OUKA
The University of Osaka
Institutional Knowledge Archive

Title	概念階層比較手法に基づくオントロジー洗練及び拡張支援システム
Author(s)	増田, 壮志
Citation	大阪大学, 2019, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/72401
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

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

The University of Osaka

博士学位論文

概念階層比較手法に基づく
オントロジー洗練及び拡張支援システム

増田　壮志

2019 年 1 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，概念階層比較手法に基づくオントロジー洗練及び拡張支援システム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，本研究を進めるにあたっての研究背景及び概要を述べる．

第 2 章では，まず，オントロジーを構成する概念及び概念間関係について概説する．次に，オントロジー洗練及び拡張支援システムに関する既存研究について述べる．以上の内容に加え，クラス分類観点の同一性，概念階層同士の相似性について述べる．

第 3 章では，上位下位関係に着目した概念階層比較手法について述べる．本章では，オントロジーを構成する基本的な関係である概念階層を上下一段毎に区切り，概念階層同士の構造を比較し，互いに相似形となるような概念追加提案を行う手法を説明する．評価実験の結果，検出された洗練箇所の約 8 割に対して，適用すべき提案を少なくとも一つ提示できることを確認した．この結果から概念階層比較によるオントロジー洗練支援の可能性が示唆された．

第 4 章では，オントロジーの分類観点毎の概念間関係階層に着目して概念階層同士を比較する手法を説明する．3 章では，比較対象が上位下位関係で局所的なものとなっていたが，兄弟概念を併せて比較する手法により，分類観点毎に概念階層を包括的に比較できるようになる．その結果，オントロジー構築者は，オントロジーの概念定義に関するより多くの情報を考慮しながら，オントロジー洗練支援システムが提示する洗練提案の可否判断が可能となる．

第 5 章では，構築途中段階のオントロジーを対象とした拡張支援手法について述べる．構築途中段階のオントロジーは，概念定義がスパースである．そのため概念及び概念間関係の定義が密であることを前提とした洗練支援手法だけでは，洗練箇所や方法を提示できない．そこで本章で述べる拡張支援手法では，概念階層の比較手法に加えて Web 上のデータから得られる関係を利用する．本システムを利用することによりオントロジー構築者は，概念間関係数が不十分なオントロジーでも結ぶべき概念間関係の提示を受けることが可能となる．さらに，JST シソーラスの生命科学分野を基に構築が進められているオントロジーを用いて評価実験を行い，オントロジー拡張支援手法の有用性を確認した．

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．最後に今後の展望について述べる．

関連発表論文

A. 学会誌掲載論文

[A1] 増田壮志, 古崎晃司,
概念階層の比較に基づくオントロジー品質向上支援手法,
人工知能学会論文誌, Vol. 32, No. 2, p. E-G71_1-10, (2017). [第 3 章]

[A2] 増田壮志, 古崎晃司, 駒谷和範,
分類観点毎の概念階層の包括的な比較に基づくオントロジー品質向上支援システムの開発,
人工知能学会論文誌, Vol. 33, No. 2, p. B-H82_1-10, (2018). [第 4 章]

B. 国際会議発表論文（査読付き）

[B1] Masuda, T. and Kozaki, K.,
A Basic Consideration on Ontology Refine Method Using Similarity Among Is-a Hierarchies,
Joint International Workshop on the Linked Data in Practice and the Practical Application of Ontology for Semantic Data Engineering, Seoul, Korea, Nov. 30, 13 pages, (2013). [第 3 章]

[B2] Masuda, T.,
Ontology Refinement and Evaluation System Based on Similarity of Is-a Hierarchies,
International Semantic Web Conference, Doctoral Consortium, Kobe, Japan, Oct. 18, 8 pages, (2016). [第 3 章]

[B3] Masuda, T., Kozaki, K. and Komatani, K.,
Ontology Refinement System for Improving Consistency of Classification Among Brother Concepts,
Joint International Semantic Technology Conference, Poster & Demo Session, Singapore, Nov. 2-4, 4 pages, (2016). [第 4 章]

[B4] Masuda, T., Kozaki, K., Kushida, T., Tateisi, Y., Watanabe, K., Matsumura, K., Kawamura, T. and Komatani, K.,
Extending A Bioscience Ontology Based on Comparison between Sibling Concepts,
International Conference on Internet Technologies & Society, Sydney, Australia, Dec. 11-13, 10 pages, (2017). [第 5 章]

C. 研究会発表

[C1] 増田壮志, 古崎晃司,
オントロジーにおける is-a 階層の相似性を利用した内容洗練方法の提案,
第 29 回セマンティックウェブとオントロジー研究会, SIG-SWO-A1203-07, 東京都, (2013). [第 3 章]

[C2] 増田壮志, 古崎晃司,
Is-a 階層間の粒度の差に着目したオントロジー内容洗練支援システムの試作と評価方法の考察,
第 34 回セマンティックウェブとオントロジー研究会, SIG-SWO-A1402-7, 東京都, (2014). [第 3 章]

[C3] 増田壮志, 古崎晃司, 櫛田達也, 建石由佳, 渡邊勝太郎, 川村隆浩, 駒谷和範, 兄弟概念の比較に基づくバイオサイエンスオントロジーのプロパティ推定手法, 第 43 回セマンティックウェブとオントロジー研究会, SIG-SWO-043-09, 神奈川県, (2017). [第 5 章]

D. 全国大会発表

[D1] 増田壮志, 古崎晃司, Is-a 階層の相似性を双方向に利用したオントロジー内容洗練方法の考察, 第 27 回人工知能学会全国大会, 3I1-5, 富山県, (2013). [第 3 章]

[D2] 増田壮志, 古崎晃司, オントロジー内の is-a 階層の相似性に基づいた内容洗練システムの開発, 第 28 回人工知能学会全国大会, 2F1-1, 愛媛県, (2014). [第 3 章]

[D3] 増田壮志, 古崎晃司, 参照関係にある is-a 階層間の相似性を利用したオントロジーの一貫性向上手法の評価, 第 29 回人工知能学会全国大会, 2M1-3, 北海道, (2015). [第 3 章]

[D4] 増田壮志, 古崎晃司, 駒谷和範, 兄弟概念の分類観点に着目したオントロジー一貫性向上手法の開発, 第 30 回人工知能学会全国大会, 1J2-1, 福岡県, (2016). [第 4 章]

[D5] 増田壮志, 古崎晃司, 駒谷和範, 分類観点に沿った概念階層全体の比較によるオントロジー品質向上手法の提案, 第 31 回人工知能学会全国大会, 3G1-2, 愛知県, (2017).[第 4 章]

目次

図目次 vii
表目次 viii
第 1 章 序論 1
1.1 研究背景 1
1.2 研究概要 5
1.3 本論文の構成 8
第 2 章 オントロジーを構成する基本的要素及び関連研究 9
2.1 オントロジー構築の一般事項 9
2.1.1 オントロジーの概念階層を構成する is-a 関係 9
2.1.2 各概念の部分や属性を定義する概念間関係 10
2.1.3 概念及び概念間関係の特殊化 12
2.1.4 オントロジーを構成する概念と概念間の関係 13
2.1.5 概念の分類観点 15
2.1.6 オントロジーの表現方法 15
2.2 オントロジー構築に関する関連研究 16
2.2.1 オントロジーの評価洗練に焦点を当てた支援 16
2.2.2 オントロジーの拡張に焦点を当てた支援 19
2.3 本研究で目指す品質の高いオントロジー 19
2.3.1 概念クラス分類観点の同一性 19
2.3.2 概念間関係階層を介して参照関係にある概念階層間の相似性 22
2.3.3 概念階層比較によるオントロジー洗練・拡張支援手法 24
第 3 章 上位下位の関係に着目したオントロジー洗練支援システム 26
3.1 上位下位の関係に着目した概念階層比較手法 26
3.1.1 上位下位関係に着目した洗練箇所検出 26
3.1.2 上下一段毎に区切った概念間関係階層を中心とした概念階層比較 27
3.1.3 上下一段毎の概念間関係階層毎の相似判定例 30
3.1.4 洗練箇所に対する洗練提案の組み合わせ 32
3.2 上位下位関係に着目した洗練支援システムの実装 35
3.3 評価実験 36
3.3.1 評価実験の目的 36
3.3.2 評価実験の概要 37
3.3.3 実験結果及び考察 39
3.3.4 洗練提案例 42
第 4 章 分類観点毎の概念間関係階層に着目したオントロジー洗練支援システム 44
4.1 分類観点毎の包括的な概念階層比較手法 44
4.1.1 分類観点毎の包括的な概念階層比較の必要性 44
4.1.2 分類観点毎の概念階層の比較による洗練箇所検出 45
4.1.3 洗練箇所に対する洗練提案の組み合わせ 49
4.1.4 洗練提案の組み合わせの限定 49
4.2 分類観点毎の包括的な概念階層に着目した洗練支援システムの実装 53
4.3 評価実験 55
4.4 実験結果及び考察 59
4.4.1 検出される洗練提案数 59
4.4.2 洗練箇所毎の評価 59
4.4.3 洗練提案の妥当性の評価 60
第 5 章 概念階層比較手法を応用したオントロジー拡張支援システム 63
5.1 拡張支援に向けた概念階層比較手法の応用 63

5.1.1 構築途中段階の概念定義がスパースなオントロジーへの対応 63
5.1.2 JST シソーラスを基にしたバイオサイエンスオントロジー 64
5.2 概念階層比較による組み合わせ候補単語の抽出 65
5.2.1 概念間関係の定義がスパースなオントロジー 65
5.2.2 同じ段数にある概念に着目した概念階層の比較 66
5.2.3 組み合わせ候補の例 68
5.3 Linked Open Data 中の共起情報を利用した順位付け 69
5.3.1 共起情報を利用した順位付けの流れ 70
5.3.2 Linked Open Data を利用した共起単語の取得 70
5.3.3 共起単語を利用した組み合わせ候補の順位付け 71
5.4 バイオサイエンスオントロジーへの適用 72
5.4.1 拡張支援システムの評価実験 73
5.4.2 拡張システムによる概念間関係の追加提案例 74
第 6 章 結論 76
6.1 研究の総括 76
6.2 残された課題と将来展望 78
参考文献 82
謝辞 87

図目次

図 1.1. オントロジーの構築段階 3
図 1.2. 各章間の関係 7
図 2.1. オントロジーを構成する概念と概念間関係 9
図 2.2. ハンドルを基本概念とする概念階層 11
図 2.3. クラス制約として参照可能な概念 12
図 2.4. クラス制約に関する文法規則違反の例 17
図 2.5. 複数の分類観点が混在している概念階層例 20
図 2.6. 分類観点が一貫していない概念階層例 21
図 2.7. 中間概念を追加し分類観点を揃えた is-a 階層の例 22
図 2.8. 相似な is-a 階層 22
図 3.1. 概念間関係階層の区切りと比較対象範囲 28
図 3.2. 相似な場合の概念と概念間関係の対応 28
図 3.3. (a)の場合に非相似である概念階層 29
図 3.4. (b)の場合に非相似である概念階層 29
図 3.5. (c)の場合に非相似である概念階層 29
図 3.6. 概念間関係追加 32
図 3.7. 基本概念＋概念間関係の追加 33
図 3.8. 被参照概念＋概念間関係の追加 34
図 3.9. 洗練箇所・提案表示インタフェース 36
図 3.10. 洗練箇所の妥当性の評価 39
図 3.11. 洗練提案の妥当性の評価 41
図 3.12. 妥当であると評価された洗練提案例 42
図 3.13. 適用すべきでないと評価された洗練提案例 43
図 4.1. 比較対象となっている概念の集合間の関係 46
図 4.2. オントロジー構築者に提示される概念階層 47
図 4.3. 同じ上位概念を持つ兄弟に限定する方法 51
図 4.4. 特殊化の段数が同じ概念に限定する方法（Basic(R_2)に着目した場合） 52
図 4.5. 洗練支援システム 54
図 4.6. 追加提案される概念及び概念間関係のプレビュー表示 54
図 4.7. 洗練提案の手法による評価の差 60
図 4.8. 適用すべきとされた洗練提案例 62
図 5.1. 薬物と薬理作用の階層間の関係 65
図 5.2. 最上位概念からの距離 67
図 5.3. 拡張支援手法の概要 69
図 5.4. 組み合わせ候補と共起語の関係 71
図 6.1. 概念の削除の影響 79

表目次

表 3.1. 相似形のパターン 30
表 3.2. 非相似のパターンと洗練提案の組み合わせ 34
表 3.3. 各オントロジーから検出された洗練箇所数 37
表 3.4. 各洗練箇所の選択数内訳 38
表 3.5. 被験者毎の洗練提案の評価の差 38
表 3.6. 1 箇所当たりの平均洗練提案数 41
表 4.1. 洗練箇所と洗練提案の対応と洗練提案数 49
表 4.2. 各被験者による評価 57
表 4.3. 各オントロジーから検出された洗練箇所・洗練提案数 58
表 4.4. 洗練箇所及び洗練提案の妥当性の評価 59
表 4.5. 見せ方の違いによる評価の差 61
表 5.1. 同じ段数にある概念に限定した場合の組み合わせ数 68
表 5.2. 薬物と薬理作用の組み合わせ候補で適切であった例 68
表 5.3. 解熱鎮痛薬と共起する単語上位 25 位 70
表 5.4. 解熱鎮痛薬の組み合わせ候補上位 10 位 72
表 5.5. 概念の組み合わせ推定結果 73
表 5.6. ACE 阻害剤と共起語する単語 74
表 5.7. ACE 阻害剤の組み合わせ候補上位 10 位 74
表 5.8. オピオイドペプチドと共起する単語 75
表 5.9. オピオイドペプチドの組み合わせ候補上位 10 位 75

第1章　序論

1.1　研究背景

オントロジーは，本来，哲学の一分野である「存在論（Ontology）」を指す言葉である．哲学分野において，オントロジーは，存在に関する体系的な理論であり，世の中に存在するものを系統立てて説明することを目的としている．一方，本研究が対象としている工学的なオントロジー（オントロジー工学）は，対象となる知識の体系化を鍵としている点では，哲学におけるオントロジーと共通である．しかし，実世界全般に関わる存在を考察の対象とするのではなく，計算機処理される特定の世界を対象としている点に違いがある．この違いは，哲学におけるオントロジーが，一般的な国語辞書のように，存在に関する説明的・記述的（descriptive）な知識を扱っているのに対し，工学的なオントロジーは「(このオントロジーに基づくモデルの世界では,）存在物を，このようなものとして扱う」という宣言的・模範的な知識を扱っているという点に現れている[岡田 02].

工学的オントロジーは，計算機内のモデルが対象とする世界に関して，その構成に必要となる概念と，それらの概念が満たすべき制約や関係を提供する．このため，同一のオントロジーでは,「あるものを，何と見なし，どのように扱うか」という点に一貫性があることが望ましい. なぜなら, データの意味的相互運用を実現するためには, オントロジーに含まれる概念及び関係の一貫性が必要不可欠な要素だからである．また，異なるオントロジーに基づいて構築されたモデルの間においても，互いのオントロジーを比較することによって，相互の違いや共通点を見いだすことができるようになる．そして，そのような比較を通して，オントロジー間のマッピングやアライメントを行うことにより，オントロジー間の関係を介したデータの変換が可能となる．これらは，データの相互運用という観点から見たオントロジーの効用であるが，情報工学や知識工学一般の観点からは,

- 合意形成の手段の提供
- 暗黙情報の明示化
- 知識の再利用性と共有性の向上
- （コンピュータ上での）知識の体系化
- 知識の標準化
- メタモデル的機能

等も，オントロジーの役割としてあげられている[溝口 05].

オントロジーは，仕様の標準化や利用ツールの普及に伴い，情報科学のみならず，様々な領域への応用が進められている．これまでにもバイオ分野[福田 07]，医療分野

[大江　10, Imai 16]でオントロジーが構築・利用されている．さらに，生物規範工学[古崎　16]，環境[Kumazawa 09]，自動運転技術[Zhao 15]などの様々な領域においても多くのオントロジーが構築され，知識システム開発の知識基盤として用いられている．

例えば，生命科学分野では，生物やその機能を多数の遺伝子や蛋白質で構成された複雑なシステムとして捉えている．こうしたシステムの理解のためには，関連分野の最新知識を統合的に理解する必要があり，研究を進めるためには効率的な情報収集が求められる．そこで GeneOntology[1]（GO）[Ashburner 00]や Open Biomedical Ontology[2]（OBO）[OBO]をはじめとしたバイオ系オントロジーが構築されている．こうしたオントロジーは，分子の機能や生体内での役割，疾患への関与などの最新の研究成果のデータを標準化し，再利用性を高め，さらに高度な研究や新たな知識の発見につなげる重要な役割を担っている．

また，医療分野では，従来から診療現場における電子化が進められてきた．それに伴い，電子化された診療情報を計算機で処理するための，方法論・知識リソースの構築に関する研究が進められてきた．こうした診療データの画像や数値データ，診療報告書や電子カルテ上に入力されたテキストデータを計算機で解析し，診療支援や統計分析を行い，新たな医学的知見を発見することが期待されている．電子データのこうした利用のためには，統計処理や知識処理の手法とともに，その処理基盤を支える用語や構造化情報を標準化することが必要である．そこで医療分野では，SNOMED-CT[3]（Systematized Nomenclature of Medicine-Clinical Terms）[SNOMED-CT]，ICD-11[4]（International Statistical Classification of Diseases and Related Health Problem-11th revision）[ICD]などのオントロジーが構築され，診療情報を記述する際の標準語彙として利用されている．また，現在も時代の変化に合わせて逐次改訂され続けている．

しかしながら，その構築には人・時間共に多くの労力が必要であり，保守や更新が困難である．なぜなら，オントロジーの構築には，オントロジーそのものに関する知識と構築経験に加え，構築対象としているオントロジーの対象領域に関する専門知識も必要とされるからである．SNOMED-CT は，現時点で医療分野の情報の網羅性において世界最大である．このオントロジーは，多くの医療分野の専門家と知識工学の研究者によって，1999 年に米国病理学会が定めた病理室病分類をベースとして構築がスタートし，2009 年の段階で，39 万の概念と 65 種類，延べ約 142 万の概念間の関係が定義されている．このように膨大なオントロジーの構築は，通常，多くのオントロ

1 http://www.geneontology.org/
2 http://www.obofoundry.org
3 https://www.snomed.org/snomed-ct
4 http://www.who.int/classifications/icd/en/

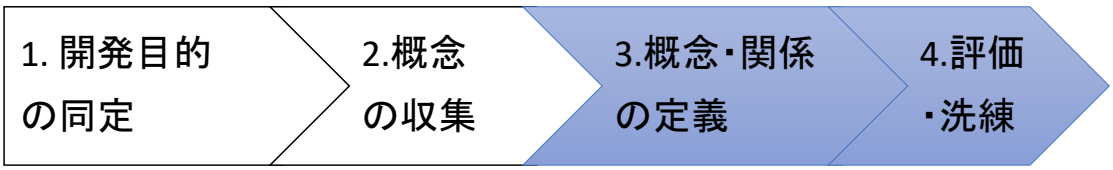

図 1.1.　オントロジーの構築段階

ジー構築の専門家と対象領域の専門家の共同・分業作業で長期間にわたって行われる．その労力の軽減のために，これまでにオントロジー構築方法論に関する様々なガイドラインや方法論，構築支援ツール，オントロジーの半自動構築に関する研究が行われてきた．

オントロジーの構築過程全体を対象とした開発ガイドライン[Uschold 96，Fernandez-Lopez 99，Staab 01，Noy 01a]では，オントロジー構築を，図 1.1.に示す 4 段階に分けて議論している．オントロジー構築は，これらの段階を行き来しながら行われるとされ，それぞれの構築段階毎に，様々な構築支援方法が提案されている．

第 1 段階の「開発目的の同定」においては，オントロジーが網羅すべき対象領域，その利用目的，利用者及び保守者が決定される．Gruninger は，オントロジーを基礎とする知識ベースが回答可能な質問である「能力質問（competency questions）」を決定し，構築後の評価で用いることを提案している[Gruninger 95]．企業活動オントロジー開発[5]に用意した能力質問の例として，

- あるゴールを達成するためにはどのような活動が遂行されなければならないか
- 未来の複数時点における活動が与えられたとき，それ以外の時点における資源と活動に関する性質は？
- あるタスクを前（またには後ろ）に移動するとどのようなことが起こるか？

等である．これらのインフォーマルな能力質問を公理記述言語として Prolog を用い，定式化している．

第 2 段階の「概念の収集」では，オントロジー構築者は，開発目的に沿ったオントロジーには，どのような概念及び概念間の関係が必要とされるのか検討し，収集する．概念及び概念間関係の収集には，

- 既存のオントロジー中の概念及び関係の利用
- 文献，専門家へのインタビュー等からの抽出

の 2 通りが存在する．

まず，既存のオントロジー中の概念及び概念間関係の利用について述べる．この方法では，ある程度構造化されたオントロジーの上に概念や関係を追加可能である．再利用可能なオントロジーライブラリとしては，Protégé Ontology Library[6]，BioPortal[7]，

[5] http://www.eil.utoronto.ca/enterprise-modelling/tove/
[6] https://protegewiki.stanford.edu/wiki/Protege_Ontology_Library
[7] https://bioportal.bioontology.org/

Ontolingua Ontology Library[8]などがある．また，DOLCE[9]（a Description Ontology for Linguistic and Cognitive Engineering）[Guarino 98]，BFO[10]（Basic Formal Ontology）[Grenon 04]，YAMATO[11]（Yet Another More Advanced Top-level Ontology）[Mizoguchi 10]などの上位オントロジーも利用される．さらに近年では，Web 上に，音楽，地理情報，書籍情報などの多種多様なデータが，Linked Open Data（LOD）として公開されている．こうした LOD は，オントロジーを記述するための共通の言語を利用して記述された半構造化情報であるため，オントロジーで用いる概念やインスタンスとして再利用可能である．

次に，専門文書やドメインの専門家へのインタビューからの抽出について述べる．LOD 上に公開されているデータは標準的な知識であるのに対して，こちらの方法では，オントロジー構築者は，より個別の知識を抽出することが可能である．例えば，看護知識は，各現場において作業内容が異なるため，各現場に固有のオントロジーを構築する必要がある．西村らは，そうした固有知識を獲得するために，現場の看護師と共に看護プロセスの構造化ワークショップを行うことで，知識を獲得している[西村 17].

第 3 段階の「概念・関係の定義」では，まず，概念及び概念の上位下位関係である is-a 関係が定義される．上位下位関係の定義方法は，主に以下の 3 種類である．

- トップダウンな方法
- ボトムアップな方法
- トップダウンとボトムアップを組み合わせた方法

トップダウンな方法では，先に最も一般的な概念を定義し，その後それらの概念を特殊化する．ボトムアップな方法では，概念階層のリーフとなるような概念を先に定義し，それらの概念をまとめたより一般的な概念を定義し，組織化していく．組み合わせた方法では，主要な概念を先に定義し，その後，その概念を特殊化または一般化していく．

概念と is-a 関係の定義の後，「プロパティ」を定義する．「プロパティ」は，“自動車”と“エンジン”のような「全体部分」関係や“自動車”と“色”のような概念の持つ「属性」関係など，「2 つの概念がどのような関係で結ばれるのか」を定義するものである．

最後 4 段階目の「洗練・評価」では，構築者は，3 段階目までで構築されたオントロジーに定義された概念及び概念間の関係に矛盾があれば修正する．また，不足している概念があれば追加，不必要な概念があれば削除する．また，1 段階目において能

8 http://www.ksl.stanford.edu/software/ontolingua/
9 http://www.loa.istc.cnr.it/old/DOLCE.html
10 http://ifomis.uni-saarland.de/bfo/
11 http://download.hozo.jp/onto_library/upperOnto.htm

力質問を設定している場合，構築したオントロジーがその質問に回答可能かを調べ，オントロジーの品質を評価する．

1.2 研究概要

本研究ではオントロジー構築における「3.概念・関係の定義」，「4．洗練・評価」を対象とした構築支援手法の開発を目指す．これらを対象としたのは，構築段階の「1．開発目的の同定」，「2．概念の収集」は，オントロジー構築者・利用者ごとに，目的や対象ドメインなどが異なるため，一般化した支援ツールの開発は困難だからである．一方，構築の第 3，4 段階の，ある程度構築が進んでいるオントロジーであれば，その構造情報が利用可能であり，より幅広いオントロジーを対象とした一般化された拡張・洗練支援手法の開発が可能である．ゆえに本研究では，第 3，4 段階目にあるオントロジーを対象とした，拡張・洗練支援手法を開発する．これらの手法によってオントロジー構築が効率化される．

3 段階目にあるオントロジーの中で，本研究が対象とするのは，概念と is-a 関係が十分定義された，構築途中段階にあるオントロジーである．こうした構築途中段階のオントロジーを入力として，全体部分関係や属性などを定義するプロパティの追加提案を出力する拡張支援システムを開発する．オントロジー構築の，「3．概念・関係の定義」では一般的に，概念と is-a 関係を定義したのち，プロパティを定義する．実際には，概念と is-a 関係の定義までを行い，プロパティに関しては未定義もしくは著しく定義数が少ないオントロジーが多い．なぜなら，プロパティの定義は，is-a 関係に比べ，対象領域やオントロジー構築に関するより深い知識を必要とするため，より複雑でコストのかかる作業であるからである．本研究では，オントロジー内の既存の定義内容を用いて，プロパティで表される関係を結ぶべき概念の組み合わせを自動的にオントロジー構築者に提案することで，概念間関係の定義の追加を支援するシステムを開発する．

4 段階目にあるオントロジーに対する洗練手法には，形式的なエラー修正と定義内容の修正がある．形式的なエラー修正は，オントロジーの記述言語に対応した推論機構を用いて機械的に処理できるため多くの先行研究がある．一方，定義内容に対する洗練手法は，オントロジー構築者に対して定義内容を分かりやすく可視化し，洗練箇所の発見を促すもの[Ohta 11b]や，メタプロパティと呼ばれる概念の性質に関するメタ的な情報を付与し，概念階層を整理するもの[Guarino09]がある．このような洗練手法を利用するには，オントロジー構築に関する知識はもとより，それぞれの手法を利用するための知識も必要である．そのため，オントロジー構築者に課する負担が，依然として大きいままであり，オントロジーの規模によっては数万の概念とその概念間の関係を確認する必要がある．

そこで本研究では，洗練段階のオントロジーを入力として，構築者に対して自動的に，定義内容の面で修正すべき箇所とそうした箇所に対する洗練提案を出力するシステムを開発する．本研究で提案する内容洗練手法によって，オントロジー構築者は，オントロジー洗練作業時にすべての概念の定義内容を再度確認する必要がなくなる．さらに構築者は，システムからオントロジーの品質向上のために修正が必要な箇所と方法や新たに追加すべき概念間関係に関する提案を受けることが可能となる．

第 3，4 段階のオントロジーに対する支援システムを開発するためには，個々の概念の定義内容ではなく，オントロジー構築におけるより一般的な基準に基づく必要がある．オントロジー構築の際に推奨される基準としては，代表的なものとして，クラス分類観点の同一性，インスタンス集合のパーティション性，本質属性の体現，上下関係の一般性等がある[溝口 06]．

本研究では，その中でも「クラス分類観点の同一性」に着目し，オントロジーの品質を向上させるための洗練・拡張提案を行うシステムの開発を目指す．分類観点とは，構築者が，オントロジー上に概念をどのように定義したいのかを表現したものである．つまり，オントロジー構築者が対象世界を分類する際にどのような性質を着目したのか，という各概念定義の核となる意味を持つものである．クラスの分類観点が一貫していない場合には，概念毎の分類観点が明示化されず，オントロジー構築者が対象世界をどのようにモデル化したのかが不明確になる．そのため，先に述べたオントロジーの役割を十分に果たすことができない．オントロジー構築は，前述の通り，多大な人的時間的コストを要求するものであるため，分類観点の一貫性を保ちながら品質の高いオントロジーを構築していくことは困難を伴う．

この分類観点の一貫性は，オントロジーの構築を通してオントロジー構築者が留意すべきことであるので，拡張・洗練支援システムの両方において提案時に守るべき基本的な指針とする．本研究では，この推奨基準に沿って構築されたオントロジーの概念階層に現れる，特徴的な構造を利用する．オントロジー中の各概念の性質や属性は，同一オントロジー中の別部分にある概念階層中の概念を参照しながら定義される．そのためクラス分類の基準が同一であり，分類観点に一貫性がある時，参照関係にある概念階層がその分類観点に沿って相似形になるといった特徴が現れる．この「参照関係にある概念階層の相似性」は，オントロジーの is-a 階層の構造上の特徴であるため，オントロジーの構築対象であるドメインには依存しない．ゆえに本洗練・拡張手法は，幅広いドメインのオントロジーに対して適用可能である．

本研究では，参照関係にある概念階層の相似性を利用した概念階層の比較手法によるオントロジーの洗練及び拡張支援システムを開発する．

3，4 章で述べる洗練支援システムでは，同一オントロジー中の参照関係にある概念階層同士を比較し，非相似な概念階層が相似形となるような概念及び概念間関係の追加提案を行う．こうした提案によって，オントロジーの定義内容の一貫性を向上させ

るような洗練方法を，オントロジー構築者に対して自動的に提示する内容洗練支援が可能となる．オントロジー構築者は個々に，異なるドメインを対象にオントロジーを構築するが，この洗練システムは，対象ドメインに依存しないため，幅広いオントロジー構築者にとって有用である．さらに，開発した内容洗練支援システムの有用性を検証するために，オントロジー構築経験者を被験者とした，システムの評価実験を行う．

5 章で述べる拡張支援システムでは，構築第 3 段階目の構築途中段階のオントロジーを対象とする．途中段階のオントロジーは，洗練段階のものと比べて，概念間の定義が疎であるため，概念間関係が密であることを前提としている洗練手法と同様の概念階層比較手法をとることはできない. そのため, 少ない概念間関係を有効に利用し，既存のプロパティと同様の関係を結ぶべき概念の組み合わせを概念階層の比較によって推定・提案する手法を開発する．このシステムの利用により，オントロジー構築者は，概念階層に含まれる全概念を考察することなく，対象オントロジー中に潜在的に存在するプロパティを，簡単に発見することが可能となる．さらにプロパティの拡張作業の更なる効率化のため，Linked Open Data と呼ばれる Web 上のリソースを用いて，推定したプロパティをより尤もらしい順に順位付けする手法も併せて開発する．さらに評価実験として，実際に開発途中の大規模なオントロジーである JST シソーラスを基にしたバイオサイエンスオントロジーに対して，本拡張支援システムを適用する．オントロジー構築作業者にシステムによる推定結果を提示し，構築者からの評価を得る．

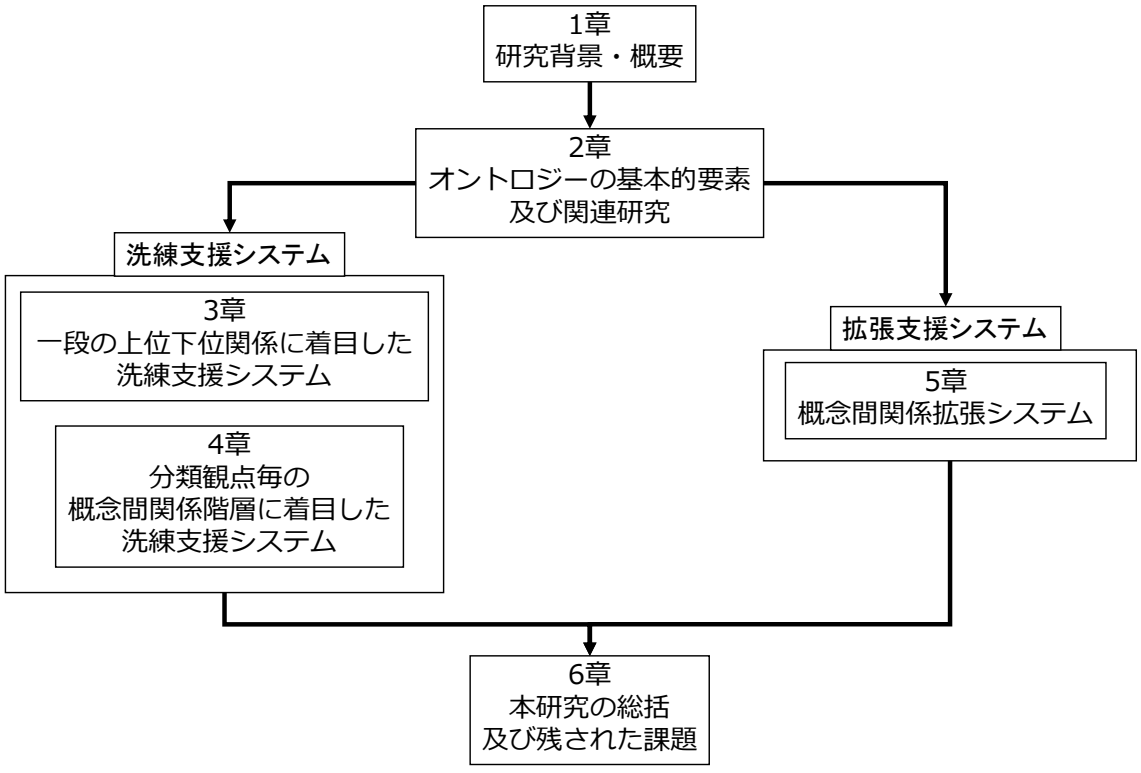

図 1. 2.　各章間の関係

1.3　本論文の構成

本論文は，全 6 章から構成されており，その内容は以下のとおりである．図 1.2.に各章間の関連を示す.

第 2 章では，まず，オントロジーを構成する基本的な要素である，「概念」と「概念間関係」について述べる．その後，関連研究として，従来研究におけるオントロジー洗練支援システム及び拡張支援システムについて述べ，本研究の位置づけを行う．さらに，オントロジー洗練・拡張システムが目指す品質の高いオントロジーが満たすべきオントロジー構築の推奨基準について述べ，本手法の基盤となっている参照関係にある概念階層同士に見られる相似性について述べる.

