでは関連づける事が難しいという問題が生じる場合がある。

例えば、2.3.1.1 節で概説したように、PATO、HPO はともに表現型オントロジーとして知られている。PATO は生物種横断的に参照できる概念を提供することを目的としており、「部品の欠損」、「濃度減少」のような汎用的な概念を体系化している。そのため、ある生物種の特定の器官で起こる概念は PATO には含まれていない。一方 HPO はヒトに特化し、人体の各器官で観察される異常 (例：冠動脈閉塞、肺高血圧) を提供している。しかし、それらを一般化した概念は持っていないため、PATO で定義されているような概念レベルの大きく異なるものをそのまま関連付けることは困難である。

ここで、大規模数の概念をカバーする事で知られる臨床用語集 SNOMED-CT や MeSH を用いれば、これらのレベルをうまくつなぎ合わせることができるのではないかと考えられる。しかし、4.6.2 節で述べたように、これらの用語集は、もともと病理学会や国立図書館が開発を行ってきたため、オントロジー理論に準拠しておらず、上位オントロジーをもたない。そのため、is-a 関係自体のエラーや、多様な視点の下、多重継承を認めた分類階層となっている。その結果、概念間の関係が非常に複雑化し、必要な概念を同定することさえ困難である。したがって、これらの用語集に概念レベルのギャップを埋める役割は期待できない。

2) 特性と属性の不分離

次に、PATO、HPO は、特性 (P) と属性 (A) がオントロジー工学的な観点において識別されておらず、その結果、PATO の is-a 階層では上位概念が複数存在することを 4.6.2 節で指摘した。このような多重継承は、PATO と他のリソースと統合する際にも問題が生じる可能性がある。例えば、HPO の「冠動脈狭窄」という用語は、冠動脈の面積が縮小した状態を指す。そこで、PATO の「面積縮小」を上位概念との関連付けが適切と考えられる。ここで、PATO 「面積縮小」は、属性「面積」と特性「サイズ縮小」の2つを上位概念として持つ。そのため計算機で自動的に統合処理すれば、面積縮小の上位概念として「面積」も関連づけられ、その結果「冠動脈狭窄 is-a 面積」のような、異常状態としては間違った is-a 関係が推論されてしまう可能性がある。

3) 概念の不明確性：疾患と異常状態の混在

外部データと統合する場合、概念自体の意味が曖昧で不明確であると、統合の際に、本来は概念的に区別されるべきものが、システム上では同一と判断される恐れがある。

臨床ドメインでは、異常状態と疾患の識別は、電子カルテシステムにおける診療記録において特に重要となる。しかし、HPO では定義内容が記載されていないものも多い。また、MeSH では、例えば「心筋虚血 (Myocardial Ischemia)」の場合、定義では「心筋の血流不足から生じる心機能の異常」というように異常状態とされているが、同義語に「虚血性心疾患 (Ischemic Heart Disease)」という疾患が入っているというように、疾患か異常状態か不明なものが多い。さらに、MeSH、HPO および SNOMED-CT はいずれも疾患と異常状態が混在した階層化が行われている。

SNOMED-CT は、近年国際的な標準医療用語として各国の病院間の電子医療情報システム等に使用されつつある。しかし、SNOMED-CT の分類階層では、異常状態、疾患に加え、「所見」など複数の上位概念をもつことを認めている。概念が疾患か異常状態か不明なまま使用されると、記録自体の信頼性が懸念される。

表現型のオントロジーである HPO の場合においても同様に、表現型とは言えない疾患概念が混在している。例えば、「冠動脈狭窄」の上位概念は、「冠動脈疾患」となって、さらにその上位には異常状態や疾患が混在する。

このように、リソースを統合するには、疾患と異常状態というような概念の混在を避けることが必要であり、臨床の場での利用に耐えうるものとするには、その違いについて明示する必要がある。

5.2.2 異常状態オントロジーによるアプローチ

本異常状態オントロジーを用いた生命科学系リソースとの統合へのアプローチの概要は以下の通りである。

まず、統合における最初の課題 1) 概念レベルの不一致については、異常状態オントロジーの 3 層構造を用いて概念層毎に既存のリソースの異常状態と関連付けることで、リソース間の概念レベルのギャップを埋めことができると考えられる。具体的には、本異常状態オントロジーの第 1 層汎用異常状態は、PATO の汎用的な概念と対応すると考えられる。そこで、実際に PATO の対応概念へ関連付けることとした。

次に、第 2 層の対象物に応じて共通に現れうる異常状態では、異常の対象となるオブジェクト (対象物) に固有の異常状態が定義されている。HPO は人体の表現型を対象としているので、本異常状態オントロジーの第 2 層の下位に位置する人体の各器官依存の概念と対応するといえる。そこで、本第 2 層下位の異常状態概念を HPO の対応する概念に関連付けることとした。MeSH についても同様の理由から、本レベルと関連付けることとした。

さらに、第 3 層では、疾患を構成する異常状態を定義しており、例えば、「虚血性心疾患における心筋虚血」、「糖尿病における高血糖」のように、疾患コンテキストに依存して定義されている。そこで、これらと対応する概念が外部リソースで存在する場合、関連付けることとした。

続いて、第二の課題である 2) 特性と属性の不分離については、本研究では、3 章で述べたように、本研究の異常状態はすべて特性と属性を分離して、特性・属性分解形式で統一表現を行っている。そこで 1) で関連付けた外部リソースについては、異常状態オントロジーの異常状態の概念定義を参照することで、すべて特性と属性を識別が可能となる。

さらに、第三の課題 3) 概念の不明確性：疾患と異常状態の混在については、4.3 節で述べたように本異常状態オントロジーの異常状態は、疾患そのものとは明確に区別していることから、異常状態については、異常状態オントロジーで定義される異常状態とマッピングし、疾患については、2.2.1 節で概説した疾患オントロジーの各疾患とマッピングすることで、両者を分離することができる。

5.3 統合の試行

前節で示したアプローチが実際に統合上の問題を解決する手段となりえるかを検討するため、具体例を用いて、異常状態オントロジーの第 1 層から第 3 層まで体系化された異常状態と、疾患オントロジーおよび外部リソースとの統合を試みた。対象は、専門医との協議の下、3 診療科 (循環器、神経内科、消化器内科) の代表的な 12 疾患を構成する異常状態 107 とその上位概念をあわせた総計 386 の異常状態を選定した。

異常状態オントロジーと外部リソースとの関連付けは、表層上ラベルの文字列完全一致で機械処理した後、その結果を 4 章で議論したオントロジー工学的な観点に基づいて精査し、属性の継承関係を考慮しながらマッピング内容を修正することで行った。

5.3.1 概念層別マッピング

異常状態オントロジーの各概念層 (第1-3層)における外部リソース (PATO, HPO, MeSH, SNOMED-CT)との統合結果を表5.1(a)に示す。異常状態オントロジーの第1層汎用異常状態134概念のうち52概念について、PATOと1対1対応で関連付けられた。また、第2層の対象物に応じて共通に現れうる異常状態では、第2層上位の2概念とPATOの用語が対応することがわかった、これは、細胞壊死、組織炎症という複数の生物種に共通の器官で表れる汎用的な異常状態であった。また、PATOは第2層の下位にある人体の器官依存の異常状態に関する概念や第3層の疾患を構成する異常状態とはマッピングされないことが確認された。

次に、第2層の人体器官依存の異常状態はHPO, SNOMED-CTと対応する概念を見いだすことができるが、第1層の汎用的な概念はHPO, SNOMED-CTとは直接対応が付けられないことが確認された。MeSHもほぼ同様の結果であった。

表 5.1 異常状態オントロジーと外部リソースとのマッピングに関する統計

(a) 異常状態オントロジーと外部リソースとの概念レベル別マッピング数

概念レベル	異常状態 オントロジー	PATO	HPO	MeSH	SNOMED- CT
レベル1	134	52	0	2	0
レベル2	145	2	27	28	17
レベル3	107	0	0	0	0
総計	386	54	27	30	17

(b) HPO および異常状態オントロジーから PATO へのマッピング内訳

	HPO		異常状態 オントロジー	
	個数	割合	個数	割合
PATOへマッピングされていたもの	14	52%	20	74%
マッピングされないもの	13	48%	7	26%
合計	27	100%	27	100%
マッピングされないものの内訳				
対応概念がPATOに存在せず	7	26%	7	26%
対応概念がPATOに存在	6	22%	0	0%

ここで、異常状態オントロジーでは、4.3.3節で述べたように、第2層上位に管状構造物や弁構造物のような汎用構造物依存の異常状態を定義している。本概念は、PATO

の汎用レベルの概念と、HPOの人体の器官固有の概念とのギャップを埋める役割を果たすことがわかった。一例を挙げると、HPO「僧帽弁閉鎖不全」は僧帽弁の閉鎖機能が不完全な状態を指し、異常状態オントロジーとは第2層下位(人体構造物依存)の「僧帽弁閉鎖不全」とマッピングされる。一方、PATOの「機能不全」は第1層の「機能不全」とマッピングされる。ここで、異常状態オントロジーでは第2層において、人体の器官固有概念の上位概念として汎用構造物依存の異常状態である「弁閉鎖(機能)不全」を設けており、両者の中間層として医療ドメイン固有の知識を複数ドメインに共通する汎用的な知識まで円滑につなぐこと役割を果たしているといえる。さらに、本概念は「三尖弁閉鎖不全」など人体における様々な弁構造物の異常との関連性を見出す役割としても機能する。

5.3.2 外部リソース間マッピングとの比較評価

HPOは近年、Open Biomedical and Biomedical Ontology(OBO) Foundry内の各種リソースと関連付けを行い、各種オントロジーとの相互運用を試みている[Köhler 14]。そこで、異常状態オントロジーと対応するHPOの27概念のうち、HPOから直接PATOへマッピングされているものがあるか調査した。その結果、27概念中14概念(52%)について直接マッピングされていることがわかった(表5.1(b))。

次に、PATOへのマッピングがされていないHPOの13概念について精査すると、13概念中6概念について、PATO内に、本来ならば上位概念としてマッピングできる概念が存在することがわかった。例えばHPO「高血圧(Hypertension)」の場合、それに対応する「圧力増加(increased pressure)」という汎用的な概念がPATOには存在し、HPO「発熱(Fever)」では、対応する概念として「温度上昇(increased temperature)」がPATOに定義されている。しかし、これらは表層上のラベルが一致しないため、概念レベルの違いを超えてHPOからPATOへ直接マッピング処理する事ができずに残っていると推察される。

