

# 日本語動詞に関する EDR 概念辞書の分析

加藤 恒昭<sup>1,a)</sup> 林 良彦<sup>2</sup>

**概要:** 含意認識やその他の推論に用いることができる語彙知識の構築についての検討の一環として、EDR 概念辞書の日本語動詞に関連する部分の構造を分析した。動詞は、その間に包摂関係、自他対応、アスペクト対応等の意味関係があり、語彙知識の有用性が大きいと考えている。分析を通じて、日本語動詞が構成する概念体系と英語動詞が構成する概念体系の交わりが比較的少ないこと、サ変動詞の意味の扱いに関連して、動詞の意味記述としては扱いにくい点があることが明らかになった。これを受けて、日本語動詞に関する語彙知識として利用しやすい部分を抜き出し、その部分と日本語 WordNet との対応づけることで、ふたつの語彙知識の融合の可能性を検討した。結果、ふたつの語彙知識の骨格となる概念体系の上位構造に考慮すべき違いがあるという感触を得た。

## 1. はじめに

含意認識やその他の推論に用いることができる語彙知識の構築を検討している。特に動詞は、語と語の間に、包摂関係、自他対応、アスペクト対応等、様々な意味関係があり、それらの系統性に配慮して、適切な表現を与えることで、推論での幅広い応用が期待できる。そのような意味関係を記述をした言語知識（語彙的オントロジ）は幾つか構築されているが、それぞれは必ずしも充分なものではなく、複数の言語資源の融合等が求められる。一方、これらの言語資源はそれぞれに設計方針を持つので、利用や融合のためには、それらの方針を十分に理解することも重要である。本稿では、このような言語知識構築の事前検討として、既存の言語知識の中でも比較的よく利用されている EDR 電子化辞書<sup>\*1</sup>の概念辞書について分析を行った結果を報告する。

EDR 電子化辞書 [6], [7], [10] は、複数の辞書の集合体で、日本語英語各 20 万語に約 40 万の概念が単語辞書によって関係づけられ、概念辞書で、それらの概念が包摂関係<sup>\*2</sup>に基づいて構造化（体系化）されることで、語彙的オントロジを構成している。概念辞書は包摂関係に加えて選択制約に相当する概念記述を含み、対訳辞書や共起辞書等にも様々

な語彙知識が記述されているが、本稿では、主に、単語と概念との関係づけ、包摂関係による概念の体系に着目する。

本稿では、まず第 2 節で、英語と日本語の単語に共通の概念の体系を関係づけている、語の品詞をまたがって共通の概念の体系を用いているという EDR 辞書（以下、誤解を招かない範囲で、概念辞書、日英単語辞書を EDR 辞書と呼ぶ）の特徴を主な原因とすると思われる概念体系の問題があることを述べ、日本語動詞のために扱いやすい体系（包摂関係の階層構造）を抽出し、その特徴を述べる。次に 3 節で、その体系と日本語 WordNet<sup>\*3</sup>の体系とを対応づけ、両者の違いを考察する。対応づけでは、統計的機械翻訳における単語アライメントを利用した手法を提案する。4 節で全体をまとめる。

## 2. 概念体系の分析

EDR 辞書の大きな特徴は日本語と英語をまたがる語彙的オントロジであるという点である。日本語と英語の単語を共通の概念の体系で関連づけるという発想は斬新で、その後の EuroWordNet [12] にも繋がると思われるが、その一方で日本語の語に関する語彙知識として捉えた場合、不必要な概念や包摂関係が数多く存在し、必ずしも見通しのよいものとは言いがたい。しかも、日本語英語の関連づけの試みが成功しているかについては、この後示すような理由から若干疑問が残る。

本節では、動詞と関連づけられた EDR 辞書の部分について、まず、日本語と英語の関連づけの現状を報告する。次に日本語の語彙的オントロジとして適当な部分を取り出

<sup>1</sup> 東京大学  
The University of Tokyo

<sup>2</sup> 早稲田大学  
Waseda University

a) kato@boz.c.u-tokyo.ac.jp

\*1 入手先 ([https://www2.nict.go.jp/out-promotion/techtransfer/EDR/J\\_index.html](https://www2.nict.go.jp/out-promotion/techtransfer/EDR/J_index.html))

\*2 文献では上下関係、kind-of 関係と呼ばれているが、本稿では包摂関係と呼ぶ。

\*3 入手先 (<http://compling.hss.ntu.edu.sg/wnja/>)

表 1 EDR 辞書の単語と概念の数

	語 (表記)	語義 (概念)	レコード (語と語義の対)
日本語動詞	25,416	17,918	52,792
英語動詞	12,727	28,320	37,605

して概観し、日本語動詞の語彙知識としての EDR 辞書の性格を考察する。

それぞれの言語の単語とその語義となっている概念の数を表 1 に示す。日本語について EDR 辞書は表記の揺れという考え方をとらず、ある表記を持つ語とその語義である概念が直接関係づけられて、ひとつのレコードとなっている。ここではそれぞれの表記を一語と数えて、その異なり数を語数としている\*4。なお、EDR 辞書では、サ変動詞化する名詞 (例えば「勉強」) は辞書項目に含まれず、それに「する」をつけた表記が、日本語動詞 (JVE) と日本語一般名詞 (JN1) の品詞を持つ語として登録されている。今回の検討では、これらは動詞に含めていない。英語については、一般的な辞書でひとつの語として扱われる項目が活用形と文法情報毎に異なるレコードとなっているが、ここでは活用形が英語動詞原型 (EVB)、英語動詞語幹 (EVSTM) のもののみを取り出し、文法情報の異なりは無視して、同一表記、同一品詞、同一概念のものをひとつずつとりだして、表記の異なり数を語数としている。