第 3 章では，洗練段階のオントロジーに対する洗練支援手法として，まずは，概念階層を構成する基本的な関係である上位下位関係に着目した概念階層比較手法について述べる．この概念階層の比較では，概念間関係を一段毎に区切り，概念階層同士の構造を比較し，互いに相似形となるような概念追加提案を行う．オントロジー構築者は，この支援手法を実装したシステムによって，洗練対象のオントロジーから洗練箇所及び洗練方法が提示されるため，洗練コストの削減が期待できる.

第 4 章では，分類観点毎に兄弟概念を併せて概念階層を比較し，包括的な概念階層の情報を構築者に提示する手法について述べる．3，4 章の手法を統合することで，オントロジー洗練作業のより一層の効率化が望める.

第 5 章では，概念階層比較手法を応用したオントロジー拡張支援システムについて述べる．第 3，4 章で述べてきた洗練支援システムとは異なり，構築途中段階のオントロジーを対象としたものである．構築途中段階のオントロジーは，概念定義がスパースであるため，概念及び概念間関係の定義が密であることを前提とした洗練支援手法では，検出・提案が不可能である．そこで，概念階層の比較に加えて，外部のデータから得られる関係を利用することで，概念間関係数が不十分な概念階層に対して，結ぶべき概念間関係を推定する手法を開発した．この拡張支援により，オントロジー構築作業の効率化が期待できる.

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．さらに残された課題及び今後の展望について述べる.

第2章　オントロジーを構成する基本的要素及び関連研究

本章では，オントロジーを構成する基本的な要素である「概念」と「概念間関係」について概説し，本論文で用いる表現方法の定義を行う．次にオントロジーの洗練・拡張についての関連研究について述べ本研究の位置づけを行う．最後に後の3，4，5 章での共通事項である，概念階層の比較について概説する．

2.1　オントロジー構築の一般事項

ここでは，オントロジーを構成する重要な要素である，概念と概念との間の関係について述べる．概念と概念の間には，以下の 2 種類の関係がある．

1. is-a 関係：概念の分類階層を表す．
2. part-of / attribute-of 関係[12]：概念の性質を定義する．

2.1.1　オントロジーの概念階層を構成する is-a 関係

is-a 関係は，概念に所属する個物（instance, individual）の包含関係を規定するものであり，オントロジーの中枢となる要素である．例えば，

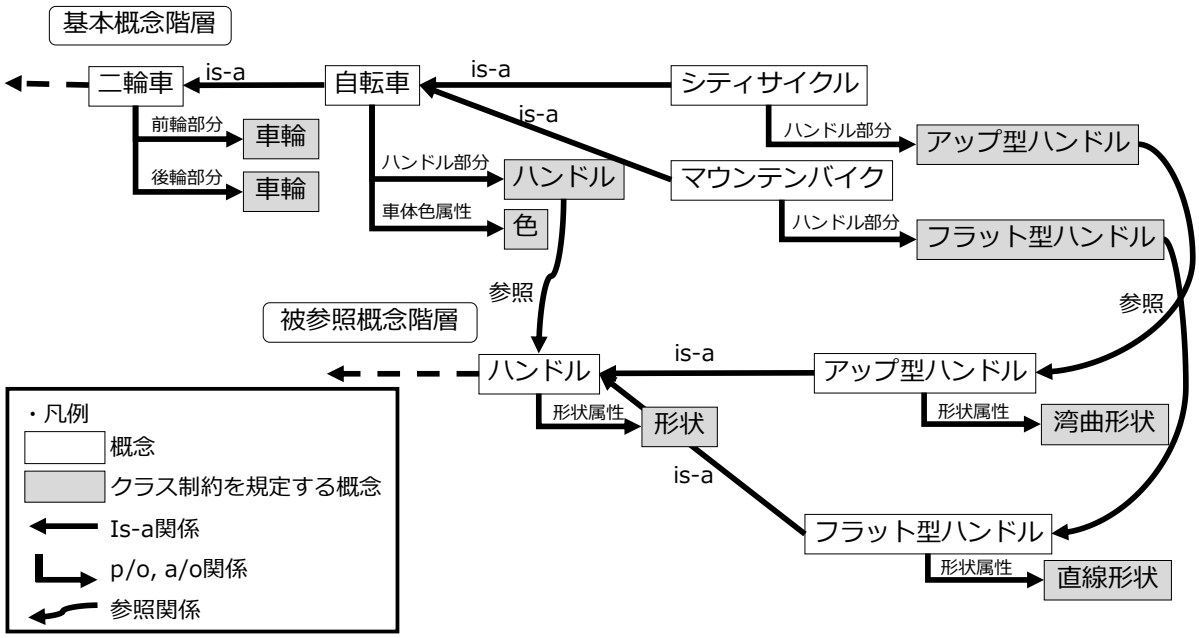

図 2.1.　オントロジーを構成する概念と概念間関係

[12] 以降の説明で，単に「概念間関係」と述べる場合，狭義の概念間関係である part-of / attribute-of 関係を指す．

<人間 is-a 哺乳類>

という表現は，“人間の外延⊆哺乳類の外延”，つまり，人間として数えられる個物の集合は，哺乳類として数えられる個物の集合に包含される，ということを意味している．is-a 関係がある 2 概念のうち，子にあたる概念を下位概念，親にあたる概念を上位概念と呼ぶ．このような関係に基づいた階層は，「is-a 階層」と呼ばれる．オントロジーは，is-a 関係を反映した概念階層である is-a 階層を中心に構成される．

is-a 階層の例を図 2.1.に示す．これは乗り物に関するオントロジーの一部である．ここでは，白長方形で示されている“二輪車”，“自転車”などが概念を表す．概念間の“is-a”と書かれた矢印は，is-a 関係を意味する．

2.1.2　各概念の部分や属性を定義する概念間関係

次に，part-of / attribute-of 関係について説明する．これらは，概念の性質を定義するために用いられる関係である．この p/o，a/o 関係は，is-a 関係が上位概念及び下位概念の 2 つの要素によって規定されていたのに対し，

- p/o，a/o によって性質が定義される「基本概念」
- p/o，a/o の関係の種類を表す「概念間関係のラベル」
- p/o，a/o の定義のために部分や属性の制約として参照される「被参照概念」

の 3 つの要素によって規定される．

part-of 関係（p/o）は，ある概念を全体物とみなしたとき，その部分となる概念（部分概念）との関係を示す．例えば，図 2.1.では，“二輪車”がその部分として“前輪部分”を持つことが，“二輪車”ノードと“車輪”ノードと p/o 関係を示す矢印記号によって表現されている．

attribute-of 関係（a/o）は，ある概念がもつ属性を表す関係である．例えば，“自転車”は何かしらの“色”を“車体色属性”として持つことが，“自転車”と“色”ノードと 2 概念間に結ばれた a/o 関係“車体色属性”の矢印によって，表現されている．

この p/o 及び a/o は，“前輪部分”や“車体色属性”のように概念間関係を表す名前を持つ．この「概念間関係のラベル」は，ある概念の下でどのような部分や属性であるかを表す．図 2.1.中の“二輪車”と“車輪”との間の p/o 関係“前輪部分”では，“車輪”は，“二輪車”を構成する部分の中でも“前輪部分”であることを表している．この概念間関係のラベルにより，p/o 及び a/o をより明示的な概念間関係として表現できる．例えば，図 2.1.の“自転車”の“ハンドル”，“二輪車”の“車輪”はどちらも全体部分関係であるが，異なる概念間関係のラベル“ハンドル部分”と“前輪部分”を持つ．“ハンドル部分”は，“自転車”にはその進行方向を制御する

ために用いられる[13]“ハンドル部分”があり，その部分には“ハンドル”が使われるとしている．一方，“前輪部分”は，“二輪車”には，車体を接地させハンドルの入力を地面に伝えるために用いられる“前輪部分”があり，その部分には“車輪”が使われるとしている．このように同じ部分を表す関係の中でもある概念の下でどのような部分であるかを明示化できる．

さらに，p/o および a/o によって，“ハンドル”や“車輪”のように，“自転車”や“二輪車”の部分となる概念の制約が定義されている．このように概念間関係によって規定された役割を担う対象の概念に関する制約を「クラス制約」と呼ぶ．図 2.1. では，“二輪車”と“車輪”の間に，“前輪部分”の概念間関係が結ばれている．この概念間関係は，“二輪車”の部分を表すと同時に，

「“二輪車”の“前輪部分”は，“車輪”でなければならない」

という制約を表している．このクラス制約は，オントロジー内の別の箇所で定義された概念を参照して，概念の部分や属性となるものの制約を規定する．このクラス制約として参照されている概念を「被参照概念」と呼ぶ．また，この概念間関係によって性質や部分が定義されている概念を「基本概念」と呼ぶ．この概念間関係のラベルと被参照概念のラベルは，“ハンドル部分”と“ハンドル”のように似たラベルとなることもあるが，前者は，この概念間関係により規定される部分や属性を規定するものであり，後者は，クラス制約として参照されている概念を指すため，それぞれ区別して扱う必要がある．

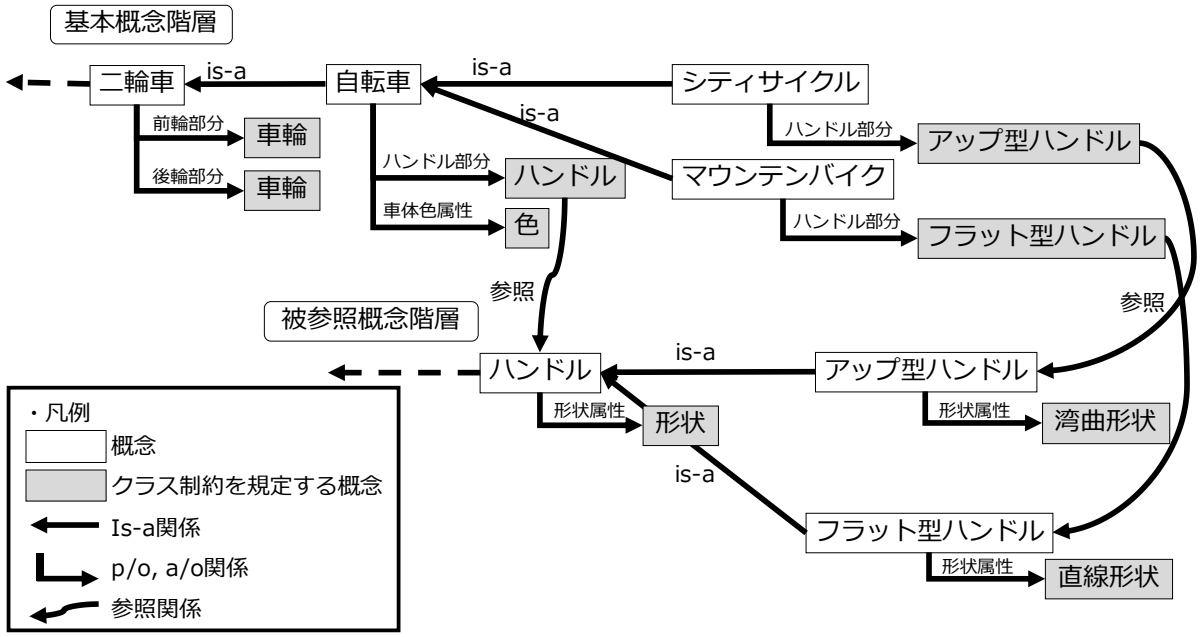

図 2.2.　ハンドルを基本概念とする概念階層

[13] ハンドル部分というラベルのみでは，「進行方向を制御するために用いられる」といった内容は厳密には定義されていない．概念間関係の厳密な定義には，各概念の定義と同様に部分や属性などの性質を記述する必要がある．ただし本研究の範囲では，ラベルのみでこうした定義が行われているとする．

図 2.1.の is-a 階層では，“自転車”，“シティサイクル”，“マウンテンバイク”の各概念は，その“ハンドル部分”を定義するために，“ハンドル”，“アップ型ハンドル”，“フラット型ハンドル”の各概念をクラス制約として参照している．これらの被参照概念は，図 2.1.下部のような is-a 階層を構成している．このように被参照概念によって構成されている is-a 階層を「被参照概念階層」と呼ぶ．それに対して，図 2.1.上部の“自転車”，“シティサイクル”，“マウンテンバイク”の「基本概念」も同様に is-a 階層を構成しており，「基本概念階層」と呼ぶ．

この「基本概念」と「被参照概念」の関係は相対的なものである．図 2.1.下部の“ハンドル”の is-a 階層は，図 2.1.上部の“自転車”から見れば被参照概念である．一方，図 2.1.下部の被参照概念階層は，“ハンドル”とその下位概念から構成された is-a 階層であり，それぞれの形状を定義するために，“形状”，“湾曲形状”，“直線形状”の概念を参照している．このときハンドルの階層は「基本概念階層」であり，形状の階層は「被参照概念階層」となる（図 2.2.）．このようにそれぞれの概念間関係毎に参照・被参照の関係は相対的に決まる．

2.1.3　概念及び概念間関係の特殊化

概念の特殊化は，その性質の範囲をより狭くして，その条件を満たす個物の集合を小さくすることである[溝口 05]．この概念の特殊化は，is-a 関係が結ばれている上位概念と下位概念の関係に相当する．また，概念間関係の特殊化は，上位概念で定義された概念間関係を，下位概念でより詳細な概念間関係にすることである．特殊化された概念間関係は，以下の 2 つの条件を満たす．

条件1.　ある概念 *C* に定義される概念間関係のクラス制約として参照できる概念 *C_ref* は，*C* の上位概念 *C_upper* に定義された概念間関係のクラス制約として参照されている概念 *C_upper_ref* の下位概念である（図 2.3.）．

条件2.　上位の概念間関係と下位の概念間関係は，同一の概念間関係のラベルを持つ．

この時，概念と同様に概念間関係にも is-a 関係が規定される．

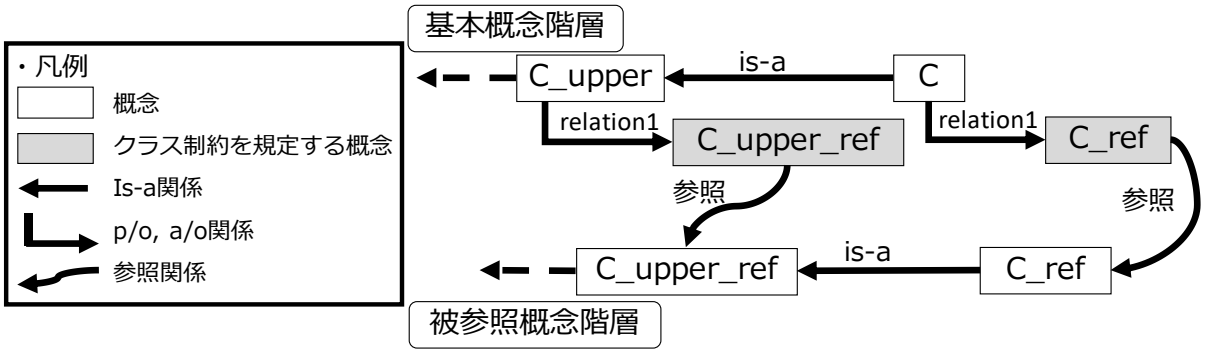

図 2.3. クラス制約として参照可能な概念

図 2.1.の is-a 階層では，“自転車”から“シティサイクル”へ特殊化されている．この時，概念間関係の三要素である，基本概念，概念間関係のラベル，被参照概念を 3 つ組<基本概念，概念間関係のラベル，被参照概念>で表すと，“自転車”の持つ概念間関係は，

<“自転車”，“ハンドル部分”，“ハンドル”>

であり，この概念間関係が規定するクラス制約として，“ハンドル”を参照している．一方，“シティサイクル”の持つ概念間関係は，

<“シティサイクル”，“ハンドル部分”，“アップ型ハンドル”>

であり，この概念間関係が規定するクラス制約として，“アップ型ハンドル”を参照している．この 2 つの概念間関係は，概念間関係のラベル“ハンドル部分”が共通であり，“アップ型ハンドル”は“ハンドル”の下位概念であるので，条件 1，2 を満たす．この例の場合は“ハンドル部分”のクラス制約として参照する概念の範囲を狭くすることで，“自転車”から“シティサイクル”へと特殊化されている．このように概念の特殊化は，クラス制約として参照する概念の範囲を限定することで行われる．

ただし，クラス制約として参照される概念は，直接 is-a 関係が定義されている必要はなく，中間概念が存在していてもよい．なぜなら is-a 関係には推移律が成り立つため，ある概念 *C* の下位概念の下位概念も *C* の下位概念であるからである．このように再帰的に is-a 関係が繰り返され，上位の概念間関係の被参照概念から下位の概念間関係の被参照概念へと繋がる経路が存在する場合，その間に概念が存在していても，条件 1 は満たされる．こうした共通のラベルを持つ概念間関係が構成する is-a 階層を「概念間関係階層」[14]と呼ぶ．

2.1.4 オントロジーを構成する概念と概念間の関係

これまでに述べてきたように，オントロジーは概念と概念間の関係によって構成される．オントロジーに含まれる概念の集合を，

$$\text{全概念の集合：}\mathbf{C} = \{C_1, C_2, C_3, \dots, C_l, \dots, C_L\} \quad (2.1.)$$

と表現する．また，is-a 関係と p/o，a/o 関係は，

$$\text{概念及び概念間関係の is-a 関係の集合：}\mathbf{I} = \{I_1, I_2, I_3, \dots, I_m, .., I_M\} \quad (2.2.)$$

$$\text{全 p/o, a/o 関係の集合：}\mathbf{R} = \{R_1, R_2, R_3, \dots, R_n, \dots, R_N\} \quad (2.3.)$$

と表現する．各添え字は互いに独立であり，*L*，*M*，*N* はそれぞれの集合の要素の総数である．

また，ある概念 *C* の一段上位および一段下位を，

[14]概念間関係階層は，原則として概念間関係のラベルにつき 1 つである．ただし，ある概念間関係のラベルを分類観点とする複数の基本概念階層が存在する場合，複数存在することもある．この場合，それぞれの概念間関係階層は別のものとして取り扱う．

$$\mathrm{Upper}(C),\ \mathrm{Lower}(C) \tag{2.4.}$$

とし，ある概念間関係 R の一段上位および一段下位を，

$$\mathrm{Upper}(R),\ \mathrm{Lower}(R) \tag{2.5.}$$

と表現する．Upper および Lower の引数は，概念 C と概念間関係 R の両方を取りうる．オントロジーの文法制約上，ある概念・概念間関係の上位は一つに限られる[15]ので Upper(C)及び Upper(R)は一つの概念である．一方，ある概念・概念間関係の下位は複数存在するので Lower(C)及び Lower(R)は概念または概念間関係の集合である．また，概念と概念間関係の間に is-a 関係は定義されない．そのため，概念 C の上位や下位に概念間関係を取ることはない.

式 2.2.の is-a 関係の各要素は，

$$I_m = \mathrm{Isa}(\text{上位},\ \text{下位}) \tag{2.6.}$$

と表現する．式 2.4.を用いると概念間の is-a 関係は，

$$I_m = \mathrm{Isa}(\mathrm{Upper}(C_l), C_l)$$

と書け，同様に式 2.5.を用いて概念間関係間の is-a 関係は，

$$I_m = \mathrm{Isa}(\mathrm{Upper}(R_n), R_n)$$

と書ける．

R は，オントロジーに定義された p/o，a/o の概念間関係の集合であるので，各R_nは，「基本概念」,「被参照概念」と「概念間関係のラベル」の計 3 要素で定義される．そこで，概念間関係のラベルの集合を

$$\text{全概念間関係ラベルの集合：}\mathbf{P} = \{P_1, P_2, P_3, \dots, P_k, \dots, P_K\} \tag{2.7.}$$

と表現し，概念間関係R_nを，

$$R_n = \mathrm{Relation}(\text{“基本概念”},\ P_k,\ \text{“被参照概念”}) \tag{2.8.}$$

と表現する．「基本概念」と「被参照概念」は，**R** の各要素R_nに対して一意に決まるので，

$$R_n\text{の基本概念} = \mathrm{Basic}(R_n) \tag{2.9.}$$

$$R_n\text{の被参照概念} = \mathrm{Const}(R_n) \tag{2.10.}$$

と表現できる．Basic()は，基本概念（Basic Concept）を意味し，Const()は，クラス制約（Class Constraint）として参照されていることを意味する．式 2.8.～2.10.を用いて，概念間関係 R_n は，

$$R_n = \mathrm{Relation}\left(\mathrm{Basic}(R_n),\ P_k,\ \mathrm{Const}(R_n)\right)$$

と表現できる．

[15]多重継承を認める場合は，複数の上位概念の存在を許すため，Upper(C)も概念の集合になるが，本研究では多重継承を認めないオントロジーを対象とする．

2.1.5 概念の分類観点

概念の特殊化は，概念の部分や属性によって定義されるクラス制約の範囲を狭くし，概念間関係によって定義される性質を限定することで行われる．そのため，オントロジー構築者がどのような観点で概念を特殊化したかは，概念間関係がどのように特殊化されたかに表されている．本研究では，この概念間関係のラベルによって表される，オントロジー構築者が概念定義時に着目した観点を「分類観点」と呼ぶ．例えば，図 2.1.では，“自転車”の，“ハンドル部分”というラベルを持つ概念間関係

Relation(“自転車”,“ハンドル部分”,“ハンドル”)

が，その下位概念の“シティサイクル”と“マウンテンバイク”では，それぞれ，

Relation(“シティサイクル”,“ハンドル部分”,“アップ型ハンドル”)

Relation(“マウンテンバイク”,“ハンドル部分”,“フラット型ハンドル”)

に特殊化されている．そのため，オントロジー構築者は，“自転車”から“シティサイクル”と“マウンテンバイク”に特殊化する際に，各概念の“ハンドル部分”に着目したことがわかる．

2.1.6 オントロジーの表現方法

ここまでに述べてきたオントロジーを構成する要素は，オントロジーを表現する枠組みやツールによって表現が異なる．ここでは，代表的なオントロジー表現方法である，OWL 形式と法造形式について述べる．

OWL（Web Ontology Language）は，2004 年に W3C（World Wide Web Consortium）による標準化されたものである．セマンティックウェブは，情報リソースにセマンティクスを付与し，コンピュータが自律的に情報を処理できるようにする技術[W3C]である．様々な知識や事実は，概念と概念間関係からなるグラフとして表現され，その記述に用いる語彙として，RDF（Resource Description Framework）がある．その語彙を定義するための言語として RDF スキーマ（RDFS）がある．OWL は，より厳密なオントロジーを定義するために提案されたものである．OWL では，概念間関係 R_n は，「プロパティ」と呼ばれ，概念間関係のラベルは「プロパティ名」と呼ばれる．また，RDFS における，rdfs : subClassOf（is-a 関係），rdfs : domain（プロパティの定義域）rdfs : range（プロパティの値域）の表現が利用可能である[神崎 04, W3C]．OWL は，その表現能力から次の 3 つに分けられる．

- OWL lite：クラスの is-a 階層や簡単な制約などの最低限の表現能力を持つ
- OWL DL：クラスの論理表現や詳細な制約記述が可能で，推論の完全性が保証される
- OWL Full：最も自由な記述が許されるが推論の完全性が保証されない

この OWL 形式のオントロジーに対応したエディタとしては，スタンフォード大学が開発した Protégé[16]が代表的である．

次に法造形式について紹介する．法造は，ロール理論を中心としたオントロジー構築理論に基づいたロール概念を正しく扱うための記述能力を持つオントロジーエディタである．ロール概念とは，ある概念が特定の概念のもとで果たす役割を捉えて概念化したものである[古崎 02]．この「ロール概念」は，2.1.4.節での式 2.7.の概念間関係のラベル **P** に相当する．例えば，図 2.1.中の概念間関係のラベルである“ハンドル部分”は，“自転車”という概念の中でのある部分の役割（ロール）を捉えて概念化した「ロール概念」である．一方，概念間関係のクラス制約として参照されている“ハンドル”は，その役割を担う対象となる概念である．このように法造上では，ある概念における役割を概念化した「ロール概念」とその役割を担う対象となる概念を明示的に書き分けることが可能である．

法造上で取り扱うオントロジーでは，概念間関係も，概念と同様にある概念間関係R_nと上位の概念間関係Upper(R_n)の間に明示的に is-a 関係が明示的に定義される．OWL 形式の場合では，明示的に概念間関係の is-a 関係が定義されることはないが，同一のプロパティ名を持つプロパティ間には，法造上での概念間関係の is-a 関係相当の関係があるとみなすことができる．

また法造のインポート機能を用いることによって，Protégé 同様に OWL 形式のオントロジーを扱うことも可能であり，さらに機能拡張のための API が提供されている．本研究では，この API をベースに洗練・拡張支援システムを開発する．

2.2　オントロジー構築に関する関連研究

オントロジー構築には，ここまでに述べてきたようなオントロジーに関する基本的な要素を考慮することに加え，構築しようとしているオントロジーのドメインに関する専門知識も必要であり，多くの時間と労力を必要とする．そのためこれまでにもオントロジー構築を支援するような関連研究が行われてきた．本節では，本研究の対象であるオントロジー構築の第 3，4 段階に対応付けて，オントロジー構築支援に関する関連研究を紹介する．

2.2.1　オントロジーの評価洗練に焦点を当てた支援

オントロジーの評価及び洗練段階に対する支援方法には大きく分けて 2 種類存在する．オントロジーの文法規則に関する形式的なエラー修正とオントロジーの内容洗

[16] https://protege.stanford.edu/

練の 2 種類である．本節では，これら 2 種類の洗練支援に関する既存研究について述べ本研究の位置づけを説明する．

2.2.1.1 文法規則に関する形式的エラー修正

文法規則に関する形式的エラーとは，オントロジーの記述言語の仕様における文法規則の違反である．例えば，概念に関する文法規則には，「概念名ラベルは重複しない」，「自身の上位概念を下位概念として指定できない」などがあり，他にも概念間関係，インスタンスに関する規則などがある．

こうした文法規則に関する形式的エラーは，その記述言語に対応した推論機構を用いて，機械的に処理できるため，Ontology Debugging として多くの既存研究で修正の対象とされた．OWL 形式の推論機構を用いたものには，Pellet[Sirin 07]や OWL のエディタである Protégé[Noy 01b]にも実装されているもの[Horridge 11]などがある．また，オントロジーエディタ“法造”[古崎 02]上にも，「整合性チェック機能」[太田 11a]として実装されている．こうしたエラーチェックの機能を応用して，オントロジーへの公理の追加やそのエラーチェックを自然言語の対話形式で行うシステム[Denaux 12]などがある．この他にも，Poveda らは，オントロジー構築時に陥りやすい文法的な間違いを収集し，その違反を検出する Ontology Pitfall Scanner（OOPS!）[Poveda-Villalon 12]を開発し，さらにオントロジーの品質診断に応用している[Poveda-Villalon 16].

形式的なエラーの例として，クラス制約違反を紹介する．これは，2.1.3.節で説明した条件 1「ある概念 *C* に定義される概念間関係のクラス制約として参照できる概念 *C_ref* は，*C* の上位概念 *C_upper* に定義された概念間関係のクラス制約として参照されている概念 *C_upper_ref* の下位概念である.」の違反である．例えば，図 2.4.上部の概念階層は，乗り物とその下位概念からなる概念階層であり，“移動空間属性”という属性の概念間関係が定義されている．この属性を定義するために，図 2.4.下部にある

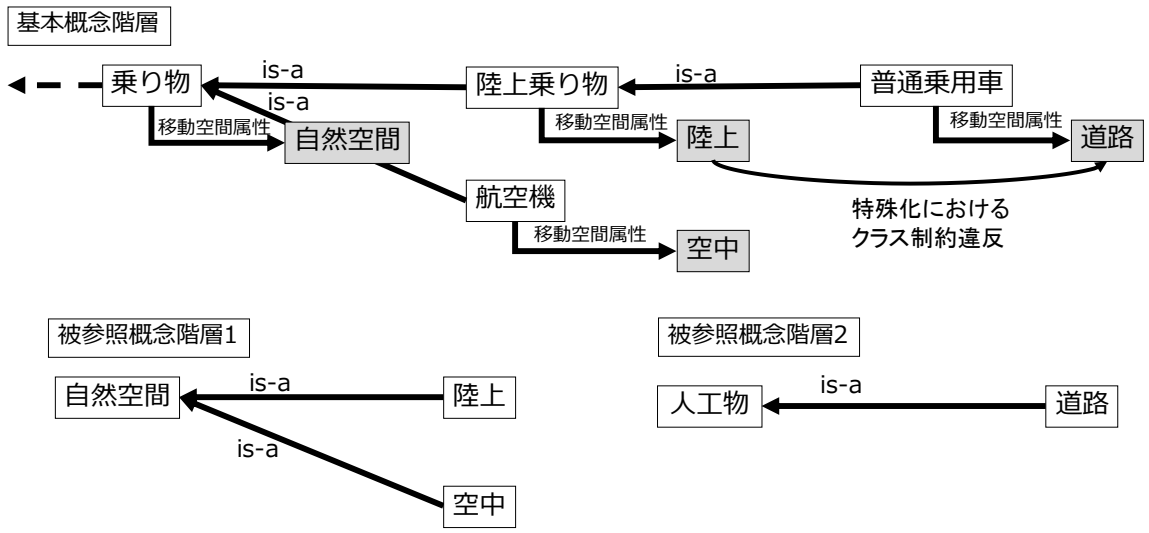

図 2.4. クラス制約に関する文法規則違反の例

2 つの被参照概念階層中の概念がこの概念間関係によるクラス制約として参照されている．いま，“乗り物”，“陸上乗り物”，“航空機”は，

$$R_1 = \text{Relation}(\text{“乗り物”}, \text{“移動空間属性”}, \text{“自然空間”})$$
$$R_2 = \text{Relation}(\text{“陸上乗り物”}, \text{“移動空間属性”}, \text{“陸上”})$$
$$R_3 = \text{Relation}(\text{“航空機”}, \text{“移動空間属性”}, \text{“空中”})$$

の 3 つの概念間関係を持ち，各関係はそのクラス制約として，被参照概念階層 1 中の概念を参照している．各関係とそのクラス制約として参照されている被参照概念は，

$$\text{Const}(R_1) = \text{“自然空間”}$$
$$\text{Const}(R_2) = \text{“陸上”}$$
$$\text{Const}(R_3) = \text{“空中”}$$

と表すことができる．ここで，“普通乗用車”は，

$$R_4 = \text{Relation}(\text{“普通乗用車”}, \text{“移動空間属性”}, \text{“道路”})$$

の概念間関係を持ち，クラス制約として参照している被参照概念は，

$$\text{Const}(R_4) = \text{“道路”}$$

であり，被参照概念階層 2 中の概念を参照している．R_1～R_4までの概念間関係は同じ概念間関係のラベルを持ち，is-a 関係を持つため，最上位の概念間関係R_1が，“自然空間”を参照している時，R_2～R_4でクラス制約として参照できる概念は，その下位概念でなければならない．しかし，“陸上”には下位概念は存在せず，“道路”は“人工物”の下位概念である．そのため図 2.4.の概念階層は，特殊化におけるクラス制約に関する違反を犯している．

こうした文法規則に関するエラー修正に関する研究は数多く行われ多くの構築ツール上に実装されているので，本研究で洗練対象とするオントロジーは，こうしたエラーは修正済みであるという前提とする．

2.2.1.2　オントロジーの内容洗練

オントロジーの内容洗練は，文法規則のエラーとは異なり，オントロジーの概念定義や特殊化の方針などオントロジーの意味的内容を洗練することである．内容洗練の例として，オントロジーの分類上の関係の妥当性を評価する OntoClean[Guarino09]やドメイン知識俯瞰のためのオントロジー探索ツール[廣田 09]を利用したオントロジー内容洗練[Ohta 11b]などがある．OntoClean で提案されている手法を用いるためには，メタプロパティと呼ばれる概念の性質に関するメタ的な情報付与が必要であり，そうした作業には，オントロジー構築に関する知識はもとより，付加的な知識も必要であり利用者に課する負担が大きい．また，Ohta らの手法は，オントロジーの内容評価や洗練を行うためには，専門家が概念間の関係性を把握し，オントロジーの定義内容を適切に理解する必要があるという点に着目している．オントロジー探索ツールを用いて概念間の関係性をドメインの専門家の視点に応じて，内容の把握しやすい形式で可

視化することで，ドメインの専門家がオントロジーの意味的な評価を行いやすくするというアプローチである．こうした既存の内容洗練支援手法を利用するためには，内容評価に対象ドメインの専門知識及び，オントロジー構築に関する深い知識と経験が不可欠である．つまり，既存の内容洗練手法では洗練箇所と方法を，構築者自らが見つける必要がある．ゆえに，洗練作業におけるコストは依然として大きいままとなっている．

2.2.2 オントロジーの拡張に焦点を当てた支援

本節では，オントロジー構築過程における第 3 段階の「概念・関係の定義」を拡張するための支援手法について述べる．概念の収集や概念の記述の拡張には，オントロジー学習という手法が代表的である．オントロジー学習とは，オントロジー構築コストを削減することを目的として，（半）自動的にオントロジーの構築を支援する試みである[森田 12]．学習には，Wikipedia や Folksonomy などの半構造化情報，データベースや既存のオントロジーなどの構造化情報が用いられ，様々な情報資源が用いられている[Osborne 15，Suchanek 07，Hoffart 13，Wu 08，玉川 10，玉川 11]．ここでは，半構造化データを利用したオントロジー学習について述べる．半構造化情報として代表的なものとして Wikipedia がある．Wikipedia は，Web 上の百科事典として，多くの語彙が収録されており，Wikipedia ユーザによって随時更新が行われているため即時更新性に優れている．また，構造化情報として，infobox，カテゴリツリー，Wikilink 等が存在する．こうした特徴を生かした Wikipedia のオントロジー化に関する研究が進められている．DBpedia[Bizer 09]では，infobox やカテゴリを利用し，半構造化情報を RDF に変換した大規模なデータベースを構築している．また YAGO[Sunchenek 08]では，Wikipedia のカテゴリと Wordnet の synet を統合することで大規模なオントロジーを構築している．

こうした手法では，獲得可能な概念や関係は，入力とする情報源に依存する．そのためオントロジーのカバー範囲を確保するためには，大規模な入力データが必要とされる．また，Wikipedia のような一般語のデータベースには存在しない専門用語や暗黙的な知識については，ドメインの専門家が後から手動で追加する必要がある．