一方、異常状態オントロジーでは、これら6概念についてはすべてPATOへマッピングしており、本異常状態オントロジーを介した場合、HPOからPATOへのマッピング数は20概念(74%)となり、マッピング率が向上することがわかった。なお、本異常状態を介してもHPOからPATOにマッピングされない残りの概念については、胸痛や吐き気など、感覚異常に関する概念であり、PATOにはそのような概念に対応する汎用概念は提供されていないことが原因である。

次に、HPOからPATOに直接マッピングされている概念について、その整合性について検証を行った。概念レベルを超えて適切にマッピングするには、その特性を継承している属性とその値まで知る必要があるが、前章で述べたようにHPO、PATOでは属性に分解されていない特性形式のみで記述されている。そのため、関連付けの根

拠が暗黙的となり、説明性に欠ける恐れがある。そこで、HPO から PATO へのマッピングされていた 14 概念について、本異常状態オントロジーを介してマッピングされた PATO 概念との一致について比較を行い、マッピングの整合性について検証した。

その結果、HPO からマッピングされた PATO の 14 概念中、3 概念について異常状態オントロジーを介して関連付けられた PATO の概念と相違がみられた。

HPO から直接 PATO に関連づけられていた概念は、

1. HPO 「動脈硬化」 → PATO 「構造的過剰」
2. HPO 「心筋壊死」 → PATO 「構造的過剰」
3. HPO 「心不全」 → PATO 「部分プロセスの減少」

の 3 つであった。これに対し、本異常状態オントロジーを介すると、PATO への関連付けとして、各々、

1. HPO 「動脈硬化」 → 「血管硬化」 → 「管の硬化」 → PATO 「硬化」
2. HPO 「心筋壊死」 → 「筋壊死」 → PATO 「壊死」
3. HPO 「心不全」 → 「臓器不全」 → PATO 「機能不全」

となる (太字は異常状態オントロジーが提供する中間概念を示す。)。

そこで、1 番目の違いについて、3 つの概念「動脈硬化」、「構造的過剰」および「硬化」の概念間の関係性について考察する。

HPO からの直接マッピングの場合、「動脈硬化」から「構造的過剰」に、突然概念レベルを超えて、ヒトの特殊化された概念から汎用概念にマッピングされている上 (図 5.1-①), 直感的にも概念が適切につながっているとは言にくい。さらに、動脈硬化が構造的過剰へマッピングされている理由についても記述されていない。PATO には「硬化」という概念が存在するが、HPO 「動脈硬化」から PATO 「硬化」へのマッピング情報は確認できなかった (図 5.1-②)。

HPO はヒトで現れる表現型を対象にしていることから、疾患コンテキストを意識した観点で概念がつくられていると考えられる。そこで、「虚血性心疾患」の「動脈硬化」の因果関係を確認したところ、「動脈硬化」→「動脈血栓形成」というように、血栓という本来なかったものが生じることがわかる(図 5.1-③)。さらに、「動脈血栓形成」について、異常状態オントロジーの階層の上位概念をたどると、第 1 層では新生成物形成となり、その上位が「構造的過剰」で、PATO の「構造的過剰」とマッピングされることがわかる(図 5.1-④)。

臨床知識では、よく使う複数の概念を 1 用語としてまとめて扱うことが多い。例えば「イレウス」という用語は「構造異常が原因で起こる閉塞」という意味内容を持つように因果関係をセットで 1 概念とする。動脈硬化についても、通常、血栓など動脈に過剰物が蓄積される事が観察されるため、「構造的過剰」にマッピングする事は、当該ドメインでは自然と考えられるかもしれない。しかし、知識統合においては関係付

けた根拠は明示すべきであり、それが分野横断的に知識を統合し、相互運用を実践するための信頼性確保に役立つ。

本研究の異常状態オントロジーでは、第1層では、材質の異常状態として「硬化」<硬さ属性 (A), 硬い (V)>を定義している。第2層では、第1層「硬化」の本質属性を継承しながら、対象物依存の概念として「管の硬化」から「血管硬化」、「動脈硬化」というように一貫性をもって汎用レベルから器官依存のレベルまで特殊化を行っている(図 5.1 中央)。そのため、本異常状態オントロジーでは、概念レベルのギャップを埋めながら横断して客観的に各々のリソースの関連づけを行うことが可能である(図 5.1-⑤)。ここで、第1層の汎用異常状態「硬化」からは PATO 「硬化」へマッピングし(図 5.1-⑥)、第2層の「動脈硬化」からは HPO の「動脈硬化」とマッピングしていることがわかる(図 5.1-⑦)。したがって、HPO から本研究の異常状態オントロジーの is-a 階層を参照すれば、自身の概念「動脈硬化」の本質属性をもつ上位概念として適切な PATO の概念「硬化」を知ることができる。さらに、HPO とマッピングされた第2層下位(人体構造物依存)と、PATO とマッピングされた概念レベル(第1層)を知ること、両者の概念レベル上の立ち位置を把握することが可能である。

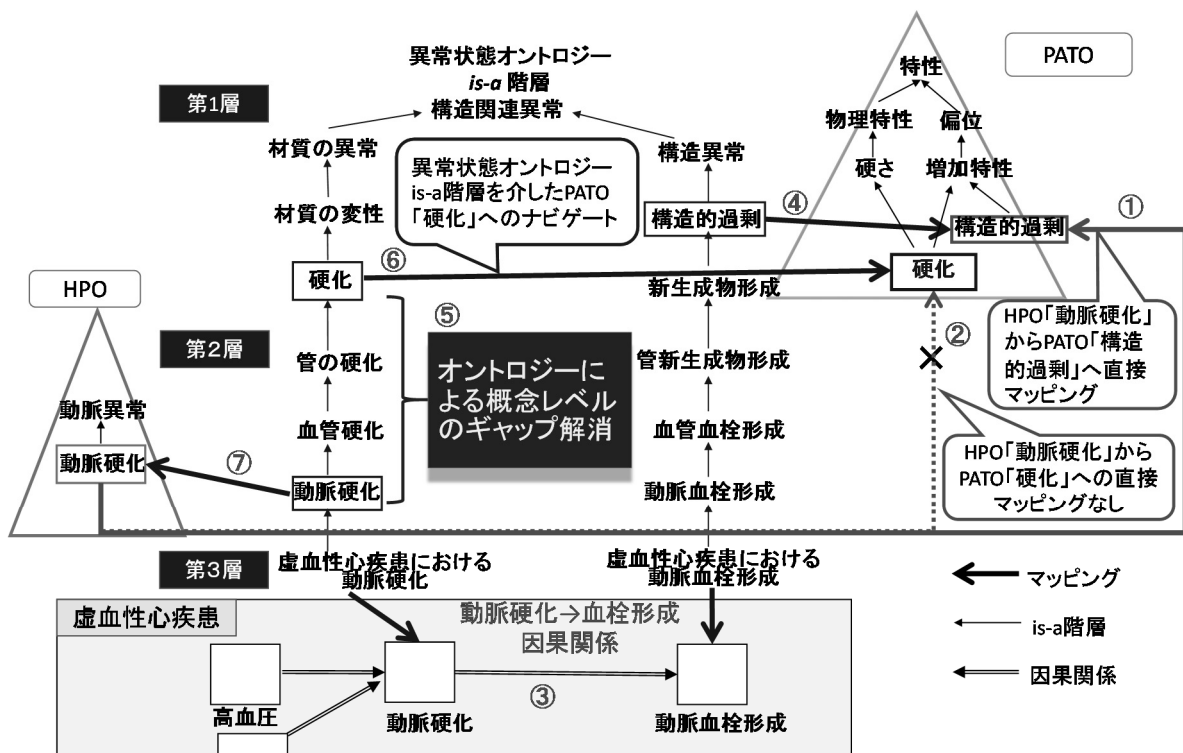


図 5.1 既存リソースと異常状態オントロジーとのマッピング

このように、本異常状態オントロジーは、すでに関連付けられたリソース間の互いの立ち位置を客観的に示すとともに、概念レベルの切り替えとコンテキストの明示により、その結びついた一対一の間接関係を多様な観点から検証することで整合性を検証し、異常状態の観点において適切な概念にナビゲートする役割を果たすことがわかった。

(1) 概念の明示による信頼性の向上

本異常状態オントロジーは特性を属性分解する表現形式をとっているため、その性質がどのようなパラメータ属性と値をもつかが明示される。例えば「狭窄」のような臨床用語も、本異常状態表現モデルでは、〈断面積 (A), 小(V)〉と属性とその値に分解されることで、概念の透明性が増すとともに、どのような汎用異常状態を上位概念としてもち、それを特殊化したのが客観的に把握可能である。その結果、自身の概念構築時における ad hoc 性を排除するだけでなく、他の外部ユーザーにとっても観点が明示されているため、意味内容が理解しやすいというメリットがあると考えられる。今回の試行実験では、特性と属性の分離により、本異常状態と関連付けたとしても、PATO のように「狭窄 is-a 面積縮小 is-a 面積」というような推論でエラーは起こる心配はないことを確認した。

Linked Data では、任意のデータのリンクを許しているが、不明瞭な概念では再利用が困難となる。概念レベルと概念の捉え方を明示でき、それをユーザーが参照できるような枠組みがあれば、なぜ関連付けられているかが解釈可能となり、信頼性が高まる。そのような対象の理解のための視点を与え、互いの関連性付与を支える骨格がオントロジーであり、安心して実践で使えるための妥当性を担保する。

異常状態オントロジーは異常状態一つ一つを対象としてオントロジーを構成している。また、疾患概念にはとらわれずに、状態としての本質に着目して一つの原理 (視点) でオントロジーが構築されている。したがって、各リソースに現れる概念と対応する本異常状態オントロジーの概念をマッピングし、本異常状態オントロジーの is-a 階層を核として参照すれば、異常状態の本質のみを捉える事が可能となる。本モデルの 3 層構造は、異なった概念レベルを統合し、リソース間のギャップを埋める役割を果たすと同時に、レベル横断的に様々なデータへのアクセスへの利便性を向上させる役割も果たすと考えられる。

