

# 日本語の用語法に見る日本人の〈被害〉の概念化: 語彙分析からオントロジーへ

黒田 航  
独立行政法人 情報通信研究機構 (NICT)

仲本 康一郎  
山梨大学

## 1 はじめに

### 1.1 目的

[3]では「襲う」のコーパス事例を基に人手で作成された階層的フレームネットワーク (Hierarchical Frame Network: HFN) を実験材料に使用して、ヒトの状況理解の一面を調査した (この用語で「フレーム」というのはヒトが理解する具体的な状況をさす)。その目的は、ヒトの知識体系の中でモノが階層化されているのと基本的に同じ仕方で状況=事態が階層化されていること、その階層を素性を使って記述できる可能性があることを実験的証拠に基づいて示すことだった。その後の研究 [4] から、[3] や関連研究 [5] で使用された「襲う」の HFN は、単に「襲う」という語の多義性の体系を記述したものである以上に、日本人が理解する〈被害の発生〉の状況をそれなりの範囲で記述すると見なせる可能性があることがわかった。この是非を調べるのが本研究の目的である。

### 1.2 方法

上で言及した「襲う」の HFN の日本人の〈被害の発生〉の概念化の全体に対する被覆率を調べるため、日英新聞記事対応付けデータ (JENAAD) [6] (p11 と pnm<sup>1)</sup>) を合わせて合計 179,252 文) から抽出した「被害者」と「犠牲者」のコーパス事例に対して、「襲う」の HFN がどれぐらいの被覆率をもっているかを評価し、次の方法で予想を確かめた:

- (1) JENAAD コーパスから KWIC 検索で「犠牲者」と「被害者」の事例を抽出 (「犠牲者」の事例数は 142, 「犠牲者」の事例数は 501)。(2) おのおのの語の事例の一つ一つに対し、まず第二著者が次の手順でタグづけを行ない、その結果を第一著者が修正した: タグづけ手順 (i) 喚起された状況の同定 (問題の「襲う」の HFN からフレーム ID (e.g. F06: 捕食目的の攻撃) を選択) する。それと平行し、(ii) 日本語文側の「被害者」か「犠牲者」の Harm-causer の同定、(iii) 「被害者」か「犠牲者」に対応する英語 (e.g., victims) の同定、(iv) 英語側で認定された状況を喚起する語群の同定を行なう。

### 1.3 結果の概要

解析結果から次の三つのことが示唆されている:

- (2) 「犠牲者」の用法 (全部) で評価の対象となった HFN の被覆率は高く、文脈から状況が正解に判別できなかった場合を除くと、100% 近い。
- (3) 「被害者」の用法で評価の対象となった HFN の被覆率は「犠牲者」の場合より低く、被覆率は 70% 程度である。特に取りこぼしが目立つのは、後述の図 1 の (1) G1Ab: 〈ヒトの個体 (群) の他個体 (群) への非攻撃的加害〉 (〈犯罪〉) の上位クラスの下位タイプ、(2) G2A=G3B: 〈事故〉=〈社会災害〉の下位タイプ、(3) D3: 〈個体が経験する非一時的な心身の異常〉の下位タイプの D3C: 〈非短期的な心身の異常〉である。特に H:

〈詐欺〉の被害や〈不正取引〉のような身体的苦痛を伴わない被害は「 $x$  が  $y$  を襲う」「 $y$  が  $x$  に襲われる」の用法を基に作成された HFN ではカバーされていないかった。

- (4) G2A=G3B: 〈事故〉=〈社会災害〉の下位類として図 1 の I: 〈第三者  $y$  の争い  $x$  への巻添え〉の場合が体系的に見落とされている。

### 1.4 HFN の改良

調査から明らかになった問題点を改善するため、[5, 3] の「襲う」の用法に基づいて作成された被害のオントロジーを、「 $y$  が  $x$  の {i. 犠牲者; ii. 被害者} 者になる」の用法を説明できるように改訂した。それが図 1 に示す体系である。

この図の緑で縁取りした、F01 から F15 までインデックスのついたノードは [5, 3] の実験に使われた「襲う」の評定実験の材料になったもので、灰色のノードは今回の改訂で補ったノードである。