2.3 本研究で目指す品質の高いオントロジー

2.3.1 概念クラス分類観点の同一性

オントロジーの内容洗練とは，概念定義の内容や特殊化の方針などオントロジーの意味的内容を洗練することである．オントロジー構築の際に推奨される基準としては，いくつか種類があるが[溝口 05]，本研究では，クラス分類観点の同一性に着

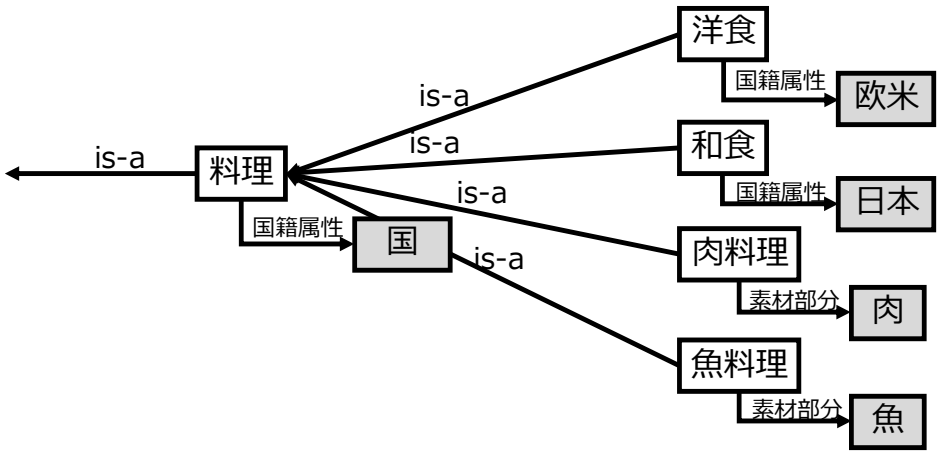

図 2.5.　複数の分類観点が混在している概念階層例

目する．なぜなら，オントロジーにおける分類観点は，オントロジー構築者が対象世界を分類する際にどのような性質に着目したのか，という各概念定義の核となる意味を持つからである．そのため，継承・特殊化される各概念の性質は，オントロジー構築者がどのような性質で分類したかが，互いに矛盾せずに明示的に表現されていることが望ましい．オントロジーの分類観点が一貫していない場合には，概念毎の分類観点が明示化されず，オントロジー構築者が対象世界をどのようにモデル化したのかが不明確になる．オントロジー構築には，多大な人的時間的コストが要求され，分類観点の一貫性を保ちながら品質の高いオントロジーを構築していくことは困難を伴うため，本研究では分類観点の同一性[溝口 05]に着目した洗練支援に着目する．

図 2.5.に複数の分類観点が混在している例を示す．この例では，“料理”の下位概念を定義する際の分類観点として“国籍属性”と“素材部分”の 2 種類存在する．こうしたオントロジーでは，1 つの概念が複数の上位概念を持ってしまう「多重継承」が発生することがある．例えば，このオントロジーに新しく下位概念として，“ビーフシチュー”を追加することを考える．“ビーフシチュー”は，“洋食”の下位概念であり，“肉料理”の下位概念でもある．このようになる理由として，“料理”の下位概念に複数の分類観点が混在しているため，新規に下位概念を追加する際に，どちらの分類観点にも該当する可能性があるからである．この「多重継承」は，多くのオントロジー構築方法論やガイドライン[Noy 01，溝口 05，Smith 07]で非推奨とされている．そのため，その概念の特殊化の際にどの性質に着目するべきであるかを常に意識しながら，その観点が混在しないように概念定義及び特殊化を行う必要がある．

また，概念の分類時には，各概念がどのように分類されるのかを明示的に示す分類観点に着目する必要がある．料理を素材や国籍で分類するための概念階層であれば，“国籍”，“材料”などは，それが変わると別の料理となる．“魚料理”は，主た

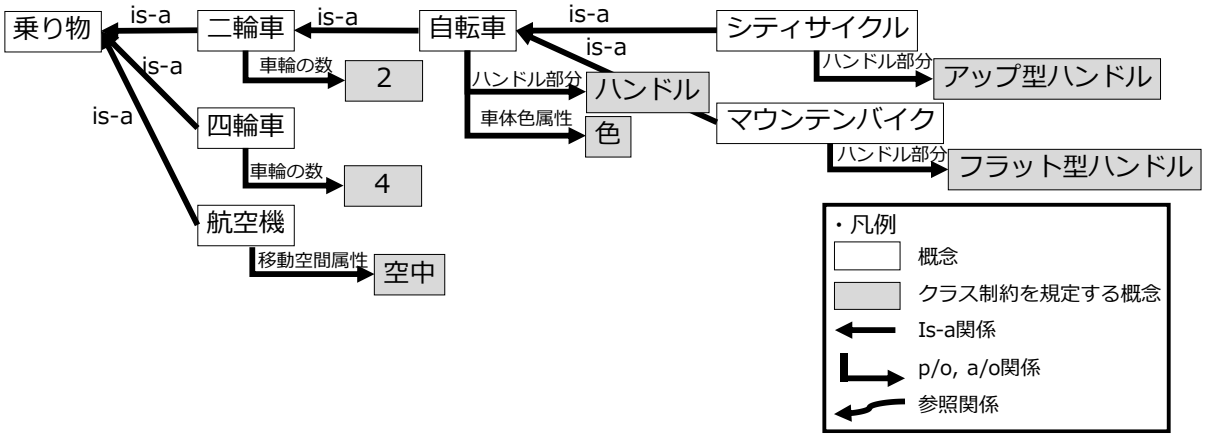

図 2.6.　分類観点が一貫していない概念階層例

る素材が“魚”である“料理”として定義される概念であり，それ以外の素材が主な場合，別の料理として定義されるべきである．一方，“利用する調理器具”，“誰が調理したか”，“いつ調理されたか”等は，調理器具，調理者，時間等の違いが，料理の分類に直接影響を及ぼさない．オントロジー構築者は，このような分類に直接関係のない性質に着目して概念を定義するべきではない．

概念階層を構築していく中で，オントロジー構築者が何に着目して概念の分類を行ったか，つまり分類観点は，概念間関係のラベルP_kに現れる．図 2.6.[17]の自転車のオントロジーでは，P_1 = “ハンドル部分”の分類観点に着目するとその概念間関係は，

$$R_1 = \text{Relation}(\text{“自転車”}, \text{“ハンドル部分”}, \text{“ハンドル”})$$

が，

$$R_3 = \text{Relation}(\text{“シティサイクル”}, \text{“ハンドル部分”}, \text{“アップ型ハンドル”})$$

$$R_4 = \text{Relation}(\text{“マウンテンバイク”}, \text{“ハンドル部分”}, \text{“フラット型ハンドル”})$$

に特殊化されている．これらの関係から，このオントロジー構築者が，上位概念から“ハンドル部分”を分類観点として着目して，“自転車”の下位概念を定義したことが分かる．品質の高いオントロジーでは，ある概念の直下の概念の集合の中で分類観点が混在しないことが望ましいとされる．

一方，分類観点が混在している特殊化の例が図 2.6.のオントロジーに存在する．自転車の上位概念である“二輪車”とその兄弟概念“四輪車”及び“航空機”では，複数の分類観点が存在する．“乗り物”から“二輪車”及び“四輪車”へは，“乗り物”の“車輪の数”に着目して特殊化されている．すなわち，この 2 つの下位概念の分類観点は，“車輪の数”であると言える．しかし，“航空機”はその車輪部分に着目して“乗り物”から特殊化されていない．ここでは，“航空機”は“移動空間属性”で分類されている．オントロジーは「共有されている概念化の形式的・明示的仕様」[Gruber 93]であるので，概念の特殊化における分類観点は，明示的でなくてはならない．しか

[17] 概念階層の見やすさを優先し，図 2.6.および図 2.7.では，説明に必要な概念間関係のみが示されている．

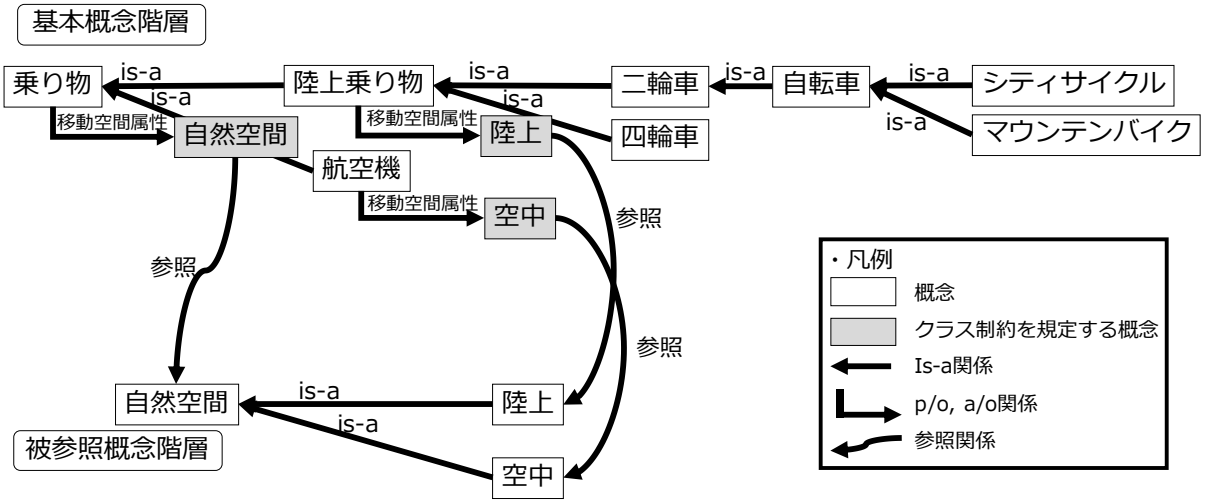

図 2.7.　中間概念を追加し分類観点を揃えた is-a 階層の例

しながら図 2.6.の is-a 階層では，分類観点が混在しているため，乗り物を 3 つの下位概念に特殊化する際に着目している性質が異なっており，分類観点が明示的ではない．このような分類観点が混在している概念階層に対しては，分類観点を統一するために，“二輪車”と“四輪車”の上位に“陸上乗り物”のような移動空間の属性を持つ中間概念を追加する洗練を行うことが望ましい．図 2.7.に中間概念“陸上乗り物”を追加し，概念クラス分類観点の同一性が満たしたオントロジーを示す．

2.3.2　概念間関係階層を介して参照関係にある概念階層間の相似性

2.3.1 節の図 2.7.の例のように分類観点が同一である概念階層は，その分類観点がP_kである概念間関係階層$\mathbf{R_{P_k}}$中の概念間関係を持つ基本概念階層と，クラス制約として参照されている被参照概念階層の両方が，互いに相似となる．$\mathbf{R_{P_k}}$は，全概念間関係の集合 $\mathbf{R}$ のうち分類観点がP_kである概念間関係の集合である．この相似条件は，2.1. 節で定義した概念及び概念間関係の表現方法を用いて，

$$\mathrm{Isa}\left(\mathrm{Basic}\left(\mathrm{Upper}(R_n)\right),\ \mathrm{Basic}(R_n)\right) \qquad (2.11.)$$

$$\mathrm{Isa}\left(\mathrm{Const}\left(\mathrm{Upper}(R_n)\right),\ \mathrm{Const}(R_n)\right) \qquad (2.12.)$$

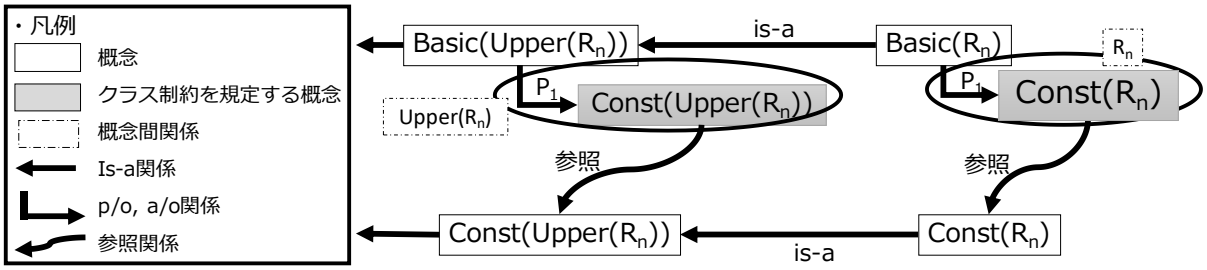

図 2.8.　相似な is-a 階層

の 2 式を満たすときである（図 2.8.）．図 2.8.の is-a 階層では，概念間関係R_nが定義されている基本概念$\mathrm{Basic}(R_n)$と，その上位の概念間関係$\mathrm{Upper}(R_n)$が定義されている基本概念$\mathrm{Basic}\big(\mathrm{Upper}(R_n)\big)$に is-a 関係が定義されているので式 2.11.を満たす．また，ある概念間関係 R_n が参照している被参照概念$\mathrm{Const}(R_n)$とその上位の概念間関係$\mathrm{Upper}(R_n)$が参照している被参照概$\mathrm{Const}\big(\mathrm{Upper}(R_n)\big)$に is-a 関係が定義されているので式 2.12.を満たす.

この時，$\mathbf{R}_{\mathbf{P_k}}$の集合が持つ基本概念および被参照概念と，それぞれの概念の間の is-a 関係から構成される概念階層は，2.1.1.節で述べたように,

- 基本概念階層（Basic Concept Hierarchy, BH）：基本概念とその is-a 階層に含まれる概念の集合からなる概念階層
- 被参照概念階層（Constraint Concept Hierarchy, CH）：被参照概念とその is-a 階層に含まれる概念の集合からなる概念階層

の 2 つである.

図 2.7.の例であれば,“移動空間属性”を定義している概念間関係は，次の 3 つである.

$$R_1 = \mathrm{Relation}(\text{“乗り物”},\text{“移動空間属性”},\text{“自然空間”})$$
$$R_2 = \mathrm{Relation}(\text{“陸上乗り物”},\text{“移動空間属性”},\text{“陸上”})$$
$$R_3 = \mathrm{Relation}(\text{“航空機”},\text{“移動空間属性”},\text{“空中”})$$

その概念間関係階層は,

$$\mathbf{R}_{\mathbf{P_k}} = \{R_1, R_2, R_3\}$$

とその間に定義される is-a 関係の集合である．いま,

$$\mathrm{Isa}(\text{“乗り物”},\text{“陸上乗り物”})$$

であり，R_1とR_2は,

$$\mathrm{Basic}(R_2) = \text{“陸上乗り物”}$$
$$\mathrm{Upper}(R_2) = R_1$$
$$\mathrm{Basic}\big(\mathrm{Upper}(R_2)\big) = \mathrm{Basic}(R_1) = \text{“乗り物”}$$

と書けるので，式 2.11.を満たす．また同様に,

$$\mathrm{Isa}(\text{“自然空間”},\text{“陸上”})$$

であり,

$$\mathrm{Const}(R_2) = \text{“陸上”}$$
$$\mathrm{Upper}(R_2) = R_1$$
$$\mathrm{Const}\big(\mathrm{Upper}(R_2)\big) = \mathrm{Const}(R_1) = \text{“自然空間”}$$

なので，式 2.12.を満たす．R_1とR_3も同様に相似条件を満たす．つまり，$\mathbf{R}_{\mathbf{P_k}}$に含まれる全ての概念間関係$R_n$において，$R_n$を持つ基本概念と，$R_n$から参照されている被参照概念が一対一対応している時，相似である．言い換えれば，概念間関係を起点として

相似判定の対象となっている基本概念階層と被参照概念階層に含まれる概念数を比較し，その数が同じであれば，相似である．

2.3.3　概念階層比較によるオントロジー洗練・拡張支援手法

本研究での洗練支援手法では，「概念間関係階層」を中心に，「基本概念階層」と，「被参照概念階層」の 3 つの階層に含まれる要素の数を互いに比較し，相似性の判定を行う．概念間関係階層は，概念間関係のラベルP_kごとに決まる概念間関係の集合$\mathbf{R}_{\mathbf{P_k}}$である．洗練支援手法の大まかな流れは，

1. オントロジーを入力
2. 概念間関係のラベルP_k毎に概念間関係階層を抽出
3. 概念間関係階層を比較しやすい範囲に分割
4. 概念間関係階層毎に，概念間関係を持つ「基本概念階層」と参照されている「被参照概念階層」を比較
5. 階層同士が非相似であれば洗練箇所として検出
6. 階層を互いに相似にするような概念の追加提案を出力

となる．

本研究では，概念間関係階層を一度に比較せず，上記 3 にあるように，概念階層の比較の前に，その中心となっている概念間関係階層$\mathbf{R}_{\mathbf{P_k}}$を分割する．分割される範囲は，以下 2 種類である．

- 上位下位間の関係に着目した概念階層比較（3 章で議論する）
- 兄弟概念間の関係に着目した概念階層比較（4 章で議論する）

3 章では，上位下位間の関係に着目した概念階層比較手法を用いて，洗練が必要な箇所の検出及び概念の追加方法の推定を行う．この手法では，概念の上位下位関係に着目するため，「基本概念階層」と「被参照概念階層」に定義された概念がどの程度詳細に定義されているかを揃えるための洗練提案となる．上下一段毎に概念間関係階層を区切り，それに伴う「基本概念」と「被参照概念」を比較することによって，各概念の特殊化の詳細度を揃えるような概念の追加提案を行い，概念階層毎に特殊化の詳細度が偏ることを避け，両方の概念階層の特殊化の詳細度がバランス良く定義されることを目指す．

4 章で述べる兄弟概念間の関係に着目した概念階層比較では，概念階層の比較範囲を上下一段の概念間関係階層から，同じ概念間関係のラベルP_kを持つ概念間関係によって構成される概念間関係階層全体とその概念間関係をもつ「基本概念階層」と参照されている「被参照概念階層」を比較対象とする．この比較範囲の拡張により，3 章での局所的な概念階層比較手法では検出できなかった兄弟概念を考慮した洗練箇所が検出可能となる．そのため，ある概念と兄弟関係にある概念同士で分類観点を揃え

る洗練提案が可能となる．さらに，オントロジー構築者に提示する概念階層が包括的なものとなり，概念の分類観点，つまり，オントロジー構築者は，概念がどのように特殊化されたのかを考慮しながら洗練提案の適用可否判断を行うことが可能となる．これらにより，上位下位方向のバランスに加え，兄弟概念間でも特殊化の一貫性を保つような洗練提案を効率よく行うことが可能となる．

これら 3，4 章の洗練支援手法を統合することによって，上位下位関係の特殊化のバランス及び兄弟概念間の特殊化のバランスの両方に一貫性のあるオントロジーとするための洗練提案が可能となる．

さらに，5 章では，概念間関係R_nの定義が不十分な構築初期段階のオントロジーに対して，新規の概念間関係の追加提案を行うような拡張支援手法について述べる．拡張支援においても，あるR_nによって関係が結ばれた「基本概念階層」と「被参照概念階層」を比較することで，概念間関係を結ぶべき概念の組み合わせを推定する．さらに，Web 上に公開されているデータリソースを用いて，その組み合わせを順位付けし，提示することで，オントロジー構築者に対して，概念間関係の拡張作業の支援を行うシステムを開発する．

第3章　上位下位の関係に着目したオントロジー洗練支援システム

本章では，概念階層を構成する基本的な関係である一段の上位下位関係に着目して概念階層を比較する手法について述べる．この手法を実装した洗練支援システムを利用することで，オントロジー構築者に対し，対象オントロジーの分類観点の同一性を向上させるような洗練提案が提示される．

3.1　上位下位の関係に着目した概念階層比較手法

本章で述べる洗練支援手法では，上下一段毎に概念間関係階層を区切り，参照・被参照の関係にある 2 つの概念階層，「基本概念階層」と「被参照概念階層」を比較し，相似性の判定を行い，比較対象となっている概念階層が相似となるような提案を行う．

3.1.1　上位下位関係に着目した洗練箇所検出

概念階層間の相似性の検出は，2 章で述べたように概念間関係により概念の性質が定義される「基本概念」の階層と，クラス制約として参照されている「被参照概念」の階層の各要素数とを比較して行う．通常，オントロジーの is-a 階層には，複数の分類観点が混在している．そのため，分類観点の同一性を保つような洗練支援を行うためには，分類観点毎に概念階層を抜き出す必要がある．そこで本研究では，概念間関係階層を利用する．なぜなら，各概念間関係階層中では，2.1.3.節の条件 2 により，特殊化の際の分類観点を表す概念間関係のラベルP_kが共通となるため，分類観点が共通しているとみなせるからである．この概念間関係階層中の概念間関係が定義されている基本概念の階層と，参照されている被参照概念の階層を抜き出すことによって，分類観点毎に相似性の判定が可能となる．

本研究では，すべての概念間関係のラベル毎に概念間関係階層を抽出し，概念階層を比較する．その後，オントロジー洗練者がどの概念間関係のラベルを洗練するかを判断する．一般的に，概念間関係のラベルは，オントロジー構築者がどのような分類観点に着目して各概念を定義したのかを示している．ただし，それらの概念間関係の中でどの分類観点が重要とされるかは，オントロジー構築者によって異なる．そのため，全ての概念間関係から，実際に洗練すべきか否かを機械的に判別することは困難である．そこで本研究では，洗練提案時にオントロジー洗練者に，どの分類観点を重要とするかを判断させる．

本章で述べる概念階層比較手法では，2.3.3.節で述べた洗練手法の大まかな流れの，「3．概念間関係階層を比較しやすい範囲に分割」の範囲を上位下位一段毎の概念間関係としている．概念階層比較の前に概念間関係階層の分割を行うので，以下のような流れとなる．

1　オントロジーを入力
2　概念間関係のラベルP_k毎に概念間関係階層を抽出
3　概念間関係階層を比較しやすい範囲に分割
3.1　概念間関係階層の最下位の概念間関係のそれぞれについて，最上位の概念間関係に至る概念の is-a 関係の経路を取り出す．
3.2　各経路中の概念間関係を上下一段ごとに区切り，比較対象箇所とする．
4　比較対象箇所に含まれる，概念間関係が定義された基本概念の階層と，クラス制約として参照されている被参照概念の階層を比較
5　階層同士が非相似であれば洗練箇所として検出
6　階層を互いに相似にするような概念の追加提案を出力

3.1.2　上下一段毎に区切った概念間関係階層を中心とした概念階層比較

ラベルP_kを持つ概念間関係階層$\mathbf{R_{P_k}}$には，複数の最下位の概念間関係がある．そのためラベルP_kを持つ最上位の概念間関係$R_{\mathrm{P_k1}}$から最下位の概念間関係の集合$\mathbf{R_{P_kN}}$の各要素へ至る is-a 関係の経路は，複数存在する．最下位の概念間関係の集合$\mathbf{R_{P_kN}}$は，$\mathbf{R_{P_k}}$の集合の要素R_nの中で，下位の関係を持たない，すなわち$\mathrm{Lower}(R_n) = \emptyset$を満たす概念間関係の部分集合であり，

ラベルP_kを持つ最下位の概念間関係の集合$\mathbf{R_{P_kN}}$：$\{R_{P_kN_1}, R_{P_kN_2}, \ldots, R_{P_kN_i}, \ldots, R_{P_kN_I}\}$と表現できる．これら$\mathbf{R_{P_kN}}$の各要素となる最下位の概念間関係から，上位となる$\mathrm{Upper}(R_n)$を最上位の概念間関係まで一段毎に再帰的に辿ることによって，最上位の概念間関係から最下位の概念間関係$R_{P_kN_i}$の各要素へ至る is-a 関係のパスが得られる．これらのパス全てに対して概念階層を比較することで，ある$\mathbf{R_{P_k}}$に属する概念間関係階層を網羅的に比較できる．

以降の議論は，ある概念間関係のラベルP_kを持った概念間関係階層$\mathbf{R_{P_k}}$の中での議論である．煩雑となることを防ぐため表記上P_kを省略し，最上位の概念間関係をR_1，その中間の概念間関係をR_nとする．相似性の判定は，この概念間関係階層の最下位から最上位までのパスを上下一段毎に区切った範囲で行う．上位下位一段毎に概念間関係階層を区切った場合，概念間関係階層は，各パスの一段毎の概念間関係階層と最上位と最下位に分けられるので，

(a)　R_nと$\mathrm{Upper}(R_n)$，但し，$\mathrm{n} \neq 1$
(b)　最上位の概念間関係R_1

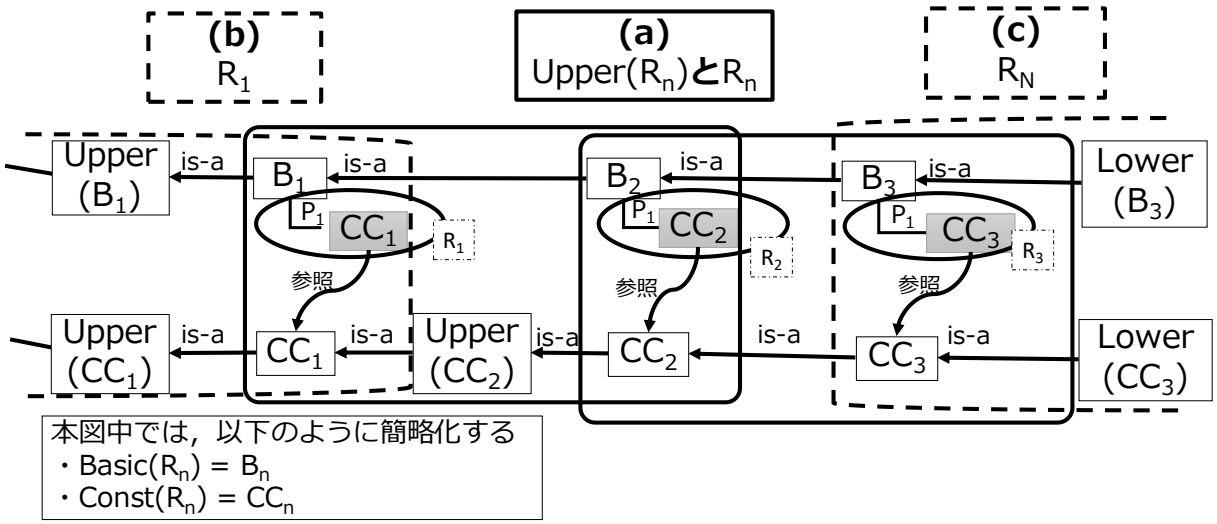

図 3.1.　概念間関係階層の区切りと比較対象範囲

(c)　　最下位の概念間関係R_{N_i}

の 3 種類に分けられる（図 3.1.）．通常，最下位概念の集合$\mathbf{R_N}$には，1 つ以上の要素が含まれる．以降の議論は，その要素R_{N_i}についてのものである．煩雑となることを防ぐために，この最下位の各概念間関係をR_Nと記述し，i についてループ処理を行うものとする．図中では，簡単のため$\mathrm{Basic}(R_n) = B_n$，$\mathrm{Const}(R_n) = CC_n$と表現している．それぞれの範囲において，概念間関係R_nを持つ基本概念階層とクラス制約として参照されている被参照概念階層が，2.3.2.節で述べた相似となる条件，式 2.11.および 2.12.を満たすかによって，相似であるかどうかを判定する．

(a)，(b)，(c)のそれぞれの場合で相似であり，式.2.11.と 2.12.の両方を満たすとき，図 3.2.にあるように，「概念間関係階層」はR_nと$\mathrm{Upper}(R_n)$，「基本概念階層」は$\mathrm{Basic}(R_n)$と$\mathrm{Basic}(\mathrm{Upper}(R_n))$，「被参照概念階層」は$\mathrm{Const}(R_n)$と$\mathrm{Const}(\mathrm{Upper}(R_n))$から構成され，基本概念と概念間関係，被参照概念と概念間関係がそれぞれ一対一に対応する．この時，それぞれの階層中の概念及び概念間関係の数が一致する．この概念及び概念間関係の数がすべての比較対象範囲で一致するとき，「基本概念階層」，「被参照概念階層」，「概念間関係階層」の 3 階層は相似である．

それぞれの比較対象箇所が非相似であるときは，

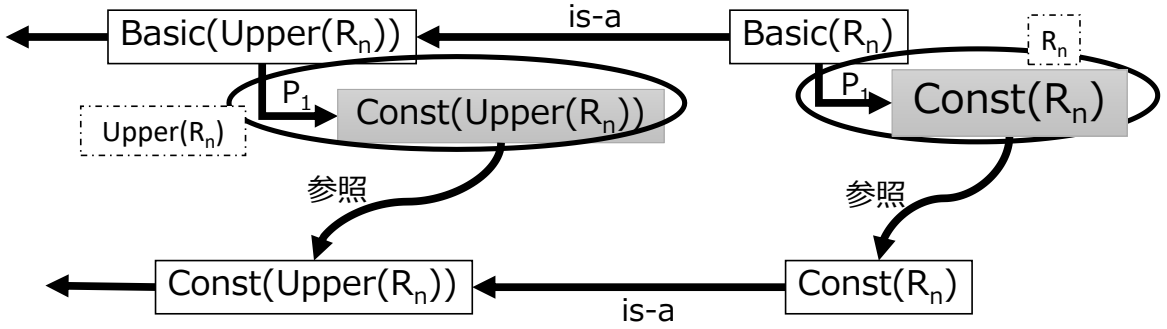

図 3.2.　相似な場合の概念と概念間関係の対応

- 「基本概念階層」に「概念間関係階層」中の概念間関係によって性質が定義されていない概念が存在する
- 「被参照概念階層」に「概念間関係階層」中の概念間関係からクラス制約として参照されていない概念が存在する

のいずれかの場合である．そのため，(a)，(b)，(c)のそれぞれにおいて相似性は，

(a) 基本概念階層中の$\mathrm{Basic}(R_n)$と$\mathrm{Basic}\big(\mathrm{Upper}(R_n)\big)$，または被参照概念階層中の$\mathrm{Const}(R_n)$と$\mathrm{Const}\big(\mathrm{Upper}(R_n)\big)$の間に中間概念が有るどうか(図3.3.)