(2) 対象物視点に基づく動的 is-a 生成

HPO では、2.3.1.1 節で示したように、表現型がどの器官や組織で観察されるかという視点に基づいて階層を構築しており、「〇〇器官の異常」(例：血管の異常)のように器官のみが特定された異常状態概念が多く存在する。本異常状態オントロジーでは、そのような概念について、異常状態オントロジー内に構築すると多重継承の恐れがあるため、現在は異常状態オントロジーで定義されていない。しかし、これらの概念は臨床上しばしば登場し、実用上必須となることがわかっている。そこで我々は、

システム利用時に、異常状態の対象物 (O) としてもつものを集め、さらに、既に開発済みの、1つのオントロジーから必要に応じて指定した視点に基づいて is-a 階層を動的に構築するという「動的 is-a」生成技術 [古崎 12] を用いて対処することとした。例えば、利用者が血管の異常について知りたい場合、本研究の異常状態オントロジーで定義された異常状態のうち、「血管」を対象物 (O) としてもつものを計算処理して集めることが可能である。さらに、それを用いて、本研究プロジェクトの解剖学オントロジーの階層と組み合わせることで、各器官の下に再分類した is-a 階層を動的に生成することも可能である (図 5.2)。

このように、部位別に生成された異常状態を用いて、対応する HPO の「動脈異常」、「冠動脈異常」というような各器官の異常という各器官の異常とのマッピングができることを確認した。

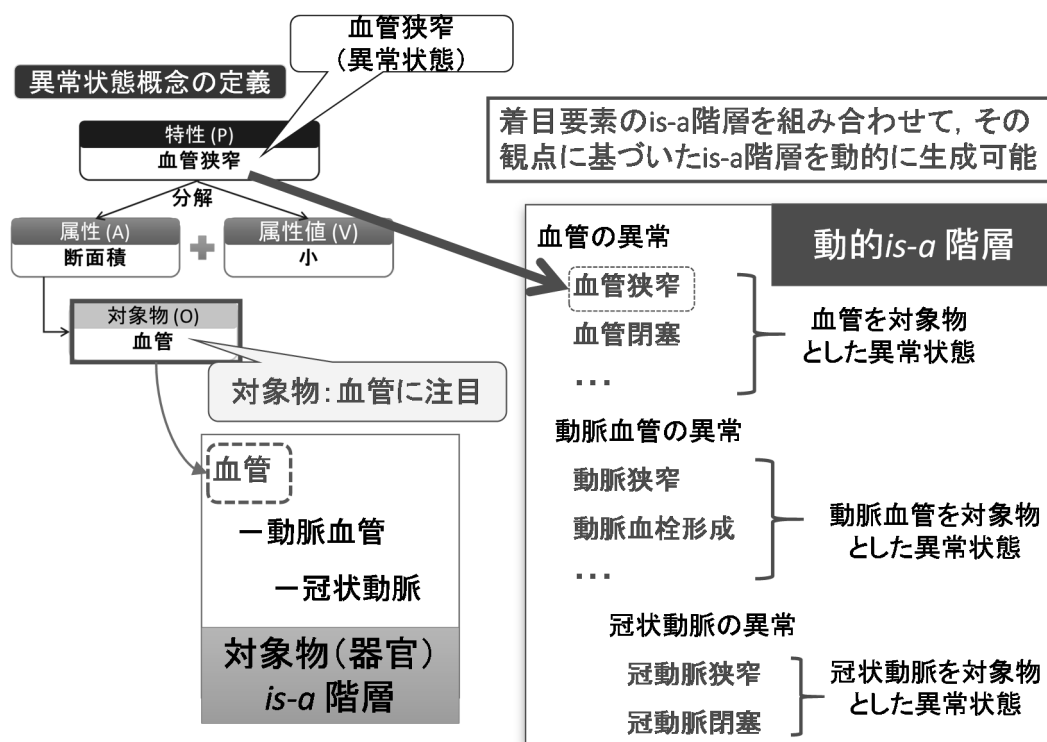


図 5.2 対象物視点の動的 is-a 階層

(3) 試行結果の考察とまとめ

試行実験では対象とした数が少数ではあるが、異常状態オントロジーは、既存リソースに対し、1) 概念階層における適切な概念レベルの付与、2) 統合に関する整合性の検証およびサポート、3) レベル横断的に適切な概念へのナビゲート、4) 統合の根拠となる本質属性の明示、などを提供できる事を確認した。これらの成果は概念数がスケールアップしても適用可能であり、知識統合基盤づくりへ貢献できると考えている。

外部リソースとのマッピングについては先行研究 [瀬川 08] で機能語彙体系間でのマッピングにおける考察と同様、概念定義の違いや、視点の違い、あるいは粒度の

違いから対応関係に非対象性が生じることがわかっている。今後、類似する分類軸によるグループマッピングに基づいた相互互換の利用を設けるなど関連付けの成功率を高める仕組みが必要となる。

5.4 生命医学系リソース間ナビゲーションシステムの設計

前節の結果から、異常状態知識の統合には、本異常状態オントロジーを骨格として用いて各種リソースと関連付けることが有効なアプローチであることがわかった。そこで、統合した結果を用いて、各種リソース間をナビゲートする応用システムを試作することとした。システム開発については Web 上で構造化されたデータを相互につながる技術として知られる Linked Data [Heath 11] を用いることとした。

5.4.1 異常状態オントロジーの Linked Data 化の概要

異常状態オントロジーは「法造」を用いて記述されている。そこで、異常状態オントロジーの Linked Data 化にあたり、法造が提供する OWL エクスポート機能を用い、ロールホルダー間の is-a 関係の表現をサポートしつつ、ロール理論の詳細表現を省略することで記述を単純化したモデルを利用した [周 08, Kozaki 07]。これにより、OWL 形式で出力された異常状態オントロジーにおいては、異常状態の is-a 階層は `rdfs:subClassOf`, `attribute-of` 関係で表されていた異常状態の対象物 (O), 副対象物 (So), 属性 (A) および属性値 (V) は、`owl:Restriction` を用いて表される。

この OWL 形式で出力された異常状態オントロジーと、3 章で述べた外部リソースのマッピング情報は、5.2.1 節で議論したように概念レベルが必ずしも一致しないことを考慮して、リソースの種類毎に `medo:hpo` や `medo:pato` といったプロパティを導入して表現した。これらのマッピング情報は、今後、新たな情報が追加されることを考慮して、異常状態オントロジーとは独立した RDF ファイルとして管理している。

5.4.2 疾患コンパスと異常状態オントロジーの連携

これまで疾患オントロジーでは、疾患を構成する異常状態の因果連鎖 (疾患連鎖) に関する知識を Linked Data 化し、外部リソースと連携して疾患情報を閲覧できる Web アプリケーション「疾患コンパス」を公開している [古崎 14]。そこで、本異常状態オントロジーの Linked Data と、疾患コンパスにおける疾患知識を関連付けることとした。

疾患オントロジーにおける疾患定義は、異常状態オントロジーで定義された異常状態概念を参照して定義されている。よって、異常状態オントロジーにおける異常状態を、疾患コンパスにおける第3層の疾患を構成する異常状態に対応づけた。技術的には、疾患コンパスが提供する API を用いて連携させることとした。

5.4.3 応用システムの試作と実装

3章と4章で述べた異常状態オントロジーの OWL ファイルおよび外部リソースとのマッピング情報を、トリプルストア に格納し Linked Data として公開した。現在、更なる統合の実現に向け、URI の設計方針や参照解決可能な公開ページの開発、データ公開の範囲など、臨床医学プロジェクト内で検討をすすめている。一方、OWL や SPARQL エンドポイントを用いた必要な情報の取得は、臨床医学のユーザーにとっては通常、あまり馴染みがないため困難であるといえる。そこで、必要なデータを容易にアクセスし、利用できるように、専用の閲覧システムを開発した (図 5.3)。

本システムは、JavaScript を用いた Web アプリケーションとして実装されている。システム上に表示される異常状態知識は、すべて、SPARQL クエリにより動的に取得されている。例えば、異常状態の定義内容を表示する際には、owl:Restriction の表現に用いられる空白ノードを含む RDF トリプルのグラフ構造を勘案した SPARQL クエリを発行している。この実装方法により、異常状態オントロジーに含まれる全ての情報を正しく取得し、表示できることが確認できている。合わせて、疾患コンパスで公開されている疾患知識や、マッピング情報を介した外部リソースとの連携も実現されている。すなわち、本システムは、統合された異常状態知識を利用した応用システム開発に向けた、基本機能を実装した試作システムと位置づけられる。

5.4.4 ナビゲーションシステムの機能

試作システムでは、入力された用語について、異常状態オントロジー内部データセットの検索、疾患コンパスへのリンク情報、および外部リソースへのリンクが表示される。図 5.3-①に示すように、「異常状態 (Abnormal state)」、「対象物 (Object)」、「副対象物 (SubObject)」、「属性 (Attribute)」、「属性値 (Attribute Value)」を条件として指定して検索ができる。検索対象は、さらに構造関連異常 (Structural Abnormality)、機能関連異常 (Functional Abnormality) などから異常状態をフィルタリングすることができる (図 5.3-②)。また、異常状態の is-a 階層の表示 (図 5.3-③) から異常状態を選択することも可能である。検索結果 (図 5.3-④) については、いずれかを選択すると図 5.3-⑤の

1) 異常状態の is-a 階層を用いた閲覧機能例

試作システムでは、異常状態オントロジーの内部データセットの検索により、is-a 階層の上位/下位概念を利用した診療科横断的に、複数の疾患で起こりうる異常状態の共通性が可視化される。例えば、「脳動脈狭窄」と「冠動脈狭窄」は神経内科と循環器内科の各領域の疾患に依存した異常状態として扱われるが、いずれの上位概念も「動脈狭窄」であり、汎用レベルでは「面積縮小」という概念をもつことが異常状態の is-a 階層で表示される。また、「ポリープ形成」の上位概念は「新生成物形成」であるが、「結石形成」、「嚢胞形成」、「腫瘍形成」も「新生成物形成」を上位概念としていることで共通するということが is-a 階層からわかり、構造体に何らかの構造物が生成されたことが、定義内容をみればわかる。このように、オントロジーの統一表現による概念定義と is-a 関係を最大限有効活用した機能となっている。

