

状況理解の単位としての意味フレームの实在性に関する研究

黒田 航 (情報通信研究機構) 中本 敬子 (京都大学教育学部) 野澤 元 (京都大学大学院)

1 研究の目的と結果

(Berkeley) FrameNet ((B)FN) は意味フレームのデータベースを構築する研究企画である [1]. [3] は (B)FN の洞察を取り入れながら日本語のための意味タグ体系 S を開発するプロジェクトを開始した, その目標は日本語の意味タグつきコーパスを開発することである. ただし, S の要となる意味フレームの概念は (統語構造やその派生等と同じく) 単なる言語学者の理論仮構物かも知れず, 心理的实在性が伴わないかも知れない. となれば (B)FN ベースの言語資源開発の意義は根本的に怪しいものとなる. この研究の目的は [2] と同様, 意味フレームの实在性の検証とそのための手法の確立である. 結果は肯定的で, 意味フレームは实在すると言える証拠を得た.

2 分析手法

2.1 言語学者による意味フレームの同定

まず二人の言語学者 (第一, 第三著者) が共同で日英対訳コーパス [4] の日本語部分から「襲う」の用例をすべて収集し, それらに (B)FN が示している仕方で意味役割を割り当て, データベース化した. この際, 用例の比喩性の区別は行わなかった. 従って, この分析はコーパスから観察可能な全用法を含んだものである. この手順は極めて専門性が高く, 本研究の中核をなす重要な過程だが, それに言及するには紙面があまりに不十分なので, 全面的に割愛する. これを通じて最終的に得られた事例は 416 例であった. このデータの意味素性表現に基づいて, それから階層構造をなす意味フレームのネットワークを同定した. その最下位意味フレームは 12 個あり, 以下のようなものである:

- (1) F01 グループ間抗争/紛争;
- F02 軍事侵略;

- F03 (強盗などの) 資源強奪;
- F04 虐待;
- F05 強姦;
- F06 動物の他個への攻撃 (捕食系);
- F07 動物の他個体への攻撃 (非捕食系);
- F08 (地震などの) 大規模な異常気象;
- F09 (高波などの) 小規模な異常気象;
- F10 疫病の流行;
- F11 活動への打撃の発生;
- F12 発病

これらの意味フレームは「襲う」という語の選択制限に反映される限り, なるべく細かく区別した. 例えば「{地震, 台風, ...} が太郎 [+human, -grouped] を襲った」が奇妙であるのに対し「{高波, 突風, ...} が太郎 [+human, -grouped] を襲った」は自然である. これは F08, F09 の区別の根拠となる. なお, この手順で同定された対象は, 自然言語処理で格フレーム [5] と呼ばれるものと, ほぼ外延が一致する.

2.2 注意

ただし HFN は「襲う」という語の意味の記述ではない. それは“ x が y を襲う”, あるいは“ y が x に襲われる”という文が理解される時に特定される客観的状況の網羅的特定である. 語から理解される状況への写像はほぼ網羅的に特例されていると言えるが, その逆の状況から表現への写像は言えない. 従って, HFN は「その人は仕事の帰りに発作に襲われた」が理解可能であるとき, その理解は F12 のタイプであることは予測するけれど, 「??その人は若くしてアルツハイマー病に襲われた」が少なからず逸脱した表現であることは説明しない. x が進行性の病気 (e.g. アルツハイマー (病), 痴呆症, 筋ジストロフィー) である場合, $\langle\langle$ 病気 x が, \langle 犠牲者 y を, 襲う \rangle という表現は一般に奇妙である. これは HFN が必ずしも「襲う」という語の使用条件を特

定しているわけではないということである。

2.3 心理実験

1 フレームにつき 3 個の文を新たに作例し、合計 36 を心理実験の提示文とした (36 文は a,b,c の 3 セットに分割した)。**実施した実験は、カード分類課題、意味素性評定課題の二種**である。後者の実験については、再現性の確認のために同一条件の実験を二つの異なる被験者集団について行った。以下に提示した文を示す (ここでは a セット文のみを示す):

- (2)F01a 二人組の強盗がその銀行を襲った;
- F02a 政治的に孤立した国が隣国を襲った;
- F03a 二人の組員が敵対する組長を襲った;
- F04a 通り魔がその小学生を襲った;
- F05a ストーカーがその女性を襲った;
- F06a サメが傷ついたイルカを襲った;
- F07a スズメバチの大群が子供たちを襲った;
- F08a 大洪水が東海地方を襲った;
- F09a 突風がテレビのリポーターを襲った;
- F10a 悪性のインフルエンザがわが国を襲った;
- F11a 大型の不況がその国を襲った;
- F12a 言いようのない不安が彼を襲った