(b) 基本概念階層または被参照概念階層中の$\mathrm{Basic}(R_1)$または$\mathrm{Const}(R_1)$に上位概念$\mathrm{Upper}(\mathrm{Basic}(R_1))$，$\mathrm{Upper}(\mathrm{Const}(R_1))$が少なくとも一つ有るかどうか（図 3.4.）

(c) 基本概念階層または被参照概念階層中の$\mathrm{Basic}(R_N)$または$\mathrm{Const}(R_N)$に下位概念$\mathrm{Lower}(\mathrm{Basic}(R_N))$，$\mathrm{Lower}(\mathrm{Const}(R_N))$が少なくとも一つ有るかどうか（図 3.5.）

で判定される．

このときに比較される概念を，

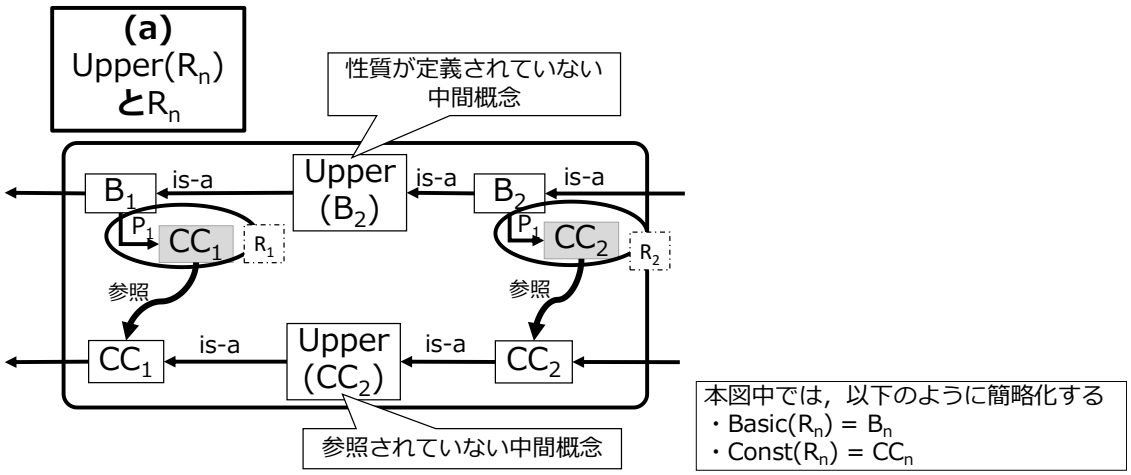

図 3.3. (a)の場合に非相似である概念階層

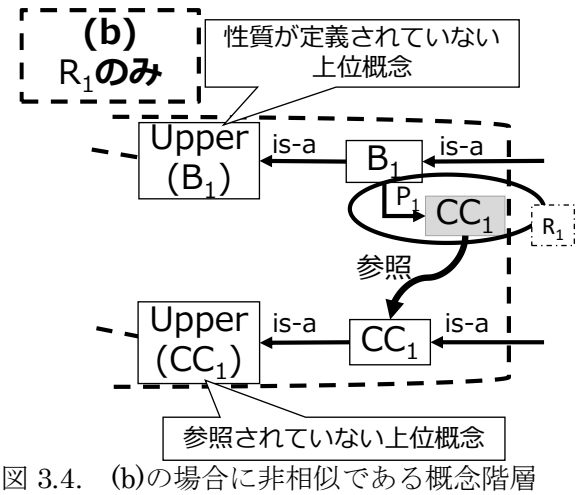

図 3.4. (b)の場合に非相似である概念階層

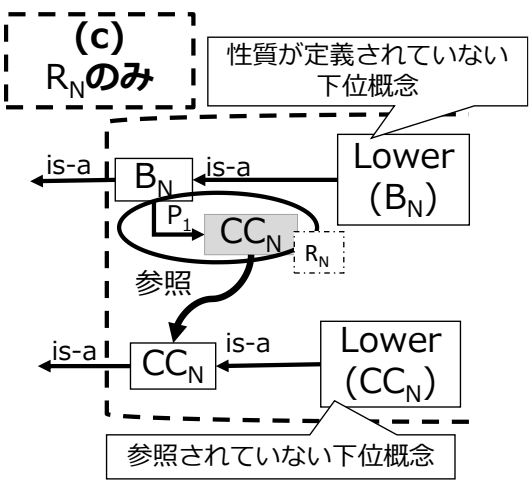

図 3.5. (c)の場合に非相似である概念階層

表 3.1.　相似形のパターン

		基本概念階層	
		$\mathrm{n}(\mathbf{B}_{nr_n}) > 0$	$\mathrm{n}(\mathbf{B}_{nr_n}) = 0$
被参照概念階層	$\mathrm{n}(\mathbf{C}_{nr_n}) > 0$	(i)非相似	(iii)非相似
	$\mathrm{n}(\mathbf{C}_{nr_n}) = 0$	(ii)非相似	(iv)相似

* $\mathrm{n}(\mathbf{B}_{nr_n})$は，基本概念階層中の中間，上位，下位概念の数
$\mathrm{n}(\mathbf{C}_{nr_n})$は，被参照概念階層中の中間，上位，下位概念の数

- $\mathbf{B}_{nr_n}$：比較範囲内の基本概念階層中の，概念間関係によって性質が定義されていない中間・上位・下位概念の集合
- $\mathbf{C}_{nr_n}$：比較範囲内の被参照概念階層中の，参照されていない中間・上位・下位概念の集合

として，n()を集合の要素数とすると$\mathbf{B}_{nr_n}$，$\mathbf{C}_{nr_n}$の概念数はそれぞれ$\mathrm{n}(\mathbf{B}_{nr_n})$，$\mathrm{n}(\mathbf{C}_{nr_n})$と表現することができる．$\mathbf{B}_{nr_n}$の添え字$\boldsymbol{nr}$は，no relation の略，$\mathbf{C}_{nr_n}$の添え字$\boldsymbol{nr}$は，not referred の略であり，それぞれの$\boldsymbol{nr}$の後ろの$\boldsymbol{n}$は，概念間関係R_nのものである．これらの表現を利用すると(a), (b), (c)のいずれの場合であっても，比較対象となる箇所毎に，

$$\mathrm{n}(\mathbf{B}_{nr_n}) = 0 \tag{3.1.}$$

$$\mathrm{n}(\mathbf{C}_{nr_n}) = 0 \tag{3.2.}$$

を満たすとき，相似であると判定できる．(a)の最下位の箇所は，$\mathrm{Basic}(R_N)$と$\mathrm{Basic}(\mathrm{Upper}(R_N))$または$\mathrm{Const}(R_N)$と$\mathrm{Const}(\mathrm{Upper}(R_N))$の中間概念なので，$\mathbf{B}_{nr_N}$，$\mathbf{C}_{nr_N}$と表記される．そのため，(c) $\mathrm{Basic}(R_N)$または$\mathrm{Const}(R_N)$への下位概念の有無を判定する場合は，それぞれの下位概念の集合を$\mathbf{B}_{nr_{N+1}}$，$\mathbf{C}_{nr_{N+1}}$とする．式 3.1.は，式 2.11.と等価であり，式 3.2.は式 2.12.と等価である．ここで，概念間関係R_nは，基本概念および被参照概念なしには定義できないため，概念間関係階層中の概念間関係の数は，基本概念の数および被参照概念の数と同数以下となることに注意する．この条件式を満たさない場合，すなわち$\mathrm{n}(\mathbf{B}_{nr_n})$または$\mathrm{n}(\mathbf{B}_{nr_n})$が，0 より大きい場合は，非相似であると判定される．

このように，(a)，(b)，(c)のそれぞれの場合において，相似であるかどうかは，各概念階層中の概念数で判定できる．そのため，表 3.1.に示すように相似・非相似のパターンを(i)～(iv)に分類することが可能である．

3.1.3　上下一段毎の概念間関係階層毎の相似判定例

本節では，前節で説明した相似性判定について，図 3.1.の模式的な is-a 階層を用いて具体例を示す．図 3.1.の is-a 階層では，概念間関係が上位からR_1，R_2，R_3と定義されており，それぞれ，

$$R_1 = \mathrm{Relation}(B_1, P_1, CC_1)$$

$$R_2 = \mathrm{Relation}(B_2, P_1, CC_2)$$
$$R_3 = \mathrm{Relation}(B_3, P_1, CC_3)$$
$$\mathrm{Isa}(R_1, R_2)$$
$$\mathrm{Isa}(R_2, R_3)$$

の関係が定義されている．ここでは，式表記が長くなるのを防ぐため図 3.1.中の簡略化した表記に従う．

まず, (a)の$\mathrm{Upper}(R_n)$とR_nからなる一段の概念間関係階層の場合を考える．この時, $\mathbf{B}_{nr_n}$は，$\mathrm{Basic}(R_n)$と$\mathrm{Basic}\big(\mathrm{Upper}(R_n)\big)$の中間概念の集合，$\mathbf{C}_{nr_n}$は，$\mathrm{Const}(R_n)$と$\mathrm{Const}\big(\mathrm{Upper}(R_n)\big)$の中間概念の集合である．(a)の例として，$R_1$と$R_2$からなる概念間関係階層を考える．この時，基本概念階層に中間概念が無く，被参照概念階層に中間概念が存在するため,

$$\mathrm{n}\big(\mathbf{B}_{nr_2}\big) = 0, \quad \mathrm{n}\big(\mathbf{C}_{nr_2}\big) = 1$$

であり，式 3.2.を満たさないため非相似であると判定される．

同じ箇所を式 2.11.及び 2.12.を用いて確認すると,

$$\mathrm{Basic}\big(\mathrm{Upper}(R_2)\big) = B_1, \ \mathrm{Basic}(R_2) = B_2$$

であり,

$$\mathrm{Isa}(B_1, \ B_2)$$

が成り立つため，式 2.11.を満たすが,

$$\mathrm{Const}\big(\mathrm{Upper}(R_2)\big) = CC_1, \ \mathrm{Const}(R_2) = CC_2$$

であり,

$$\mathrm{Isa}(CC_1, \ CC_2)$$

が成り立たないため，式 2.12.を満たさない．ゆえに非相似であると判定される．これは中間概念数での判定結果と一致する．

次に，R_2とR_3からなる概念間関係階層を考える．この時，基本概念階層，被参照概念階層の両方に中間概念が無いため,

$$\mathrm{n}\big(\mathbf{B}_{nr_3}\big) = 0, \quad \mathrm{n}\big(\mathbf{C}_{nr_3}\big) = 0$$

であり，式 3.1.及び 3.2.を満たすため相似であると判定される．

同じ箇所を式 2.11.及び 2.12.を用いて確認すると,

$$\mathrm{Basic}\big(\mathrm{Upper}(R_2)\big) = B_1, \ \mathrm{Basic}(R_2) = B_2$$

であり,

$$\mathrm{Isa}(B_1, \ B_2)$$

が成り立ち，式 2.11.を満たす,

$$\mathrm{Const}\big(\mathrm{Upper}(R_2)\big) = CC_1, \ \mathrm{Const}(R_2) = CC_2$$

であり,

$$\mathrm{Isa}(CC_1, \ CC_2)$$

も成り立ち，式 2.12.を満たし，相似であると判定される．これは中間概念数での判定結果と一致する．

次に(b)の最上位の概念間関係のみ場合を考える．この時，$\mathbf{B}_{nr_n}$は，$\mathrm{Basic}(R_n)$の上位概念の集合，$\mathbf{C}_{nr_n}$は，$\mathrm{Const}(R_n)$の上位概念の集合である．(b)の例として，図 3.1.のR_1に着目する．この時，

$$\mathrm{Upper}(B_1) \neq \emptyset$$
$$\mathrm{Upper}(CC_1) \neq \emptyset$$

であり，両方に上位概念が存在するため，式 3.1.及び 3.2.は，

$$\mathrm{n}(\mathbf{B}_{nr_1}) > 0, \mathrm{n}(\mathbf{C}_{nr_1}) > 0$$

であるため満たさない．ゆえに非相似であると判定される．

最後に(c)の最下位の概念間関係のみの場合を考える．この時，$\mathbf{B}_{nr_n}$は，$\mathrm{Basic}(R_n)$の下位概念の集合，$\mathbf{C}_{nr_n}$は，$\mathrm{Const}(R_n)$の下位概念の集合である．(c)の例として，図 3.1.の最下位の概念間関係R_3に着目する．この時，

$$\mathrm{Lower}(B_3) \neq \emptyset$$
$$\mathrm{Lower}(CC_3) \neq \emptyset$$

であり，両方に下位概念が存在するため，

$$\mathrm{n}(\mathbf{B}_{nr_4}) > 0, \mathrm{n}(\mathbf{C}_{nr_4}) > 0$$

となり，式 3.1.及び 3.2.を満たさないため非相似であると判定される．

3.1.4　洗練箇所に対する洗練提案の組み合わせ

各概念階層中の中間・上位・下位概念の数でパターン分けした洗練箇所に対して提示する洗練提案について述べる．ここでは，検出された洗練箇所に対する概念及び概念間関係の追加のみを考え，概念及び概念間関係の削除は考えないこととする．なぜなら洗練対象のオントロジーから概念を削除した場合，そのオントロジーを用いたシステムが扱うことのできる概念のカバー範囲が狭くなるからである．例えば，医療に利用されるオントロジーから概念を削除した場合，代替となる概念を追加しなければ，削除された概念についての検索や推論が不可能となる．このような不具合の発生が懸念されるため，本研究では，概念及び概念間関係の追加提案のみを考える．

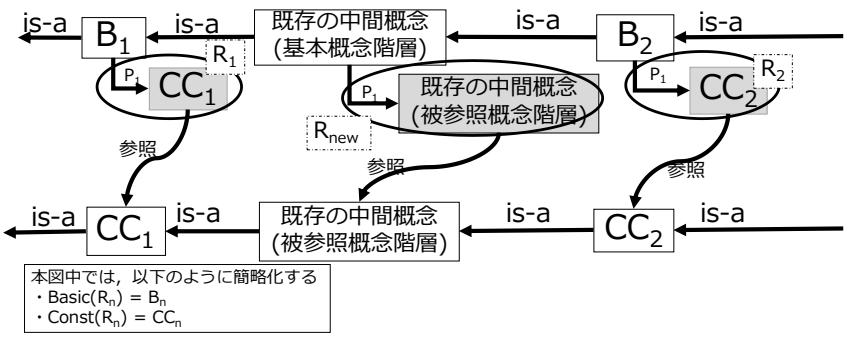

図 3.6.　概念間関係追加

3.1.2.節で述べたように，本手法では，概念間関係階層を上位下位一段に区切り，概念階層同士を比較するため，概念階層が非相似であるとして検出される洗練箇所には，表 3.1.に示すように基本概念階層か被参照概念階層の少なくとも一方に$\mathbf{B}_{nr_n}$，$\mathbf{C}_{nr_n}$が存在する．このような概念及び概念間関係の階層を概念及び概念間関係の追加によって相似にするには，他よりも概念又は関係数の少ない階層に対して概念及び概念間関係を追加すればよい．また，すべての概念間関係は，必ず基本概念および被参照概念を持つ．そのため，非相似であるとして検出される箇所に概念間関係が単独で存在することはなく，常に概念間関係よりも基本概念または被参照概念の方が多くなる．ゆえに，基本概念や被参照概念のみの追加提案は存在しない．よって，提案される洗練方法は次の 3 種類である．

方法1.　概念間関係追加

方法2.　基本概念の追加＋概念間関係追加

方法3.　被参照概念の追加＋概念間関係追加

本研究では，基本概念階層と被参照概念階層の両方へ同時に概念を追加する方法は提案しない．例えば，図 3.3.の場合であれば，「$\mathrm{Upper}(B_2)$とB_2の中間概念B_{middle}を追加し，$\mathrm{Upper}(CC_2)$とCC_2の中間概念CC_{middle}を追加し，B_{middle}に新たな概念間関係を追加してクラス制約としてCC_{middle}を参照する．」といった提案は許可しない．なぜなら，両方の概念階層への概念追加を許す場合，既存の概念の中間，上位，下位への概念追加は，既存の概念の存在とは，無関係に行うことができるからである．そのため，全ての概念間に概念追加が提案されることになる．

方法 1 では，基本概念中の既存の概念に被参照概念階層中の概念を参照した概念間関係，

$$\mathrm{Relation}\left((\mathbf{B}_{nr_n}\text{の要素のいずれか}), P_k, (\mathbf{C}_{nr_n}\text{の要素のいずれか})\right)$$

を追加することが提案される．そのため，方法 1 は，$\mathrm{n}(\mathbf{B}_{nr_n}) > 0 \wedge \mathrm{n}(\mathbf{C}_{nr_n}) > 0$の時，すなわち，基本概念階層と被参照概念階層の両方に，$\mathbf{B}_{nr_n}$または$\mathbf{C}_{nr_n}$が一つ以上存在する場合に提案できる．

図 3.6.に方法 1 の概念間関係の追加の例を示す．これは(a)の場合の例であり，基本概念階層と被参照概念階層が概念間関係階層に比べて詳細に特殊化された中間概念

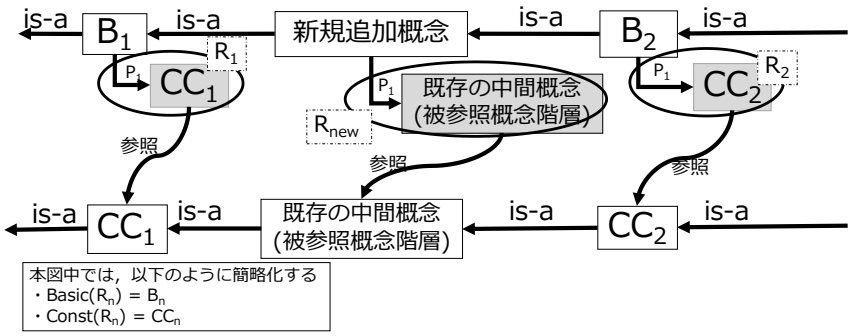

図 3.7.　基本概念＋概念間関係の追加

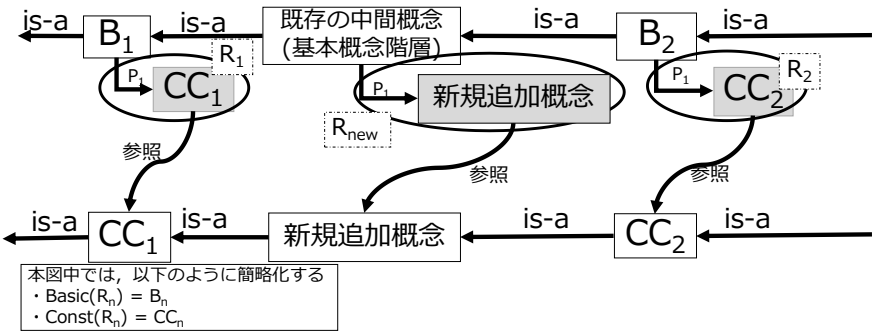

図 3.8.　被参照概念＋概念間関係の追加

がある．このような場合，基本概念階層の既存の中間概念にこの概念間関係階層に共通な概念間関係のラベルP_1を持つ新たな概念間関係R_{new}の追加が提案され，新規に追加した概念間関係のクラス制約として被参照概念階層中の既存の中間概念を参照することが提案される．

方法 2 は，基本概念階層に新たな概念を追加し，その概念に被参照概念階層中の概念を参照した概念間関係，

$\mathrm{Relation}\big((\text{基本概念階層に新規に追加する概念}), P_k, ((\mathbf{C}_{nr_n}\text{の要素のいずれか})\big)$

を追加することが提案される．ゆえに，被参照概念階層中に，クラス制約として参照するための$\mathbf{C}_{nr_n}$の要素が一つ以上必要である．すなわち，この方法 2 は，$\mathrm{n}(\mathbf{C}_{nr_n}) > 0$の時に適用可能である．また，方法 2 は，基本概念階層に新たな概念を追加し，その概念への概念間関係追加であり，既存の概念の有無は関係ないため，基本概念階層の要素数$\mathrm{n}(\mathbf{B}_{nr_n}) = 0$の場合であっても提案可能である．

方法 2 の「基本概念の追加＋概念間関係の追加」の例を示す（図 3.7.）．これは，図 3.1.のR_1とR_2からなる一段の概念間関係階層を中心に比較した部分の抜粋である．これは表 3.1.の(iii)の$\mathrm{n}(\mathbf{B}_{nr_n}) = 0 \wedge \mathrm{n}(\mathbf{C}_{nr_n}) > 0$に相当し，非相似である．この場合，被参照概念階層にのみ（$\mathbf{C}_{nr_n}$の要素）が存在し，他に比べて詳細に特殊化されている．このような場合，基本概念階層と概念間関係階層に，図 3.7.のように基本概念階層に“新規追加概念”及び，新たな概念間関係R_{new}の追加が提案され，新規に追加した概念間

表 3.2.　非相似のパターンと洗練提案の組み合わせ

		(2)洗練提案		
		方法 1 概念間関係の追加	方法 2 基本概念+概念間関係の追加	方法 3 被参照概念+概念間関係の追加
(1) 非相似な is-a 階層のパターン	(i) $\mathrm{n}(\mathbf{B}_{nr_n}) > 0 \wedge \mathrm{n}(\mathbf{C}_{nr_n}) > 0$	○	○	○
	(ii) $\mathrm{n}(\mathbf{B}_{nr_n}) > 0 \wedge \mathrm{n}(\mathbf{C}_{nr_n}) = 0$	×	×	○
	(iii) $\mathrm{n}(\mathbf{B}_{nr_n}) = 0 \wedge \mathrm{n}(\mathbf{C}_{nr_n}) > 0$	×	○	×

関係のクラス制約として既存の中間概念を参照することが提案される．このとき，“新規追加概念”の概念名は，構築者が考える必要がある．

最後に方法 3 は，被参照概念階層に新たな概念を追加し，その概念を基本概念階層中の既存の概念で参照した概念間関係，

$\mathrm{Relation}\left(\left(\left(\mathbf{B}_{nr_n}\text{の要素のいずれか}\right), P_k, \left(\text{被参照概念階層に新規に追加する概念}\right)\right)\right)$

を追加することが提案される．ゆえに，基本概念階層中に，概念間関係を追加する先の$\mathbf{B}_{nr_n}$の要素が一つ以上必要である．すなわち，この方法 3 は，$\mathrm{n}\left(\mathbf{B}_{nr_n}\right) > 0$の時に提案可能である．また，方法 3 は，被参照概念階層に新たな被参照概念を追加し，その概念を概念間関係のクラス制約として参照する提案であり，被参照概念階層中の既存の概念の有無は関係ないため，被参照概念階層の要素数$\mathrm{n}\left(\mathbf{C}_{nr_n}\right) = 0$の場合であっても提案可能である．

方法 3 の「被参照概念の追加＋概念間関係の追加」の例を示す（図 3.8.）．これは表 3.1.の(ii)の $\mathrm{n}\left(\mathbf{B}_{nr_n}\right) > 0 \wedge \mathrm{n}\left(\mathbf{C}_{nr_n}\right) = 0$で非相似形な場合に相当する．この場合，基本概念階層にのみ既存の中間概念が存在し，他に比べて詳細である．このような場合，方法 2 とは逆に被参照概念階層と概念間関係階層に，“新規追加概念”と新たな概念間関係R_{new}の追加が提案され，新規に追加した概念間関係R_{new}のクラス制約として“新規追加概念”の参照が提案される．この“新規追加概念”名は，構築者が考える．

ここまで述べてきた(1)洗練箇所と(2)洗練提案の関係について述べる．洗練提案は，比較対象となっている 3 つの is-a 階層のうちで少ない箇所，すなわち特殊化の粒度が粗い箇所に対して行われる．そのため，表 3.1.の(i)～(iii)の分類に応じて適用可能な洗練方法が決まる．表 3.2.は，表 3.1.中の分類(i)～(iii)と，方法 1～方法 3 の対応を示したものである．また，洗練箇所毎に複数の方法によって洗練提案が提示される．さらに$\mathrm{n}\left(\mathbf{B}_{nr_n}\right)$, $\mathrm{n}\left(\mathbf{C}_{nr_n}\right)$によって概念間関係の追加先及び参照先の組み合わせ数が増えるため，1 つの洗練箇所につき複数の洗練提案が提示される．

3.2 上位下位関係に着目した洗練支援システムの実装

3.1.節で述べた洗練手法を評価するために実験用のシステムを試作した．本システムは，4 つのモジュールから構成されている．洗練箇所推定モジュール，洗練提案推定モジュール，画像描画モジュール，洗練箇所・提案表示インタフェースの 4 つである．各モジュール概要を，システムの全体的な処理の流れに沿って，以下に示す．

処理1　洗練箇所推定モジュールが，入力されたオントロジーから概念間関係階層の一覧を取得し，分類観点毎に洗練箇所を検出する．

処理2　処理 1 の検出結果を入力として，洗練提案推定モジュールが各洗練箇所に対する洗練提案を推定する．

処理3　画像描画モジュールが，処理 1 及び 2 の結果を入力として，それらの内容をオントロジー構築者が把握しやすい表現で図示した画像を生成する．

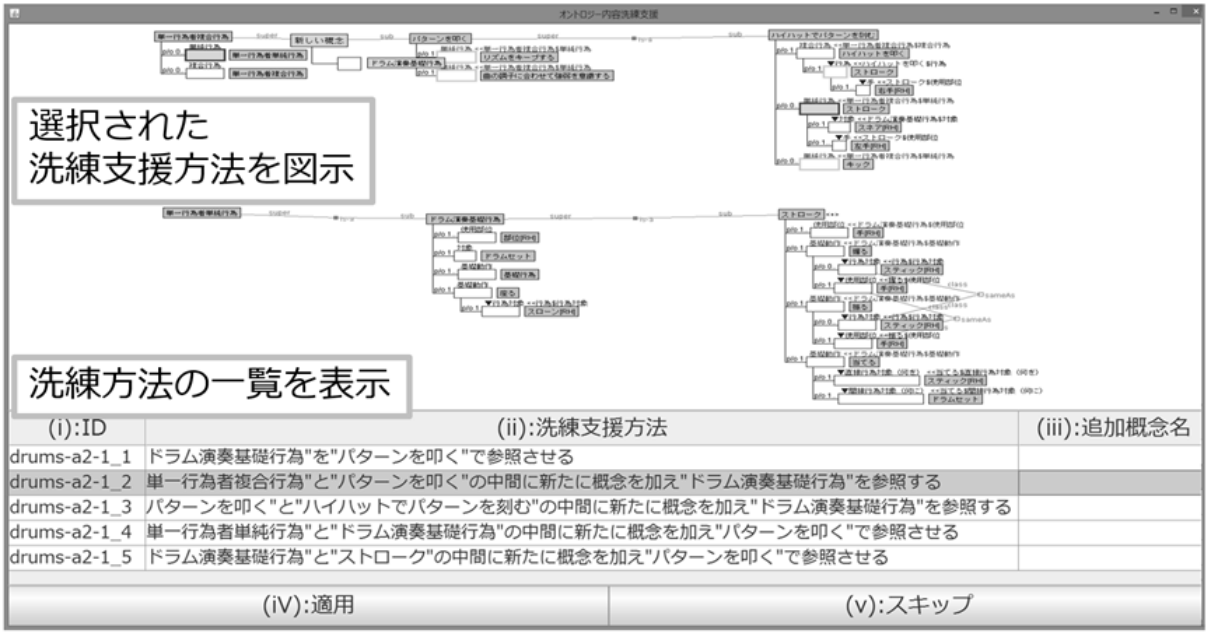

図 3.9.　洗練箇所・提案表示インタフェース

処理4　処理 1 から 3 の処理結果が，洗練箇所・提案表示インタフェース（図 3.9.）上でオントロジー構築者に提示される．構築者は，提示された洗練支援方法を見て，洗練提案を評価する．

図 3.9.に洗練箇所・提案表示インタフェースの表示例を示す. この画面の下部には，その洗練箇所に提示される洗練提案が一覧表示され，この一覧からユーザが選択した洗練提案がこの画面の上部に図示される．

3.3　評価実験

3.3.1　評価実験の目的

開発したオントロジー内容洗練支援システムの評価実験について述べる．本評価実験では, システムがオントロジー構築者に提示する洗練提案を以下の 2 点で評価した．

- 検出された洗練箇所の妥当性
- 洗練箇所に対する洗練提案の妥当性

1 点目は，対象オントロジーからシステムが検出した洗練箇所自体が，修正すべき箇所として適切であるかという点での評価である．本手法で洗練箇所として検出されるのは，特殊化の詳細度の一貫性の崩れている箇所である．これはオントロジー構築におけるクラス分類の推奨基準を基に得られた仮説である．従って，オントロジーの文法規則違反に概念定義上の矛盾が生じている箇所ではないため，修正は必須ではない．そのため，まず洗練箇所自体の検出の妥当性を評価する．

表 3.3.　各オントロジーから検出された洗練箇所数

分類	全概念数	R_nとUpper(R_n)			R_1			R_N			全体合計
		a(i)	a(ii)	a(iii)	b(i)	b(ii)	b(iii)	c(i)	c(ii)	c(iii)	
ドラム演奏	136	7	16	17	37	0	0	75	33	60	245
列車	158	12	5	11	49	0	0	43	13	21	154
交通	252	19	2	13	77	0	0	14	27	36	188
初学者合計	546	38	23	41	163	0	0	132	73	117	587
サッカー	262	12	5	16	99	0	0	20	35	73	260
サステナビリティ	1927	7	4	39	257	0	0	16121	344	4926	21698
乗り物	122	6	4	1	30	0	0	140	24	39	244
熟練者合計	2311	25	13	56	386	0	0	16281	403	5038	22202
全体合計	2857	63	36	97	549	0	0	16413	476	5155	22789

2 点目は，洗練箇所に提示される洗練提案に対する評価である．本手法では，1 箇所の洗練箇所に対して，1 つ以上の洗練提案が提示される．そこで，それらの洗練提案のうち，それぞれの洗練提案の適切さを評価する．

3.3.2　評価実験の概要

実験の被験者は，計 9 人である．システム利用者のオントロジーの構築経験の影響を調査するために，システムの利用に最低限必要な構築経験を持つ学生から，豊富なオントロジー構築経験を持つ研究者まで幅広い被験者を所属研究室の内外から募った．具体的な内訳は，筆者の所属する研究室から特任助教 1 名，大学院生 2 名，学部 4 回生 1 名，研究室外のオントロジー構築経験者が 5 名である．

評価実験に使用したオントロジーは，対象領域や構築者の熟練度の偏りを少なくするために 6 つ利用した．具体的には，熟練者が構築した，サッカーオントロジー[溝口 06]，サステナビリティサイエンスオントロジー[Kumazawa09]，乗り物オントロジー[溝口 06]の 3 つと，本研究室に配属された直後の学部 4 回生の学生が初めて構築した，ドラム演奏オントロジー，列車オントロジー，交通オントロジーの 3 つの計 6 つである．これらのオントロジーから 1 つ当たり 25 箇所ずつの洗練箇所を評価用に抽出した．表 3.3.に各オントロジーから検出された洗練箇所数を示す．

表 3.4.　各洗練箇所の選択数内訳

分類	R_nとUpper(R_n)			R_1			R_N		
	a(i)	a(ii)	a(iii)	b(i)	b(ii)	b(iii)	c(i)	c(ii)	c(iii)
個数	54	24	24	12	0	0	12	12	12

9 名の被験者には，これら 6 つのオントロジーから検出された全ての洗練箇所から各 25 箇所ずつ，計 150 箇所を無作為に抜き出したものを評価させた．各洗練箇所に対して，被験者 3 人からの評価を得るために，一人当たり 50 箇所を割り当てた．被験者への各洗練箇所の提示は，対象となるオントロジーの順番，ドメインの難易度などの影響を抑えるため，提示順をランダムにした．

実験に使用した洗練箇所の内訳を表 3.4.に示す．事前に行った予備実験では，R_nと$\mathrm{Upper}(R_n)$をもとに概念階層を比較する分類 a が他の分類に比べ提案の精度が高かった．そこで分類 a については，より詳細に調査するために他に比べ多くの洗練箇所を評価させることにした．分類 b(ii)、b(iii)は，検出数が 0 となっているが，それぞれの概念階層の最上位のどちらか一方のみが存在しない場合である．これは、理論上は考えられるが，実際のオントロジーでこのような概念階層が構築されることはほとんどない．

これらの洗練箇所に対する評価は，各箇所に複数提示される洗練提案を評価することで行う．被験者には，洗練提案に対して次の 3 段階と評価外の 4 通りの評価を与えるように指示した．

- ◎：最も適用が妥当であると判断した洗練提案
- ○：適用することが妥当であると判断した洗練提案
- ×：適用すべきではないと判断した洗練提案
- 評価外：提示された洗練箇所に定義間違いや矛盾が存在する

最も適用が妥当であると判断される洗練提案は，洗練箇所毎に 1 つ選択することが可能である．◎と○に分けて評価しているのは，優先順位等の分析時に利用可能であると考えたからである．

洗練箇所の妥当性の判断は，各箇所に提示される複数の洗練提案のうち 1 つ以上に◎または○が与えられた場合，その洗練箇所の検出が妥当であるとした．但し，評価外が選択された場合，その洗練箇所に提示される洗練提案は，評価外であるとした．なぜなら，オントロジー自体の誤りや定義上の矛盾は，本手法の洗練対象外であるからである．

表 3.5.　被験者毎の洗練提案の評価の差

		評価者								
		A	B	C	D	E	F	G	H	I
分類	a	27%	11%	31%	52%	23%	87%	25%	11%	16%
	b	15%	5%	4%	5%	0%	92%	3%	16%	18%
	c	38%	8%	19%	17%	7%	58%	20%	15%	20%

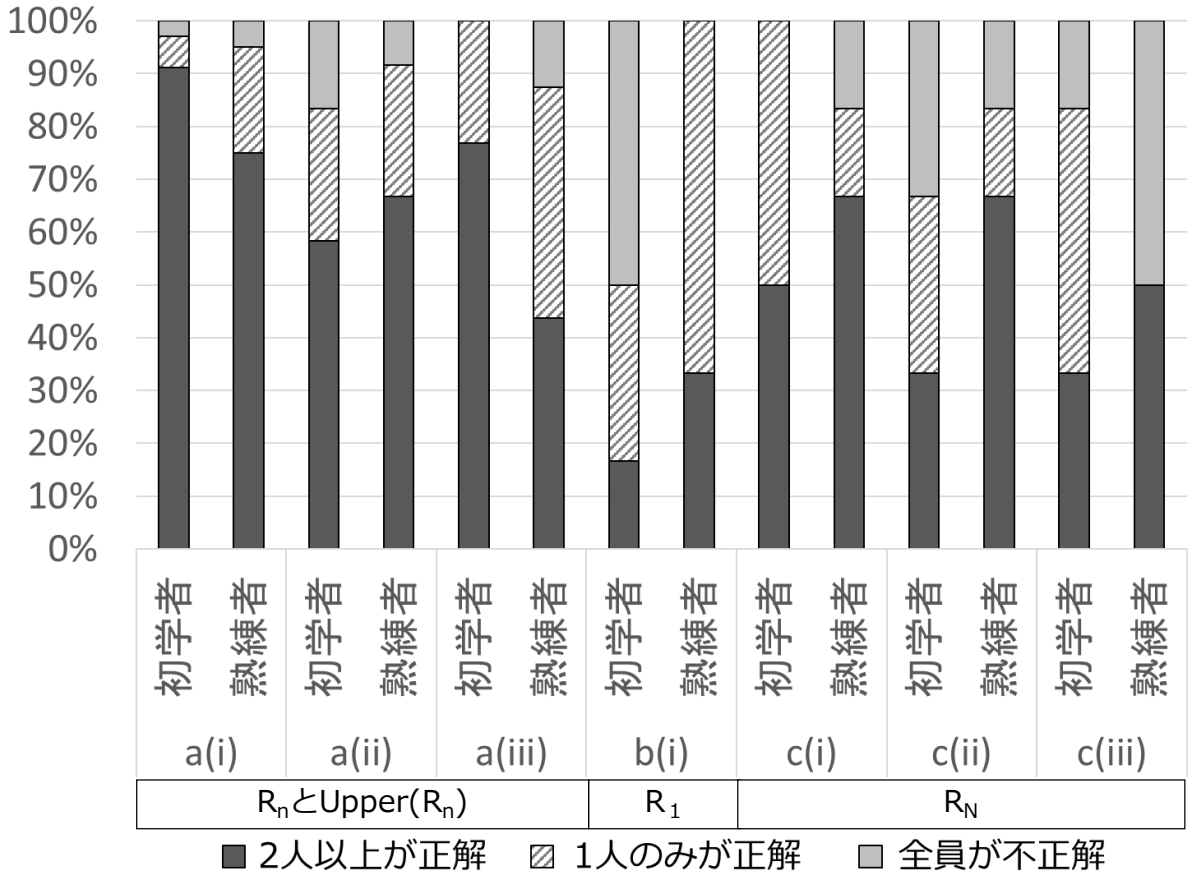

図 3.10.　洗練箇所の妥当性の評価

3.3.3　実験結果及び考察

本実験では，150 の洗練箇所へそれぞれ 3 人から評価を得るように 9 人の被験者を割り振った．被験者間で評価の付与傾向が大きく異なる場合，今回のように複数の被験者による評価方法では，洗練箇所及び提案の評価結果の信頼性が損なわれることになる．そこで被験者毎に与えられた全提案中に妥当であると評価した割合を計算したところ表 3.5.の結果となった．表 3.5.に示すように評価者 F が他に比べ極端に正解とした割合が多かった．そこで付与傾向の大きく異なる評価者 F を除く 8 人での評価をもって実験結果とした．この 8 人から各洗練提案に与えられた評価の一致度を示すκ値[Fleiss 71]を計算したところ，0.75 となり被験者間の特性の違いによる評価のぶれは少ないと考えられる．

3.3.3.1　検出された洗練箇所の妥当性

図 3.10.に，検出された洗練箇所の妥当性に関する評価結果を示す．ここでは，a(i)-1，a(i)-2，a(i)-3 は，a(i)としてまとめられており，分類毎に初学者の構築したオントロジーと熟練者の構築したオントロジーで分けられている．洗練箇所の妥当性の判断は，各箇所に提示される複数の洗練提案のうち 1 つ以上に◎または○が与えられた場