## 2.1 概念体系の問題点

(日本語) 動詞に関連づけられた概念の殆どは包摂階層において [事象] の下位に位置する (以下では概念を [] で囲って表現する)。ここで [事象] は、EDR における最上位 (包摂階層の根となる) の概念 [概念] の娘で、[物事] [位置] [時] 等の姉妹である。日本語動詞に関連づけられた概念でこの位置にないもの、つまり包摂階層を上位に辿っていて [事象] に到達しないものは 126 件であった。EDR 辞書は包摂階層の多重継承を許しているため包摂階層で複数の母を持つ場合があるが、そのうちのひとつが [事象] に到達しない場合もこの数に含めている。

これらについて分析したところ、20 件は上位の概念が指定されていない、あるいは誤った記述がされている (「色めき立つ」の語義のひとつ (0e5450) の上位概念が [嚙窩] となっている等) という作業上のミスと思われるものであった。

残りの大多数で見られるのは、「売り出す」に [販売のやり方からみた販売] [商法]、「縫う」に [縫いの種類でみた縫うこと] [縫い方]、「冴える」に [人間の能力・才能] [力] の語義を与えるようなものである。これらには、次の方針が関連すると思われる。

\*4 同じ表記で同じ概念識別子を持つが、文法情報が異なるために複数のレコードとなっている場合もあるが、その数は英語に較べると少ない。

前述のように、EDR 辞書はサ変動詞化する名詞を辞書項目としておらず、例えば、「戦争」は含まれず「戦争する」だけが存在し、その語義は [事象] を上位とする [行為] と、[物事] を上位とする [事柄] や [出来事] との 2 面性を持つ。これは設計方針と思われ、サ変動詞についての言及はないが、文献 [6] では、属性概念で修飾された概念ということで、「美人」が持つ、美しい人間と、美しい人であるという属性との多面性を多重継承で表現することが述べられている。「戦争」という辞書項目をおかないとすれば、「戦争する」が [行為] と [事柄] の 2 面性を持つことは自然と思われる。ただし、この対応は必ずしも系統的ではないようで、例えば「勉強する」「結婚する」「誕生する」等は、[事象] としての語義しか記述されていない。

この方針で「安売りする」「刺繍する」「才気煥発」等の多面性が扱われ、これらと同じ上位概念にまとめられた動詞 (「売り出す」「縫う」「冴える」) の語義についても、同様の多面性が与えられたものと推察される。

たぶん関連して、「食い余す」に「食べ残し」の語義、「弾ます」に「歓声」の語義が与えられているような例もある。繰り返しになるが、全てのサ変動詞の多面性が系統的に扱われているわけでないため、これらは実際問題として例外に属する。加えて、[事象] を上位に持たず、そうではない語義だけが記述されているものもみかけられた。

この他、「係る」を [文法用語] とするようなメタレベルの分類も 6 件みられた。

## 2.2 日本語動詞のための概念体系

これらの問題を持つ語や概念を、動詞の語義は [事象] の下位に位置するものとみなして、修正した。問題のある上位しか持たない概念はその概念自体を除去、多重継承で一部が不適切な上位概念となっている場合は、それに繋がる関係を削除した。これにより、[事象] を根とし、すべての葉に日本語動詞が関連づけられている概念体系を得ることができる。

抽出された概念体系 (階層構造) の概念である節点の数は 19,668 であり、葉となっているものが 17,140、葉でないものが 2,528 である。葉の全てと葉でない 1,797 節点のとの計 18,937 概念に日本語動詞が関係づけられている。英語を含めた動詞に関連する EDR 辞書の概念数は、46,562 であるため、その 40% 強が日本語動詞と関わっている。[事象] を上位としない日本語動詞の語義に関連する概念を削除したこともあるが、もし日本語と英語の同じ規模で全く重なりがない概念体系が融合している場合でも日本語に関する部分はその 50% になるはずであるため、この数字はかなり小さいものである。

上位構造をみると、動詞についての根である [事象] には EDR 辞書全体では 14 概念を娘に持つのに対し、抽出された階層構造では 5 つの娘 [移動] [行為] [変化] [現象]

- 30f7e4 事象
  - 30f7e5 現象 (P0)
    - 4438e6 phenomenon
    - 30f7e6 自然現象 (P1)
    - 30f7ea 静物に関する現象 (P2)
    - 30f7f4 生理現象 (P3)
    - 30f7ff 社会現象
    - 3aa947 人間に関する現象 (P4)
    - 3d017d 事象
    - 444d1c いろいろな現象
  - 30f801 移動 (M0)
    - 30f802 空間移動 (M1)
    - 30f826 所有権の移動 (M2)
    - 30f832 情報の移動 (M3)
    - 3f96ae 方向性でみた移動 (M4)
    - 444d41 いろいろな移動
  - 30f83e 行為 (A0)
    - 0e5a01 (人が) 活動する
    - 1fa1a2 人間の活動
    - 30f83f 身体的活動 (A1)
    - 30f863 感情活動 (A2)
    - 30f8af 行為や物事の進行にかかわる活動
    - 444d1a いろいろな人間活動 (A3)
    - 444d85 遊びやスポーツ
    - 444d96 自身行為 (A4)
    - 444dd8 対象行為 (A5)
  - 3f9856 変化 (C0)
    - 2625a7 firm\_up
    - 1f7703 物がいつの間にか別の物と摩り替わる
    - 1f99d7 (別の人物の役に) 成り切ることができる
    - 2625a7 break
    - 30f90d 数量や程度の変化 (C1)
    - 30f911 開始, 終了, 継続など, 物事の経過 (C2)
    - 3aa95c 関係の変化 (C3)
    - 444cdd 変わる
  - 3aa963 状態 (S0)
    - 30f7c8 物事に対する評価
    - 30f9a4 関係 (S1)
    - 3bc006 ある状態である
    - 3f9871 性状・性向 (S2)
    - 3f98f6 存在状態 (S3)
    - 444d1e いろいろな状態