2) 異常状態の構成要素を用いた検索例

異常状態の構成要素である対象物 (O)、属性 (A)、および属性値 (V)を用いれば、観点に応じた検索を行うことができる。例えば、ある器官で起こった異常状態を検索したい場合、対象物 (O)に探したい器官を入力すれば、検索結果として、器官依存の異常状態が得られる。検索例として、「冠動脈」を入力すると、「冠動脈閉塞」、「冠動脈壁肥厚」、「冠動脈狭窄」、「冠動脈攣収縮」が結果として表示される。

次に、特定の属性に注目し、例えば「長さ(A)」に関する異常状態を集めよというクエリについては、「管内径縮小」などの他、「冠動脈壁肥厚」、「心室壁菲薄化」、「胃粘液層薄化」など長さのロール概念「厚さ」も含めた結果を診療科横断的に取得することが可能である。

3) 外部リソースとの統合による関連情報の取得例

今回対象とした 12 疾患で検証した外部リソース 4 つとマッピングを行ったものについては、OBO Foundry ライブラリ内各種オントロジーの Linked Data サーバーブラウザである Ontobee [Xiang 11, Ontobee] または外部リソースが提供するサイトの対応データにアクセスすることが可能である。例えば、弱力/力低下には、定義内容表示ペイン内に PATO の ID が記載されているが、これは PATO の URI へリンクがされており、Ontobee のブラウザ上で関連づけられた各種データを閲覧できる。同様に、試作システムの定義内容表示ペインの MeSH ID をクリックすると、文献検索サービスの PubMed [PubMed] サイトの該当用語の情報にアクセスすることが可能である。例えば、心筋虚血の例を示す (図 5.4)。異常状態オントロジーの「心筋虚血」状態は、MeSH タームの心筋虚血 (MeSH ID:D17202) がマッピングされている。マッピング先の NCBI では PubMed で「心筋虚血 (Myocardial Ischemia)」がアノテーションされている生命医学の関連文献を検索が可能となっており、363000 件の心筋虚血に関する文献情報の取得ができる (2015 年 5 月 21 日調べ)。

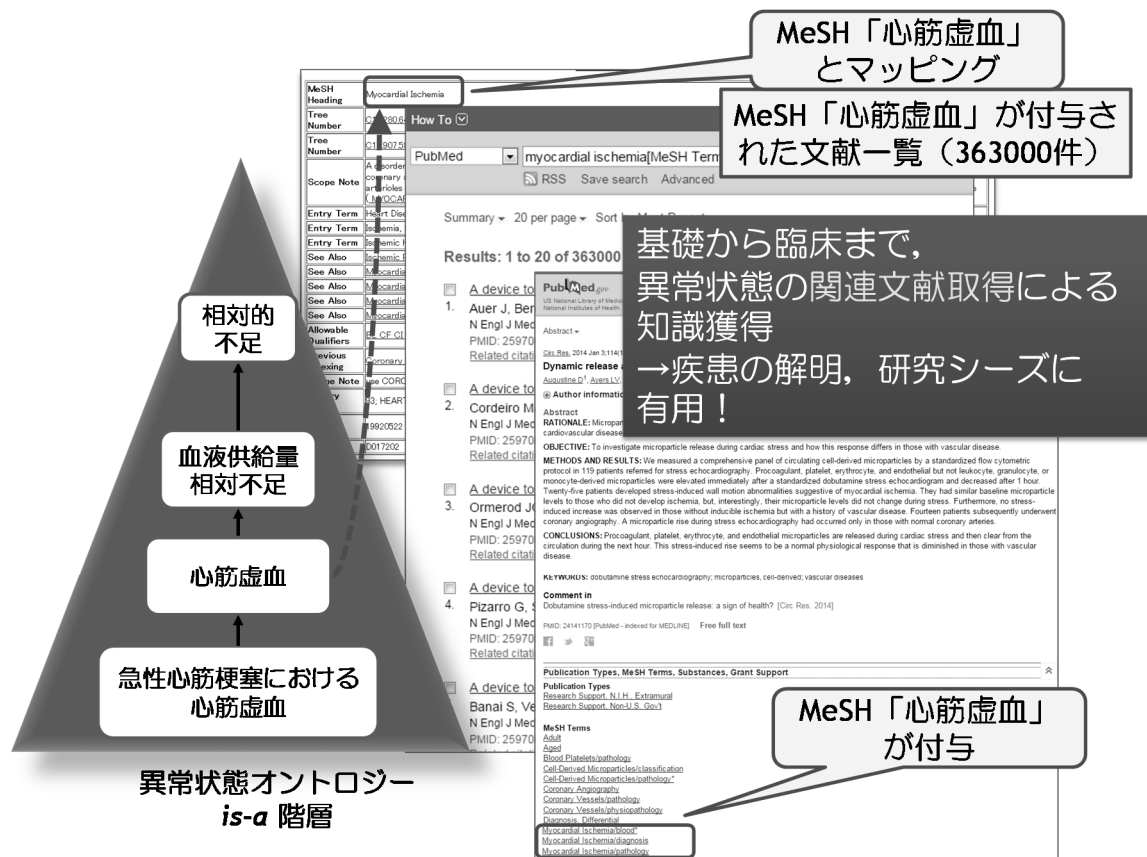


図 5.4 文献統制語彙集の連携による文献情報の活用例

このように、ある異常状態について、生命医学系の関連文献と統合することで、他の実験動物での知識など、基礎から臨床までの最新知見を知ることができるので、研究シーズの発掘にも有用である。

4) 疾患コンパスとの連携による知識獲得

試作システムでは、疾患コンパスとの連携により、疾患定義より異常状態が上流か下流に位置するかによって、疾患の原因となるのか、結果として現れる異常かが疾患コンパスの図から把握可能である。例えば、異常状態「高血圧」は複数疾患へ関連付けられている。そのうち、疾患定義に高血圧を含むものは循環器内科の「高血圧症」とその下位疾患の他、腎臓内分泌内科の「良性腎硬化症」でも疾患定義に高血圧を含み、その疾患を発症した患者には共通して高血圧持続が見られることが、疾患コンパスでは明示される。また、疾患「大動脈瘤」では、高血圧は上流に位置し、高血圧が一因となり「大動脈粥状硬化」を引き起こすことがわかる。一方「腎動脈瘤」疾患では、高血圧は下流に位置し、「腎動脈に発生した動脈瘤」が起こった結果、高血圧は症状として現れる。このように、本システムでは、同じ一つの異常状態とされていても、原因となる場合と結果となる場合を、各疾患での出現ごとに差異が明示できるので、

早期治療の対象とすべきか、あるいは対症療法とすべきかなど疾患に応じた治療指針の確立にも貢献できると考えている。

次に、複数の疾患を比較することで、因果関係の共通性も認識できる。例えば、虚血の場合、循環器内科の疾患「心筋梗塞」と神経内科の「虚血性脳血管障害」では、心臓と脳において部位は異なるものの、途中虚血までのパス (因果連鎖)はほぼ同じであることがわかった(図 5.5)。いずれも、狭窄、閉塞などの構造的異常がおこるか、攣

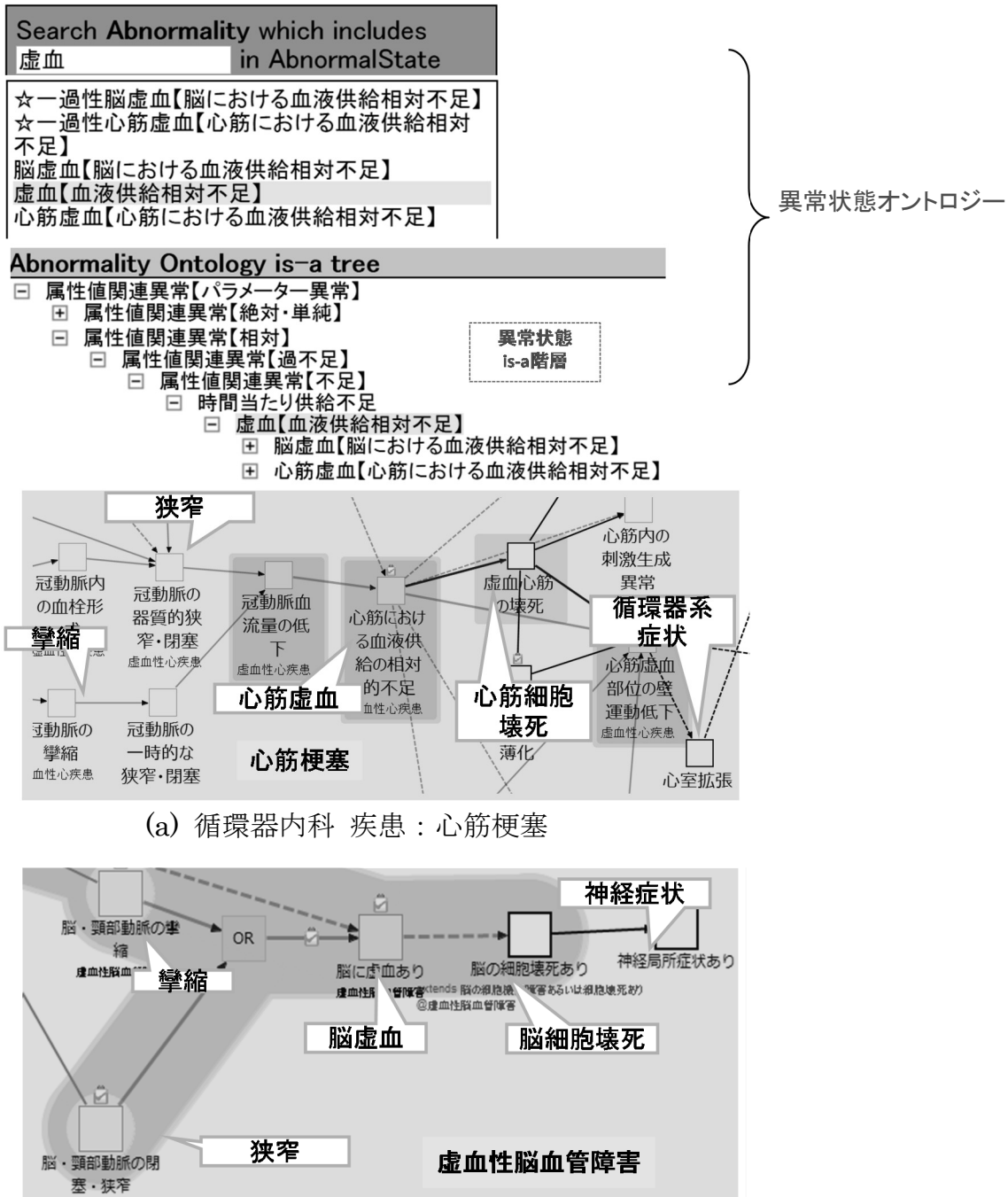


図 5.5 異常状態オントロジーと疾患コンパスの連携例

縮によって過収縮が起こり、(血流量が低下し), 虚血がおこり, 虚血後の壊死が生じるといふ因果をたどることが共通する. その後, 心筋壊死か, 脳の細胞の壊死か部位によって下流が大きく異なることが疾患コンパスの因果関係によりわかる. 前者は, 心臓の心室の壁運動低下から循環系に様々な症状を引きおこすが (図 5.5 (a)), 後者は脳から神経に影響を及ぼし, 神経が支配する領域で麻痺など, 循環器の場合とは異なる固有の状態を引きおこす (図 5.5(b)).