合，正解箇所であるとした．各洗練箇所は，2 人以上の異なる実験参加者によって評価されるため，

- 2 人以上が正解箇所とした
- 1 人のみが正解箇所とした
- 全員が不正解とした

の 3 段階の評価とした．

洗練箇所の検出の妥当性の評価結果を，表 3.1.で示した概念間関係階層の分類毎に示す．この結果，a.一段の概念間関係階層，c.最下位の概念間関係，b.最上位の概念間関係，の順に妥当性が高い．特に分類 a(i)では，約 70%～90%の洗練箇所に対して 2 人以上の評価者が正解箇所とし，高い精度で洗練箇所を検出できた．その理由としては，a の一段の概念間関係階層，つまり，R_nと$\mathrm{Upper}(R_n)$の上下一組の 2 つの概念間関係と他の 2 階層を比較しているため，b，c に比べ比較点が多いために提案の正確性が向上しているからであると考えられる．また，c の最下位の概念間関係を比較の基準とした場合の正解率が，b の最上位の概念間関係の場合よりも高くなっている．この理由としては，最上位の概念間関係である b で検出される洗練箇所に定義されている概念は，“もの”，“実在物”等の非常に抽象度の高い概念であり，新たに概念や概念間関係を追加することが困難だからである．

さらに分類 b，c では，熟練者の構築したオントロジーの方が，2 人以上から正解とされた箇所が多い．この理由としては，熟練者の構築したオントロジーでは，上位概念から下位概念に特殊化する際に定義の詳細度の過度な飛躍が少ないからであると考えられる．初学者は，無計画に概念階層の構築を進めるために，上位概念と下位概念の間で定義内容が飛躍することが多い．そのため被験者は，提案された箇所に適切な中間概念を追加することが困難であり不正解と判断したと考えられる．

図 3.10.の結果から，分類 a(i)では，70%以上，分類 a，c では，検出された洗練箇所の 30%以上に対して少なくとも 1 つ妥当であると評価された洗練提案ができることが分かる．この結果から，概念階層の特殊化の詳細度の一貫性が崩れている箇所を，洗練提案すべき箇所として検出できることが確認できた．

3.3.3.2　　提案した洗練提案の妥当性

次に各洗練箇所に提案される洗練提案に対する評価について述べる．今回の実験では，各洗練提案につき計 3 人からの評価を得た．各洗練提案の評価は 3 人の多数決によって以下のようにした．

- ◎が 2 人以上：最も妥当
- ○が 2 人以上：妥当
- ×が 2 人以上：不適当

但し，評価者が 2 人の場合は，

- ◎◎：最も妥当

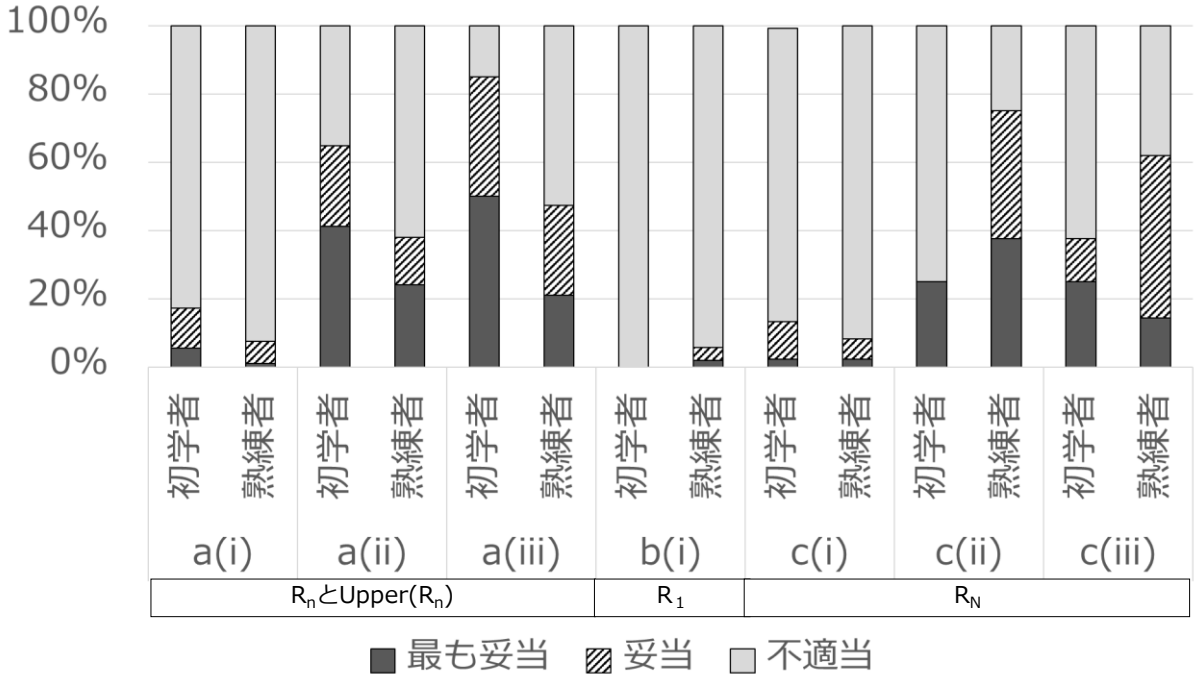

図 3.11.　洗練提案の妥当性の評価

- ◎◯：妥当
- ◎×：妥当
- ◯◯：妥当
- ◯×：不適当
- ××：不適当

とした．結果を図 3.11.に示す．

図 3.11.の洗練提案毎の評価でも，先ほどの洗練箇所の評価と同様に，a.一段の概念間関係階層，c.最下位の概念間関係R_N，b.最上位の概念間関係R_1，の順に妥当と評価された割合が高い．最も提案精度に対して高い評価を得た a(iii)では，初学者のオントロジーに対する洗練提案のうち 85%の提案の適用が妥当であると評価された．また分類 c(ii)，c(iii)は，a(ii)，a(iii)と同程度の妥当性となっている．このことより，a，c に対して優先的に洗練提案を行うことで効率的に品質向上作業を行うことができると考えられる．

初学者と熟練者の構築したオントロジー毎の比較では，それぞれのオントロジーから洗練提案の妥当性の傾向に差があることが分かった．初学者のオントロジーでは分類 a(i)~(iii)の提案の妥当性が高くなっていた．この理由として，初学者は，概念の分類観点や参照関係に対する考察が浅く，上位下位関係を中心に構築を進めていくからであると考えられる．こうした概念階層では，概念間関係の定義が不十分であり，基

表 3.6.　1 箇所当たりの平均洗練提案数

分類	R_nとUpper(R_n)			R_1	R_N		
	a(i)	a(ii)	a(iii)	b(i)	c(i)	c(ii)	c(iii)
個数	15	1.9	1.5	12.8	21.7	1.3	2.4

本概念階層と被参照概念階層のどちらか一方のみが詳細に定義されている傾向がある．そのため既存の概念を利用した概念間関係の追加，中間概念の追加が提案される分類 a(i)~(iii)の妥当性が高くなったと考えられる. 一方, 熟練者のオントロジーでは，分類 c(ii)，c(iii)の提案の妥当性が高かった．熟練者は，上位から下位まで一貫して，本質的な分類観点を用いた特殊化を行っている一方で，末端の概念は抽象度が高いままとなっていることが多い．抽象度の高い概念については，それを特殊化した下位概念の追加の余地が考えられやすい．そのため，末端の概念に下位概念追加を提案する分類 c において，妥当であると判定されやすい洗練提案が多く提示された可能性がある．

洗練提案の中で，a(i)，b(i)，c(i)は洗練提案の妥当性が他に比べて極端に低い．この理由としては, 表 3.6.に示すように他の分類に比べて洗練提案の数が多く, その結果, 他に比べて不適当な提案の割合が増加したからである. これらは表 3.1.に示すように, 基本概念階層，被参照概念階層の両方に中間，上位，または下位概念が存在する時, (i)$\mathrm{n}\left(\mathbf{B}_{nr_n}\right) > 0 \wedge \mathrm{n}\left(\mathbf{C}_{nr_n}\right) > 0$の場合である．このとき，表 3.2.の(ii)，(iii)よりも，比較対象となっている概念階層に含まれる中間概念数が多い．概念・概念間関係の追加先は，着目している概念階層の中間概念の組み合わせ数だけ候補があるため，概念・概念間関係の追加候補先の組み合わせが多くなる．そのため a(i)，b(i)，c(i)は，他に比べて提案数が多く，相対的に妥当性が低くなったと考えられる．

3.3.4　洗練提案例

3.3.4.1　　妥当であると評価された洗練提案例

図 3.12.にシステムによって検出・提案された洗練提案が妥当であると評価された例を示す．この例は，分類(a)のR_nと$\mathrm{Upper}(R_n)$からなる概念間関係階層を介して参照関係にある概念階層を比較したもので，$\mathrm{n}\left(\mathbf{B}_{nr_n}\right) > 0 \wedge \mathrm{n}\left(\mathbf{C}_{nr_n}\right) = 0$の場合，つまり基本概念階層のみが他の階層よりも詳細に定義されているものである．この場合，表 3.2.にあるように，方法 3 の被参照概念及び概念間関係の追加提案として，

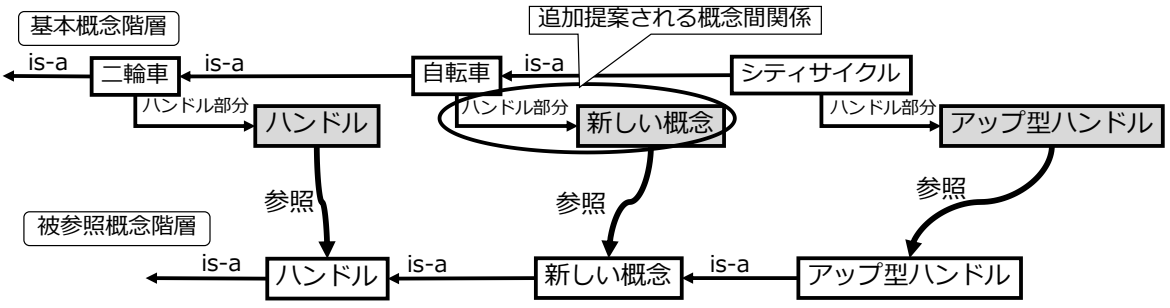

図 3.12.　妥当であると評価された洗練提案例

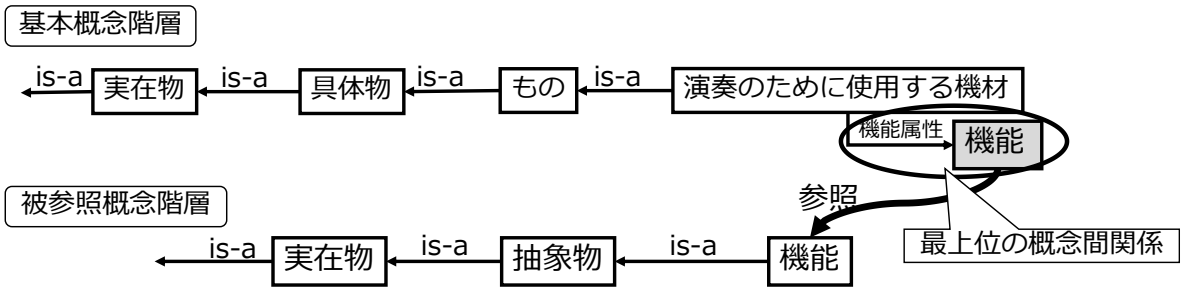

図 3.13.　適用すべきでないと評価された洗練提案例

「“ハンドル”と“アップ型ハンドル”の中間に新しい概念を追加し，“自転車”に追加した“ハンドル部分”というラベルを持つ概念間関係のクラス制約として，その新しい概念を参照する」

といった提案が提示される．追加される新しい概念の内容としては，“シティサイクル”での“ハンドル部分”のラベルを持つ概念間関係のクラス制約として参照されている概念が，“アップ型ハンドル”となっているので，その上位概念“自転車”の“ハンドル部分”として“自転車用ハンドル”等の追加が考えられる．被験者は，この概念及び概念間関係の追加提案を適用することは，妥当であると評価した．

3.3.4.2　　適用すべきでないと評価された例

次に，洗練提案が不適切であると評価された例を図 3.13.に示す．この例は，分類(b)最上位の概念間関係R_1を介して参照関係にある基本概念階層及び被参照概念階層を検出したものであり，その両方が概念間関係階層よりも詳細に特殊化されている．つまり，$\mathrm{n}(\mathbf{B}_{nr_n}) > 0 \wedge \mathrm{n}(\mathbf{C}_{nr_n}) > 0$であるので，表 3.2.から方法 1～3 全ての洗練提案が提示される．例えば，

「“もの”に“機能属性”というラベルの概念間関係を追加し，そのクラス制約として“抽象物”を参照する」

「“具体物”に“機能属性”というラベルの概念間関係を追加し，そのクラス制約として参照するための新規概念を被参照概念階層に追加する」

など，他の同様な箇所に対しても，“実在物”，“具体物”等の抽象度の高い概念階層への概念追加や“実在物”や“抽象物”のクラス制約としての参照が提案される．それらは，オントロジーの中でも抽象度が高い概念であるが故に概念追加や参照関係に対する考察が非常に困難な箇所である．そのため被験者は，この洗練箇所に対する全ての洗練提案が不適切であると評価した．

第4章　分類観点毎の概念間関係階層に着目したオントロジー洗練支援システム

本章では，分類観点毎に概念間関係階層を包括的に比較し，非相似の判定を行う洗練支援手法について述べる．この比較手法により，洗練提案の可否判断時にオントロジー構築者が利用できる情報が増える．さらに，これまで別々の洗練箇所として検出されていた箇所を統合して提示できる．そのため，本章での提案システムの利用によって，オントロジーの洗練作業がより効率化される．

4.1　分類観点毎の包括的な概念階層比較手法

本節では，分類観点毎に概念間関係階層全体と参照関係にある基本概念階層と被参照概念階層を比較し，非相似の判定を行う洗練支援手法について述べる．

4.1.1　分類観点毎の包括的な概念階層比較の必要性

まず，3 章で開発した洗練支援手法の問題点を整理する．3 章で述べた上位下位の概念間の関係に着目した概念階層比較手法では，概念間関係階層を上下一段ごとに区切るため，構築者に対して提示される洗練箇所の範囲が局所的となる．オントロジー中に定義されている概念は，その周辺の概念や概念間関係によって定義されるため，局所的な概念階層だけでは，構築者がどのような観点や目的で概念を定義したのかが不明確である．ゆえに，各階層に含まれる概念及び概念間関係を包括的に考察しなければ，オントロジー構築者以外の人にとっては，その概念階層がどのような分類観点に沿って特殊化されているか，どのような性質に着目して分類したのかを理解することは困難である．構築者自身が洗練作業を行うこともあるが，大規模なオントロジーの場合，全ての概念階層を把握することは困難である．そのため，洗練時には一部の概念階層だけを考察するのではなく，包括的に概念階層を考察することが必要となる．この点に関しては，3 章での評価実験の際に行われた洗練システムに対する自由記述アンケートにおいて，

- オントロジーの全体が見えないと判断しにくい．
- オントロジーは全体の階層構造からそのオントロジーがどのような目的で作られたかが理解できると思う．ある箇所の中間概念が妥当か判断かは，全体物

をみて評価できると考えており，今回の実験のように一部だけ切り取って判断せよといわれても判定が難しいものが多かった.

等のコメントが挙げられたことからも解決すべき問題である.

そこで，本章では概念階層の比較対象として，分類の観点に沿った概念間関係階層全体を基準に，基本概念階層全体と被参照概念階層全体を比較する手法を開発する．こうした比較対象の拡張によって，さらに，洗練提案の際にオントロジー構築者に対して分類観点毎に包括的な概念階層に関する情報を提示できる．さらに，比較対象の概念が，3 章の手法では，上位・中間・下位概念であったのに対し，4 章の比較手法では，それぞれの階層の兄弟概念も比較対象となる．これにより，これまで提案できなかった兄弟概念に対する概念及び概念間関係の追加提案が可能となる．

本章で述べる概念階層比較手法では, 2.3.3.節で述べた洗練手法の大まかな流れのうち，3 の「概念間関係階層の分割」を行わないため,

1　オントロジーを入力
2　概念間関係のラベルP_k毎に概念間関係階層を抽出
3　概念間関係階層毎に概念間関係を持つ「基本概念階層」と参照されている「被参照概念階層」を比較
4　階層同士が非相似であれば洗練箇所として検出
　4.1　洗練箇所に含まれるR_n毎に範囲を限定し，洗練提案を生成
5　洗練箇所毎に，4 で生成した洗練提案を出力

となる.

4.1.2　分類観点毎の概念階層の比較による洗練箇所検出

分類観点毎の概念階層全体に着目した概念階層の比較手法では，共通の分類観点P_kを持つ概念間関係の集合からなる概念間関係階層$\mathbf{R_{P_k}}$を介して，参照・被参照の関係にある 2 つの概念階層「基本概念階層」「被参照概念階層」を比較する．この場合の相似性の判定は，$\mathbf{R_{P_k}}$の全ての概念間関係R_nに対して，2 章で述べた条件，式 2.11 および 2.12.を満たすかどうかで行われる．つまり，概念間関係階層中の全ての概念間関係と,「基本概念階層」,「被参照概念階層」中の全ての概念が完全に一対一対応しているとき相似である．この場合，比較対象となっている 3 つの概念階層は完全に相似となる．ゆえに概念間関係階層$\mathbf{R_{P_k}}$を介して参照関係にある概念階層が相似であることの条件は，概念及び概念間関係の集合を,

- $\mathbf{R_{P_k}}$：P_kを分類観点として持つ概念間関係R_nの集合
- $\mathbf{B_{P_k}}$：$\mathbf{R_{P_k}}$を持つ基本概念の集合$(\mathrm{Basic}(R_n)|R_n \in \mathbf{R_{P_k}})$
- $\mathbf{CC_{P_k}}$：$\mathbf{R_{P_k}}$から参照されている概念の集合$(\mathrm{Const}(R_n)|R_n \in \mathbf{R_{P_k}})$

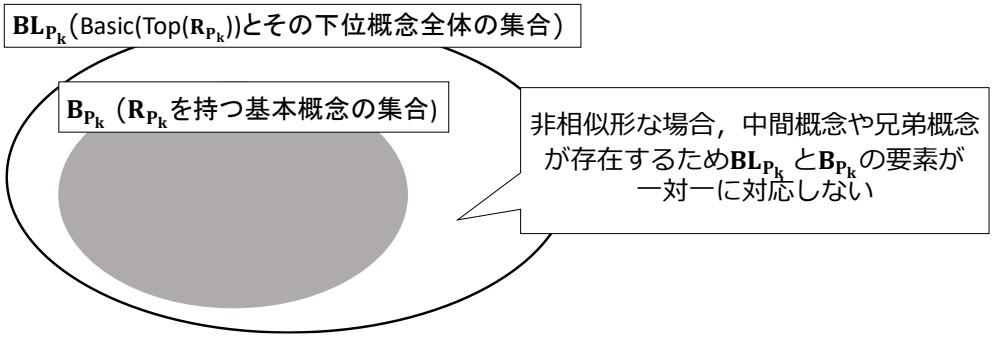

図 4.1.　比較対象となっている概念の集合間の関係

と表現すると，それぞれの集合の要素の数が，

$$\mathrm{n}\left(\mathbf{B_{P_k}}\right) = \mathrm{n}\left(\mathbf{CC_{P_k}}\right) = \mathrm{n}\left(\mathbf{R_{P_k}}\right) \tag{4.1.}$$

を満たし，それ以外の基本概念，被参照概念が存在しない場合である．ここで，概念間関係R_nは，基本概念および被参照概念なしには定義できないため，概念間関係階層の要素の数$\mathrm{n}\left(\mathbf{R_{P_k}}\right)$は，$\mathbf{R_{P_k}}$を持つ基本概念の数$\mathrm{n}\left(\mathbf{B_{P_k}}\right)$，および$\mathbf{R_{P_k}}$から参照されている概念の数$\mathrm{n}\left(\mathbf{CC_{P_k}}\right)$と同数以下となることに注意する．

一方，非相似な場合，

- Top($\mathbf{R_{P_k}}$)：$\mathbf{R_{P_k}}$の最上位の概念間関係
- $\mathbf{BL_{P_k}}$：Basic(Top($\mathbf{R_{P_k}}$))とその下位概念全体の集合
- $\mathbf{CCL_{P_k}}$：Const(Top($\mathbf{R_{P_k}}$))とその下位概念全体の集合

とすると，比較対象となっている基本概念階層中の概念の集合間の関係は，図 4.1.に示すように，

$$\mathbf{BL_{P_k}} \supset \mathbf{B_{P_k}} \tag{4.2.}$$

という包含関係になる．$\mathbf{BL_{P_k}}$には，概念間関係が定義されていない中間概念又は兄弟概念が存在する．中間概念は，3 章と同じく基本概念階層であれば，概念間関係を持つ概念を上位または下位に持つ概念であり，被参照概念階層であれば，概念間関係から参照されている概念を上位または下位に持つ概念である．兄弟概念は，共通の上位概念を持つ概念を指すが, 4 章では共通の上位概念を is-a 関係の再帰を含めて考える．

被参照概念階層でも同様に，非相似となる場合，概念の集合関係は，式 4.2.と同様に，

$$\mathbf{CCL_{P_k}} \supset \mathbf{CC_{P_k}} \tag{4.3.}$$

となる．

このように，非相似となる場合，$\mathbf{B_{P_k}}$は$\mathbf{BL_{P_k}}$の部分集合となり，$\mathbf{CC_{P_k}}$は$\mathbf{CCL_{P_k}}$の部分集合となる. ゆえに, 次の 2 式のいずれかを満たすとき, 非相似であると判定される．

$$\mathrm{n}\left(\mathbf{BL_{P_k}}\right) > \mathrm{n}\left(\mathbf{B_{P_k}}\right) \tag{4.4.}$$

$$\mathrm{n}\left(\mathbf{CCL_{P_k}}\right) > \mathrm{n}\left(\mathbf{CC_{P_k}}\right) \tag{4.5.}$$

なぜなら，図 4.1.のベン図で示す通り，$\mathbf{BL_{P_k}}$に含まれる要素の数が$\mathbf{B_{P_k}}$より多い，つまり，概念間関係階層$\mathbf{R_{P_k}}$中の概念間関係を持たない概念が存在するからである．こうした場合に非相似形であると判定される．

基本概念階層および被参照概念階層において概念間関係階層を介して参照関係にある概念階層の概念数は，これらがお互いに相似の関係にある場合，$\mathbf{BL_{Pk}}$および$\mathbf{CCL_{Pk}}$において，概念間関係を持たないまたは参照されていない中間概念及び兄弟概念の数は，それぞれ，

$$\mathrm{n}\left(\mathbf{BL_{P_k}}\right) - \mathrm{n}\left(\mathbf{B_{P_k}}\right)$$
$$\mathrm{n}\left(\mathbf{CCL_{P_k}}\right) - \mathrm{n}\left(\mathbf{CC_{P_k}}\right)$$

となる．

4 章の手法では，概念間関係階層を上下一段毎に分割せず，概念間関係階層全体を比較範囲として相似性を判定するため，

- $\mathbf{BL_{P_k}}$：基本概念階層の最上位概念 Basic(Top($\mathbf{R_{P_k}}$))とその下位概念全体
- $\mathbf{CCL_{P_k}}$：被参照概念階層の最上位概念 Const(Top($\mathbf{R_{P_k}}$))とその下位概念全体

に対する洗練提案が同時に提示される．その結果，オントロジー構築者は，分類観点毎に関連した周囲の概念階層を同時に考慮しながら提案の可否判断ができるようになる．

例えば，上位下位関係に着目した概念階層比較手法では，図 4.2.の概念階層の場合，「B_2[18]」に，「CC_2」をクラス制約として参照した概念間関係を追加する」といった提

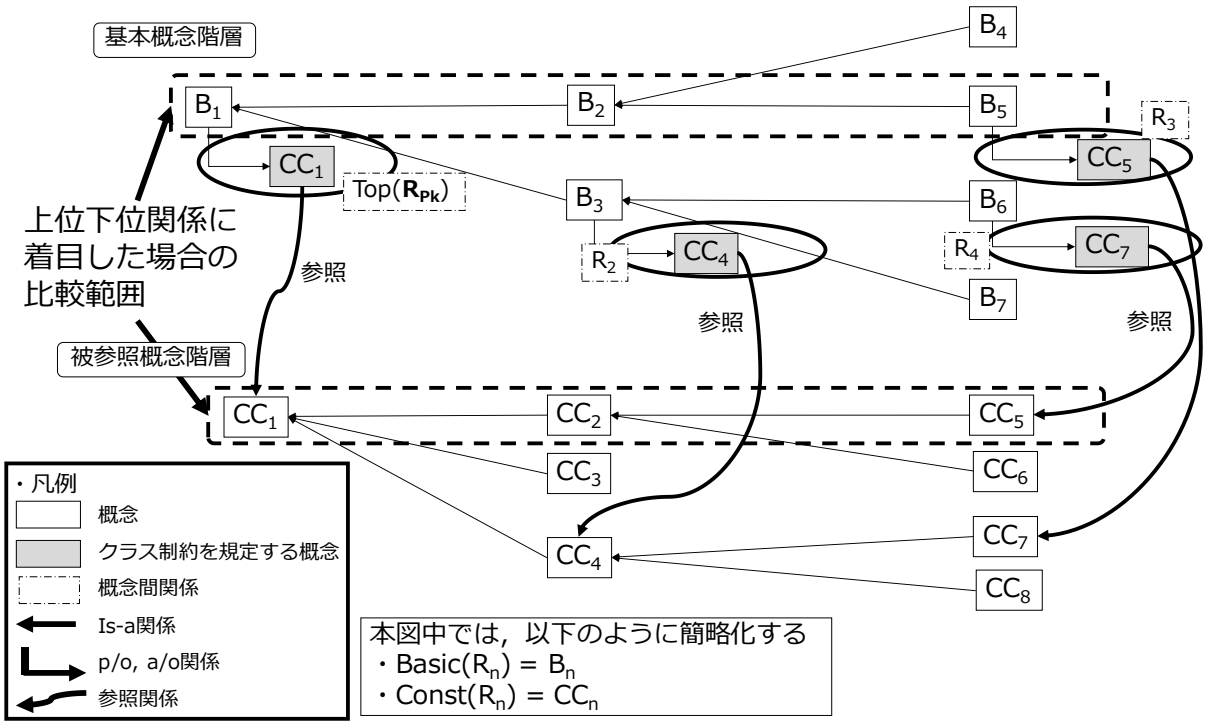

図 4.2.　オントロジー構築者に提示される概念階層

[18]図 4.2.中の各概念ラベルの添え字の数字は，概念を区別するために上位から順に便宜上つけている．

案がなされる．図中では，簡単のため$\mathrm{Basic}(R_n) = B_n$，$\mathrm{Const}(R_n) = CC_n$と表現している．この時，3 章で述べた手法の場合，上下一段毎に概念間関係階層を区切るので，例えば，

$$\mathrm{Isa}(\mathrm{Upper}(R_3),\ R_3)$$

で区切った場合，オントロジー構築者には，図中の破線内の概念階層が提示される．そのため，同じ分類観点を用いて定義された兄弟概念が存在するにもかかわらず，この局所的な概念階層中に定義された概念及び概念間関係から，構築者は，洗練提案の適用の可否を判断しなければならなかった．さらに，それぞれの概念には兄弟概念があるにも関わらず，それらに対する概念間関係や概念の追加提案は，別々の洗練箇所として検出されるため，同時に提案されない．

一方，分類観点毎に概念階層を比較する手法では，概念間関係階層を区切らないため，同じ分類観点を持つ概念間関係の集合，

$$\{\mathrm{Top}(\mathbf{R_{P_k}}),\ R_2, R_3, R_4\}$$

からなる概念間関係階層を中心に，各概念間関係によって性質が定義されている基本概念の集合

$$\mathbf{B_{P_k}} = \{B_1, B_3, B_5, B_6\}$$

と差集合

$$\mathbf{BL_{P_k}} - \mathbf{B_{P_k}} = \{B_2, B_4, B_7\}$$

からなる$\mathbf{BL_{P_k}}$の全ての概念が比較対象となる．被参照概念階層も同様に，

$$\mathbf{CCL_{P_k}} = \{CC_1, CC_2, CC_3, CC_4, CC_5, CC_6, CC_7, CC_8\}$$

が比較対象となる．そのためオントロジー構築者は，先に述べた「“B_2”に，“CC_2”をクラス制約として参照した概念間関係を追加する」を，より多くの情報から判断することができる．具体的には，基本概念階層中の“B_2”は，“B_3”を兄弟概念として持ち，“R_2”がクラス制約として“CC_4”を参照していることが分かる．そのため構築者は，「“B_2”の兄弟概念“B_3”が“CC_4”を参照しているので，“B_2”で“CC_4”の兄弟概念である“CC_2”を参照することは適切である」という判断ができる．

また，図 4.2.の概念階層の場合，“B_2”，“B_3”の下位には，“B_5”，“B_6”があり，それぞれ被参照概念階層中の“CC_5”，“CC_7”を参照している．この時，“B_4”，“B_7”に対して，“CC_6”，“CC_8”を参照するような提案も同時に提示されるようになる．

このようにシステムが分類観点毎に概念間関係階層全体を比較して洗練箇所を検出するため，オントロジー構築者は，洗練対象の概念に対する包括的な情報を考慮して，洗練提案の適用可否を判断できる．

表 4.1.　洗練箇所と洗練提案の対応と洗練提案数

洗練方法	適用可能な洗練箇所分類	洗練提案数
方法 1	分類 4.1.	{n(**BL**)-n(**B**)}* {n(**CCL**)-n(**CC**)}
方法 2	分類 4.1.，4.2.	{n(**CCL**) –n(**CC**)}
方法 3	分類 4.1.，4.3.	{n(**BL**) – n(**B**)}

4.1.3　洗練箇所に対する洗練提案の組み合わせ

本システムでは，概念及び概念間関係の追加提案を行うので，オントロジー構築者に対して示される洗練方法は，第 3 章の時と同様に次の 3 種類となる．

方法1　概念間関係の追加

方法2　基本概念階層への概念追加＋概念間関係の追加

方法3　被参照概念階層への概念追加＋概念間関係の追加

方法 1～3 の洗練提案は，3 章での洗練提案と同様に概念階層数の少ない概念階層への概念追加を行う．4 章では，式 4.4.と 4.5.の少なくとも一方を満たすとき非相似であると判定される．

例えば，方法 2 の基本概念階層への概念追加+概念間関係の追加は，被参照概念階層に，概念間関係によるクラス制約として使用されていない中間概念，もしくは兄弟概念がある場合にのみ提案可能である．なぜなら，提案される基本概念階層への概念追加は，被参照概念階層でより多く定義されている概念を参照する概念の追加が提案されるからである．このように洗練箇所の概念階層の構造によって，適用可能な洗練方法が異なる．取りうる洗練箇所の分類を以下に示す．また，洗練方法と洗練箇所分類の対応及び，洗練提案数を表 4.1.に示す．

分類4.1.　$\mathrm{n}(\mathbf{BL}_{\mathbf{P_k}}) > \mathrm{n}(\mathbf{B}_{\mathbf{P_k}}) \wedge \mathrm{n}(\mathbf{CCL}_{\mathbf{P_k}}) > \mathrm{n}(\mathbf{CC}_{\mathbf{P_k}})$の時

分類4.2.　$\mathrm{n}(\mathbf{BL}_{\mathbf{P_k}}) > \mathrm{n}(\mathbf{B}_{\mathbf{P_k}}) \wedge \mathrm{n}(\mathbf{CCL}_{\mathbf{P_k}}) = \mathrm{n}(\mathbf{CC}_{\mathbf{P_k}})$の時

分類4.3.　$\mathrm{n}(\mathbf{BL}_{\mathbf{P_k}}) = \mathrm{n}(\mathbf{B}_{\mathbf{P_k}}) \wedge \mathrm{n}(\mathbf{CCL}_{\mathbf{P_k}}) > \mathrm{n}(\mathbf{CC}_{\mathbf{P_k}})$の時

4.1.4　洗練提案の組み合わせの限定

分類観点毎に概念階層を比較する場合，1 つの概念間関係階層を介して参照関係にある概念階層中の概念数が多いため，洗練箇所 1 箇所あたりに提示される洗練提案の数が膨大になることがある．例としてサッカーオントロジー[溝口 06]の“行為”に関する基本概念階層とそれらの概念が“場所”のクラス制約として参照している“場所”の被参照概念階層を考える．今，“行為”は，

$$\mathrm{Top}\left(\mathbf{R}_{\text{場所}}{}^{19}\right) = \mathrm{Relation}(\text{“行為”},\text{“場所”},\text{“場所”})$$

となる最上位の概念間関係を持つ.“行為”の下位概念の集合$\mathbf{BL}_{\text{場所}}$の要素の中で$\mathbf{R}_{\text{場所}}$の要素を持たない概念数は,

$$\mathrm{n}(\mathbf{BL}_{\text{場所}}) - \mathrm{n}(\mathbf{B}_{\text{場所}}) = 40$$

と多い．なぜなら，“行為”の下位には，選手の行為，審判の行為，また，走る，蹴るなどといった基本的な行為も幅広く定義されているからである．また，“場所”の下位概念の集合からなる被参照概念階層中のクラス制約として参照されていない概念数は,

$$\mathrm{n}(\mathbf{CCL}_{\text{場所}}) - \mathrm{n}(\mathbf{CC}_{\text{場所}}) = 3$$

と少ないが，被参照概念階層には，“場所”の下位に“サッカーフィールド”，“競技場”，“領域”，“サッカー競技場”のみが定義されており，“サッカーフィールド”が“サッカー”の“場所”のクラス制約として参照されている．今，分類 4.1.の

$$\mathrm{n}(\mathbf{BL}_{\mathbf{P_k}}) - \mathrm{n}(\mathbf{B}_{\mathbf{P_k}}) > 0 \wedge \mathrm{n}(\mathbf{CCL}_{\mathbf{P_k}}) - \mathrm{n}(\mathbf{CC}_{\mathbf{P_k}}) > 0$$

にあたるので，4.1.3.節で述べたように方法 1，2，3 の洗練提案が提示される．洗練提案数は，表 4.1.の式を用いて計算すると,

$$(40 \times 3) + 40 + 3 = 163(\text{個})$$

となる．この例では，被参照概念階層の兄弟概念のうち，分類観点として着目した概念間関係$\mathbf{R}_{\text{場所}}$で参照されていない概念数は比較的少なく 3 個であった．しかし，被参照概念階層中の概念数が，基本概念階層と同程度である場合，提案数は，その組み合わせで増加するため，洗練箇所 1 か所あたり千以上の洗練提案が提示される．