## 2 考察と議論

### 2.1 改良前の被害のオントロジーの被覆率

すでに言及したように「 $y$  が  $x$  に襲われる」「 $x$  が  $y$  を襲う」の用法<sup>2)</sup>に基づいて構築した状況の体系で「被害者」の用例で取りこぼしが目立つのは、次の 4 つである:

- (5) a. G1Ab: 〈ヒトの個体 (群) への非攻撃的加害〉の下位タイプ (e.g., 〈犯罪〉)  
b. G2A=G3B: 〈事故〉=〈社会災害〉の下位タイプ  
c. H2Ab: 〈非武力抗争〉の下位タイプ (e.g., H2Ab1: 〈権力争い〉)  
d. I: 〈第三者武力的抗争の巻添え〉の下位タイプ  
e. D3C: 〈個体が経験する非短期的な心身の異常〉の下位タイプ

特に (5a) は体系的に「襲う」の用法から落ちていたことが、被害者の用例の網羅的分析から明らかになった。これは襲うの語の意味を考えると、驚くことではない。それよりも驚くべきことは、これらの他のクラスでは取りこぼしが少ないことである。これは後述の「襲う」のメタファ儿的用法の広がりを考えないと理解しがたい結果である。

F08, G1Ab, I が「襲う」や「襲われる」では表わされない場合だと考えると、「犠牲者」の被覆率は  $100\% = (142-6-0)/(142-6) = 97\%$  から  $(142-6-47)/(142-6)$  の間である (推定では 80% ぐらい)。これに対し「被害者」の用例には F08, G1Ab, I がおのおの、71 例、23 例、9 例 (合計 103 例) ある。G1A (213 例) に関しては G1Aa か G1Ab かの区別が明白でない場合である (そのうちの G1Ab 相当分は「襲う」では表現されない)。従って、「襲う」の用法に基づいて作成されたもとの HFN の被覆率は、 $79\% = (501-6)-103/(501-6)$

<sup>2)</sup> 本稿では詳しくは説明しないが、一般に受動形の方がメタファeric 用法への容認度が高く、中には受動性でしかい得ない被害のタイプも存在する (e.g., ?\*痙攣が少女を襲った)。これは受動性が被害者の視点を表わしやすいという特徴と無関係ではない。

<sup>1)</sup> p11 と pnm は対応づけが「1 対 1」と「 $n$  対  $m$ 」という意味である。



表1 「被害者」で言及されていた状況の分布 (重複アリ): \*は明示的にコーディングしていないことを表す

状況 index	犠牲者	被害者
F01	28	53
F02	50	23
F03	4	12
F04	2	24
F05	1	61
F06	0	1
F07	0	0
F08	47	71
F09	0	1
F10	22	2
F11	1	0
F12	0	21
F13	0	65
G1A	*	213
G1Ab	*	23
I	*	9
A	1	16
UNSPEC	6	6

と  $36\% = (501-6)-(103+213)/(501-6)$  の間である (推定では  $60\%$  ぐらいではないかと思われる)<sup>3)</sup>。

### 2.1.1 「被害者」と「犠牲者」の用法の違い

(2)と(3)から、〈被害〉とタイプと「襲う」の用法パターンに相関があることがわかる。実際、「犠牲者」より「被害者」の方が用法が広く、タグづけ結果での分布を見ると「犠牲者」は「被害者」下位類であると言える (日本語の「犠牲者」には「被害者」にはない〈回復不可能な被害を被った〉(典型的には「死んだ」)という含意がある)<sup>4)</sup>。二つの語の違いは2.2.4でも言及する。

### 2.2 「襲う」の HFN 改良

図1にある HFN は [5, 3] で想定されていた HFN を、(4)の問題を解消するように、改訂した結果である (ただし(3)で明らかになった G1Ab の下位分類は完全に補われてはいない。これは今後の課題としたい)。この改良によって、主に「襲う」で言及可能な〈被害〉の概念化の大分類が明確になったと思われる。