さらに, 本応用システムにより, 本来, 互いに関係があるが現在の臨床では知識が互いに関連付けられていない概念同士について, それらの関係性を発掘でき, 知識を獲得できることがわかった(図 5.6). 例えば, 「心不全」については, 臨床では主に循環器内科の専門医が扱っている知識といえるが, 本システムでは診療科横断的に疾患で起こりうる異常状態を汎用連鎖として把握可能である. その結果, 循環器内科以外の診療科の情報として, 例えばアレルギーリウマチ内科においても, 全身性強皮症では「自己免疫」が原因で「心不全」が起こるといふことを提供できる. また, 腎臓内科の後天性腎動静脈瘻という疾患においても, 「腎動静脈短絡」という本来つながっていない動脈と静脈がつながってしまうことが原因で, 心臓の機能が不全となる状態である「心不全」が結果として引き起こされるということが明示できる. これらは, 循環器内科の教科書にはあまり記述されず, 各々の診療科としての各疾患固有の知識として通常扱われている. さらに, 腎動静脈瘻は希少疾患として知られていることから, 「腎動静脈短絡」が原因で心不全が起こることが見逃されやすいといえる. 本システムで複数疾患を横断して, このような起こりうる原因を発見できるということは, 臨床においても非常に有用であると考えられる. 従来のように, 診療科毎に個別に扱っていた疾患固有の異常状態に対し, 本研究では, 疾患で起こる原因とそこから進行する病理学的プロセス, その結果おこる症状について診療科横断的に把握可能とする. 本成果は, 臨床の場の疾患像を適切に把握することを支援するという点においてその意義は大きい.

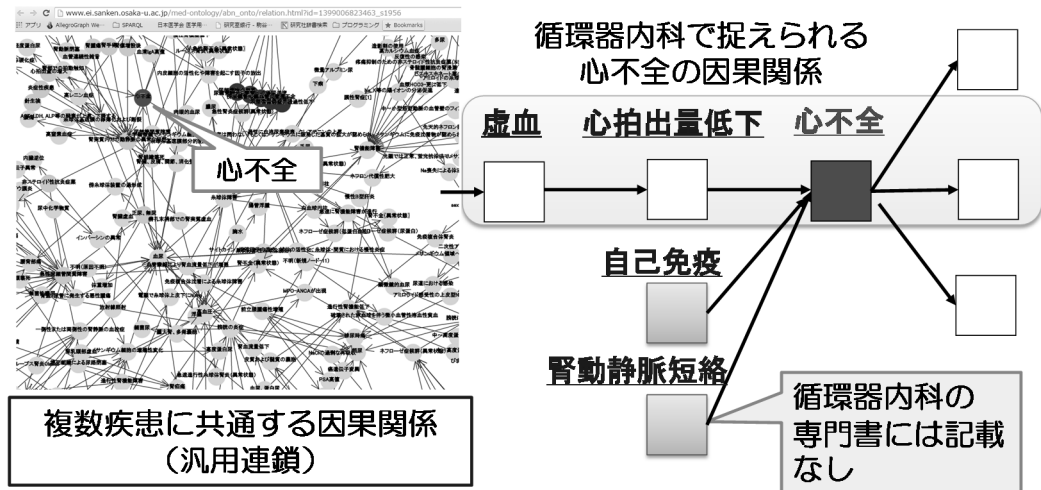


図 5.6 心不全の起こりうる原因, 結果の例

5.5 知識統合基盤の全体像

本節では、異常状態オントロジーを媒介とした各種リソース間ナビゲーションによる異常状態に関する知識統合基盤の全体像を示すとともに(図 5.7), その意義について考察する.

5.5.1 生命医学系リソース間ナビゲーションとその意義

本節では、外部の生命医学系リソースとの統合におけるその意義について考察する. 異常状態オントロジーは、汎用的な概念から疾患固有の概念までカバーしており、基礎研究の段階から医療の実践へと橋渡しを支援する役割が期待される.

本研究では、異常状態オントロジーを介して、生物種をヒトに特化した表現型 HPO と複数の生物種で利用できる汎用的な表現型 PATO との統合を試みた. 今後、さらに他の実験動物との関連付けを進めたいと考えている. 生物種を超えたマウスやラットの表現型とヒトの表現型との対応については、属性 (A) とその値 (V) のセットが一致した上で、対象物 (O) である解剖構造物同士についても類似性を確かめなければならない. 現在、マウス、ラットの表現型データベースの統合が試みられているので [枘屋 13], それらを利用し、Linked Data 化が進めば、生物種横断的なマッピングが可能になり、ヒトの疾患で起こる各種症状からそれに対応する実験動物で観察される概念までを統合化した知識を提供でき、所望の情報へナビゲートできると期待される (図 5.7).

また、前節で示したように本研究では MeSH タームとの統合により、生命医学の関連文献情報が入手できる. 例えば、患者で観察されたある症状に関する文献を生物種横断的に収集すれば、それにより新たな知見を取得できる可能性がある. 別の例として、ある疾患について注目した異常状態に関する文献を疾患横断的に集め、既に記述した因果関係について新たな原因を推論するなどの知識獲得にも有用である. さらに基礎研究で得られた知見を希少疾患の解明に役立てるなど、本オントロジーを骨格として統合する事で基礎から臨床へさらなる研究の発展に役立つ可能性がある.

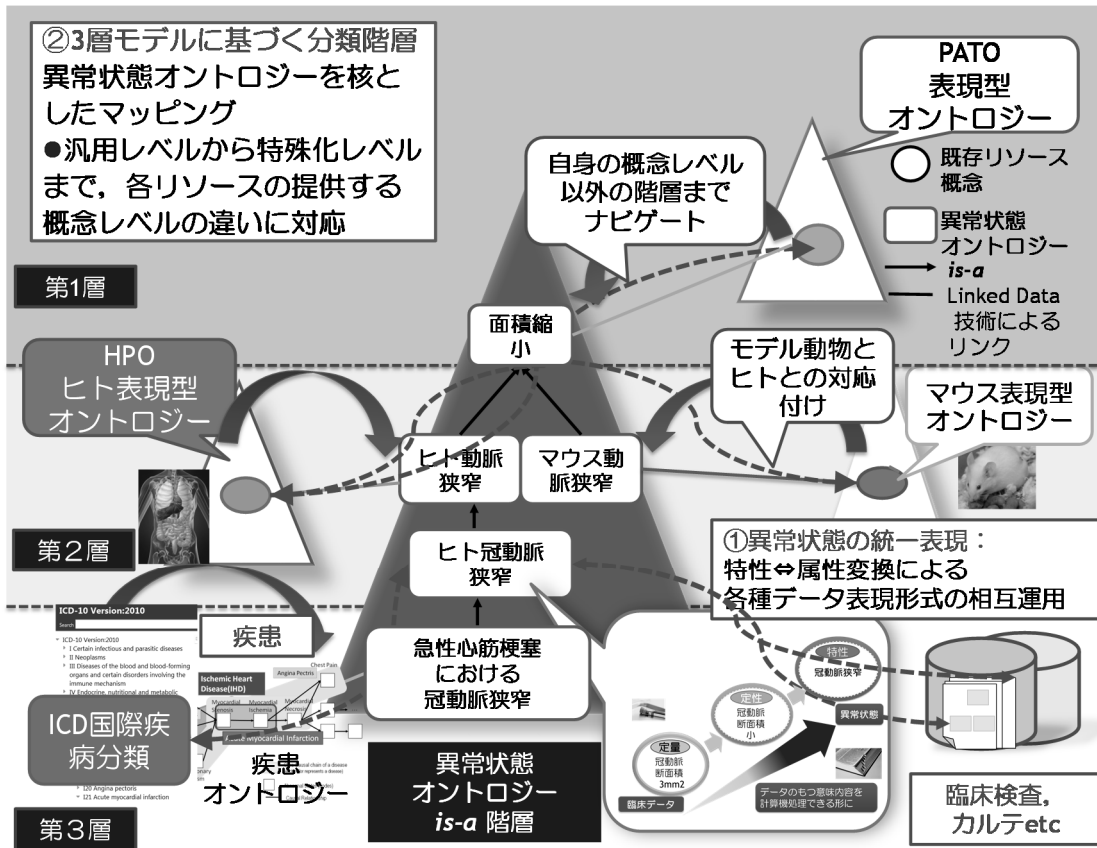


図 5.7 異常状態に関連する知識統合基盤の全体像

5.5.2 臨床的な意義

異常状態オントロジーによる関連知識統合により、臨床では、以下の点において有用であると考えられる。

まず、臨床医が利用する場合、本異常状態オントロジーでは診療科毎の疾患固有の異常状態の把握に加え、第2層では人体構造物固有の異常状態の共通性が把握できる。そのため、例えば、代謝内分泌科の糖尿病と循環器内科の虚血性心疾患に共通して起こる異常状態を解析したり、解剖学オントロジーとの連携により特定の器官で生じた異常状態を診療科横断的に把握したりすることが可能であり、治療方法の検討に有用であると考えられる。さらに、診療記録に活用し、薬剤情報と連携すれば投薬の際に副作用としてどのような異常状態が起こったかなど診療科横断的に調査することが可能になると期待される。