このような膨大な洗練提案の提示を一度に行うことは，非現実的である．そこで，これらの洗練提案のうち，どの提案を優先的に表示するかを考える必要がある．提示される洗練提案を少なくするためには，比較対象となる概念を限定すればよい．比較対象を限定すれば，洗練提案の組み合わせが減少するからである．

本章では，比較対象となる概念の限定方法として，“同じ上位概念を持つ兄弟概念に限定する方法”と“特殊化の段数が同じ概念に限定する方法”を提案する．なぜなら本洗練支援を開発する上で着目した分類観点の同一性を満たすオントロジーでは，ある概念の下位概念は同じ分類観点で特殊化することが推奨されているからである．この基準を満たす時，最上位概念から同じ回数特殊化された概念は，概念の定義の詳細度が統一されているべきである．そのため同じ上位概念を持つ兄弟概念もしくは特殊化の段数が同じ概念が，概念間関係を追加する先の基本概念及びそのクラス制約として参照される被参照概念の候補である可能性が高いといえる．

[19]ここでは，着目している分類観点を明示するために，添え字を概念間関係のラベルとしている．

また，多くの概念を含む大規模な is-a 階層では，ある一つの分類の観点のみで is-a 階層全体が構成されることは珍しく，複数の分類観点を順番に用いて is-a 階層が構成されることが多い. 例えば, 乗り物のオントロジーの場合では, まず“移動空間属性”, 次に“車輪部分”, “動力源部分”のように異なる分類観点を順番に用いて特殊化される．洗練提案の組み合わせを，兄弟概念や特殊化の段数が同じ概念から生成することで，現実的な提案数に保ちつつ，全組み合わせの中から追加すべき洗練方法を抽出して提案できると考えられる．

4.1.4.1 同じ上位概念を持つ兄弟概念に限定する方法

洗練時に追加提案される概念を，同じ上位概念を持つ兄弟概念に限定する方法について説明する．ここで追加される概念は，概念間関係を追加する対象となる基本概念と概念間関係からクラス制約として参照される被参照概念である．この時，洗練箇所は，着目している概念間関係階層を介して参照関係にある基本概念階層と被参照概念階層全体であるが，システムが洗練提案を生成する際に概念及び概念間関係の追加先，クラス制約として参照する概念の候補を限定することで 1 箇所あたりの洗練提案数を減らす. 図 4.3 に, 同じ上位概念を持つ兄弟概念に限定した場合の範囲の例を示す. $\mathrm{Basic}(R_2)$と同じ上位概念$\mathrm{Basic}(R_1)\left(=\mathrm{Upper}\left(\mathrm{Basic}(R_2)\right)\right)$を持つ兄弟概念に限定した場合，図 4.3 中の実線で囲まれた範囲で，概念・概念間関係の追加先，クラス制約として参照する概念の組み合わせを提示することになる．図中のC_{basic_1}，C_{const_1}は，概念間関係を持たない基本概念および参照されていない被参照概念階層に便宜上通し番号をつけたものである．ある$\mathbf{R}_{\mathbf{P_k}}$に含まれる任意の概念間関係$R_n$毎に，基本概念階

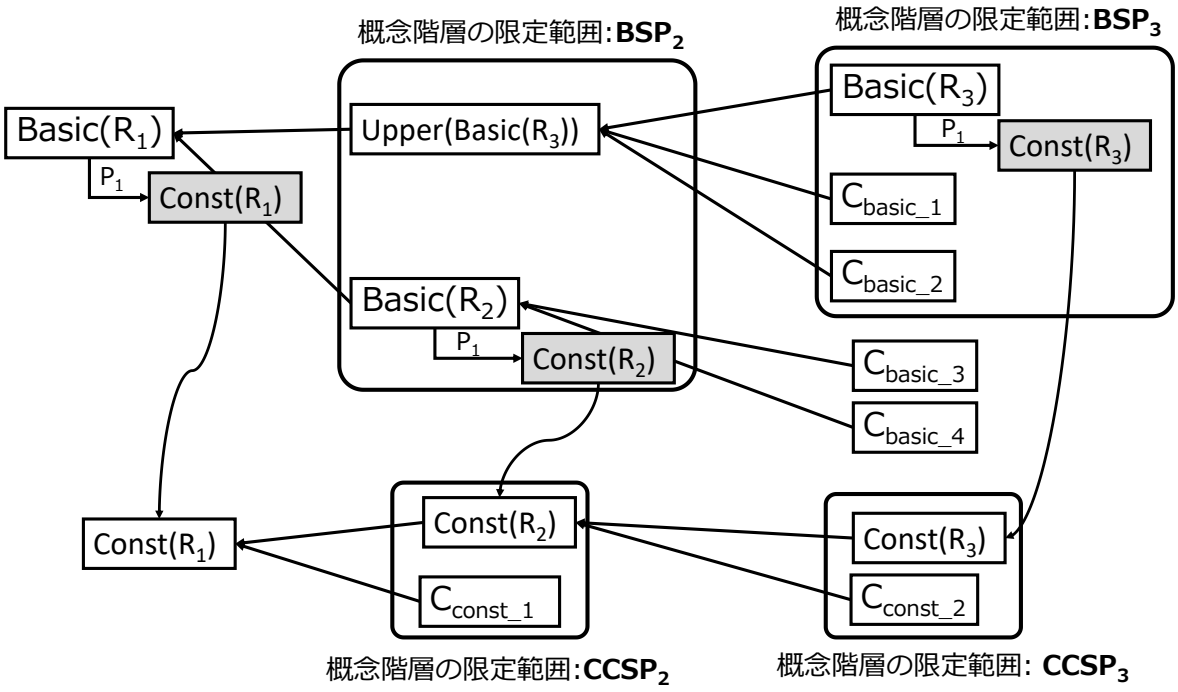

図 4.3.　同じ上位概念を持つ兄弟に限定する方法

層における兄弟概念の集合$\mathbf{BSP_n}$および，被参照概念階層における兄弟概念の集合$\mathbf{CCSP_n}$[20]を以下のように表現する．

$$\mathbf{BSP_n} = \mathrm{Lower}\left(\mathrm{Upper}\left(\mathrm{Basic}(R_n)\right)\right) \tag{4.6.}$$

$$\mathbf{CCSP_n} = \mathrm{Lower}\left(\mathrm{Upper}\left(\mathrm{Const}(R_n)\right)\right) \tag{4.7.}$$

この範囲にある概念同士で洗練提案の組み合わせを生成する．例えば，図 4.3.中のR_3は，

$$\mathrm{Relation}\left(\mathrm{Basic}(R_3), P_1, \mathrm{Const}(R_3)\right)$$

と定義されており，同じ上位概念を持つ概念の範囲は，$\mathbf{BSP_3}$ と $\mathbf{CCSP_3}$ である．この範囲では，

「“C_{basic_1}”に概念間関係P_1を追加し，そのクラス制約として“C_{const_2}”を参照する」
「“C_{basic_2}”に概念間関係P_1を追加し，そのクラス制約として“C_{const_2}”を参照する」

のように，範囲内の概念を用いた洗練提案が生成さる．“$\mathrm{Const}(R_2)$”や“C_{const_1}”を用いた洗練提案は，この範囲内の概念には提示されない．これらは，R_2に関係する範囲内のみで組み合わせが行われ洗練提案が生成される．

最終的に，あるP_kを持つ$\mathbf{R_{P_k}}$を介して参照関係にある概念階層に対して提示される洗練提案は，このR_n毎に生成される洗練提案の和集合となる．ただし，1 つの限定範囲に複数の概念間関係R_nがある場合は，重複する範囲を省く．なぜなら，洗練提案を生成する際の概念間関係の追加先，クラス制約の参照先の候補概念が重複して選ばれるからである．

4.1.4.2　特殊化の段数が同じ概念に限定する方法

2 つ目は，特殊化の段数が等しい概念，つまり is-a 階層を最上位概念から等距離にある概念に限定する方法である．この方法も洗練提案を生成する際の概念の組み合わせ数を減らし，洗練箇所当たりの提案数を制限することが目的である．Is-a 関係にあ

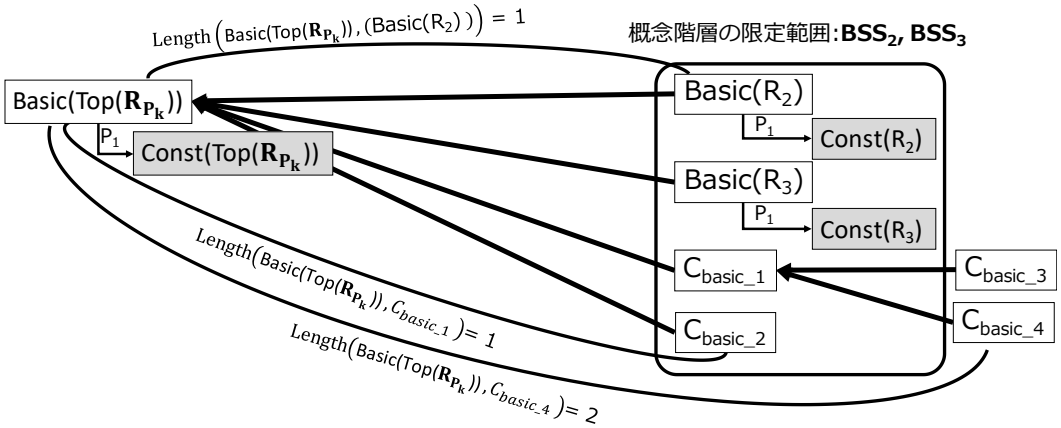

図 4.4.　特殊化の段数が同じ概念に限定する方法（$\mathrm{Basic}(R_2)$に着目した場合）

[20] SP は，Same Parent を意味する．

る任意の 2 概念について上位概念をC_{super}，下位概念をC_{sub}とすると，何段階特殊化されているかを，概念間の距離とみなし，

$$\mathrm{Length}\left(C_{super}, C_{sub}\right) \quad (4.8.)$$

と表現する．例えば，図 4.4.のBasic(R_2)は，Basic(Top($\mathbf{R}_{\mathbf{P_k}}$))から 1 段特殊化されているので，

$$\mathrm{Length}\left(\mathrm{Basic}(\mathrm{Top}(\mathbf{R}_{\mathbf{P_k}})), \mathrm{Basic}(R_2)\right) = 1$$

と書ける．特殊化の段数が同じで，洗練提案の生成時に組み合わせの候補となる概念は，同じく特殊化の段数が 1 である“Basic(R_2)”，“Basic(R_3)”，“C_{basic_1}”，“C_{basic_2}”の 4 つの概念である（図 4.4.実線範囲内）．その他の概念は，

$$\mathrm{Length}\left(\mathrm{Basic}(\mathrm{Top}(\mathbf{R}_{\mathbf{P_k}})), C_{basic_3}\right) = \mathrm{Length}\left(\mathrm{Basic}(\mathrm{Top}(\mathbf{R}_{\mathbf{P_k}})), C_{basic_4}\right) = 2 \neq 1$$

なので，システムが洗練提案を生成する際には，除外される．被参照概念階層も同様に対象となる概念階層を限定する．

これを一般化すると，基本概念階層の場合，ある概念Basic(R_n)に着目した時の限定範囲の集合 $\mathbf{BSS_n}$[21]は，Basic(Top($\mathbf{R}_{\mathbf{P_k}}$))とその下位概念の集合である$\mathbf{BL}_{\mathbf{P_k}}$のうち，Basic(Top($\mathbf{R}_{\mathbf{P_k}}$))からの距離が，Basic(Top($\mathbf{R}_{\mathbf{P_k}}$))とBasic($R_n$)の距離

$$\mathrm{Length}\left(\mathrm{Basic}\left(\mathrm{Top}\left(\mathbf{R}_{\mathbf{P_k}}\right)\right), \mathrm{Basic}(R_n)\right)$$

と等しくなる要素の集合である．被参照概念階層も同様に，ある概念Const(R_n)に着目した時の限定範囲の集合 $\mathbf{CCSS_n}$ は，Const(Top($\mathbf{R}_{\mathbf{P_k}}$))とその下位概念の集合である$\mathbf{CCL}_{\mathbf{P_k}}$のうち，Const(Top($\mathbf{R}_{\mathbf{P_k}}$))からの距離が，Const(Top($\mathbf{R}_{\mathbf{P_k}}$))とConst($R_n$)の距離

$$\mathrm{Length}\left(\mathrm{Const}\left(\mathrm{Top}\left(\mathbf{R}_{\mathbf{P_k}}\right)\right), \mathrm{Const}(R_n)\right)$$

と等しくなる要素の集合である．

こちらの方法の場合も，同じ上位概念を持つ兄弟概念に限定する方法と同様に，分類観点毎に提示されるのは，$\mathbf{R}_{\mathbf{P_k}}$に含まれる$R_n$毎に組み合わせに用いる概念が限定された洗練提案の集合となる．ただし，同じ上位概念を持つ兄弟概念に限定した場合と同様に，1 つの限定範囲に複数の概念間関係R_nがある場合は，重複を省く．例えば，図 4.4.では，概念階層の限定範囲内にR_2とR_3が存在するが，$\mathbf{BSS_2}$ と $\mathbf{BSS_3}$ は，同じ範囲となるので洗練提案が同じになる．そのため重複を除くため，R_2のみを考慮する．

4.2 分類観点毎の包括的な概念階層に着目した洗練支援システムの実装

上位下位概念を比較する手法と，本章で述べてきた分類観点毎の概念階層全体を比較する手法の両方に対応した品質向上支援システムを開発した．ただし，上位下位概念を比較する手法での全ての提案ではなく，その中でも最も提案に対して高い評価を

[21]SS は，Same Step を意味する．

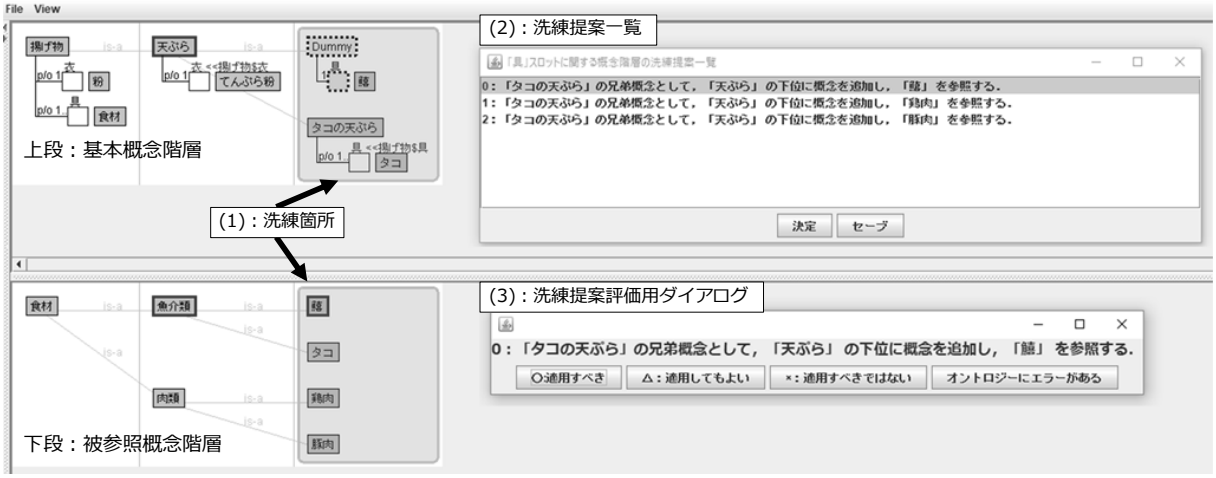

図 4.5.　洗練支援システム

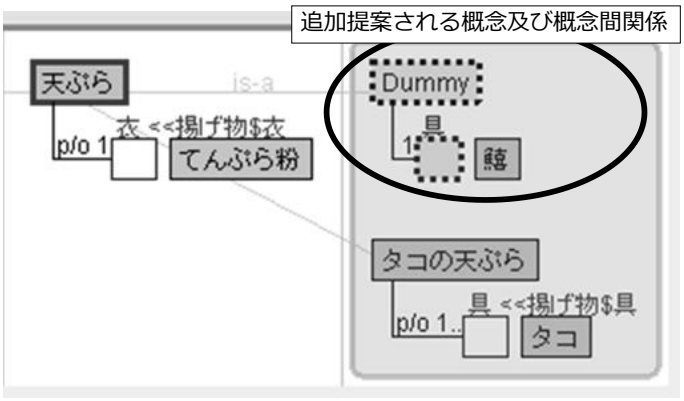

図 4.6.　追加提案される概念及び概念間関係のプレビュー表示

得たR_nとUpper(R_n)で概念間関係階層を区切る手法（図 3.11.中の a（i），a（ii）a（iii））を利用した．本手法は，基本概念階層中のBasic(R_n)とBasic$\big(\mathrm{Upper}(R_n)\big)$，または被参照概念階層中のConst$(\mathrm{R}_n)$とConst$\big(\mathrm{Upper}(R_n)\big)$の間に中間概念が有るどうかで相似かどうかを判定するものである．この手法を利用した理由は，他の手法は 3.3.3.2.節で述べたように，提案が妥当と判定された割合が低かったからである．特に最上位の概念間関係を中心に概念階層を比較する手法は，0~4%と低かった（図 3.11.中の b（i））．

洗練支援システムは，以下の 4 つのモジュールから構成される．

- 概念間関係階層抽出モジュール
- 概念階層描画モジュール
- 洗練箇所推定モジュール
- 洗練提案推定モジュール

これらのモジュールから構成されるシステム全体の処理の流れを以下に示す．

処理1　洗練対象としたオントロジーをシステムに入力すると，概念間関係階層抽出モジュールによって，対象オントロジーから概念間関係のラベルP_k毎に概念間関係階層が抽出され，それぞれの最上位の概念間関係Top$\big(\mathbf{R}_{\mathbf{P_k}}\big)$一覧が表示される．

処理2　オントロジー構築者は，表示された最上位の概念間関係一覧から，洗練したい概念間関係階層を選択する．

処理3　処理 2 の選択結果を入力として，概念階層描画モジュールが，基本概念階層と被参照概念階層を描画する．また，上段と下段の概念階層中に参照関係があり，洗練提案が提示される段数をハイライト表示する（図 4.5 の基本概念階層及び被参照概念階層の上位から 3 階層目）．

処理4　オントロジー構築者は, 処理 3 でハイライトされている 2 箇所から選択し，洗練提案の生成方法を(i)同じ上位概念を持つ兄弟概念に限定する方法，(ii)特殊化の段数が同じ概念に限定する方法，(iii)上位下位概念を比較する手法の 3 種類から選択する．

処理5　処理 4 でオントロジー構築者が選択した，洗練箇所及び洗練提案の生成方法を入力として，洗練提案推定モジュールが洗練提案一覧（図 4.5.中の(2)）を生成し，構築者に提示する．

処理6　オントロジー構築者は，洗練提案一覧に示されたものから 1 つを選択すると洗練プレビュー（図 4.6.は，図 4.5.の当該部分を拡大表示したもの）が表示される．この概念階層を参考に，各洗練提案に対する評価を図 4.5.中の(3)のダイアログから選択する．

4.3　評価実験

本研究で開発したオントロジー内容洗練支援システムの評価実験について述べる．本評価実験では，システムがオントロジー構築者に提示する洗練提案を以下の 3 点で評価した．

1. 概念間関係階層全体の比較によって新たに検出された洗練箇所，洗練提案数
2. 洗練提案の妥当性の評価
3. 3 章での提示方法と 4 章での提示方法での評価の比較

1 点目は，概念間関係階層全体の比較で，新たに提案可能となった洗練箇所，洗練提案数を評価する．また，1 箇所当たりの洗練提案数を見ることで，提案の組み合わせの限定が有効であったかを評価する．

2 点目として，洗練提案毎にその適用が妥当かどうかを評価し，4 章での洗練支援手法による提案がどの程度有用であったのかを評価する．

3 点目は，3 章で述べた一段の上位下位関係に着目した場合の概念階層の提示方法と 4 章で述べた分類観点全体を提示する方法とで評価の比較を行う. 4 章での手法は，オントロジー構築者に提示する情報をより包括的なものにするために関連する概念階層全体を見せている．この提示方法の違いによるオントロジー構築者の評価を比較することが目的である．

実験の被験者は，計 17 人である．システム利用者の構築経験の影響を調査するために，システムの利用に最低限必要な構築経験を持つ大学院生 5 名と，オントロジーの構築・利用に関する研究に携わっている研究者 12 名に協力を得た．被験者は，全員筆者の所属研究室外から募った．

評価実験に使用したオントロジーは，全部で 5 種類である．熟練者が構築した，サッカーオントロジー[溝口 06]，乗り物オントロジー[溝口 06]の 2 つと，本研究室に配属された直後の学部 4 回生の学生が初めて構築した，ドラム演奏オントロジー，列車オントロジー，交通オントロジーの 3 つの計 5 つである．これは，3 章での実験で利用したオントロジーからサステナビリティサイエンスオントロジーを除いたものである．その理由としては，このオントロジーは，他のオントロジーと比べて規模が桁違いに大きく，提示される洗練箇所や提案の比較対象として不適当であると考えたからである．

17 名の被験者には，これら 5 つのオントロジーから検出された全ての洗練箇所から計 108 の洗練箇所を抜き出し評価を依頼した．洗練箇所は，1 箇所当たりの洗練提案数が 100 を超えない範囲で，無作為に抽出した．1 箇所当たりの提案数が 100 を超えると評価者の負担が大きすぎるためである．5 人の学生は，5 人中 4 人には一人当たり 65 箇所ずつ，残り一人は 64 箇所の評価を依頼した．残りの 12 人の熟練者に対しては，一人当たり 27 箇所の評価を依頼した．この割り当てにより，初心者３人からの評価と熟練者 3 人からの評価が洗練箇所ごとに 2 種類得られる．被験者への各洗練箇所の提示は，対象となるオントロジーの順番，ドメインの難易度などの影響を抑えるため，提示順をランダムにした．

洗練箇所の妥当性の判断は，各箇所に提示される複数の洗練提案のうち 1 つ以上に◎または○が与えられた場合，その洗練箇所の検出が妥当であるとした．評価実験では，提示される洗練提案の評価を目的としているため，洗練箇所と組み合わせの限定方法は予め指定された状況から評価を始める．そのため評価実験時には，4.2.節で紹介したシステムを実験用に調整した実験用システムを用いた．評価実験時の被験者のシステム利用手順は，具体的には以下の流れで進められる．

手順1　洗練システムを起動

手順2　作業一覧より，まだ評価していない洗練箇所を選択

手順3　洗練提案の一覧（図 4.5 中の（2））とその概念階層が表示されるので，各洗練提案をクリックし，ダイアログ（図 4.5 中の（3））に従い評価

手順4　洗練提案一覧の全提案に評価を終えたら，データを保存し手順 2 に戻り，他の洗練箇所を選択．（被験者には，図 4.5 中の（3）にあるように，各洗練提案に対して次の 3 段階と評価外の，4 通りの評価を与えるように指示した．）

- ◎：適用すべきであると判断した洗練提案
- ○：適用してもよいと判断した洗練提案

- ×：適用すべきではないと判断した洗練提案
- エラー：提示された洗練箇所に定義間違いや矛盾が存在するため，評価の対象外とする

評価実験では，洗練提案毎に初心者と熟練者の各 3 人から 2 種類の評価が得られるため，初心者と熟練者のそれぞれについて洗練提案の評価を決定する．3 人による評価を表 4.2.に示すように統合してそれぞれの最終評価とする．

表 4.2.　各被験者による評価

1 人目	2 人目	3 人目	洗練提案の評価
◎	◎	◎	◎：適用すべき
◎	◎	○	◎：適用すべき
◎	◎	×	◎：適用すべき
◎	◎	エラー	◎：適用すべき
◎	○	○	○：適用してもよい
◎	○	×	○：適用してもよい
◎	○	エラー	○：適用してもよい
◎	×	エラー	○：適用してもよい
○	○	○	○：適用してもよい
○	○	×	○：適用してもよい
○	○	エラー	○：適用してもよい
○	×	エラー	○：適用してもよい
○	×	×	×：適用すべきではない
×	×	×	×：適用すべきではない
×	×	エラー	×：適用すべきではない
◎	×	×	×：適用すべきではない
◎	エラー	エラー	エラー：オントロジーにエラーがある
○	エラー	エラー	エラー：オントロジーにエラーがある
×	エラー	エラー	エラー：オントロジーにエラーがある
エラー	エラー	エラー	エラー：オントロジーにエラーがある

表 4.3.　各オントロジーから検出された洗練箇所・洗練提案数

評価対象のオントロジー		総概念数	概念間関係階層数	洗練箇所数							洗練提案数													
											(a) 4 章の手法のみで提案される				(b) 3，4 章の手法のどちらでも提案される				(C) 3 章の手法のみで提案される			1 箇所当たりの提案数			
				同じ親に限定		同じ段数に限定		3 章の手法		合計	同じ親に限定		同じ段数に限定		同じ親に限定		同じ段数に限定		合計			合計				
				基本	被参照	基本	被参照	基本	被参照		基本	被参照	基本	被参照	基本	被参照	基本	被参照		基本	被参照		平均値	最小値	最大値	中央値
初心者	ドラム演奏	136	16	14	14	14	12	7	9	70	687	785	497	907	59	72	23	68	222	38	20	58	3156	8	318	1	31
	鉄道	158	10	8	8	8	4	5	3	36	69	74	107	78	8	10	13	12	43	21	4	25	396	3	66	1	6
	交通	252	18	31	27	27	17	17	13	132	736	874	387	2062	54	68	24	30	176	223	33	256	2635	7	221	1	18
熟練者	サッカー	262	18	27	25	23	17	13	5	110	376	324	547	2281	24	12	26	67	129	64	12	76	1680	6	104	1	14
	乗り物	122	8	11	10	10	7	6	7	51	60	57	185	112	4	8	13	28	53	85	26	111	578	5	108	1	10
合計		930	70	91	84	82	57	48	37	399	1928	2114	1723	1531	149	170	99	205	623	431	95	526	8445	6	318	1	18

表 4.4.　洗練箇所及び洗練提案の妥当性の評価

評価対象のオントロジー		洗練箇所毎の評価				洗練提案毎の評価			
		◎以上		○以上		◎以上		○以上	
		初心者	熟練者	初心者	熟練者	初心者	熟練者	初心者	熟練者
初心者	ドラム演奏	12/18 (67%)	2/18 (11%)	18/18 (100%)	15/18 (83%)	19%	1%	76%	34%
	鉄道	10/18 (56%)	5/18 (28%)	18/18 (100%)	15/18 (83%)	10%	3%	47%	24%
	交通	11/18 (61%)	7/18 (39%)	18/18 (100%)	17/18 (94%)	12%	10%	76%	51%
熟練者	サッカー	13/24 (72%)	7/24 (29%)	23/24 (96%)	19/24 (79%)	13%	3%	66%	32%
	乗り物	13/30 (43%)	4/30 (13%)	28/30 (93%)	22/30 (73%)	7%	2%	64%	22%
合計		59/108 (50%)	26/108 (24%)	105/108 (97%)	88/108 (81%)	13%	4%	68%	33%

4.4　実験結果及び考察

4.4.1　検出される洗練提案数

表 4.3.に，各オントロジーから検出された洗練箇所数と洗練提案数を示す．各オントロジーの概念間関係階層数は，10～20 程度で，洗練箇所数や洗練提案数はその数によって増えていることがわかる．ただし，洗練提案数は，含まれる概念の組み合わせによって増加するため，概念間関係階層や洗練箇所数の増加と線形に増加するのではなく，概念間関係階層数が増えると爆発的に増加していることがわかる．また，オントロジー毎の 1 箇所当たりの洗練提案数を見ると，最大値は多いものでは，1 箇所当たり 300 を超える提案がなされている．一方，中央値はどのオントロジーでも一桁となっている．平均値は最も多いドラム演奏でも 30 程度である．このことから，最大値の 3 桁を超えるような洗練箇所は少数であり，全体として，数個～数十個程度が大勢を占めている．

4.4.2　洗練箇所毎の評価

表 4.4.の左側に，洗練箇所毎の評価結果を示す．今回の評価実験には，初心者のオントロジーから各 18 箇所，熟練者の構築したオントロジーのうち，サッカーからは 24 箇所，乗り物からは 30 箇所の計 108 か所を無作為に抽出した．表 4.4.の結果から，洗練箇所毎に，◎または○，つまり“適用すべき”または“適用してもよい”と評価された洗練提案が 1 つ以上存在するものは，初心者で 97%，熟練者で 81%であり，8 割以上の洗練箇所に対して，1 つ以上の適切な洗練提案を提示することができた．

4.4.3　洗練提案の妥当性の評価

4.4.3.1　　洗練提案の分類毎の評価

本章で提案した洗練支援システムによって提示される洗練提案を
(a) 4 章での洗練支援手法のみで提案されるもの
(b) 3 章の手法と 4 章の手法の両方で提案されるもの
(c) 3 章の手法のみで提案されるもの
の 3 つに分けて比較した結果を図 4.7.に示す.

図 4.7.の(a)と(c)を比較した場合について述べる．◎“適用すべき”と，○“適用してもよい”の合計，つまり適用が妥当であると判断された割合は，初心者では，(a)が 68%であり，(c)が 83%である．熟練者では，(a)が 34%であり，(c)が 38%である．初心者及び熟練者の両方で，(a)が(c)よりも低くなっているが，(a)，(c)のどちらも初心者では 6 割以上，熟練者からの評価でも 3 割以上の洗練提案が妥当であるとされた．初心者よりも熟練者の方が，より少ない概念で効率的に概念階層を構築し，実際の利用を考慮した概念を追加する傾向にあるため，この差が生まれたと考えられる．洗練(a)は(c)よりも，初心者が妥当であると評価した割合が 15%，熟練者が妥当であると評価した割合が 4%低くなっている．一方，表 4.3.の手法毎の提案数の合計を見ると，(a)では 7296 個，(c)では 526 個であり，(a)の 4 章での手法によって新たに 10 倍以上提案されていることが確認できる．10 倍以上の新しい提案をしながらも妥当であると評価された洗練提案の割合を大幅に落としてはいないことが分かる．

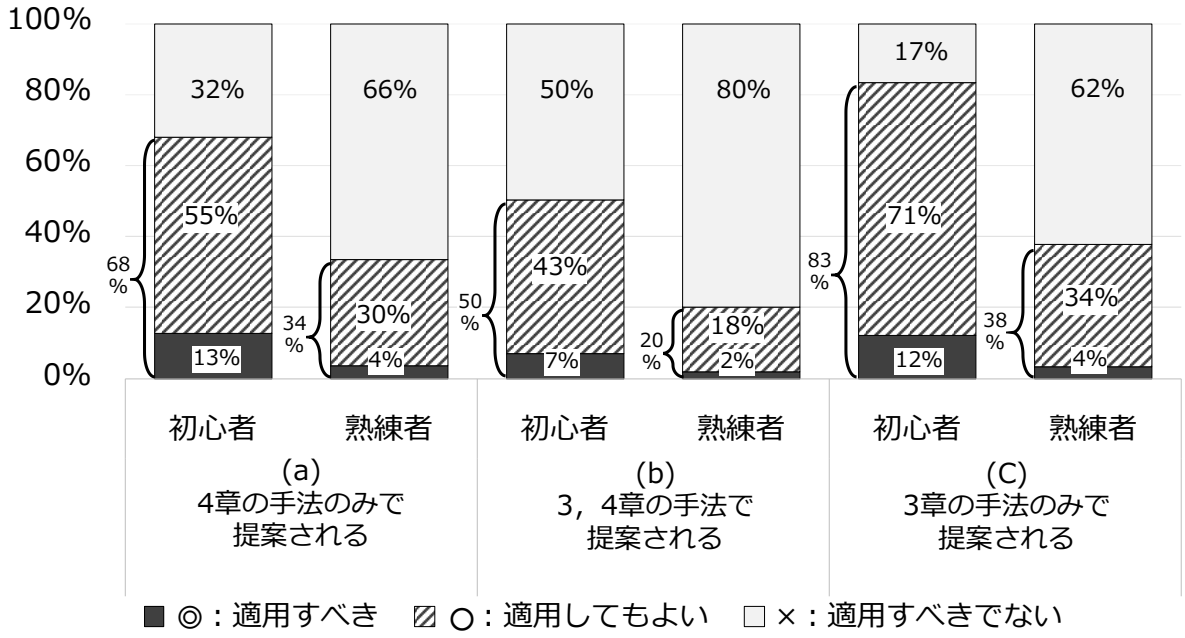

図 4.7.　洗練提案の手法による評価の差

次に，(a)4 章の手法のみで推定される提案と(b)3 章と 4 章で共通する提案を比較した結果について述べる．図 4.7.では，(a)の○“適用してもよい”以上の評価は，初心者で 68%，熟練者で 34%であり，(b)では，初心者で 50%，熟練者で 20%であり，(b)の方が妥当であると評価された提案の割合が低い．4 章の手法のみでの提案では，基本概念階層の最下位にある概念に対しての概念間関係の追加，または被参照概念階層の最下位にある概念のクラス制約としての参照が多く提案される．末端にある概念は，下位概念を持たないため，提案の適用判断時に考慮すべき概念が少ない．そのため，追加先の上位概念と下位概念の両方を考慮する必要のある中間概念の追加よりも判断が容易である．ゆえに，被験者は提案が妥当であると評価しやすい傾向にあり，今回の実験では 4 章の手法のみで推定された提案の方が，3 章と 4 章の両方に共通する提案よりも提案の妥当性が高くなったと考えられる．

概念階層の見せ方の違いによる洗練提案に対する評価の違いを表 4.5.に示す．この表には，左に 3 章での洗練提案の評価実験の結果が示されており，右には同じ方法で検出された洗練箇所を分類観点毎に概念階層を包括的に見せた場合の評価が示されている．ここでは，3 章，4 章で共通して提示された洗練提案に対する評価を集計している．左では，「◎：適用すべき」と「○：適用してもよい」の合計は，初心者では 22%，熟練者では 12%となっていた．一方，本章で提案した包括的な提示方法では，同じ洗練提案であるにも関わらず，「◎：適用すべき」と「○：適用してもよい」の合計は，初心者では 83%，熟練者では 37%となっていた．この結果より，洗練提案の適用可否の判断を行うためには，オントロジー構築者に対して局所的な情報だけを提示するのではなく，洗練部分に関する包括的な情報を提示することが有効であることが確認できた．