図 1 日本語動詞についての上位階層

[状態]に絞られる。[事象]の孫は抽出されたもので36であるのに対し、全体では1,202、抽出された5つの娘だけでも1,019の娘を持つ。抽出された上位の階層を図1に示す。概念の名称には、EDR辞書において概念を参照する概念識別子と、日本語説明文、それが無い場合は英語名称等、を用いている。以下、スペースの都合で、表等において概念識別子の代わりに、本図中の行末の括弧内に示した略号を用いる場合がある。

抽出された体系と英語動詞に関係づけられた概念との関係として、英語動詞の語義となっている概念から包摂関係を幾つ辿ると、抽出された体系中の概念に到達するかを表2に示す。「未整理」は[MISC]という概念を上位としており、抽出された階層の概念に到達しない。0の列の括弧内の数字はその概念が日本語動詞の語義となっている内数であり、同じ概念を語義として持つということで関連づけられている日本語と英語の動詞は語義にして日本語動詞全体に対して14%程度であることがわかる。

これらの数値から、日本語動詞の語彙オントロジと英語動詞のそれとは上位の構造等、かなり性格が異なり、その対応づけも困難であったと判断される。日本語動詞につい

ての知識源としてEDR辞書を見た場合、サ変動詞の多義性の問題を残しているが、[事象]を根とする図1に示したような上位構造を持つ体系を扱った方が見通しがよいと考えられる。

### 3. 日本語 WordNet との対応づけ

EDR辞書から抽出した日本語動詞の概念体系(本節ではこの体系をEDR辞書と呼ぶ)の特徴を明らかにし、他の言語資源との融合の可能性を検討するために、日本語 WordNet[4]との対応づけ(アライメント)を検討した。日本語 WordNetはPrinceton大学で開発された英語 WordNetの概念体系を構成するsynsetに日本語単語を関係づけたものである。ここで、synsetとは、同義語の集合により語義を表現したものでEDR辞書における概念に相当する。以下、EDR辞書概念と日本語 WordNetのsynsetをまとめて概念と呼ぶ場合がある。WordNetは様々な応用で用いられている代表的な言語知識である。包摂関係だけでなく含意関係等、様々な意味関係が記述されており、それとの対応づけは様々な示唆を与えてくれるものと期待できる。

#### 3.1 手法

語彙的オントロジの対応づけでは、そこに含まれる個々の概念どうしの対応づけを考えることが多いが、ここでは、お互いの構造の違い等に関する考察の材料としたいという動機から、概念を関連づける包摂関係も考慮に入れることを考えた。それぞれの体系において、包摂関係にないという意味で独立した(ただし多重継承の関係で下位概念を共有している場合はある)概念の集まりを一定の基準で選び、それを対応づけの対象とする。以下、選ばれた概念を対象概念と呼ぶ。

一般の対応づけでは、クラスの名前、どのような属性を持っているか等の内部構造、他のクラスとの関係つまりオントロジの構造、インスタンスの共通性等が手掛かりとして用いられる[2]。語彙的オントロジにおいては、クラスである概念の説明文(語釈文, gloss)も名前と同様の手掛かりとなる。ここでは、オントロジの構造の違い等を考察したいので、インスタンスの共通性を手掛かりとしてアライメントを行う(Step 1)。この場合のインスタンスはその概念をもつ語(表記)である。提案する手法では、それぞれの概念の対について方向性を持った2つの指標が得られるので、それを基に1対nの概念の対応づけを行う(Step 2)。本来のアライメントであれば、整合性を考慮して体系全体を対応づけるべきであるが、今回はそれはおこなわず、結果の一部である個別の対応について考察を行うにとどまっている。

##### 3.1.1 概念の選択

EDR辞書における動詞は事象を根としたひとつのDAGを構成している。原則として、この構造の4段目にある概

表 2 英語動詞と日本語動詞の体系との距離

関係数	0	1	2	3	4	5	>5	未整理
概念数	2,735 (2,576)	8,755	4,959	5,542	1,724	588	335	3,682

念を対象概念とした。

EDR 辞書の体系は語義となる概念をグループ化することで上位の概念を作成してことで作られる [7]。そのようなして得られた概念は中間ノードと呼ばれる。このような中間ノードが更に上位の中間ノードにまとめられるが、その際そのレベルで特殊で雑多な概念でありそれ以上の下位概念を持たない末端概念もあわせて上位の中間ノードにまとめられるとされている。動詞についてみると実際には、雑多な概念がひとつの中間ノードにまとめられて、その後、他の中間ノードとあわせて上位の中間ノードにまとめられていることも多い。図 1 において [いろいろな現象] [いろいろな移動] 等の概念はこのような雑多な末端概念をまとめた中間ノードである。このことは多数の末端概念を娘としていることから推測される。末端概念をまとめあげた中間ノードとそうでない中間ノードは娘の数で区別され、前者は娘が 50 を越えるが、後者たかだか数十に留まる。一方、[変化] における [物がいつの間にか別の物と摩り替わる] 等は末端概念で、それらが直接他の中間ノードとまとめあげられている。

なお、EDR 辞書のこのような体系の構成は、語義と関係づけられた末端概念を階層的に分類していることに相当し、階層中の全ての概念 (synset) に語義が対応づけられている WordNet とは思想を異にしている。いずれの場合もこれは原則であって、必ずしも厳密にそうになっているわけではないが、その違いは利用において意識されるべきであろう。