5.4.4 節で述べたように、疾患オントロジーで定義された異常状態の因果関係を一般化することで、診療科横断的に疾患で起こりうる異常状態の因果関係を汎用連鎖として取得可能である。ここで、各疾患の因果における異常状態を一般化するには、本研究で提案した統一表現モデルによって各異常状態の記述を統一することが鍵となる。たとえ、専門の診療科では知られている疾患であったとしても、希少疾患の場合、その知識が診療科内に特化した表現形式では、他の診療科での情報取得は困難となる。オントロジー理論に基づく厳格な定義に基づいた異常状態オントロジーによる統一表現と疾患オントロジーが連携することではじめて、診療科横断的に異常状態を統合し、複数疾患で起こりうる因果の可能性をすべて導出することができるといえる。従来、臨床医のもつ背景知識や教科書等で、疾患固有の知識として断片的に把握されてきた異常状態について、診療科横断的に異常状態間で起こりうる因果関係（汎用連鎖）の取得を達成した成果は、国内外含めて既存研究に類をみないものであり、臨床医学の知識の進歩への貢献は大きいといえる。

次に、外部リソースの活用の点において、疾患については、WHO が提供する国際疾病分類 ICD(International Classification of Diseases) [ICD] との関連付けを予定している。本分類は、国際死因分類に関する統計の他、近年は各種統計調査や医療機関における診療録の管理等に使われている。現在は第10版 (ICD 10) であり、2017年改定予定の第11版 ICD11 では、疾患にその「原因因子 (Causal Properties)」を記述する予定となっているので、将来、それらの中の異常状態とマッピングできればと考えている。ICD10/11 の各疾患データと疾患オントロジーの疾患概念を対応づけることで、関連付けを介して注目した疾患でどのような異常状態の因果連鎖が起こっているかを把握できる。また、本分類は、国際死因分類に関する統計に用いられているため、例えば、国レベルで重点的に死亡率の高い疾患でおこる異常状態について、さらにその症状と原因の関係を把握することで早い段階からの治療介入による医療費のコスト削減などが期待される。

異常状態オントロジーでは HPO と関連付けているが、HPO は遺伝病データベースである OMIM (Online Mendelian Inheritance in Man) [Hamosh 05] に関する情報がある。そこで、OMIM の情報を活用することで、原因遺伝子などの情報も取得できれば遺伝子レベルでどのような異常を引き起こし、その結果身体に異常が現れるか解析可能となり、疾患の解明に役立つと考えられる。

臨床では、患者に観察される症状を適切に把握する上で、その対象となる人体各器官の情報が欠かせない。そこで、異常状態の構成要素である対象物(O)については、外部では著名な FMA (Foundational Model of Anatomy ontology) [Rosse 03] の人体解剖構造物の各概念と関連付けなければならない。また、本研究グループの解剖学オントロジーとの連携も進めており、各疾患で起こりうる異常状態について、隣接臓器や血管の接続情報などを知ることができることから、影響する器官の状態予測や解析にも有用であると考えられる。さらに、検査データについては、例えば、国際的には 2.3.1.2 節で紹介したような LOINC や国内では医療情報開発センターの提供する MEDIS 標準マスター [MEDIS 03] が検査に関する標準用語やコードの提供を行っており、将来これらと統合すれば、患者群の病態と検査データとの比較解析、相関分析等に有用であると考えられる。異常状態オントロジーでは、検査データから教科書で記述されている異常状態までを体系的に把握できることから、疾患を理解するための医療従事者への教育支援への利用にも有用であると考えられる。

5.6 結言

本章では、異常状態オントロジーを媒介とした生命医学系リソース間ナビゲーションを達成するための応用システムについて提案し、ナビゲート可能であることを確認した。

まず、既存リソースの異常状態関連知識統合のための課題をオントロジー工学の観点から述べるとともに、その解決法として異常状態オントロジーの概念レベルに応じた外部リソースとのマッピングについて提案した。さらに、統合実現に向けて異常状態オントロジーの応用として、Linked Data 化を行い、知識統合に向けたシステム開発の取り組みについてのべた。

既存のリソースとの統合が成功すれば、異常状態オントロジー自身のデータ拡充にもつながるといえる。しかし、表層ラベルのみの一致で関連付けを行うと、意味内容は一致しないことも多い。また、データ源にアクセスしてもその概念定義や属性が明示されていない場合があり、概念階層構造にエラーが生じている可能性もある。そこで、データに健全性を与え、知識統合の骨格となるのがオントロジーである。オントロジーは知識を明示し、客観的かつ一貫性のある本質属性を捉えた厳格な階層構造により、質の高い概念構造を保有する。オントロジーの Linked Data 化は、誰もが簡単

にオントロジーにアクセスし、参照するために有効な手段となる。Linked Data 化したオントロジーを骨格として参照すれば、汎用レベルから専門ドメインまでの知識を自由に横断して、各種リソースの関連データについて信頼性のある情報取得が可能となる。本システムは汎用から疾患固有の知識まで概念レベルに応じて関連知識の統合を可能とし、基礎と臨床における知識の橋渡しの役割が期待される。

今回の統合試行においては3診療科を横断した知識を把握することができたが、今後スケールアップすれば、より多くの診療科を横断した複数の疾患の異常状態の共通性が明らかになると考えている。応用システム開発の取り組みでは、従来、関連付けがなされずに各診療科に分散していた異常状態の概念同士を結びつけ、これまで隠れていた知識を発掘するとともに、臨床の専門家が持つ疾患のイメージに即した意味内容を明示し、提供できることを示した。

第6章 結論

本章では、本研究で得られた成果を総括し、今後の展望について述べる。

6.1 研究の総括

本研究では、疾患記述において必要な異常状態の意味内容を表現し、理論的に構造化するための異常状態オントロジーの構築について提案した。さらに、異常状態オントロジーの応用研究として、既存のリソースにおける異常状態に関する知識を統合するためのシステムの理論的な枠組みについて提案した。

異常状態を適切に処理するには、二つの課題、すなわち異常状態の多様性への対処、診療科横断的な異常状態の共通性の把握が存在する。さらに、既存のリソースに存在する異常状態関連知識の獲得という課題においては、情報の散在という問題がある。そこで、これらの課題解決に向けたアプローチとして、前者2つについてはオントロジーの構築において、特に1) 異常状態の統一表現、および2) 異常状態の階層的モデリングを検討し、後者の課題については、3) 異常状態オントロジーを媒介とした生命医学系リソース間ナビゲーションについて取り組んだ。その結果、疾患を理解するための異常状態の体系化における理論的枠組みを確立するとともに、既存のリソースを統合し、疾患記述に必要な情報を獲得する枠組みを実現し、臨床における有用性を確認した。

第2章では、従来研究に対し、本研究の位置づけを議論した。まず、臨床医学オントロジーで開発されている疾患オントロジーおよび解剖学オントロジーについて、本研究の異常状態オントロジーとの関係性について示した。次に、従来研究について、異常状態を含む生命医学系オントロジーと、異常状態の統一記述に必要となる特性、属性や属性値などの基本的な概念を扱う上位オントロジーを対象として、本研究の3つの目的の観点から問題点を指摘し、本研究の位置づけを明らかにした。

第3章および4章は、本研究の提案する異常状態オントロジーの構築について述べた。

第3章ではオントロジー構築において、最初の課題である異常状態多様性に対処するため、その基礎となる1) 異常状態の統一表現モデルの提案とその記述方法について述べた。

提案モデルでは、まず臨床における様々な異常状態に対して YAMATO の性質の表現に基づき、大きく定量的な表現、定性的表現、特性表現の3種類に分類できることを明らかにした。次に、疾患を理解するために必要な異常状態に関する考察から、異常状態を特性表現と対応付けて捉え、概念定義することが必要であることを議論した。さらに、特性・属性分解の手法を用い、特性を属性とその値に分解することにより、

具体値をもつ定量的な情報との関係性の明示を可能とした。これにより、検査データから疾患記述で示される異常状態まで多様な異常状態に対し、一貫性をもって統一的に記述できることを示した。加えて、上位オントロジーでは対応が困難であった比やメタ属性の表現についても一貫性を保持しつつ、柔軟に対応可能であることを確認した。

次に、医療機関で蓄積されている検査データ等の観測データと、疾患の記述に見られる異常状態のような概念レベルの高い概念に対して、定量的表現と特性を、定性表現を介して変換することで、両者を相互互換できることを示した。

最後に、関連研究との比較を通し、本提案手法の有用性について考察を行った。その結果、既存研究の表現では、特性記述が中心であり、特性を属性と属性値に分解できる形式をもたず、具体性に欠けるため、検査データへの対応が困難であることを示した。本提案モデルでは特性と属性が分離されており、データとの互換が可能である点において、実用においても有用であると期待される。

第4章では、診療科横断的な異常状態の共通性把握の困難性の課題解決に向けて、2) 疾患記述における異常状態の階層的モデリングを提案した。

まず、3章で提案した統一的表現モデルを基礎として、疾患を理解するために要求される概念層についてオントロジー工学理論に基づき考察した結果、3層構造による階層化を提案した。本提案枠組みでは、汎用的な異常状態から、対象物依存の異常状態、および各診療科の疾患固有の異常状態の3層に大別し、一般的な知識から臨床の高度な専門知識まで一貫性をもって定義可能であることを示した。また、第3層では、疾患そのものと疾患における異常状態とを識別可能であることを確認した。

次に、3層モデルの有用性と異常状態オントロジー構築の意義について議論した。提案モデルでは、概念の共通性を計算機処理可能であることから、疾患オントロジーと組み合わせれば、診療科横断的に複数疾患で起こりうる異常状態を把握できることを示した。その結果、これまで各診療科のみで扱われていた知識を統合し、診療科横断的な知見の獲得に貢献できると期待される。

さらに、関連研究との比較では、まず、関連研究における、複数観点に基づいた分類による異常状態と疾患の混在した階層構造や多重継承が引き起こす臨床上的の問題点を指摘した。次に、本研究のオントロジー工学理論に基づいた厳格な組織化がもたらす診療情報の計算機処理における優位性について議論した。多くの既存の用語集は、オントロジー工学が発展する前から構築されて来たため、臨床上の複雑かつ高次の知識を扱う上では、理論的にエラーのない体系化が必要とされる。本研究の異常状態オントロジーの構築では、3章で提案した統一的表現を用いて各異常状態を記述し、さらに汎用から疾患固有のレベルまでを理論に基づいて体系化していることから、これら既存の用語集に対しての骨格としての知識を提供し、参照概念の提供としての役割を果たす。

第5章では、散在したリソースからの関連情報取得という課題へ向けた解決手法として、本研究で提案する異常状態オントロジーを応用し、3) 異常状態オントロジーを媒介とする生命医学系リソース間ナビゲーションを目的とした応用システムの提案を行った。