4.4.3.2 洗練提案例

図 4.8.に，◎“適用すべき”とされた洗練提案例を示す．この時，破線部分に示す概念間関係，

Relation(“キック”，“基礎動作”，“当てる”)

の追加が提案されている．図 4.8.は，ドラム演奏のオントロジーから得られた，概念間関係のラベルが“基礎動作”である概念間関係階層の一部である．実際の概念階層は，上段(基本概念階層)に 14 個，下段(被参照概念階層)に 16 個の概念からなる概念

表 4.5. 見せ方の違いによる評価の差

	1 段の階層のみの局所的な提示			概念階層の包括的な提示		
	◎	○	×	◎	○	×
初心者	9%	13%	78%	12%	71%	17%
熟練者	4%	8%	89%	3%	34%	62%

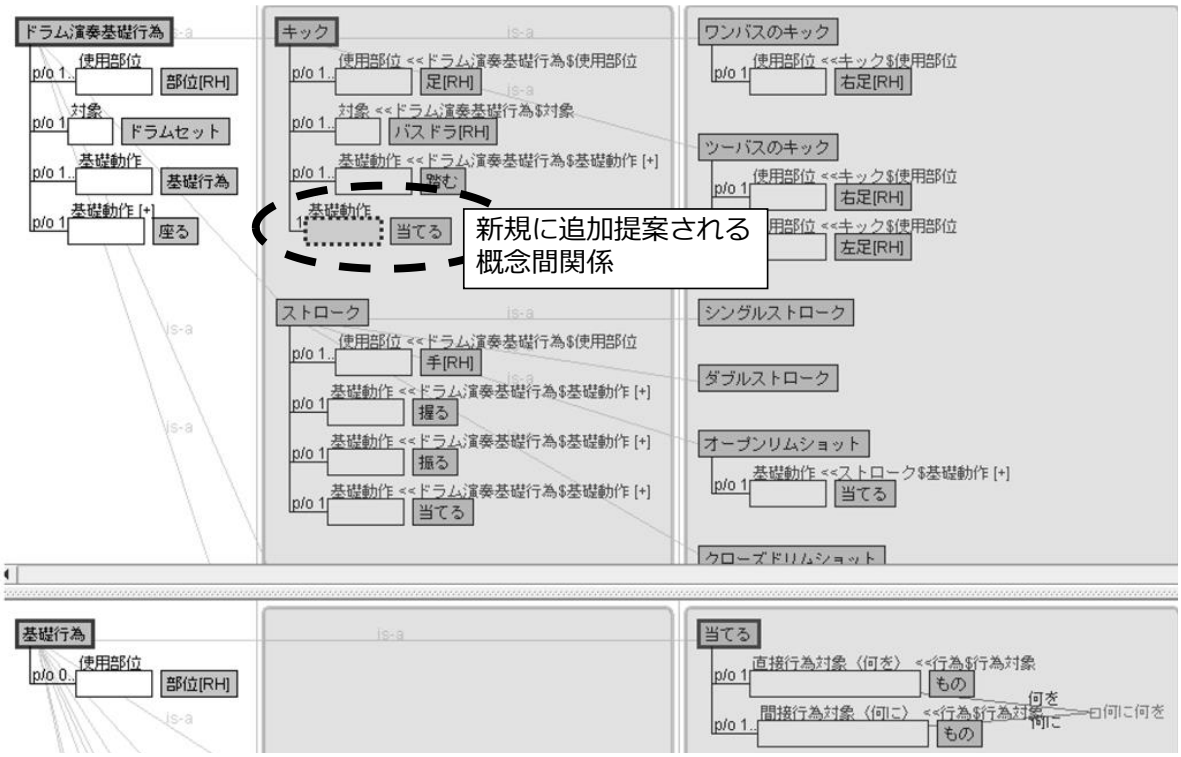

図 4.8.　適用すべきとされた洗練提案例

階層である．オントロジー構築者は，表示インタフェース上をスクロールすることで概念階層全体を見ることができる．

この概念階層に対して“適用すべき”とされた洗練提案は,「“当てる”を“キック”の“基礎動作”のクラス制約として参照する.」であり，図 4.8.中で点線円に囲まれた概念間関係が追加提案される．このとき，足を用いて楽器を演奏する“キック”に“当てる”が“基礎動作”として追加される．その結果，“キック”は，“踏む”動作だけではなく，“バスドラ”という楽器に何かしらの器具を“当てる”ことであると，その定義がより詳細化された．ゆえに被験者は，この提案を適用することが妥当であると判断したと思われる．“キック”は兄弟概念として“ストローク”を持つため，この洗練提案の適用判断時にその定義内容を参考にすることが可能である．兄弟概念の“ストローク”では,“当てる”が基礎動作のクラス制約として参照されているため，その兄弟概念であり，同様の動作を表現する“キック”でも“当てる”を基礎動作のクラス制約として参照させることが妥当であると判断できる．

第5章　概念階層比較手法を応用したオントロジー拡張支援システム

本章では, オントロジー拡張支援システムについて述べる. 本論文での拡張支援は, 対象オントロジーの概念間に新たな概念間関係を追加提案することである. 本拡張支援手法は大きく 2 ステップに分けて行われる. まず, 3, 4 章で述べてきた洗練支援システムと同様に参照関係にある概念階層を比較し, 概念間関係を追加すべき概念の組み合わせ候補を抽出する. 次に, Web 上のリソースを用いて概念の組み合わせ候補をランキングする. この拡張支援システムを, 実際に構築途中段階にある「JST シソーラスを基にしたバイオサイエンスオントロジー」に適用し, 手法の評価を行う.

5.1　拡張支援に向けた概念階層比較手法の応用

本節では, 構築途中段階のオントロジーに対しても適用可能な概念階層比較手法について述べる.

5.1.1　構築途中段階の概念定義がスパースなオントロジーへの対応

3, 4 章では, 構築の洗練・評価段階にあるオントロジーに対し, 洗練支援のための提案を行うシステムを開発した. 本章では, 構築途中段階のオントロジーに対して, 既存の 2 概念間に結ぶべき概念間関係を提案する, 拡張支援システムを開発する.

オントロジー拡張支援手法でも, 概念階層を比較する基本的考え方は共通である. 3, 4 章で述べたオントロジーの洗練支援手法では, オントロジーの各概念の性質を表す概念間関係が密に定義されており, 概念間の参照関係が十分に存在していることを前提としていた. しかしながら, 概念間関係の定義には, 概念や is-a 関係に比べ, 対象ドメインの専門知識やオントロジー構築の知識・経験がより多く必要とされる. そのため構築途中段階のオントロジーに対する支援手法として, 概念間関係を追加提案するようなツールの開発が望まれている.

このような構築途中段階のオントロジーでは, 概念間関係の定義が未だ途中段階であるため, 概念間関係階層が十分に構成されていない. そのため, 概念間関係の is-a 階層を利用している 3, 4 章で述べた洗練支援手法をそのまま用いることができない. なぜなら基本概念階層と被参照概念階層の比較は, 概念間関係を介して相対的に定義される参照・被参照の関係を利用していたためである.

本章では，このような概念間関係の定義がスパースな構築途中段階のオントロジーから，新たに結ぶべき概念間関係を推定し，オントロジー構築者に提案する概念間関係の拡張支援手法を開発する. この拡張支援手法は, 大きく 2 ステップで行われる.

ステップ1　既存の概念間関係 R_n を利用して，概念間関係を結ぶべき概念の組み合わせの候補（「組み合わせ候補」と呼ぶ）を抽出する.

ステップ2　組み合わせ候補を Linked Open Data（LOD）を用いてランキングする.

本拡張手法によって，オントロジー中の全概念間の組み合わせを考慮することなく，概念間関係を結ぶべき概念の組み合わせを絞り込むことが可能となるため，オントロジー構築作業をより効率化できる.

5.1.2　JST シソーラスを基にしたバイオサイエンスオントロジー

本研究では，構築途中段階のオントロジーの具体事例として，JST シソーラスを基にしたバイオサイエンスオントロジーを用いた．JST シソーラスは，約 24,500 の科学技術に関する幅広い語彙からなるシソーラスであり，ライフサイエンス，基礎化学，物理，機械，工業化学，環境公害分野などの分野から構成されている．こうした語彙は，上位下位関係によって階層化され，語彙間には，“関連語（RT）”関係が結ばれている．RT は，p/o，a/o 関係に相当するが，概念間関係のラベルを持たず区別されていない．現在，櫛田らによって JST シソーラスのライフサイエンス分野をオントロジー化する試みが進められている[櫛田 17]．シソーラスのオントロジー化にあたり，これまで単に RT として定義されていた概念間関係が，ライフサイエンスドメインの専門家の手により，“Function-of”，“Product”，“Phenotype”，“Provider”などの 31 種類の概念間関係に置き換えられた．この RT の置換により，各概念がどのような分類の観点によって定義・特殊化されたかが明示化された.

しかしながら，基となった JST シソーラスには，全体の概念数に比べ RT 関係の定義数が少ない．ゆえに，その RT を置換した現状のバイオサイエンスオントロジーに存在する概念間関係の数も少ない. JST シソーラスのオントロジー化の次段階として，概念間関係の数を増やし，オントロジーの持つセマンティクスを増やすことが求められている.

本研究では，このバイオサイエンスオントロジーを例として，概念間関係を結ぶべき組み合わせ候補を抽出し，それらをランキングする手法を開発する．また，本研究においては，31 の関係の種類の中から“Function-of”関係に着目し，“薬物”，“蛋白質・ペプチド”とそれぞれの“作用”との間に“Function-of”を結ぶ組み合わせ候補を探す.“Function-of”関係，“薬物”，“蛋白質・ペプチド”に着目した理由としては，生命科学分野において重要な概念及び概念間関係であるため，優先的に追加すべきだ

からである．また，この“Function-of”関係は，RT から置換された概念間関係 1005 個のうち，352 個と最も多いことからも優先的に定義すべき概念であることが分かる．

5.2 概念階層比較による組み合わせ候補単語の抽出

5.2.1 概念間関係の定義がスパースなオントロジー

図 5.1.に，バイオサイエンスオントロジーにおける薬物と薬理作用の概念階層の模式図を示す．薬物の下位概念の一つ“解熱鎮痛薬”には，薬理作用の下位概念の一つ“解熱作用”とラベルが“Function-of”である概念間関係が結ばれており，

R_n =Relation(“バソプレシン”，“Function-of”，“昇圧作用”)

として定義されている[22]．このように薬物の下位概念は，薬理作用の下位概念の階層中の概念を参照して，それぞれの薬物がどのような機能を持つかが定義されている．しかし，ラベルが“Function-of”である概念間関係は，この 2 階層間全体で 12 個しか定義されていない．今，薬物の階層には，142 概念，薬理作用の階層には 96 概念が存在することを考えると，薬物と薬理作用間の概念間関係の定義はスパースである．

本章で提案する拡張支援手法では，既存の概念間関係（ここでは“薬物”と“薬理作用”間の“Function-of”関係）を参考にして新たな概念間関係を追加することを考える．この時，単純に“薬物”と“薬理作用”の概念階層に含まれる概念間の取り得る組み合わせの全てをオントロジー構築者に提案した場合，その数は膨大なものとなる．図 5.1.の薬物と薬理作用の階層の場合，薬物には 142 概念，薬理作用には 96 概念が存在するので，その組み合わせは 13,632 通りである．そのため，関係のある概念階層に含まれる概念の全部の組み合わせを単に提示するだけでは，候補数が多すぎるため拡張作業の支援には不適切である．

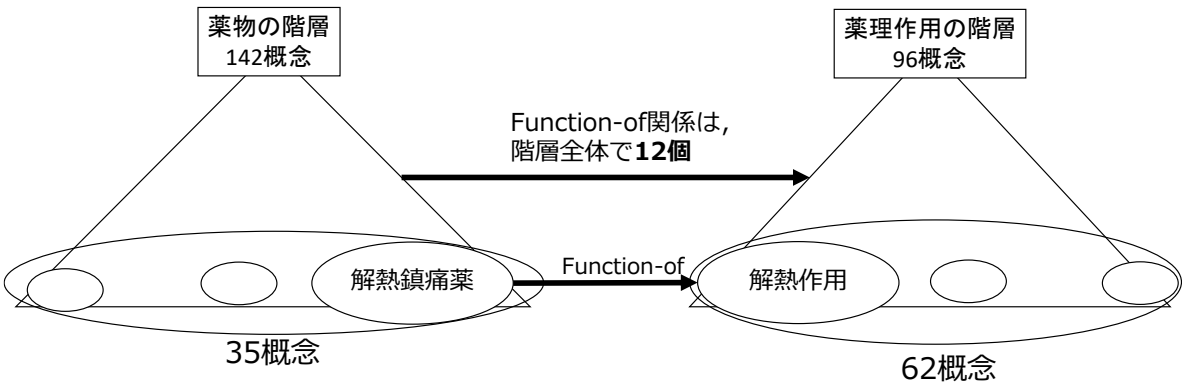

図 5.1. 薬物と薬理作用の階層間の関係

[22]図 5.1.では，概念間に Function-of の矢印が引かれているが，“解熱鎮痛薬”に定義された概念間関係から“解熱作用”がクラス制約で参照されていることを示す．

5.2.2　同じ段数にある概念に着目した概念階層の比較

既存の概念間関係を参考にして，概念間関係がスパースなオントロジーから，適切に概念間関係を結ぶべき概念の組み合わせ候補を抽出するために，4.1.4.2.節の“特殊化の段数が同じ概念に限定する方法”を利用する．本章では，対象としているバイオサイエンスオントロジーの Function-of 関係に着目しているので, この概念間関係を持つ概念と同じ階層にある概念同士を組み合わせ候補として抽出する．このように比較対象の概念を制限することで，組み合わせ数の爆発を抑えながら，概念の特殊化の詳細度が近い，より概念間関係を結ぶべき概念の組み合わせを抜き出すことが可能となる．

概念間関係を結ぶべき概念の組み合わせを抜き出すために，これまでの手法と同様に概念間関係R_nを介して参照関係にある概念階層の比較を行う．本章で述べる拡張支援では，概念間関係の is-a 関係を用いないため，定義されている概念間関係の数が少ない構築途中段階のオントロジーに対しても利用できる．概念間関係を持つ概念$\mathrm{Basic}(R_n)$とクラス制約として参照されている概念$\mathrm{Const}(R_n)$を，3，4 章の時と同様に定義する．

また，今回対象としているバイオサイエンスオントロジーは，薬物や蛋白質などの専門領域毎の複数の概念階層から構成されている．そのため，概念階層毎に複数の最上位概念が存在する．概念 *C* が所属する概念階層の最上位概念を，

$$\mathrm{DomainTOP(C)} \tag{5.1.}$$

と表現すると, 概念間関係を持つ概念$\mathrm{Basic}(R_n)$とクラス制約として参照されている概念$\mathrm{Const}(R_n)$が所属する概念階層の最上位概念は，それぞれ，

$$\mathrm{DomainTOP}\big(\mathrm{Basic}(R_n)\big) \tag{5.2.}$$

$$\mathrm{DomainTOP}\big(\mathrm{Const}(R_n)\big) \tag{5.3.}$$

と表現できる．なお，このように領域毎に複数の概念階層が存在することは，バイオサイエンスオントロジーに限られることではなく，多くのオントロジーで定義されている．これらの領域内での最上位概念は，個々のオントロジー構築者の定義に委ねられる．

また, $\mathrm{DomainTOP}\big(\mathrm{Basic}(R_n)\big)$および, $\mathrm{DomainTOP}\big(\mathrm{Const}(R_n)\big)$の下位概念の集合を，$\mathbf{BL_n} = \{BL_a | a = 0, \dots, A\}$，$\mathbf{CL_n} = \{CL_b | b = 0, \dots, B\}$とする．A，B は，各集合に含まれる要素の数である．各階層の最上位概念と各概念との特殊化の段数は，式 4.8.を用いて

$$\mathrm{Length}\big(\mathrm{DomainTOP}\big(\mathrm{Basic}(R_n)\big),\ BL_a\big)$$

と書ける．あるR_nを持つ概念と$\mathrm{DomainTOP}\big(\mathrm{Basic}(R_n)\big)$の距離は，

$$l_{basic}(n) = \mathrm{Length}\left(\mathrm{DomainTOP}\big(\mathrm{Basic}(R_n)\big),\ \mathrm{Basic}(R_n)\right)$$

と表現でき，$\mathrm{Basic}(R_n)$と同じ距離にある$\mathbf{BL_n}$の部分集合は，

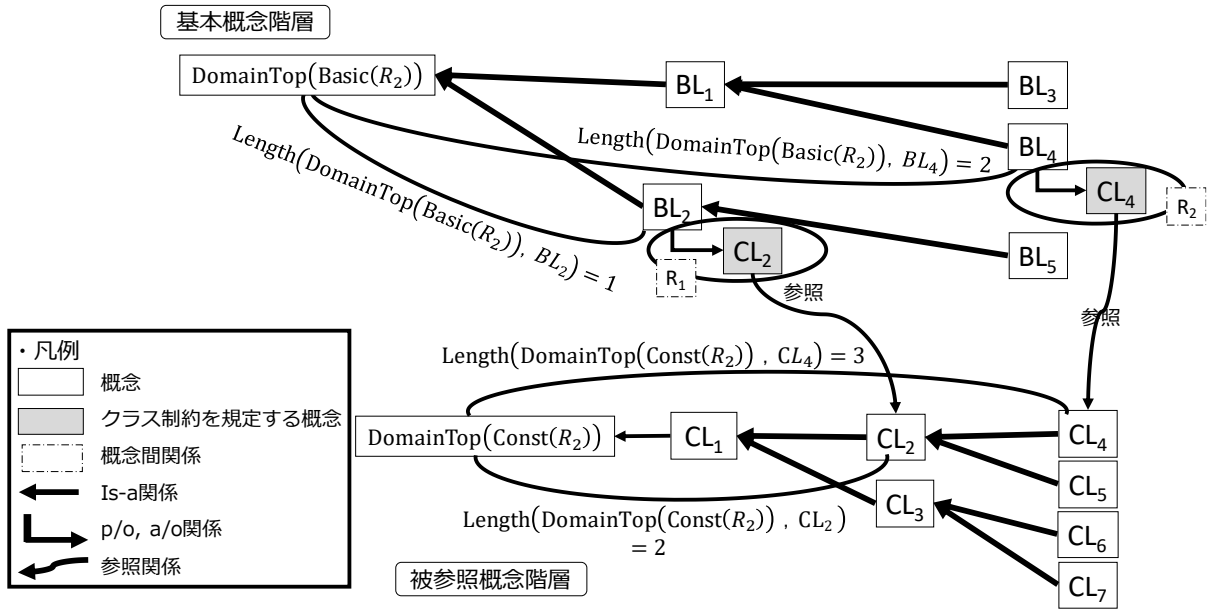

図 5.2.　最上位概念からの距離

$$\mathbf{BL_{cand}}(\mathrm{n}) = \{\mathrm{BL_a} | \mathrm{Length}\left(\mathrm{DomainTOP}\left(\mathrm{Basic}(R_n)\right),\ \mathrm{BL}_a\right) = l_{basic}(\mathrm{n})\} \qquad (5.4.)$$

と書ける．同様に，被参照概念階層についても，あるR_nから参照されている概念とDomainTOP(Const(R_n))の距離は，

$$l_{const}(n) = \mathrm{Length}\left(\mathrm{DomainTOP}\left(\mathrm{Const}(R_n)\right),\ \mathrm{Const}(R_n)\right)$$

と書ける．Const(R_n)と同じ距離にある$\mathbf{CL_n}$の部分集合は，

$$\mathbf{CL_{cand}}(\mathrm{n}) = \{\mathrm{CL_b} | \mathrm{Length}\left(\mathrm{DomainTOP}\left(\mathrm{Const}(R_n)\right),\ \mathrm{CL_b}\right) = l_{const}(\mathrm{n})\} \qquad (5.5.)$$

と書ける．この$\mathbf{BL_{cand}}$(n)，$\mathbf{CL_{cand}}$(n)の各要素を組み合わせ候補として抽出する．

例えば，図 5.2.のR_2の場合，基本概念Basic(R_2) = BL_4とその最上位概念DomainTop(Basic(R_2))との距離は，

$$l_{basic}(2) = \mathrm{Length}\left(\mathrm{DomainTop}\left(\mathrm{Basic}(R_2)\right),\ BL_4\right) = 2$$

なので，

$$\mathrm{Length}\left(\mathrm{DomainTop}\left(\mathrm{Basic}(R_2)\right),\ BL_3\right) = 2$$
$$\mathrm{Length}\left(\mathrm{DomainTop}\left(\mathrm{Basic}(R_2)\right),\ BL_5\right) = 2$$

である$\{BL_3,\ BL_5\}$が組み合わせ候補となる．被参照概念階層では，Const(R_2) = CL_4とDomainTOP(Const(R_2))の距離は，

$$l_{const}(2) = \mathrm{Length}\left(\mathrm{DomainTOP}\left(\mathrm{Const}(R_2)\right),\ \mathrm{Const}(R_2)\right) = 3$$

なので，

$$\mathrm{Length}\left(\mathrm{DomainTOP}\left(\mathrm{Const}(R_2)\right),\ CL_5\right) = 3$$
$$\mathrm{Length}\left(\mathrm{DomainTOP}\left(\mathrm{Const}(R_2)\right),\ CL_6\right) = 3$$
$$\mathrm{Length}\left(\mathrm{DomainTOP}\left(\mathrm{Const}(R_2)\right),\ CL_7\right) = 3$$

表 5.1.　同じ段数のある概念に限定した場合の組み合わせ数

	薬物/蛋白質の階層における概念数		作用の階層における概念数		取りうる組み合わせの総数	
	概念階層全体	同じ段数に限定	概念階層全体	同じ段数に限定	概念階層全体	同じ段数に限定
薬物	142	35	96	62	13,632	2,170
蛋白質	26	11	96	72	2,496	792

である$\{CL_5,\ CL_6,\ CL_7\}$が組み合わせ候補となる．この概念階層の比較によって抽出された$\mathbf{BL_{cand}}(2) = \{BL_3,\ BL_5\}$と$\mathbf{CL_{cand}}(2) = \{CL_5,\ CL_6,\ CL_7\}$の各要素の組み合わせが，概念間関係 R_2 と同様の概念間関係を結ぶ組み合わせ候補である．

表 5.1.に“薬物”と“薬理作用”，“蛋白質”と“作用”の階層に含まれる“Function-of”を持つ概念間関係をもとに，組み合わせ候補を抽出した結果を示す．薬物の階層中の解熱鎮痛薬と同じ段数にある概念は 35 個，薬理作用の階層中では，解熱作用と同じ段数にある概念は 62 個と限定することができた．その結果，取りうる組み合わせ数は, 既存の概念間関係を含む概念階層全体では, 142*96=13,632 個だったものが，概念間関係を持つ概念と同じ階層に限定した場合，35*62=2,170 個となった．また，薬物と薬理作用の概念階層の比較に加え蛋白質とその作用の概念階層で同様の概念階層比較による組み合わせ候補の抽出を行った．この場合も同様にオントロジー構築者に提示する組み合わせ候補数を 2,496 から 792 に減らすことができた．

5.2.3　組み合わせ候補の例

薬物と薬理作用の組み合わせ候補の中で適切である例を表 5.2.に示す．表 5.2.中で既存の関係列が矢印となっているものは，オントロジー中に既存の概念間関係である．

表 5.2.　薬物と薬理作用の組み合わせ候補の中で適切であった例

薬物の候補		既存の関係	薬理作用の候補	
1	解熱鎮痛薬	→	1	解熱作用
2	コリン作動薬	→	2	ニコチン作用
		→	3	ムスカリン作用
3	冠拡張薬	→	4	カルシウム拮抗作用
4	マクロライド系抗生物質		5	抗微生物作用
5	脂質低下薬		6	脂質低下作用
6	散瞳薬		7	散瞳作用
7	縮瞳薬		8	縮瞳作用
8	抗精神病薬		9	抗鬱作用
…	…		…	…

それぞれの組み合わせ候補中の概念を見ると，概念間関係が結ばれていない，“マクロライド系抗生物質”と“抗微生物作用”，“脂質低下薬”と“脂質低下作用”等の，概念間関係を追加するのに適当な概念の組み合わせがあることが確認できる．また，この組み合わせ候補を限定した結果をドメインの専門家であるバイオサイエンスオントロジーの構築者に提示したところ，新規の概念間関係の追加を考える際に十分な示唆を与えるものであると評価された．

組み合わせ候補となる概念を既存の概念間関係R_nのBasic(R_n)，Const(R_n)と同じ距離にある概念に限定することで，オントロジー構築者が考慮しなければならない取りうる概念の組み合わせ数を減らすことが可能となった．しかしながら，この組み合わせ候補は，35 個ある薬物毎に 62 個の作用が候補として提案される．オントロジー構築作業の効率化のためには，依然として考慮すべき組み合わせ数が多い．そこで，概念階層の比較によって抽出された組み合わせ候補を順位付けする方法を考える．

5.3 Linked Open Data 中の共起情報を利用した順位付け

本節では，拡張手法のステップ 2 である，組み合わせ候補の抽出結果の順位付け手法について述べる．順位付けには，Linked Open Data（LOD）として Web 上に公開されているライフサイエンス分野の論文集の単語の共起情報が含まれるデータベースを用いる．この順位付けにより，オントロジー構築者に対して構築作業時に組み合わせ候補の中から概念間関係を結ぶべきものを優先的に提示できるようになるため，より効率的な概念間関係の追加の判断が可能となる．

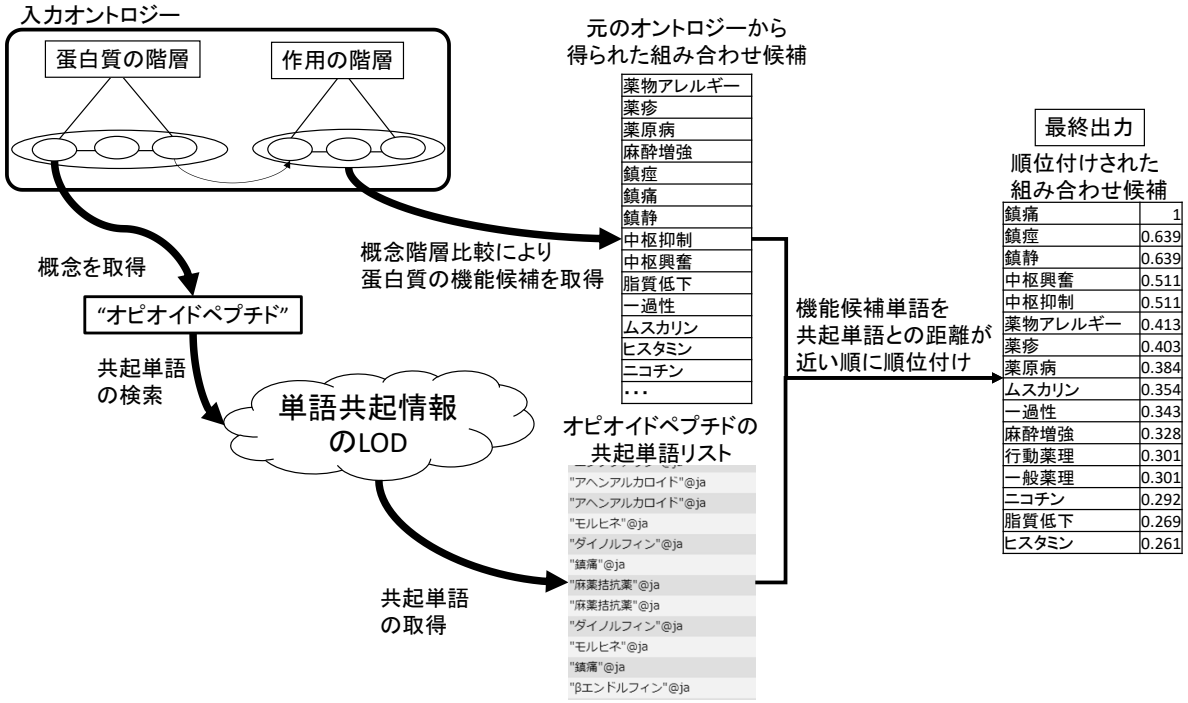

図 5.3. 拡張支援手法の概要

5.3.1　共起情報を利用した順位付けの流れ

共起情報を利用した順位付けの手順は以下のとおりである．

1　対象オントロジーから概念間関係を抽出.
2　抽出された概念間関係の要素R_n毎に，基本概念階層から組み合わせ候補$\mathbf{BL_{cand}}(n)$，被参照概念階層から組み合わせ候補$\mathbf{CL_{cand}}(n)$を抽出.
3　以下の手順を$\mathbf{BL_{cand}}(n)$の要素BL_aごとに繰り返す.
　3.1　単語共起情報の LOD から，BL_aと共起する単語の集合を取得.
　3.2　$\mathbf{CL_{cand}}(n)$とBL_a共起単語群間の全組み合わせで単語間距離を計算.
　3.3　単語間距離のスコアが高い順に$\mathbf{CL_{cand}}(n)$を並べ替え，提示.