EDR 辞書の対象概念は、原則として階層の 4 段目にある概念 (図 1 で葉となっている概念の娘達) としたが、この 4 段目の概念が末端概念で 3 段目にある概念がそれをまとめたものである場合と、3 段目の概念自体が末端概念である場合は、それを利用するようにした。結果、221 の概念が選ばれている。

WordNet の少なくとも動詞群はひとつの synset を根とする構造を構成しておらず、比較的浅い DAG が複数集まった森となっている。開発時に synset は品詞と論理的な分類に基づいて 45 分類され、動詞は、表 3 に示すような、29 から 43 の番号を与えられた 15 分類に分類されており、これが最も粗い分類ということになる<sup>\*5</sup>。以下、この分類を大分類と呼ぶ。それぞれの大分類に含まれる synset もひとつの DAG をなしているわけではなく、これらの大分類の

<sup>\*5</sup> WordNet の synset と SUMO の上位オントロジーの対応づけが行われており [8]、それを用いればこれらに構造を与えることができる。

表 3 日本語 WorNet の動詞大分類

番号	名称: 説明
29	body : grooming, dressing and bodily care
30	change : size, temperature change, intensifying, etc.
31	cognition : thinking, judging, analyzing, doubting
32	communication : telling, asking, ordering, singing
33	competition : fighting, athletic activities
34	consumption : eating and drinking
35	contact : touching, hitting, tying, digging
36	creation : sewing, baking, painting, performing
37	emotion : feeling
38	motion : walking, flying, swimming
39	perception : verbs of seeing, hearing, feeling
40	possession : buying, selling, owning
41	social : political and social activities and events
42	stative : being, having, spatial relations
43	weather : raining, snowing, thawing, thundering

中で根となっている、つまり、上位概念を持たない synset が 559 概念ある。日本語 WordNet の対象概念はこの 559 の synset とした<sup>\*6</sup>。

### 3.1.2 対応づけ

一般性の高い動詞ということで、計算機可読辞書 Lexeed[5] に含まれる動詞辞書項目の表記をインスタンスとして用いる。Lexeed には 3,588 動詞が含まれるが、これらのうち、EDR 辞書に含まれるものは 3,168 動詞、日本語 WordNet に含まれるものは 2,250 動詞、両方に含まれるものは 2,212 動詞であったので、この 2,212 動詞を用いる。

EDR 辞書、日本語 WordNet それぞれについて、各表記について、語義となっている概念の上位概念であるような対応づけ対象概念を得る。多義性のためにひとつの表記は複数の概念を語義とし、それらの語義も多重継承のために複数の対象概念を上位概念とするので、ひとつの表記に対して複数の対象概念からなる列が得られる。この集まりをコーパスと呼ぶ。このコーパスにおいて、ひとつの表記に与えられた概念数の平均は EDR 辞書で 4.36、日本語 WordNet で 3.77 であった。

前述したような基準で選択してはいはいるが、その下位概念やそれらと関係づけられる語の数等、対象概念の大きさは必ずしも均一ではない。そのため、作成したコーパスでの出現回数も大きく異なる。表 4 にそれぞれの対象概念を出現数の多い方から 10 件並べたものを示す。日本語 WordNet の synset はオフセット値に 'v' を付した識別子、

<sup>\*6</sup> ひとつの synset (v02422663) では包摂関係が巡回しており、それも対象概念に含めた。

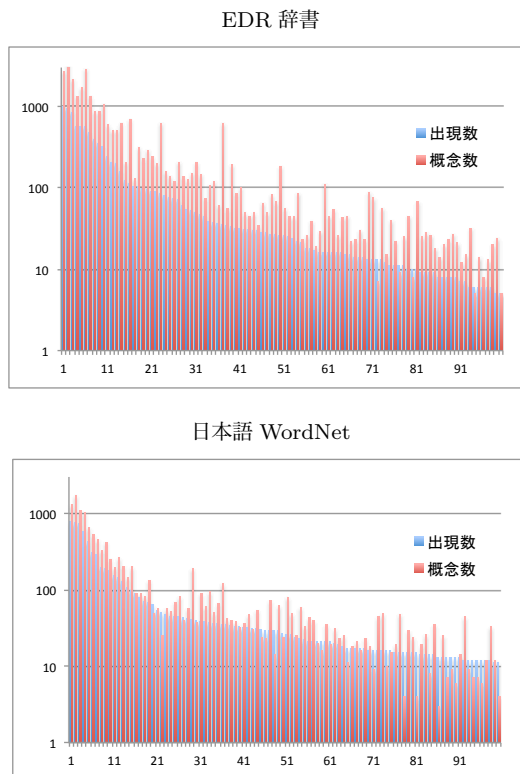


図2 対応づけ対象概念の順位と出現数および下位概念数

関係づけられている英語単語、英語説明文の一部で示している。大分類は番号とその名称を与えている。図2は出現数に基づく順位とコーパスでの出現数、それぞれの体系での（直接間接あわせた）下位概念数を上位100件について示したものである。

**Step 1** EDR 辞書と日本語 WordNet の対象概念をそれぞれの言語の語、同じ表記に対するの対象概念の列を対訳関係にある文（単語の列）の対とみなして、対訳コーパスと考える。これに対して、統計的機械翻訳における単語アライメントを行う。実装は GIZA++<sup>\*7</sup>を用いた。両方向のアライメントを行い、翻訳確率を求め、これを対応の指標とする。