#### 2.2.1 領域の重なりによる分類の曖昧性の表現

「襲う」「犠牲者」「被害者」の用法に基づいて構築した〈被害〉のオントロジーは図1に示す通りである。これはラティス構造をしており、事例の分類は排他的であることを保証しない (これはオントロジーが一意的なものであっても、ヒトの反応が一意的ではなくなる可能性を生じさせる)。しかし、以下の点でヒトの反応に合った構造を与えていると考えられる。

- (6) a. 〈被害〉の領域の基本的な分類は A の下位類としての C 領域と D 領域の区別によって与えられるが、これらは G2 の領域を共有している。  
 b. C の下位類の G1 と G2 は G1Aa=G2B を共有している。

<sup>3)</sup> G1A の事例で元の HFN で特定されていないのは例えば〈監禁〉、〈誘拐〉、〈拉致〉、〈恐喝〉、〈過失〉などであった。

<sup>4)</sup> 英語では「被害者」と「犠牲者」は特に区別はなく victim(s) と言われることを英語側の文章を調べて確認してある。これは英語では victim(s) は「被害者」に対応する generic term であり、日本語の「犠牲者」に対応する語がない可能性を示唆する。

- c. D の下位類としての G2 と G3 は G2A=G3B を共有している。

- d. 最大の分類は A の下位タイプ群と B の下位タイプ群の区別によって与えられるが、この区別がどれぐらい現実味をもつかは、個人差があるように思われる (具体的には D3 の帰属先が D か B か G3 かは個人差がある可能性を否定できない)。

(6c) は〈事故〉=〈社会災害〉のクラスの挙動の特殊性を説明すると思われる (これは人為性/非人為性の区別が難しく、中間的な振るまいをすることが予測されている (これは [3] 実験の結果でも示唆されていた)。この理由があるため、このクラスは概念化の際の個人差や視点の違いが現われやすいクラスだと想像される。

一般に異なる二人の個人が同一のオントロジーを内在化している保証はないが、違っている部分よりは共通している部分の方が多いだろう (そうでなければ話が通じるのが不思議である)。だとすれば、どこで違いが生じ、どこで生じないのかを特定できれば、オントロジーの非一意性は問題にならないと思う。

厄介なのは視点の違いである。同一のオントロジーを内在している場合にも、視点の違いによって見かけは違うオントロジーをもっているように「見える」可能性がある (これは実際、多重継承をしているノードの分類に関して特に起こりそうなことである)。これが厄介なのは、それが実際に起こっているのか否かを判別しづらいつころにある。これに対しては、行動データから得られる事実と記述的なモデルの特徴を、Formal Concept Analysis [1] などに基づいた強力な支援ツールを使ってうまく整合させることしか解決はないだろう。

#### 2.2.2 概念メタファーの働きの明確化

図1の改訂でメタファー画像 [2] の役割が [4] で論じられて以上の明確になった。

「犠牲者」「被害者」のメタファーは後で別に取り上げることにして、「襲う」に特有なメタファーで主要なものは、(7)に挙げる4つである:<sup>5)</sup>

- (7) a. MM1a: 〈〈攻撃の相手は獲物〉〉  
 b. MM1b: 〈〈攻撃の相手は(外)敵〉〉  
 c. MM2: 〈〈被害は天災による被害〉〉  
 d. MM3: 〈〈病気(の症状)はS〉〉

MM1a と MM1b を一緒にした MM1 は「? \*x は y の敵対心の餌食になった」のような不適切な具現化を許し、過剰般化である (「餌食」は F06 を喚起する語、「敵対心」は F01 を喚起する語である)。MM1 と MM1b の違いは、経由するのが G1D か G1C かの違いとして説明可能である。同様に MM2 は、E1B を経由する MM2a と E3B を経由する MM2b に細分化されるが、この違いは被害者の集団性を反映したものである。

S の候補を絞りこむことは難しいが、G3: 〈自然災害〉か G2=D2: 〈人為災害〉だと想定するのが適切だろう (例えば S を G1B だと考えると、MM3 は MM3a: 〈病気(の症状)は敵〉と MM3b: 〈病気(の症状)は捕食者〉に区別が可能なはずだが、その粒度のメタファー表現は見当たらない)。