まず、既存のリソースの現状として、関連研究のオントロジーが提供する概念レベルの不一致や、4章で述べた特性と属性の識別の欠如や複数観点の混在が統合時において不整合を生じるおそれがある問題を議論した。そこで、異常状態オントロジーを骨格として、各リソースの概念層に応じて統合するという解決アプローチを提案した。

次に、異常状態オントロジーの概念層別に、既存のリソースとの統合を試み、本アプローチの有効性を確認した。具体的には、既存のリソースの一対一の関係に対し、3層構造を用いた概念レベルの切り替えとそのコンテキストを明示することでマッピングの整合性を検証し、その結果、関連付けられたリソース間の互いの立ち位置を示し、異常状態の観点において適切な概念にナビゲートできることを例証した。

さらに、統合結果を用いた応用システム開発の取り組みについて述べた。そして、Linked Data 技術を用いた本システムの導入により、公開されている外部リソースとの連携を行い、生命医学において異常状態の関連知識を取得できることを示した。続いて、疾患オントロジーとの連携では、複数疾患の異常状態の因果関係を把握可能なことから、疾患で起こりうる原因からその進行するプロセス、そしてその結果引き起こされうる症状について診療科横断的に取得できることがわかった。また、臨床で見逃されやすい希少疾患の異常状態に関する知識も提供できることを示した。このように、本システムは、オントロジー理論に基づき、診療科横断的な異常状態の共通性を把握可能とするとともに、各診療科で理解されている疾患像についても適切に捉え、明示することができる。

従来は、疾患の把握は高度な専門知識を必要とするため、診療科毎に分断されていた。そして、各診療科の専門医が背景知識でその意味を解釈し、理解してきた概念は暗黙的であった。本研究では、それらの意味内容を明らかにし、一般化することで、診療科独立で共有できる概念を提供する役割を果たし、暗黙的な概念に透明性を与えた。その結果、本研究は、孤立した各診療科の知識を統合し、診療科横断的に臨床においてその全体像を把握することを可能にしたが、このような試みは未だ存在していないことから、世界初の成果といえる。

6.2 残された課題と将来展望

本研究によって、疾患を理解するための異常状態の意味内容の記述について理論上の枠組みは達成されたといえる。本節では、残された課題と将来展望について述べる。

今後の課題としては、まず、オントロジーの拡充という課題がある。本研究では、臨床における疾患を理解するために、細胞レベル以上の人体の知識を中心としている。ゲノム解析などで遺伝子の機能に関する情報がかなり解明されつつある現在、遺伝子や分子レベルの情報も疾患の理解には重要となる。そこで、例えば分子生物学者やゲノミクス関連の研究者と連携し、さらなる疾患の理解を深めたいと考えている。

次に、スケールアップという課題がある。これについては研究グループの属する臨床医学オントロジーの構成員である臨床医の協力を得て行なわれる予定である。理論的枠組みに基づいて、7診療科の異常状態を統一表現し、約13000の異常状態という大規模な概念が組織化されるには、多くの時間がかかる。しかし、生み出される成果により臨床上の知識をこれまでにない様々な観点で解析が可能になることから、その臨床医学分野への貢献度は大きいと期待される。

別の課題として構築したオントロジーのメンテナンスをどのようにするかという課題があるが、臨床医学オントロジープロジェクト全体で取り組むべき今後の課題としたい。例えば診療情報管理士による体制整備の他、システムのバージョン管理等、システム支援も併せて検討する必要がある。また、オントロジーについては引き続き、オントロジー工学者と臨床医が協力し、質の向上を目指す予定である。

さらに、積極的なオープン化の推進も課題として挙げられる。現在は臨床医学オントロジープロジェクトの中で本研究が行われている。今後は、国際的な生命系のコミュニティである OBO foundry と協力するなど、生命医学系の分野全体として異常状態に関する統合的な知識を享受する必要があると考えている。

また、統一表現モデルで異常状態の記述を行うことで、複数の医療機関と協力し、標準データとして、疾患記述において意味内容についてもゆれのない情報共有に向けた提供も検討したい。

最後に今後の展望を考察する。まず、臨床医学における本研究の貢献として、本研究では診療科横断的な異常状態の共通性を把握できることから、臨床における隠れた深い関係を見出し、疾患のさらなる理解のための新知識の発見につながると期待される。従来、臨床では疾患を診断の立場から個々の患者に観察される状態として、診療記録で管理されてきた。本研究では、専門医が理解している疾患の原因となる異常状態の捉え方、そしてその疾患像を描くまでのプロセスを計算機上に反映させることができるといえる。本研究は、疾患の解明に向けた臨床医学知識基盤の構築に貢献すると期待される。

次に、生命医学系では、基礎と臨床の間には深い谷が存在するとされており、その間を埋めるためのトランスレーショナルリサーチが模索されている。本研究では、オントロジー工学理論に基づき客観的な視点で、基礎から臨床まで多様な異常状態を見つめなおすことで、基礎と臨床のギャップを埋める橋渡しの役割を担うことができ

ると考えている。その結果、疾患とそれを取り巻く知識を関連付けることで、疾患のより一層深い理解に貢献すると期待され、新知識の発見にもつながると期待される。

臨床知識は、従来、他分野の技術者からは把握困難とされてきた。本研究ではそのような専門ドメインの知識を、客観的な視点で、理論に基づき一般性のある概念から特殊化していくことで、工学系の研究者でもその概念が理解しやすい形で提供している。序文で述べたように、医療に限らず工学などでも同様の問題を抱えている。本研究は、疾患の理解にとどまらず、医療分野に限らず、複雑なシステムの故障を扱うような他の分野でも適用可能な理論枠組みであると考えている。また、本研究の5章で提案した応用研究は、Web上の散在する膨大な情報から必要な関連知識のみを取り出し、自身の技術に再利用するという側面においては、情報化社会での知識の統合支援を考える上で一つの方向性を示すものといえる。

本研究を通して得られた成果は、対象とする問題をどのように向き合い、内容を同定していくかという基礎的な考察からはじまり、実用に耐えうる応用開発に導くまでの基盤理論として、医療をはじめ様々な分野での問題解決に貢献すると期待できる。

参考文献

- [Adams 09] Adams, N., Cannon, E., and Murray-Rust, P., ChemAxiom -An Ontological Framework for Chemistry in Science, Nature Precedings, <http://dx.doi.org/10.1038/npre.2009.3714.1>, (Last accessed on 2015/05/30), (2009).
- [Ashburner 00] Ashburner M., Ball C., Blake J., Botstein D., Butler H., Cherry J. et al., Gene ontology: Tool for the Unification of Biology. The Gene Ontology Consortium., Nat Genet., Vol.25, pp.25-29, (2000).
- [ASTM] American Society for Testing and Materials (ASTM), <http://www.astm.org>, (last accessed on 2015/05/29).
- [Bioportal] National Centers for Biomedical Computing (NCBO), Bioportal, <http://bioportal.bioontology.org>, (Last accessed on 2015/05/30), (2005).
- [Côté 10] Côté R., Reisinger F., Martens L., Barsnes H., Vizcaino JA., and Hermjakob H., The Ontology Lookup Service: bigger and better, Nucleic Acids Res. 38 (Web Server issue), W155-160, (2010).
- [Dolin 11] Dolin R., Alschuler L., Approaching Semantic Interoperability in Health Level Seven, J. Am. Med. Inform. Assoc. Vol.18, No.1, pp.99-103, (2011).
- [Gkoutos 05] Gkoutos, G., Green, E., Mallon, A., Hancock, J., and Davidson, D., Using Ontologies to Describe Mouse Phenotypes, Genome Biol., Vol. 6, R8, (2005).
- [Grenon 04] Grenon P., Smith B., and Goldberg L., Biodynamic Ontology: Applying BFO in the Bio-medical Domain, Stud Health Technol Inform., Vol.102, pp.20-38, (2004).
- [Gruber 93] Gruber. T., A translation approach to portable ontologies, Knowledge Acquisition, Vol.5, No.2, pp.199-220, (1993)
- [Guarino 98] Guarino N., Some Ontological Principles for Designing Upper Level Lexical Resources, In Proceeding of International Conference on Lexical Resources and Evaluation 1998, Granada, Spain, pp.527-534, (1998).
- [Hamosh 05] Hamosh A., Scott A., Amberger J., Bocchini C., and McKusick V., Online Mendelian Inheritance in Man (OMIM), a Knowledgebase of Human Genes and Genetic Disorders, Nucleic Acids Res., Vol. 1, No. 33 (Database issue), D514-517, (2005).
- [Heath 11] Heath T. and Bizer C., Linked Data: Evolving the Web into a Global Data Space. Synthesis Lectures on the Semantic Web: Theory and Technology, Morgan & Claypool, (2011).
- [法造] 大阪大学産業科学研究所, 法造, <http://www.hozo.jp/hozo/>, (last accessed on 2015/05/30).