**Step 2** 出現数の多い対象概念は複数の対象概念と対応づけられることが考えられる。ある概念  $e_i$  と対応づけられる概念  $f_j$  の集まり  $CS_{e_i}$  を得るために、図3に示す手順を行う。ここで、 $N_{e_i}$  は  $e_i$  のコーパスでの出現回数、 $P_{f_j}^{e_i}$  を  $e_i$  から  $f_j$  への翻訳確率とする。  $hm(p, r)$  は、 $p$  と  $r$  の調和平均で、 $p, r$  いずれかが0の場合は0となる。一見複雑であるが、双方向の翻訳確率の調和平均が増加するあいだ、それを最大にする対象概念を加えていくという貪欲法によって対応づける対象概念の集まりを求めている。

### 3.2 結果と考察

Step1の対応づけにおいて、出現数が10以上の概念どうしの対応で、双方向の翻訳確率の調和平均が大きい上位20

<sup>\*7</sup> 入手先 (<http://www.statmt.org/moses/giza/GIZA++.html>)

件を表6（第9頁）に示す。それぞれの概念の名称から、このレベルでは、おおよそ妥当な対応づけが行われているのが分かる。EDR 辞書における「関係」と日本語 WordNet における [42 stative], 「所有権の移動」と [40 possession] 等、上位の概念どうしても系統的な対応関係があると想像させる結果も出ている。ただし、翻訳確率の値は全般に低く、必ずしもよい対応を示しているわけではない。

Step2の対応づけは、一方の大きな概念に他方の複数の概念を対応づけるものである。表4に示したそれぞれの体系の大きい概念についての対応づけの結果を表5に示す。出現数が5以上の概念を対応づけの候補としている。一方の大きな対象概念が他方のいくつを対象概念と対応づけられたかを上位概念と大分類毎に示している。太字で示した数字はその中に Step 1 で最もよい対応を示した概念が含まれていることを示す。EDR 辞書の「対人行為」は日本語 WordNet の [38 motion] に属する概念と最もよい対応を示し、それを含めて21の概念と対応づけられ、そのうち、7つが [41 social] に属し、4つが [33 competition] に属し…ということを表している。

例えば、EDR 辞書の「方向の決まっている移動」は [35 motion] に属する概念と主に対処づけられているし、日本語 WordNet の [be] が「存在状態」をはじめとして、「状態」に属する概念と対応づけられている等、この対応づけも一応の傾向は示しているが、一方のある概念に対応する概念の集まりはその分類と一致しているとは言い難く、むしろ散らばってしまっている。例えば、「対人行為」と対応づけられた複数の概念がすべて [41 social] の下位概念であれば双方の概念体系の対応づけが期待できるが、そのような状況ではない。ふたつの概念体系の上位部分は異なる指針に基づいて設計されていると推測される。

具体例を幾つか見てみる。EDR 辞書の「対人行為」と対応づけられた [38 motion] に属する概念は [move displace] であるが、その対応や、[41 social] や [33 competition] の下位概念が対応づけに多く含まれることは、予想される「対人行為」の特徴と合致している。同様に、「情報の発信」に [32 communication] [39 perception] [41 social] の下位概念が対応づけられていることも、情報発信の幾つかの側面を捉えているように見える。[39 perception] は多少奇異な感じがするが、対応づけられている下位概念は [show] [hide conceal] である。そのような対応はあるものの、一方で「対人行為」「情報の発信」とまとめられた概念が、他方では異なる観点から複数に大分類されている。同様に、日本語 WordNet の [change] と対応づけられている EDR 辞書の「現象」の下位概念は「静物の質的变化」「静物の外観の変化」「温度の変化」「色彩の変化」「湿度の変化」「混じる」（以上、「静物に関する現象」の娘）、「動物の生理現象」「生命体の機能が異常な状態になる」「生物の外観の変化」「生物の命が消えること」「生命体の機能が正

表 4 出現数の多い対応づけ対象概念

EDR 辞書				
概念	上位概念	出現数	下位概念数	
30f8dd 対人行為	A5 対象行為	1005	2664	
444dd9 ものを対象とする行為	A5 対象行為	934	4110	
3aa94b 方向の決まっている移動	M4 方向性でみた移動	851	2145	
3f96e6 情報の発信	M3 情報の移動	568	1319	
30f8af 行為や物事の進行にかかわる活動	A0 行為	563	1693	
30f87e 対物活動	A5 対象行為	558	2873	
3aa94a 具体物を移動する	M1 空間移動	479	1309	
444d9d ものともとの関係にかかわる行為	A5 対象行為	395	865	
3aa949 主体の移動	M1 空間移動	347	871	
44485b 体の動作を行う	A1 身体的活動	318	1056	

日本語 WordNet				
概念	大分類	出現数	下位概念数	
v00109660 change (undergo a change; become different in essence; losing one's or its original nature)	30 change	774	1295	
v00126264 change alter modify (cause to change; make different; cause a transformation)	30 change	749	1704	
v02367363 act move(perform an action, or work out or perform (an action))	41 social	741	1083	
v01850315 move displace (cause to move or shift into a new position or place, both in a concrete and ...)	38 motion	586	1027	
v01617192 make create (make or cause to be or to become)	36 creation	424	659	
v01835496 travel go move locomote (change location; move, travel, or proceed, also metaphorically)	38 motion	305	526	
v00628491 think cogitate cerebrate (use or exercise the mind or one's power of reason in order to ...)	31 cognition	289	451	
v01831531 move (move so as to change position, perform a non-translational motion)	38 motion	197	329	
v02220461 transfer (cause to change ownership)	40 possession	188	420	
v02604760 be (have the quality of being; (copula, used with an adjective or a predicate noun))	42 stative	180	252	