(7)に挙げた他には、MM4: 〈〈人為的災害は動物による被害〉〉、MM5: 〈〈非意志体による被害は意志体による被害〉〉、MM6: 〈〈原因が内在する被害は原因の外在する被害〉〉のようなメタファーが潜在するが、これらは過剰般化なし

<sup>5)</sup> これらは「y が x に襲われる」に特有なメタファーではなく、y が x の {i. 被害; ii. 犠牲者} 者になる」と共有されるほどは一般的なものである。とはいえ語彙の差は存在するので、その説明にはより詳細な検討が必要である。

で定式化することはできない。

### 2.2.3 共通の上位ノードによる概念メタファーを認可

図1に示したメタファーの体系化=分析は、元領域  $S$  から先領域  $T$  へのメタファーの成立に  $S$  と  $T$  に共通の上位領域が不可欠であると想定している。この基準は「襲う」が使われるメタファー表現に限らず、概念メタファーで説明される事例一般に対して有効であると思われる。例を一つ挙げるなら、図1に示したオントロジーは「議論は戦争」=「Argument is War」メタファー [2] の成立条件について次のような説明が与えられる。

図1の体系は F02: 〈戦争〉とは H2Aa: 〈(グループ間の)武力抗争〉の下位類としている。H2: 〈(利害の対立による)争い〉はより一般的な概念である (F02 以外の H2 の下位クラスには H2Ab や H2B などがある)。(議論) (argument) は利害の対立がある場合に起こる状況なので、H2 の下位クラスである。これは利害の対立がグループ間のものか個人間のものかの基準で二つの場合に分けることができる。この区別は図1の MM7 と MM8 に対応する (ここで元領域は E1Aa: 〈武力抗争〉であり、実際には F02: 〈戦争=軍事侵略〉ではないことに注意されたい<sup>6)</sup>)。「議論は戦争」という概念メタファーの存在は、〈議論〉と〈戦争〉(正確には〈武力抗争〉)とがいずれも H2: 〈利害の対立に基づく争い〉の下位タイプであり、〈武力抗争〉に H2 に対する代表性があることを想定すれば説明できる<sup>7)</sup>。

### 2.2.4 「 $x$ の犠牲/被害者」のメタファー

図1の一部に示したように、(8)を使ったメタファー表現が成立するが、「襲う」を使ったメタファー表現が成立しない場合がある (e.g., 彼は派閥政治の {i. 犠牲 (?); ii. ?被害 (\*者)} になった vs. \*彼は派閥政治に襲われた)。つまり次がメタファー表現になる場合もある:

(8)  $y$  は  $x$  の {i. 犠牲; ii. 被害} 者 (である)

だが、その確率は大きく異なる。「犠牲者」の 142 例のうち 8 例が明らかにメタファーであった。これに較べると「被害者」のメタファー用法は少なく、次のような例だけがかかるうじてメタファーと認められるだけだった:

(9) 環境訴訟は、自然を“被害者”とするケースも多いため、だれを原告とするかや訴えの利益をどう主張するかなど、通常の訴訟にはない難しさがある。

一般に「 $x$ の被害者」と「 $x$ の犠牲者」のメタファー用法は、同一の  $x$  に関して容認度が異なる<sup>8)</sup>。それは (i) 被害者の方が用法が広く、(ii) 「襲う」の用例に基づいて構成された被害の体系化が「被害者」の用例よりも「犠牲者」の用例に対して大きな被覆率をもっていたことから伺える。

### 2.3 F13: 〈心身の異常による被害〉の位置づけ

科学的見地からすると奇妙な F13: 〈心身の異常〉の位置づけを行なっている。現代的知識では、病気の原因の多くは病原体であり、それは生命体の一種であると了解されているが、これは図1に示したオントロジーで試案の B の部分と整合しにくい (e.g., 病気は自然災害か?)。筆者らはこれが奇妙であることを知らないわけではないが、図1が日本人の概念化の体系=オントロジーを表わしているもので

<sup>6)</sup> こう考えないと〈議論〉には〈戦争〉が含意する〈侵略〉の意味が転写されないことを説明するのに補助仮説が必要になる。

<sup>7)</sup> とはいえ、上位クラスの存在が過剰般化に導く場合も存在するので、すべてが説明できるわけではない。例えば図1は「G2A は G1」(via C) というメタファーが成立する可能性を許すが、それが実現されているか否かを仮説から独立に判断することは難しい。