- [ICD] World Health Organization (WHO), International Classification of Diseases (ICD), <http://www.who.int/classifications/icd/en/>, (Last accessed on 2015/05/30).
- [IHTSDO] International Health Terminology Standards Development Organisation (IHTSDO), <http://www.ihtsdo.org>, (last accessed on 2015/05/29).
- [Imai 09] Imai T., Kou H., Zhou J., Kozaki K., Mizoguchi R., and Ohe K, Japan Medical Ontology Development Project for Advanced Clinical Information Systems, In Proceedings of 10th International HL7 Interoperability Conference 2009 (IHIC2009), pp.42-46, (2009).
- [加地 05] 加地 大介, オントロジー構築のための実在論的方法論, 人工知能学会誌, Vol. 20, No.5, pp.595-603, (2005).
- [Köhler 14] Köhler S., Doelken S., Mungall J. et al., The Human Phenotype Ontology Project: Linking Molecular Biology and Disease through Phenotype Data, *Nucleic Acids Res*, Vol. 42 (Database issue), D966-974, (2014).
- [Kozaki 07] Kozaki K., Sunagawa E., Kitamura Y. and Mizoguchi R., Role Representation Model Using OWL and SWRL, In Proceedings of 2nd Workshop on Roles and Relationships in Object Oriented Programming, Multiagent Systems, and Ontologies, Berlin, July 30-31, pp.39-46, (2007).
- [古崎 02] 古崎 晃司 , 來村 徳信 , 佐野 年伸, 本松 慎一郎 , 石川 誠一 , 溝口 理一郎, オントロジー構築・利用環境「法造」の開発と利用 : 実規模プラントのオントロジーを例として, 人工知能学会論文誌, Vol. 17, No.4, pp.407-419, (2002).
- [古崎 10] 古崎晃司, 国府裕子, 今井健, 大江和彦, 溝口理一郎, 実践的オントロジー構築技法—臨床医学オントロジーを例として—, 2010年度人工知能学会全国大会 (第24回)資料集, 1B5-1 (2010).
- [古崎 12] 古崎 晃司, 日原 圭佑, 溝口 理一郎, 視点に基づく is-a 階層の動的生成, 人工知能学会論文誌, Vol.27, No.3, pp.235-244, (2012).
- [古崎 14] 古崎晃司, 山縣友紀, 国府裕子, 今井健, 大江和彦, 溝口理一郎, 医療知識基盤の構築に向けた疾患オントロジーの Linked Open Data 化, 人工知能学会論文誌, Vol. 29, No. 4, pp.396-405, (2014).
- [梶屋 13] 梶屋啓志, 古崎晃司, 溝口理一郎, コンテキストに依存した定性値を扱う生物表現型統合データベースの試作, 2013年度人工知能学会全国大会 (第27回)資料集, 3I1-2, (2013).
- [McDonald 03] McDonald C., Huff S., Suico J., Hill G., Leavelle D. et al. , LOINC, a Universal Standard for Identifying Laboratory Observations: A 5-Year Update, *Clin. Chem.*, Vol.49, No.4, pp.624-633, (2003).

- [MEDIS 03] 医療情報システム開発センター(MEDIS), J-MIX (電子保存された診療録情報の交換のためのデータ項目セット), <http://www2.medis.or.jp/master/jmix/>, (Last accessed on 2015/05/29), (2003).
- [メルク] メルクマニュアル 18 版, マラリア, <http://merckmanual.jp/mmpej/sec14/ch186/ch186g.html>, (Last accessed on 2015/05/30).
- [MeSH] The National Library of Medicine (NLM), Medical Subject Headings (MeSH), <http://www.nlm.nih.gov/mesh/meshhome.html>, (Last accessed on 2015/05/29), (1999).
- [Mizoguchi 10] Mizoguchi R., YAMATO: Yet Another More Advanced Top-level Ontology, In Proceedings of the Sixth Australasian Ontology Workshop, pp.1-16, (2010).
- [Mizoguchi 11] Mizoguchi R., Kozaki K., Kou H., Yamagata Y., Imai T., Waki K. and Ohe K., River Flow Model of Diseases, In Proceedings of the 2nd International Conference on Biomedical Ontology, Buffalo, USA, July. 28-30, CEUR-WS.org, Vol.833, (2011).
- [溝口 97] 溝口理一郎, オントロジー工学序説 : 内容指向研究の基盤技術と理論の確立を目指して, 人工知能学会誌, Vol.12, No.4, pp.559-569, (1997).
- [溝口 05] 溝口理一郎, オントロジー工学, オーム社, (2005).
- [OBO] The Open Biological and Biomedical Ontologies (OBO), <http://www.obofoundry.org/ontologies.shtml>, last accessed on 2015/05/30).
- [小川 01] 小川聡, 井上博, 佐藤徹, 標準循環器病学, 医学書院, pp.256, (2001).
- [大江 10] 大江和彦, 今井健, 臨床医学知識処理を目指した医療オントロジー開発, 人工知能学会誌, Vol.25, No.4, pp.493-500, (2010).
- [太田 11] 太田 衛, 古崎 晃司, 溝口 理一郎, 実践的なオントロジー開発に向けたオントロジー構築・利用環境「法造」の拡張 — 理論編 — 臨床医学知識処理を目指した医療オントロジー開発, 人工知能学会論文誌, Vol.26, No.2, pp.387-402, (2011).
- [OLS] European Ontology Lookup Service (OLS), <http://www.ebi.ac.uk/ontology-lookup/>, (Last accessed on 2015/05/30).
- [O'Neil 95] O'Neil M., Payne C., and Read J., Read Codes Version 3: a User Led Terminology, Methods Inf. Med., 1995 Vol.34, No. (1-2), pp.187-192, (1995).
- [Ontobee] Ontobee, <http://www.ontobee.org>, (last accessed on 2015/05/30).

- [OWL] W3C OWL Web Ontology Language Reference, <http://www.w3.org/TR/owl-ref/>, (last accessed on 2015/05/30).
- [OWL 2] OWL 2 Web Ontology Language Document Overview (Second Edition), <http://www.w3.org/TR/owl-overview/>, (last accessed on 2015/05/30).
- [PubMed] National Institutes of Health's National Library of Medicine (NIH/NLM), PubMed, <http://www.ncbi.nlm.nih.gov/pubmed>, (Last accessed on 2015/05/29).
- [Rosse 03] Rosse C. and Mejino J., A Reference Ontology for Biomedical Informatics: the Foundational Model of Anatomy, *J. Biomed. Inform.*, Vol. 36, pp.478-500, (2003).
- [Scheuermann 09] Scheuermann R., Ceusters W., and Smith B., Toward an Ontological Treatment of Disease and Diagnosis, *Summit on Translational Bioinformatics 2009*, pp.116-120, (2009).
- [Schulz 07] Schulz S., Suntisrivaraporn B., and Baader F., SNOMED CT's Problem List: Ontologists' and Logicians' Therapy Suggestions, *Stud. Health Technol. Inform.* Vol.129 (Pt 1), pp. 802-806, (2007).
- [Schulz 13] Schulz S., Balkanyi L., Cornet R., and Bodenreider O., From Concept Representations to Ontologies: a Paradigm Shift in Health Informatics? *Healthc. Inform. Res.*, Vol.19, No.4, pp.235-242, (2013).
- [瀬川 08] 瀬川翔, 大久保公則, 笹嶋宗彦, 來村徳信, 溝口理一郎, 機能オントロジーマッピングのガイドラインに関する一考察 : 2つの機能語彙体系間でのマッピングを例として, 2008年度人工知能学会全国大会論文集 (第22回), 3J2-4, (2008).
- [清野 12] 清野裕他, 日本糖尿病学会糖尿病診断基準に関する調査検討委員会, 糖尿病の分類と診断基準に関する委員会報告 : 国際標準化対応版, *糖尿病*, Vol. 55, No.7, pp.485-504, (2012).
- [周 08] 周俊, 国府裕子, 古崎晃司, 溝口理一郎, OWLを用いたロール概念表現モデルに関する考察, 2008年度人工知能学会 (第22回), 3G2-1, (2008).
- [Smith 07] Smith, B., Ashburner M., Rosse C., Bard J., Bugand W. et al., The OBO Foundry: Coordinated Evolution of Ontologies to Support Biomedical Data Integration, *Nature Biotechnology*, Vol. 25, pp. 1251–1255, (2007).
- [SNOMED-CT] International Health Terminology Standards Development Organization (IHTSDO), SNOMED-CT., <http://www.ihtsdo.org/snomed-ct/>, (last accessed on 2015/05/29).

- [Stearns 01] Stearns M., Price C., Spackman K., and Wang A., SNOMED Clinical Terms: Overview of the Development Process and Project Status, In Proceedings of American Medical Informatics Association (AMIA) Symp. 2001, pp.662-666 (2001).
- [Whetzel 11] Whetzel P., Noy N., Shah N., Alexander P., Nyulas C., Tudorache T., and Musen M., BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic Acids Res.*, Vol. 39(Web Server issue), W541-545, (2011).
- [WHO 99] World Health Organization (WHO), Definition, Diagnosis and Classification of Diabetes Mellitus and its Complications: Report of a WHO Consultation. Part 1, Diagnosis and Classification of Diabetes Mellitus, Geneva, World Health Organization, (1999).
- [Xiang 11] Xiang Z., Mungall C., Ruttenberg A., and He Y., Ontobee: A Linked Data Server and Browser for Ontology Terms, Ontology, In Proceedings of the 2nd International Conference on Biomedical Ontology, Buffalo, USA, July. 28-30, CEUR-WS.org, Vol.833, (2011).

謝辞

学位論文の執筆にあたり、終始懇切なる御指導と貴重な御教示を賜りました大阪大学産業科学研究所 駒谷和範教授に深く感謝致します。

本研究に関して貴重な御教示を頂きました大阪大学産業科学研究所 鷲尾隆教授，ならびに、直接に御助言と御鞭撻を頂いた大阪大学産業科学研究所古崎晃司准教授に深く感謝致します。

そして、本研究の全過程を通して、終始懇篤なる御指導，御鞭撻を賜りました北陸先端科学技術大学院大学 溝口理一郎特任教授に深厚なる謝意を表します。

博士後期課程において、御指導と御教授を賜りました大阪大学大学院工学研究科電気電子情報工学専攻滝根哲哉教授，北山研一教授，馬場口登教授，三瓶政一教授，井上恭教授に謝意を表します。

医療分野での研究を通して長年蓄積された経験から、暖かく丁寧な御指導，御助言下さいました東京大学 大江和彦教授，並びに今井健講師に心より謝意を表します。

また、臨床の専門知識をご提供頂き，本研究に対する御助言を賜りました東京大学附属病院桜井亮太氏，寺田さとみ氏，大友夏子氏，林亜紀氏，松村貴由氏，脇嘉代氏に心より感謝致します。

同じく、医療情報に関する専門知識および、提案モデルの実践の場を御提供頂き，本研究に対する貴重な御助言を賜りました東京大学病院企画情報運営部 河添悦昌氏，梶野正幸氏，並びに疾患生命工学センター篠原恵美子氏に厚く感謝の意を表します。

そして、本研究のオントロジー工学理論において、貴重な討論とご助言を賜りました，立命館大学 來村徳信教授に心より謝意を表します。

さらに、貴重な討論，激励を頂きました大阪大学産業科学研究所 武田龍助教，笹嶋宗彦特任研究員に深く感謝致します。また，日ごろよりお世話になりました駒谷研究室事務補佐員 本藪千鶴子氏に深く感謝致します。

本研究を進める中，公私にわたり御支援・御協力を頂き，常に有益な助言を頂いた，大阪大学産業科学研究所 溝口研究室および駒谷研究室の諸氏，並びに卒業された先輩諸氏に感謝致します。