$$Cands \leftarrow \{f_1, \dots, f_n\}$$

$$CS_{e_i} \leftarrow \{\}$$

$$f_{max} \leftarrow \operatorname{argmax}_{f_j \in Cands} hm(P_{f_j}^{e_i}, P_{e_i}^{f_j})$$

$$M_{update} \leftarrow hm(P_{f_{max}}^{e_i}, P_{e_i}^{f_{max}})$$

**Repeat**

$$CS_{e_i} \leftarrow CS_{e_i} \cup \{f_{max}\}$$

$$Cands \leftarrow Cands - \{f_{max}\}$$

$$M_{crnt} \leftarrow M_{update}$$

$$f_{max} \leftarrow \operatorname{argmax}_{f_j \in Cands} hm(P_{f_j}^{e_i} + \sum_{f \in CS_{e_i}} P_f^{e_i}, \frac{P_{e_i}^{f_j} N_{f_j} + \sum_{f \in CS_{e_i}} P_{e_i}^f N_f}{N_{f_j} + \sum_{f \in CS_{e_i}} N_f})$$

$$M_{update} \leftarrow hm(P_{f_{max}}^{e_i} + \sum_{f \in CS_{e_i}} P_f^{e_i}, \frac{P_{e_i}^{f_{max}} N_{f_{max}} + \sum_{f \in CS_{e_i}} P_{e_i}^f N_f}{N_{f_{max}} + \sum_{f \in CS_{e_i}} N_f})$$

**Until**  $M_{update} \leq M_{crnt}$

図 3 複数概念との対応づけ (Step 2) の手順

常な状態になる] (以上, [生理現象] の娘) であり, いずれも変化と関連していると推測されるが, これらは [現象] の下位概念であって, EDR 辞書では [変化] とは異なるものと整理されている. このように概念どうしの対応づけは適切と思われる場合も多いのに対し, 対応づけの対象概念

よりも上位において, その構造の設計指針が異なることで, 一方のある概念が他方の体系のまとまった部分として対応づけられることが少なくなっているように観察される.

もちろん, 個々の概念の対応づけに問題がないわけではない. 例えば, EDR 辞書の [対人行為] に [42 stative] に

表 5 大きい概念と複数概念の対応づけ

EDR 辞書概念	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43
30f8dd 対人行為			2	2	4		2		2	1			7	1	
444dd9 ものを対象とする行為	1	3	1	1			2	2	1		1		2		
3aa94b 方向が決まっている移動	1										5	1			
3f96e6 情報の発信				5								2		1	
30f8af 行為や物事の進行にかかわる活動	1	3					1	1				1		6	3
30f87e 対物活動	2						5	1							1
3aa94a 具体物を移動する	1	1				1	1				1	1	1		
444d9d ものともとの関係にかかわる行為	1	1					2	1		1		2	1	2	
3aa949 主体の移動						1					3			3	
44485b 体の動作を行う	2				1		5			1	1				1

日本語 WordNet 概念	現象			移動				行為					変化			状態				
	P1	P2	P3	P4	M0	M1	M2	M3	M4	A1	A2	A3	A5	C1	C2	C3	S0	S1	S2	S3
v00109660 change	6	5	1							1				2	3	1	1		1	
v00126264 change alter modify	1	1												3	1				1	
v02367363 act move							1													
v01850315 move displace					1					1						1				
v01617192 make create		1										1				1			1	
v01835496 travel go move locomote	1			1	1			1							2					
v00628491 think cogitate cerebrare										2	1									
v01831531 move	1				2			1												
v02220461 transfer						1														
v02604760 be				1														2	2	2

属する概念が、日本語 WordNet の [change] には [身体的活動] がひとつ対応づけられているが、これらは、そのアスペクトからは考えにくい対応である。原因を見てみると、[対人行為] に対応づけられた [42 stative] に属する概念は [necessitate ask ... demand] であり、この対応は両者に「求める」の語が関係づけられており、求めるという行為と必要としているという状態が対応づけられたものと推測される。また、[change] に対応づけられている [身体的活動] は [寝る] で、「寝る」がもつアスペクトの多義性「10分で寝た (状態変化)」「10時間寝た (活動)」が関係していると考えられる。

これらは語の共通性で語彙的オントロジを対応づける場合の興味深い課題と考えるが、概念と語の関連づけ、翻訳の問題等、様々な要因が絡むことに加えて、膨大な個別の対応づけをそれぞれ分析する必要があり、系統的な考察は行えていない。

### 3.3 方法についての考察

本稿で用いた対応づけ手法は、そこに属するインスタンスの共通性でクラスの対応づけを行う手法と考え方を同じくし、概念のインスタンスとしてそれを語義 (の上位概念) とする語 (の表記) を用いた。対応づけには統計的機械翻訳における単語アライメントを利用したが、この他に、一方のある概念  $e_i$  に関連づけられている語の集まり  $WS_{e_i}$  と他方のある概念  $f_j$  に関連づけられている語の集まり  $WS_{f_j}$

の重なり大きさ、 $\frac{|WS_{e_i} \cap WS_{f_j}|}{|WS_{e_i}|}$ ,  $\frac{|WS_{e_i} \cap WS_{f_j}|}{|WS_{f_j}|}$  を指標とすることが考えられる。多義性や多重継承によりある対象概念に同じ語が複数回関係づけられるので、その扱いによって幾つかの変種が考えられる。

単純に語の集まりを集合と考えると、そのような重複を無視した場合の結果を表 7 (第 10 頁) に示す。表 6 に示したのと同様、出現数が 10 以上の概念どうしの対応で、それぞれの語数を母数とした共通語数の割合の調和平均の大きいものを並べている。半数程度が表 6 のものと重複しており、おおよそ同じ傾向となっていることが伺える。