以上の手順を，概念間関係R_n毎に繰り返し実施する．

図 5.3.は，蛋白質・ペプチドの一つである“オピオイドペプチド”と Function-of 関係を結ぶべき概念の推定の具体的な流れを示したものである．まず，バイオサイエンスオントロジーから概念間関係の集合を抽出する. 概念間関係の一つ Function-of 関係に関して，$\mathbf{BL_{cand}}(n)$，$\mathbf{CL_{cand}}(n)$を抽出する．$\mathbf{BL_{cand}}(n)$の要素BL_aとして“オピオイドペプチド”に着目し，ライフサイエンス分野の単語共起情報の LOD から“オピオイドペプチド”と共起する単語の集合を取得する．取得した“オピオイドペプチド”との共起単語群間と$\mathbf{CL_{cand}}(n)$の全組み合わせで単語間距離を計算する. 最後に単語間距離のスコアが高い順に$\mathbf{CL_{cand}}(n)$を並べ替え，提示する．

5.3.2　Linked Open Data を利用した共起単語の取得

ライフサイエンス論文中の単語の共起情報を利用した組み合わせ候補単語の順位付け手法について述べる．組み合わせ候補の順位付けのために，Web 上に公開されているデータリソースの利用を考える．本章で利用したのは，Life Science Dictionary

表 5.3.　解熱鎮痛薬と共起する単語上位 25 位

アセトアミノフェン	治療
アスピリン	治療法
アセトアミノフェン	シクロオキシゲナーゼ
発熱	治療用途
抗炎症剤	シクロオキシゲナーゼ
非ステロイド性抗炎症薬	頭痛
鎮痛薬	糖質コルチコイド
シクロオキシゲナーゼ	シクロオキシゲナーゼ
鎮痛薬	抗菌薬
インドメタシン	炎症
プロスタグランジン	筋肉痛
サリチル酸塩	血小板凝集抑制薬
イブプロフェン	

（LSD）Resource Description Framework（RDF）Data Portal[23][LSD-RDF]である．これは，ライフサイエンス辞書（LSD）を RDF と呼ばれるデータ記述方法で表現したデータベースであり，LOD として公開されている．LSD は，PubMed[24][PubMed]において公開されている論文抄録を解析し，Medical Subject Headings（MeSH）[25][MeSH]に準拠して類義語を整理したシソーラスである．PubMed は，米国国立医学図書館（United States National Library of Medicine，（NLM））が，提供する医療分野の文献情報データベースである．このデータベースには，2700 万件以上の文献データが収録されており，タイトル，著者名，雑誌名，抄録などの文献の書誌情報の検索が可能である．MeSH は，NLM が提供している生命科学シソーラスであり，PubMed のデータベースを管理する際に用いる用語を規定している．

本研究では，LSD RDF Data Portal を利用し，概念階層比較によって得られた$\mathbf{BL_{cand}}(\mathrm{n})$の要素$BL_a$と共起する用語の集合を MeSH から取得する．共起単語は，LOD 検索の際に単語の出現頻度と逆文書頻度の 2 つの指標に基づいて計算される，tf-idf 順の上位 50 件を利用する．BL_aと共起する用語の集合を$\mathbf{BC}_{BL_a}$とする．表 5.3.に“解熱鎮痛薬”と共起する用語の上位 25 件を示す．これらの共起語には，“鎮痛薬”，“発熱”等の，“解熱鎮痛薬”が持つべき“鎮痛”，“解熱”などの作用と部分一致する語の存在が確認できる．

5.3.3　共起単語を利用した組み合わせ候補の順位付け

次に共起単語の集合$\mathbf{BC}_{BL_a}$を利用した$\mathbf{CL_{cand}}(\mathrm{n})$の順位付け手法について述べる．$\mathbf{CL_{cand}}(\mathrm{n})$の順位付けには，単語間距離を用い，共起単語と近い単語に高いスコアを付ける．なぜなら，共起単語には，“解熱鎮痛薬”の作用として適切な概念のラベルと部

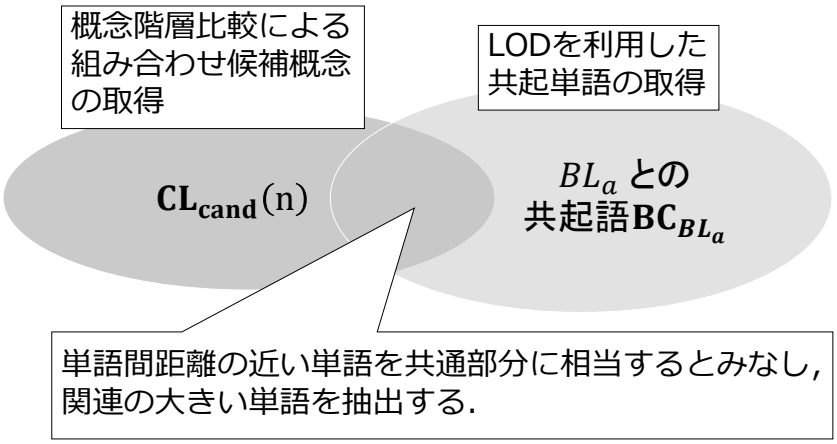

図 5.4.　組み合わせ候補と共起語の関係

[23] http://lsd.dbcls.jp/portal/
[24] https://www.ncbi.nlm.nih.gov/pubmed/
[25] https://www.nlm.nih.gov/mesh/

表 5.4.　解熱鎮痛薬の組み合わせ候補上位 10 位

組み合わせ候補	共起単語	一致度
鎮痛	鎮痛	1.00
鎮痙	鎮痛	0.64
鎮静	鎮痛	0.64
体温効果	体温	0.62
ムスカリン	アスピリン	0.61
体温上昇薬	体温	0.55
中枢興奮	中枢神経系	0.51
中枢抑制	中枢神経系	0.51
抗鬱	抗菌薬	0.48
解熱	発熱	0.47

分一致する“鎮痛薬”，“発熱”等の語が含まれ，その一致度が高い概念は，共起元になっている薬物や蛋白質との関連が大きいと考えられるからである．

図 5.4.に組み合わせ候補単語と LOD からの共起単語の関係を示す．$\mathbf{CL_{cand}}(\mathrm{n})$と$\mathbf{BC}_{BL_a}$の共通部分$\mathbf{CL_{cand}}(\mathrm{n}) \cap \mathbf{BC}_{BL_a}$を取得することで，

(作用の関連する概念の集合) ∩ (BL_aと共起する用語)

を得ることができる．$\mathbf{CL_{cand}}(\mathrm{n})$のうち，この共通部分に含まれる概念は，$BL_a$と概念間関係を結ぶべきである可能性が高いといえる．

この時，共通部分の取得に単語間距離を利用する．その理由は，組み合わせ候補概念と LOD 中の共起単語間での表記ゆれに対応するためである．この単語間距離の計算には，Java のライブラリである Apache Lucene[26]を利用し，レーベンシュタイン距離，ジャロウィンクラー距離，文字 Ngram 距離の平均値を単語間距離とした．

表 5.4.に解熱鎮痛薬の組み合わせ候補単語の順位付け結果の上位 10 位を示す．組み合わせ候補の最上位に既存の関係がある“鎮痛”が来ており，また，概念間関係はなかったが，新しく概念間関係を結ぶべきであると考えられる“解熱”も 10 位に推定された．解熱鎮痛薬の場合，その作用の組み合わせ候補は 62 概念であるので，順位付けによって，概念間関係を結ぶべき概念を上位候補として推定することによって構築作業の効率化に寄与するといえる．

5.4　バイオサイエンスオントロジーへの適用

本節では，これまで述べてきた拡張手法を実際に構築途中段階にあるオントロジーに対して適用した結果について述べる．本実験では，構築途中段階にあるオントロジーとしては，5.1.2.節で紹介した JST シソーラスを基に構築が進められているバイオサイエンスオントロジーを用いた．このオントロジーから，“Function-of”関係を結ぶべき既存の概念の組み合わせを推定し，その評価を行う．

[26] https://lucene.apache.org/

5.4.1 拡張支援システムの評価実験

本評価実験では，システムが推定する組み合わせ候補の順位付けについて評価した．本章の実験では，バイオサイエンスオントロジーの中でも蛋白質と薬物と“Function-of”の概念間関係を結ぶべき作用を適切に順位付けできているかを評価する．そこで実験にあたり，実際にバイオサイエンスオントロジーを構築しているドメインの専門家に正解データの作成を依頼した．正解データの作成範囲は，概念階層比較により検出された薬物 35 概念と薬理作用 62 概念間に結ぶべき関係と，蛋白質 11 概念と作用 72 概念間に結ぶべき関係である．この正解データとシステムによる推定結果を比較することで，本手法の推定結果の妥当性を評価する．

表 5.5.に実験結果を示す．まず 1 行目の蛋白質の結果である．蛋白質の場合，11 個ある蛋白質毎に 72 個の作用の候補が存在するため, 取り得る組み合わせは全部で 792 ある．その内，ドメインの専門家によって概念間関係を結ぶべきとされた組み合わせは 13 ある．LSD-RDF の共起情報を利用することによって，7 つの関係を各蛋白質による順位付けで 5 位以内に推定でき，30 位以内であれば 11 個の関係を推定できた．

次に，薬物についてであるが，こちらは組み合わせ候補数が多い分，正解の組み合わせ数も増えている．この場合は 60 個中 20 個を 5 位以内に推定することが可能であった．

拡張支援手法の評価実験で組み合わせ候補の推定を行った薬物 35 種のうち，薬物の 14 種には, LSD にエントリが存在しないため検索できず, 順位付けできなかった．具体例としては，“ペプチド系抗生物質”，“抗腫瘍抗生物質”などの抗生物質の下位概念と“硝酸化阻害剤”，“土壌殺菌剤”などの農業殺菌剤の下位概念である．農業殺菌剤は，ライフサイエンス論文を集めた PubMed では扱っていないため，今後，分野によって検索する LOD を適宜変更する必要がある．ただ薬物の場合も LSD 上にエンティティが存在し推定できたものに関しては，上位に適切な薬理作用を提案できることが確認できた．

本章で提案する拡張支援手法は，オントロジー構築者であるバイオサイエンスドメインの専門家に，既存の概念から概念間関係を結ぶべき組み合わせを抽出し提示することが目的である．そのため，明らかに関係を結ぶべきではない単語であれば，容易に判断可能であるので，recall を重視した評価となっている．

表 5.5.　概念の組み合わせ推定結果

	全組み合わせ候補数	正解組み合わせ数	5 位以内で推定された数	30 位以内で推定された数
蛋白質	11*72=792	13	7(53%)	11(85%)
薬物	35*62=2170	60	20(33%)	32(53%)

5.4.2　拡張システムによる概念間関係の追加提案例

一つ目の例として，先ほどの“解熱鎮痛薬”と同じく薬物の一つである“アンギオテンシン変換酵素阻害剤”と“Function-of”の関係を結ぶべき薬理作用の候補を順位付けした例を紹介する．“アンギオテンシン変換酵素阻害剤”（ACE 阻害剤）は，“解熱鎮痛薬”と同じ段数に存在する兄弟概念の一つである．概念階層比較によって抽出される組み合わせ候補の薬理作用の集合は，解熱鎮痛薬と同じ 62 個の概念である．一方，LOD から得られる共起単語群は，“ACE 阻害剤”との共起となるので異なる集合が得られる．表 5.6.に共起単語の一部を示す．この薬物は，末梢血管を拡張し，血圧を下げる作用を持つが，共起語の中にも“血圧”，“降圧薬”，“血管拡張”などの語が見られる．次に，表 5.7.に組み合わせ候補の単語を共起単語との一致度で順位付けを行ったものを示す．表中の灰色セルが正解とされた組み合わせである．組み合わせ候補の中で, ACE 阻害剤の作用として関係を結ぶべきとされた“血管拡張”と“降圧”が上位に推定されている. 一方“血管収縮”という逆の作用を示す語も上位であるが，この順位付けは，ドメインの専門家による概念間の関係の追加作業時の支援を目的に

表 5.6.　ACE 阻害剤と共起する単語

血圧
心筋梗塞
アンジオテンシン
死亡率
カルシウムチャンネル遮断薬
アンジオテンシン
降圧薬
降圧薬
プラセボ
レニン・アンジオテンシン
エナラプリル
腎疾患
血管拡張薬

表 5.7.　ACE 阻害剤の組み合わせ候補上位 10 位

組み合わせ候補	LOD 中の共起単語	一致度
血管拡張	血管拡張薬	0.85
降圧	降圧薬	0.75
利尿	利尿薬	0.75
カルシウム拮抗	カルシウムチャンネル遮断薬	0.57
血管収縮	血管拡張薬	0.51
昇圧	血圧	0.47
気管支拡張	血管拡張薬	0.44
抗不安	心不全	0.41
血糖降下	血圧	0.40
血糖上昇	血圧	0.40

表 5.8.　オピオイドペプチドと共起する単語

アヘンアルカロイド
アヘンアルカロイド
モルヒネ
ダイノルフィン
鎮痛
麻薬拮抗薬
麻薬拮抗薬
ダイノルフィン
モルヒネ
鎮痛
βエンドルフィン

表 5.9.　オピオイドペプチドの組み合わせ候補上位 10 位

組み合わせ候補	LOD 中の共起単語	一致度
鎮痛	鎮痛	1.00
鎮痙	鎮痛	0.64
鎮静	鎮痛	0.64
中枢興奮	中枢神経系	0.51
中枢抑制	中枢神経系	0.51
薬物アレルギー	薬物治療	0.41
薬疹	薬物治療	0.40
薬原病	薬物治療	0.38
ムスカリン	エンケファリン	0.35
一過性	痛覚過敏	0.34

しているため，この例のように明らかに関係を結ぶべきではない単語であれば，容易に判断可能であると考えられる．

次に蛋白質の一つである“オピオイドペプチド”と“Function-of”の関係を結ぶべき作用を順位付けした例を紹介する．“オピオイドペプチド”は，“バソプレシン”という蛋白質と同じ段数に存在する兄弟概念である．バイオサイエンスオントロジー上で“バソプレシン”は，

Relation(“バソプレシン”，“Function-of”，“昇圧作用”)

の概念間関係が定義されている．この“昇圧作用”には，72 の兄弟概念があり，概念階層比較によって組み合わせ候補として抽出される．“ACE 阻害剤”の時と同様に，“オピオイドペプチド”の共起単語を取得すると，表 5.8.の単語群が得られる．“オピオイドペプチド”は，“opium（阿片）”に由来する言葉であり，“モルヒネ様ペプチド”という意味で用いられる．ゆえに，“アルカノイド”や“モルヒネ”のような麻酔薬，“鎮痛”などの機能が共起単語として得られる．表 5.8.の共起単語と“オピオイドペプチド”の組み合わせ候補である 72 の概念を単語間距離で順位付けすると表 5.9.が得られる．組み合わせ候補の 1 位は，“鎮痛”であり，“オピオイドペプチド”と“Function-of”関係を結ぶべき作用として提案される．この組み合わせは，ドメインの専門家によって概念間関係を結ぶべき組み合わせであるとされたものである．

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，オントロジー構築の 4 段階のうち，「3.概念・関係の定義」及び「4.評価・洗練」段階にあるオントロジーに対して，概念・概念間関係の洗練及び拡張支援提案を行うシステムを開発した．本システムは，オントロジー構築の効率化に大きく寄与するものである．

本システムの実装に向けて開発した手法では，「分類観点の同一性」という概念定義の際に推奨される構築基準に着目した．概念の is-a 関係は，概念の同一性を決める性質が継承される時，つまり，概念を分類する際に着目した性質の一貫性が保たれている時に，許される関係である．そのため，各概念に対してオントロジー構築者がどのような性質で分類したかを，互いに矛盾せずに定義するためには，この「分類観点の同一性」を満たす概念階層を構築する必要がある．

本研究では，「分類観点の同一性」を向上させるオントロジーの洗練提案を行う支援手法を開発した．本手法の開発にあたり，この分類観点の同一性を満たすオントロジーの概念階層には，参照関係にある概念階層が互いに相似になるという性質を利用した．相似な概念階層の検出のために，同じラベルを持つ概念間関係によって参照関係にある概念階層同士を比較し，非相似であれば，互いに相似となるような概念の追加提案を行う洗練・拡張支援手法を開発した．

第 2 章では，まずオントロジーの概念階層を構成する「概念」と「概念間関係」を定義し，それぞれを定義する際に守らなければならない制約について述べ，オントロジー構築についての導入をした．さらに，洗練支援手法には，文法的エラー修正と内容洗練があることを示し，多くの研究やツール開発が進んでいる文法的エラー修正ではなく，本研究は，内容洗練を対象とした研究であることを示した．また，既存の内容洗練手法では洗練箇所と方法を，構築者自らが見つける必要があるという手法上の限界を明らかにし，本研究では自動的に洗練箇所と方法を提案する手法を開発することを述べた．さらに，3 章以降で共通に利用する，「分類観点の同一性」についての導入と，それを満たす概念階層に成り立つ「概念階層の相似性」について述べた．

第 3 章，4 章では，本研究で提案する概念階層の相似性を利用した洗練支援手法について述べた．本提案手法では，洗練箇所の検出に概念階層の相似性を利用し，参照関係にある概念階層同士を比較し，非相似であれば洗練対象箇所としている．その比較方法として，

- 一段の概念間関係の階層に着目する手法（第 3 章）
- 分類観点毎の包括的な概念関係階層に着目する手法（第 4 章）

の 2 種類を提案した．そのため，洗練支援手法を第 3 章と第 4 章の 2 章に分けて述べた．

第 3 章では，一段の概念間関係の階層に着目する手法について述べた．この概念階層の比較によって，参照関係にある概念階層の特殊化の粒度の一貫性が崩れている洗練箇所の検出が可能となった．さらに，そうした洗練箇所に対する洗練提案として，概念及び概念間関係を追加する方法について述べ，洗練箇所の分類と洗練方法の対応について述べた．さらに，本洗練手法を実装した洗練システムの動作について述べた．その結果，システムが検出した洗練箇所のうち，上位下位一段の概念間関係を基に概念間階層の比較を行う分類(a)では，その 90%に対して少なくとも一つの正しい洗練提案ができることを確認した．また，洗練提案毎の評価結果を見ると，約 50%が妥当な洗練提案とされた．この評価実験の結果から，本手法での「概念階層の相似性」に基づく，概念階層比較による自動的な洗練箇所の検出及び洗練方法の提案手法が，オントロジーの洗練手法に有効であることが確認できた．

第 4 章では，分類観点毎の包括的な概念関係階層に着目する手法について述べた．本手法では，同じ分類観点を持つ概念間関係階層全体を区切らずに，特殊化が行われている概念階層を比較する．この手法により，兄弟概念に対する洗練提案が可能となった．さらに，オントロジー構築者に提示される情報が包括的になり，より効率的に洗練提案の可否判断が可能となった．ここでは，比較対象の概念を次の 2 種類(i)同じ上位概念を持つ兄弟概念，(ii)特殊化の段数が同じ概念，に限定した．なぜなら，同じ上位概念を持つ概念や，最上位概念から同じ回数だけ特殊化された概念は，概念の特殊化が同程度の粒度となっているべきであるので，概念や概念間関係の追加先として適切であるからである．評価実験の結果，比較対象の概念の限定により，1 箇所当たりの提案数が，平均十数個から数十程度に抑えられた．また，実験用に無作為に抽出した洗練箇所の 89%に対して少なくとも一つの妥当な洗練提案がされると評価され，洗練提案の 51%が，適用が妥当であると評価された．さらに，第 3 章での手法で提案されていたものを第 4 章でのシステム上で提示したところ，包括的な概念階層を洗練システム上で提示した効果により，その評価が向上した．以上の結果より，概念間関係階層全体の包括的な概念階層比較が，洗練箇所の検出，提案，提示方法の面でオントロジー構築者にとって有用なものであることが確認できた．

第 5 章では，概念間関係の定義が不十分であるオントロジー構築の「概念・関係の定義」段階のオントロジーを対象とした拡張支援手法について述べた．本拡張支援手法は大きく 2 ステップに分けて行われる．1 ステップ目の組み合わせ候補の抽出では，概念間関係のある概念の兄弟概念は同程度特殊化されているという性質を利用し，それぞれの階層の最上位概念から等距離にある概念を組み合わせ候補とした．2 ステッ

プ目では，抽出された組み合わせ候補を順位付けする．Life Science Dictionary（LSD）-RDF から，基本概念階層中の組み合わせ候補の概念と共起する単語群を抽出し，被参照概念階層中の組み合わせ候補単語を共起単語群との単語間距離が近い順に並び替える．各ステップの手法を検証するために，ライフサイエンスの専門家に実際に必要とされる概念間関係のデータの作成を依頼し，本提案手法によって提示される候補と比較した．その結果，蛋白質・薬物ともに概念間関係を結ぶべき機能を上位に提示できることを確認した．

3 章~5 章までに述べてきた洗練及び拡張支援手法によってオントロジー構築者は，対象オントロジーから自動的に洗練箇所及び各箇所への洗練方法，または構築途中のオントロジーに追加すべき概念間関係の候補の提示を受けることが可能となった．ここで対象としているオントロジーは，構築の第 3，4 段階にあり，文法的なエラーは，既存研究の手法で修正されたものである．こうした支援によって，対象オントロジーの「分類観点の同一性」の品質向上作業を効率化することが可能となった．従来研究では，本研究で対象としている「分類観点の同一性」のような定義内容の品質向上案を自動的に提案するような試みは未だ存在せず，本研究の意義はそこにあるといえる．

6.2　残された課題と将来展望

本研究によって，オントロジーの「分類観点の同一性」に対する洗練・拡張提案がその相似性という性質を用いた概念階層比較によって可能であることが確認された．本節では，残された課題と将来展望について述べる．

残された課題としては，削除を含めた洗練提案の拡張，他の構築推奨基準への対応がある．前者は，洗練作業の利便性向上のための機能拡張であり，後者は「分類観点の同一性」が担保された後に考察可能な基準である．また，将来展望としては，拡張支援手法の他のドメインへの適用が考えられる．

まず，概念及び概念間関係の削除も含めた洗練提案の拡張について述べる．本研究での提案手法では，概念及び概念間関係の追加提案のみに限定して，手法の開発を進めた．なぜならオントロジー中に定義されている概念及び概念間関係は，多い方がオントロジーのカバーできる範囲が広がりより利用可能性が高くなると考えたからである．しかしながら，洗練提案には，概念や概念間関係の削除も当然考えられる．オントロジーの利用目的や構築したいシステムの要求仕様によっては，不必要な概念や関係はできるだけ削除されていることが望ましいということも十分に考えられるからである．

この点に関しては，提案手法の一部が利用可能であると考えている．今回の提案手法では，概念定義が詳細な方に合わせて，概念を追加する提案を行っていたが，同様の概念階層比較手法で，概念定義が疎な階層に特殊化の粒度を揃えるといった概念や概念間関係の削除が提案可能である．ただし，概念や概念間関係の削除は，比較対象

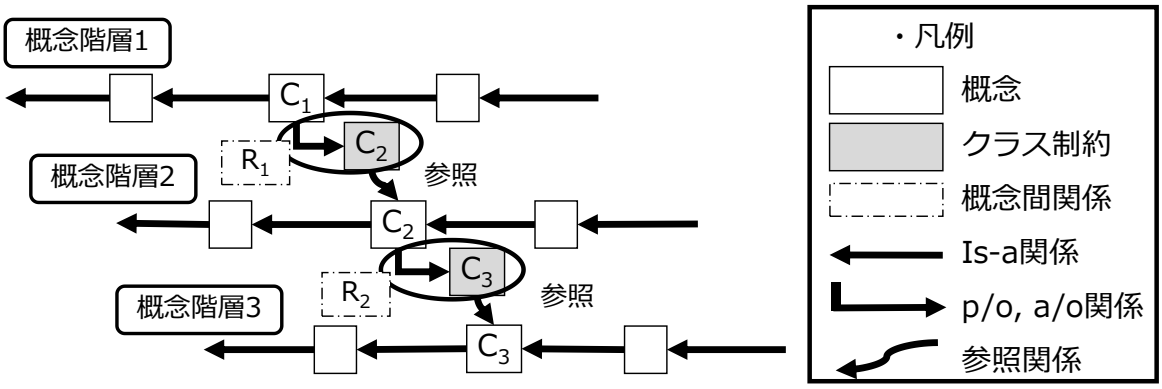

図 6.1.　概念の削除の影響

となっている箇所だけではなく，他の箇所への影響がある．例えば，ある基本概念階層と被参照概念階層を比較し，基本概念階層中の概念を削除した場合を考えてみる．図 6.1.の概念階層 2 と概念階層 3 を比較し，「概念階層 2 中の概念C_2を削除する」という提案を考える．概念C_2を削除すると，概念C_2に定義された概念間関係R_2も同時に削除される．この時，R_2から参照されている概念階層 3 中の概念C_3は，影響を受けない．一方，概念階層 1 中の概念C_1は，概念間関係R_1を持つ．そのクラス制約として参照されている概念C_2が削除されるため，概念C_2は，定義された概念の性質を失うことになる．このように概念の削除提案を洗練支援手法に組み込む場合，複数の概念への影響に関する考察を進めていく必要がある．

2 点目は，他の構築推奨基準への対応がある．オントロジーの構築推奨基準には，今回対象とした「分類観点の同一性」の他にも，「インスタンス集合のパーティション性」，「本質属性の体現」，「上下関係の一般性」等がある．本研究で対象とした「分類観点の同一性」は，他の基準を考慮する前に担保しておくべき基盤である．また，この「分類観点の同一性」は，これまで困難であるとされてきたドメイン非依存の自動的な洗練提案を実現する基準である．本研究では，様々なオントロジーに対しても一定の成果を挙げられたことに意義がある．今後，「分類観点の同一性」が担保されたオントロジーに対して，その他の推奨基準を適用可能か，考察を進める必要がある．

3 点目は，拡張支援手法の他ドメインへの適用である．今回，拡張対象としたオントロジーは，JST シソーラスを基にしたバイオサイエンスオントロジーの薬物・蛋白質に関する概念階層であった．しかしながら，バイオサイエンスオントロジーには，他にも生物，細胞，植物，遺伝子などの約 50 の階層が存在し，そうした概念にも今回対象とした機能を表す“Function-of”の関係を結ぶべき概念が存在するため，これらの概念階層に対しても利用可能であることが期待できる．こうした概念階層に対しても本研究で提案した手法を適用することで，JST シソーラスのオントロジー化の効率化に貢献できると考えている．

しかしながら，本拡張支援では，概念と is-a 関係といくらかの概念間関係が定義されているオントロジーを構築途中段階として拡張対象としている．そのため，既存の概念間関係が存在しないオントロジー，is-a 関係が定義されていない段階のオントロジーには対応できない．こうしたオントロジーに対しては，Wikipedia などの半構造化情報を用いたオントロジー学習などの研究成果が利用可能である．こうした研究成果と本研究の成果を併せ，より広い範囲の構築段階のオントロジーに対する拡張支援手法の開発に繋げることが考えられる．

また，バイオサイエンスオントロジー以外にも構築途中のオントロジーは，無数に存在している．そうしたオントロジーに対しても利用可能となるように，「法造」などのオントロジー構築ツールに本研究で開発した洗練・拡張システムを実装し，公開することで，人的・時間的コストを要求するオントロジー構築の効率化に貢献すると期待できる．

参考文献

[Ashburner 00] Ashburner, M., Gene Ontology: Tool for the Unification of Biology, Nature Genetics, Vol. 25, pp. 25-29, (2000).

[Bizer 09] Bizer, C., The Emerging Web of Linked Data, IEEE Intelligent Systems, Vol. 24, No. 5, pp. 87-92, (2009).

[Denaux 12] Denaux, R., Thakker, D., Dimitrova, V., and Cohn, A. G., Interactive Semantic Feedback for Intuitive Ontology Authoring, In Proceedings of the 7th International Conference on Formal Ontology in Information Systems, Graz, Austria, (2012)

[Fernandez 99] Fernandez-Lopez, M., Gomez-Perez, A., Sierra, J. P., and Sierra, A. P., Building a Chemical Ontology Using Methontology and the Ontology Design Environment, IEEE Intelligent Systems, Vol. 14, pp. 37-46, (1999).

[福田 07] 福田賢一郎, 生命科学におけるオントロジーとその利用(<特集>バイオインフォマティクスと人工知能の新たなインタラクション), 人工知能学会誌, Vol. 22, No. 1, pp. 70-76, (2007).

[Fleiss 71] Fleiss, J. L., Measuring Nominal Scale Agreement Among Many Raters, Psychological Bulletin, Vol. 76, No. 5, pp. 378-382, (1971).

[Grenon 04] Grenon, P., Smith B., and Goldberg L., Biodynamic Ontology: Applying BFO in the Bio-Medical Domain, Studies in Health Technology and Informatics, 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).

[Gruninger 95] Gruninger, M. and Fox, M. S., Methodology for the Design and Evaluation of Ontologies, In Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing, Montreal, Canada, (1995).

[Guarino 98] Guarino, N., Some Ontological Principles for Designing Upper Level Lexical Resources, In Proceeding of the First International Conference on Language Resources and Evaluation, Granada, Spain, (1998).

[Guarino09] Guarino, N. and Welty, C. A., An Overview of Ontoclean, Handbook on Ontologies, pp. 201-220, (2009).

[廣田 09] 廣田健, 古崎晃司, 溝口理一郎, ドメイン知識俯瞰のためのオントロジー探索ツールの開発, 第 23 回人工知能学会全国大会論文集, No. 2I3-2, (2009).

[Hoffart 13] Hoffart, J., Suchanek, F., Berberich, K., and Weikum, G., YAGO2: A Spatially and Temporally Enhanced Knowledge Base from Wikipedia, Artificial Intelligence, Vol. 194, pp. 28-61, (2013).

[Horridge 11] Horridge, M., Justification Based Explanation in Ontologies, PhD Thesis, University of Manchester, (2011).

[法造] 大阪大学産業科学研究所，法造，http://www.hozo.jp/hozo/，(last accessed on 2018/01/01).

[ICD] World Health Organization (WHO), International Classification of Diseases (ICD), http://www.who.int/classifications/icd/en/, (Last accessed on 2018/01/01).

[Imai 16] Imai, T., Shinohara, E., Kajino, M., Sakurai, R., Ohe, K., Kozaki, K., and Mizoguchi, R., An Ontological Framework for Representing Topological Information in Human Anatomy, In Proceedings of the International Conference on Biomedical Ontology and BioCreative, Corvallis, USA, (2016).

[神崎 04] 神崎正英，セマンティックウェブのための RDF/OWL 入門，森北出版株式会社, (2004).

[古崎 02] 古崎晃司，來村徳信，池田満，溝口理一郎，「ロール」および「関係」に関する基礎的考察に基づくオントロジー記述環境の開発，人工知能学会論文誌, Vol. 17, No. 3, pp. 196-208, (2002).

[古崎 16] 古崎晃司，來村徳信，溝口理一郎，生物規範工学オントロジーと Linked Data に基づくキーワード検索，人工知能学会論文誌, Vol. 31, No. 1, p.LOD-A_1-15, (2016).

[Kumazawa 09] Kumazawa, T., Saito, O., Kozaki, K., Matsui, T., and Mizoguchi, R., Toward Knowledge Structuring of Sustainability Science Based on Ontology Engineering, Sustainability Science, Vol. 4, No. 1, p. 99, (2009).

[櫛田 17] Kushida, T., Kozaki, K., Tateisi, Y., Watanabe, K., Masuda, T., Matsumura, K., Kawamura, T., and Takagi, T., Efficient Construction of a New Ontology for Life Sciences by Sub-classifying Related Terms in the Japan Science and Technology Agency Thesaurus, In Proceedings of the International Conference on Biomedical Ontology, Newcastle, UK, (2017).

[LSD-RDF] Database Center for Life Science, Life Science Dictionary-Resource Description Framework (LSD-RDF), http://lsd.dbcls.jp/portal/, (last accessed on 2018/01/01).

[MeSH] The National Library of Medicine (NLM), Medical Subject Headings (MeSH), http://www.nlm.nih.gov/mesh/meshhome.html, (last accessed on 2018/01/01), (1999).

[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).

[溝口 06] 溝口理一郎, オントロジー構築入門, オーム社, (2006).

[Mizoguchi 10] Mizoguchi, R., YAMATO: Yet Another More Advanced Top-level Ontology, In Proceedings of the Sixth Australasian Ontology Workshop, Adelaide, Australia, (2010).

[西村 17] 西村悟史, 大谷博, 畠山直人, 長谷部希恵子, 福田賢一郎, 來村徳信, 溝口理一郎, 西村拓一, 現場主体の“知識発現”方法の提案, 人工知能学会論文誌, Vol. 32, No. 4, p. C-G95_1-15, (2017).

[Noy 01a] Noy, N. F., and McGuinness, D. L., Ontology Development 101: A Guide to Creating Your First Ontology, Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI, (2001).

[Noy 01b] Noy, N. F., Sintek, M., Decker, S., Crubezy, M., Fergerson, R. W., and Musen, M. A., Creating Semantic Web Contents with Protege-2000, IEEE Intelligent Systems, Vol. 16, No. 2, pp. 60-71, (2001).

[OBO] The OBO Foundry, The Open Biological and Biomedical Ontologies (OBO), http://www.obofoundry.org/, (last accessed on 2018/01/01).

[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).

[Ohta 11b] Ohta, M., Kozaki, K., and Mizoguchi, R., A Quality Assurance Framework for Ontology Construction and Refinement, In Proceedings of the Atlantic Web Intelligece Conference, Fribourg, Switzerland, (2011).

[岡田 02] 岡田光弘, オントロジーの哲学的・論理学的背景, レクチャーシリーズ「哲学と AI における対象世界モデリング」〔第 1 ~ 4 回〕人工知能学会誌, Vol. 17, No. 2, 3, 4, 5, (2002).

[Osborne 15] Osborne, F., and Motta, E., Klink-2: Integrating Multiple Web Sources to Generate Semantic Topic Networks, In Proceedings of the International Semantic Web Conference, Bethlehem, Pennsylvania, USA, (2015).

[OWL] W3C, OWL Web Ontology Language Reference, http://www.w3.org/TR/owl-ref/, (last accessed on 2018/01/01).

[Poveda 12] Poveda-Villalon, M., Suarez-Figueroa, M. C., and Gomez-Perez, A., Validating Ontologies with OOPS!, Knowledge Engineering and Knowledge Management Lecture Notes in Computer Science, Vol. 7603, pp. 267-281, (2012).

[Poveda 16] Poveda-Villalón, M.P., Ontology Evaluation: A Pitfall-Based Approach to Ontology Diagnosis. Ph.D. Thesis, UPM Madrid, (2016).

[PubMed] National Institutes of Health's National Library of Medicine (NIH/NLM), PubMed, http://www.ncbi.nlm.nih.gov/pubmed, (last accessed on 2018/01/01).

[Sirin .07] Sirin, E., Parsia, B., Grau, B. C., Kalyanpur, A., and Katz, Y., Pellet: A Practical OWL-DL Reasoner, Web Semantics: Science, Services and Agents on the World Wide Web, Vol. 5, No. 2, pp. 51-53, (2007).

[Smith 07] Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L. J., Eilbeck, K., Ireland, A., Mungall, C. J., Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, S. A., Scheuermann, R. H., Shah, N., Whetzel, P. L., and Lewis, S., 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, https://www.snomed.org/snomed-ct, (last accessed on 2018/01/01).

[Staab 01] Staab, S., Schnurr, H. P., Studer, R., and Sure, Y., Knowledge Processes and Ontologies, IEEE Intelligent Systems, Vol. 16, pp. 26-34, (2001)

[Suchanek 07] Suchanek, F. M., Kasneci, G., and Weikum, G., YAGO - A Core of Semantic Knowledge Unifying WordNet and Wikipedia, In Proceedings of the 16th International Conference on World Wide Web Alberta, Canada, (2007).

[Suchanek 08] Suchanek, F. M., Kasneci, G., and Weikum, G., Yago: A Large Ontology from Wikipedia and Wordnet, Web Semantics: Science, Services and Agents on the World Wide Web, Vol. 6, No. 3, pp. 203-217, (2008).

[玉川 10] 玉川奨, 桜井慎弥, 手島拓也, 森田武史, 和泉憲明, 山口高平, 日本語 Wikipedia からの大規模オントロジー学習, 人工知能学会論文誌, Vol. 25, No. 5, pp. 623-636, (2010).

[玉川 11] 玉川奨, 森田武史, 山口高平, 日本語 Wikipedia からプロパティを備えたオントロジーの構築, 人工知能学会論文誌, Vol. 26, No. 4, pp. 504-517, (2011).

[Uschold 96] Uschold, M. and Gruninger, M., Ontologies: Principles, Methods and Applications, Knowledge Engineering Review, Vol. 11, No. 2, pp. 93-155, (1996).

[W3C] World Wide Web Consortium (W3C), https://www.w3.org/2001/sw/, (last accessed on 2018/01/01).

[Wu 08] Wu, F., and Weld, D. S., Automatically Refining the Wikipedia Infobox Ontology, In Proceedings of the 17th International Conference on World Wide Web, Beijing, China, (2008).

[Zhao 15] Zhao, L., Ichise, R., Mita, S., and Sasaki, Y., Core Ontologies for Safe Autonomous Driving, In Proceedings of the International Semantic Web Conference, Bethlehem, Pennsylvania, USA, (2015).

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 駒谷和範教授に心より厚く感謝の意を表します.

本研究に関して貴重な御教示を賜り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賜りました大阪大学産業科学研究所 古崎晃司准教授に心より深く御礼申し上げます.

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井上恭教授，三瓶政一教授，滝根哲哉教授，馬場口登教授，丸田章博教授，宮地充子教授に心より御礼申し上げます.

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心より深厚なる謝意を表します.

本研究の評価実験の遂行にあたり，実験への参加を快諾してくださり，貴重なお時間を割いてくださりましたオントロジー構築研究者の皆様及び大阪市立大学瀬田研究室の学生諸氏に心より感謝申し上げます.

また，本研究で提案した，拡張支援システムの実践の場として JST シソーラスに基づくバイオサイエンスオントロジーを御提供頂き，本研究に対する貴重な御助言を賜りました. バイオサイエンスデータベースセンター 櫛田達矢様，建石由佳様，科学技術振興機構 川村隆浩様，渡邊勝太郎様に心より厚く御礼申し上げます.

Visiting researcher として暖かく迎えてくださり，オントロジー工学に関する研究及び実践的な構築研究を通し，オントロジー洗練支援手法に対する御教示を賜りました Bradford 大学 Dhaval Thakker 准教授に心より感謝の意を表します.

また，Bradford 大学への短期滞在を支援してくださり，5 年間のプログラムにおいても，丁寧かつ熱心な御指導，御支援を賜りました大阪大学超域イノベーション博士課程プログラムの先生方，事務職員の方々，学生諸氏に心より深く感謝致します.

そして，本研究のオントロジー工学理論において，貴重な討論と御助言を賜りました，立命館大学 來村徳信教授に心より御礼申し上げます.

さらに，貴重な討論，激励を賜り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心より厚く感謝致します．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 谷端紀久子氏，本薗千鶴子氏に心より感謝致します.

本研究を進める中，公私にわたり御支援・御協力を頂き，常に有益な御助言を賜りました，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心より感謝申し上げます.