<sup>8)</sup> 元を正せば、「犠牲」と「被害」はかなり性質の異なる語である。「 $y$  が  $x$  の犠牲になる」という言い方はあるが、「\* $y$  が  $x$  の被害になる」という言い方はない。「 $x$  が  $y$  に被害を与える」と言えるが、「\* $x$  が  $y$  に犠牲を与える」とは言えない。

あれば、それが「科学的に妥当」な知識を表わす必要はないと考えている。これはオントロジーは何のために何のために必要か?という問題に直結するので、ここで図1が含意する〈病気〉の概念化の特殊性について考察するのは意味のないことではないだろう。

問題は、病気の原因が病原体という生命体の一種であるという概念化が、一般的な現代日本人にとってどれほど現実感を伴っているか、である。人類が病気の原因に関する「科学的」な知識を獲得したのは、実に最近のこと(具体的にはパスツール以来)である。それまで数(十)万年の歴史の中で、病気の原因は科学的な意味では「誤って」理解されて来たことを意味する。だが、病気の例に限らず、人類の数(十)万年に渡る長い長い歴史の中で、ヒトが「科学的に妥当」という意味で「正しい」知識を得たのは過去 100 年程度であるし、科学的知識の普及した今でも、いわゆる「非科学的知識」に基づいて行動して(かつ、日常的に不便を感じないで)いる人々の数は決して少なくない。別の言い方をすれば、人類は概念化を始めてから驚くほど長い間、世界の至るところで、驚くほど「無知蒙昧」であったし、それは今でも変わっていないとも言える。この事実を考えると、〈病気〉というカテゴリーが現代日本人に科学的な観点から「誤って」概念化されている可能性があることは、特に驚くことではないように思う<sup>9)</sup>。

だが、これにはリスクもある。認知科学の目標がヒトの知識構造を明確にすることだとすれば、最終的に記述される知識構造は科学的ではないかも知れない<sup>10)</sup>。ということになるからである。これが望ましいことが否かは研究者の立場によって異なるだろう。

## 3 課題と展望

今回は「襲う」の用例に基づいて作成された〈被害〉のオントロジーの拡張を「犠牲者」と「被害者」のコーパス事例を扱えるように拡張することを試みただけで、その結果の妥当性の評価は行なっていない。将来的には実験的手法を通じてそれを行いたい。更に今回は未着手の G1Ab の下位分類の明示化を行いたい。最終的にはその延長上に〈被害〉のオントロジーを完成できれば良いと考えている。

## 参考文献

- [1] G. Ganter, B. Stumme and R. Wille. *Formal Concept Analysis: Foundations and Applications*. Springer, Berlin/Heidelberg, 2005.
- [2] G. Lakoff and M. Johnson. *Metaphors We Live By*. University of Chicago Press, 1980. [邦訳: 『レトリックと人生』(渡部昇一ほか訳). 大修館.]
- [3] 中本敬子, 黒田航, 野澤元. 素性を利用した文の意味の心内表現の探索法. 認知心理学研究, Vol. 3 (1), pp. 65–81, 2005.
- [4] 黒田航. 概念メタファーの体系性, 生産性はどの程度か? 日本語学, Vol. 24, No. 6, pp. 38–57, 2005.
- [5] 黒田航, 中本敬子, 野澤元, 井佐原均. 意味解釈の際の意味フレームへの引きこみ効果の検証: “ $x$  が  $y$  を襲う” の解釈を例にして. 日本認知科学会 第 22 回大会 発表論文集, pp. 253–55 (Q-38), 2005.
- [6] 内山将夫, 井佐原均. 日英新聞記事および文を対応付けるための高信頼性尺度. 自然言語処理, Vol. 10, No. 4, pp. 201–220, 2003.

<sup>9)</sup> 図には示していないが、B が〈原因不明の被害〉だとすると、他の事例には〈超常現象による被害〉(e.g., 〈霊障〉)が含まれる。

<sup>10)</sup> これはいわゆる形式オントロジーに方法論的な限界がありえることも示唆する。