これ以外の変種として、Lexeed に含まれるものだけでなく、EDR 辞書と日本語 WordNet に共通する語 (動詞) すべてを用いることも考えられる。この場合、語数は 8,143 語、ひとつの表記に与えられた概念数の平均は EDR 辞書で 3.42、日本語 WordNet で 2.58 である。この場合も、対応づけの傾向は大きく変わらない。

EDR 辞書と WordNet とのアライメントは、しばらく以前に幾つかの試みがある [9], [11]。当時は WordNet に日本語が関連づけられていなかったため、本稿とは逆に EDR 辞書の英語部分を取り出されている。また概念体系の考察という本稿の関心とは異なり、1対1の対応づけが目的とされている。とはいえ、利用されているのはそれぞれの概念に対応づけられた語の共通性で、それに加えて説明文の重複や、母娘の語の重複が考慮されて、対応づけが求められている。得られた対応がどのようなものかは述べられて

いないが、前述のように EDR 辞書と WordNet では概念と語との関連づけの方針が異なるので、それがどう影響しているかが興味深い。本稿での対応づけではその概念の下位概念に関連づけられたすべての語が考慮されている。

最近では、Hayashi[3] が多言語資源の活用の観点から、EDR 辞書の日本語部分と WordNet との対応を行っている。ここでの WordNet は英語の語彙資源であり、日本語単語の情報は用いられていない。異なる言語による説明文の類似度を機械翻訳を利用して求め、意味情報が注釈されたコーパスから語の共通性を計算し、それら 2 種類の情報を組み合わせることで対応づけを行っている。利用している説明文はその概念のものだけでなく母娘のものも含まれるので、それを通じて構造の情報も考慮される。

Asanoma[1] は WordNet と日本語語彙大系のアライメントを行っているが、そこではラベルの文字列の重複が利用されている。

#### 4. おわりに

EDR 概念辞書について、日本語動詞に関連する構造を分析した。日本語動詞が構成する概念体系と英語動詞が構成する概念体系の交わりが比較的少ないこと、サ変動詞の意味の扱いに関連して、動詞の意味記述としては扱いにくい点があることを明らかにし、日本語動詞に関する語彙知識として利用しやすい部分を抜き出した。その部分と日本語 WordNet とを対応づけることで、ふたつの語彙的オントロジの上位構造に配慮すべき違いがあることを示唆した。

含意認識や推論に用いることができる語彙知識の構築を目的として、貴重な既存知識である EDR 概念辞書を有効に活用すべく、今後もその分析を進めていきたい。

**謝辞** 本研究の一部は次の科研費の助成を受けている。ご支援に感謝する。挑戦的萌芽研究「形式の異なる語彙知識の相互運用の試み」(課題番号:15K12873, 代表者:加藤恒昭) および基盤研究(B)「意味の対応づけによる辞書資源群の多言語セマンティック Web 化」(課題番号:25280117, 代表者:林良彦)。

#### 参考文献

- [1] Naoki Asanoma : Alignment of ontologies: WordNet and Goi-Taikei. *NAACL Workshop on WordNet & Other Lexical Resources*, pp. 89–94, 2001.
- [2] Jérôme Euzenat and Pavel Shvaiko : *Ontology Matching (Second Edition)*. Springer, 2013.
- [3] Yoshihiko Hayashi : Blending two kinds of semantic relatedness for cross-language matching of lexical concepts. *Proc. of 10th International Conference on Terminology and Artificial Intelligence (TIA2013)*, pp. 35–42, 2013.
- [4] Hitoshi Isahara, Francis Bond, Kiyotaka Uchimoto, Masao Utiyama and Kyoko Kanzaki : Development of the Japanese WordNet. *Proc. of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, pp.2420–2423, 2008.

- [5] NTT コミュニケーション科学基礎研究所監修: 基本語データベース. 学習研究社, 2008.
- [6] 荻野 考野, 中尾 由雄, 小笠原あゆみ, 長澤 陽子: 日本電子化辞書研究所における概念体系. *情報学基礎*, 1993-FI-032, pp. 27-34, 1993.
- [7] 荻野 考野, 中尾 由雄, 長澤 陽子, 小笠原あゆみ: EDR 電子化辞書における概念体系. *言語処理学会第 1 回年次大会*, pp.197-200, 1995.
- [8] Adam Pease and Christiane Fellbaum : Formal ontology as interlingua; the SUMO and WordNet linking project and global WordNet. *Ontology and the Lexicon*, Cambridge University Press. 2010.
- [9] Satoshi Sekine, Kiyoshi Sudo and Takano Ogino : Statistical Matching of Two Ontologies. *Proc. of ACL SIGLEX99 Workshop: Standardizing Lexical Resources*, 1999.
- [10] 徳永 健伸 : 辞書と情報処理. 単語と辞書, 岩波講座 言語の科学 3, pp. 155–190, 1997.
- [11] Masao Utiyama and Koiti Hasida : Bottom-up alignment of ontologies. *IJCAI-97 Workshop on Ontologies and Multilingual NLP*, 1997.
- [12] Piek Vossen : EuroWordNet: A multilingual database of autonomous and language-specific wordnets connected via an Inter-Lingual-Index. *International Journal of Lexicography*, Vol. 17, No. 2, pp. 161–173, 2004.



表 6 高い対応づけが得られた概念の対

EDR 辞書概念	上位概念	出現数	日本語 WordNet 概念	大分類	出現数	翻訳確率→	翻訳確率←	平均
30f99d 優劣	S1 関係	14	v02673965 excel stand_out surpass	42 stative	15	0.661	0.374	0.478
3aa949 主体の移動	M1 空間移動	347	v01835496 travel go move locomote	38 motion	305	0.443	0.407	0.424
3f96b4 所有権の主体への移動	M2 所有権の移動	89	v02206619 take	40 possession	33	0.242	0.744	0.365
3f96e6 情報の発信	M3 情報の移動	568	v02367363 act move	41 social	741	0.476	0.288	0.359
3f96d8 所有権の主体からの離脱	M2 所有権の移動	90	v02220461 transfer	40 possession	188	0.616	0.251	0.356
3aa94a 具体物を移動する	M1 空間移動	479	v01850315 move displace	38 motion	586	0.381	0.285	0.326
444d9d ものとのとの関係にかかわる行為	A5 対象行為	395	v01354673 connect link tie link_up	35 contact	145	0.211	0.574	0.309
444dd9 ものを対象とする行為	A5 対象行為	934	v00126264 change alter modify	30 change	749	0.262	0.348	0.299
3f98fa 種類の近さ	S1 関係	24	v02666239 differ	42 stative	13	0.191	0.554	0.284
3aa95d 位置関係の変化	C3 関係の変化	123	v01205696 touch adjoin meet contact	35 contact	44	0.203	0.423	0.274
444d9a 自身のみの感情活動	A4 自身行為	31	v02422663 restrain keep hold_back	41 social	38	0.274	0.242	0.257
30f9ae 適合する	S1 関係	18	v02664769 equal be	42 stative	36	0.407	0.182	0.252
3f96e7 情報の受信	M3 情報の移動	241	v02106506 perceive comprehend	39 perception	45	0.138	0.661	0.228
444cca ある場所における生活, 宿泊, 居住活動	A4 自身行為	44	v02655135 be	42 stative	90	0.414	0.153	0.224
444d9b 対象によって引き起こされる感情活動	A2 感情活動	54	v01771535 feel experience	37 emotion	43	0.237	0.208	0.222
444d23 存在の変化にかかわる現象	C2 開始, 終了, 継続など, 物事の経過	80	v00426958 disappear vanish go_away	30 change	17	0.152	0.387	0.218
3f96a5 生物の命が消えること	P3 生理現象	36	v00426958 disappear vanish go_away	30 change	17	0.141	0.474	0.217
444d20 位置の変化を伴わない, 空間内の変化	P2 静物に関する現象	60	v01831531 move	38 motion	197	0.587	0.131	0.214
3f98fa 種類の近さ	S1 関係	24	v02664769 equal be	42 stative	36	0.395	0.146	0.213
3f96b4 所有権の主体への移動	M2 所有権の移動	89	v02210855 get acquire	40 possession	107	0.268	0.166	0.205

表 7 概念に関係づけられた語の重なりに基づき対応づけ

EDR 辞書概念	上位概念	出現数	日本語 WordNet 概念	大分類	出現数	一致/EDR	一致/JWN	平均
30f99d 優劣	S1 関係	11	v02673965 excel_stand_out surpass	42 stative	11	0.455	0.455	0.455
3f96e6 情報の発信	M3 情報の移動	374	v02367363 act move	41 social	451	0.492	0.408	0.446
3aa949 主体の移動	M1 空間移動	226	v01835496 travel go move locomote	38 motion	190	0.398	0.474	0.433
3aa94a 具体物を移動する	M1 空間移動	331	v01850315 move displace	38 motion	345	0.438	0.420	0.429
444dd9 ものを対象とする行為	A5 対象行為	579	v00126264 change alter modify	30 change	430	0.349	0.470	0.400
30f8dd 対人行為	A5 対象行為	631	v02367363 act move	41 social	451	0.338	0.472	0.394
444dc6 物や財産や物の権利を取得する	M2 所有権の移動	24	v01380638 gather garner pull_together	35 contact	18	0.333	0.444	0.381
3ce767 ある物事が起こる	C2 開始, 終了, 継続など, 物事の経過	24	v02623529 become	42 stative	19	0.333	0.421	0.372
3f96d8 所有権の主体からの離脱	M2 所有権の移動	76	v02220461 transfer	40 possession	119	0.461	0.294	0.359
3f96b4 所有権の主体への移動	M2 所有権の移動	73	v02206619 take	40 possession	22	0.233	0.773	0.358
30f87e 対物活動	A5 対象行為	376	v00126264 change alter modify	30 change	430	0.380	0.333	0.355
30f87e 対物活動	A5 対象行為	376	v01850315 move displace	38 motion	345	0.335	0.365	0.350
1fa1a2 人間の活動	A0 行為	14	v02528380 fail go_wrong miscarry	41 social	10	0.286	0.400	0.333
444d9b 対象によって引き起こされる感情活動	A2 感情活動	48	v01771535 feel experience	37 emotion	37	0.292	0.378	0.329
3f9842 終わる	C2 開始, 終了, 継続など, 物事の経過	25	v02609764 end stop finish terminate cease	42 stative	24	0.320	0.333	0.327
3f9843 生物の命が消えること	P3 生理現象	31	v00426958 disappear vanish go_away	30 change	13	0.226	0.538	0.318
30f8dd 対人行為	A5 対象行為	631	v01850315 move displace	38 motion	345	0.246	0.449	0.318
3aa94b 方向が決まっている移動	M4 方向性でみた移動	528	v01835496 travel go move locomote	38 motion	190	0.212	0.589	0.312
444dd9 ものを対象とする行為	A5 対象行為	579	v00109660 change	30 change	486	0.278	0.331	0.302
444d9d ものとの関係にかかわる行為	A5 対象行為	259	v00126264 change alter modify	30 change	430	0.402	0.242	0.302