「動詞、主語句、目的語句の三つの組み合わせのみが意味フレームを特定する」という FN の重要な洞察を反映させるため、名詞句の組み合わせは乱化してある。

素性評定課題は、被験者に意味特徴に関する判断を「1: まったくそう思わない」から「5: 強くそう思う」までの五段階で評定するよう求める手法である。使用された 15 の評定項目は以下の通りである:

- (3) a. 襲い手は生物である;
- b. 襲い手は止むを得ず襲いかかった;
- c. 被害は直接感じられる;
- d. 襲う相手はあらかじめ決まっていた;
- e. 襲い手は気象現象である;
- f. 襲い手は道具を使って被害を与える;
- g. 襲い手の存在は感じ取れる;
- h. 被害の規模は個人/個体を越える;
- i. 襲い手には生命感がある;
- j. 襲い手は集団をなしている;
- k. 被害の受け手は死ぬこともある;
- l. 襲い手は目に見える;
- m. 被害は意図されたものである;
- n. 襲い手は人間である;
- o. 襲い手はあらかじめ襲う準備をしていた

2.4 心理実験結果の多変量解析

次に第一段階の人手解析と行動指標データが多変量解析の結果がどれほど一致するかどうかを検討した。外れ値を除去し、値を平均化したデータをクラスタ分析、非計量多次元尺度法 (MDS)、主成分分析 (PCA) で解析した。いずれの方法でも最下位フレームの分離に関して概ね同じ結果が得られている (素性評定課題で相関の高い 5 項目は除去した)。ここでは素性評定実験の結果を階層クラスタ分析 (Ward 法) し、その最適クラスタ化 [11 あるいは 12 個] での 36 文の帰属集団 G01-G11 を示す (*は項目の帰属先が予想外だったことを示す)。

- (4) a. G01 = F01,F03[暴行]: { G01.1 = F03[強奪]: {01a, 03a, 01b, 01c}, G01.2 = F01[抗争/紛争]: {02a, 02b, 02c, 03c} },
- b. G02 = F04[虐待]: {04a, 04b, 04c},
- c. G03 = F03 の下位フレーム [暴動]: 03b},
- d. G04 = F05[強姦]: {05a, 05b, 05c},
- e. G05 = F06[動物の襲撃 (捕食)]: {06a, 06a},
- f. G06 = F07[動物の襲撃 (非捕食)]: {07a, 07b, 06c*, 07c},
- g. G07 = F08,F09[異常気象の発生 (超個人規模 1)]: {08a, 08b, 09b, 09c},
- h. G08 = F11[活動への打撃: 異常事態の発生 (超個人規模 2)]: {09a*, 11a, 11b, 11c},
- i. G09 = F10[疫病の流行: 異常事態の発生 (超個人規模 3)]: {10a, 10b, 08c*, 10c},
- j. G10 = F12[発症: 異常事態の発生 (個人規模 1)]: {12a, 12c},
- k. G11 = F12[発症: 異常事態の発生 (個人規模 2)]: {12b}

なお、他の手法でもほぼ同様の集団化が得られている。

3 結果の評価と結論

(A) 言語学者によるフレームの階層的記述というの課題、心理実験としての (B) カード分類課題、(C) 意味素性評定課題の三つは互いに独立した課題であるが、PCA、階層的クラスタ分析、MDS の結果の任意の二つをそれぞれ比較しても、(心理実験の常識を超えるほど) 高い一致が得られた。これは (B、C) の結果の再現性の確認のために実験を追加しても維持された。これが示唆することは、**表面的に異なる (A、B、C) の課題が同一内容の課題で、同一の**

対象が記述されているということである。

以上の結果から、次のように結論する: (1) 意味フレームは心理的に実在し, (2) 言語データの人手分析 (あるいは格フレームの自動獲得) から同定可能であり, (3) FN は到達可能な目標を目指している。と同時に, (4) (結果の相互検証の効果から) 意味素性評定課題はフレーム同定のために有効な手法である。カード分類の課題依存性や被験者の無意識の均衡化の傾向を考えると, 意味素性評定課題の課題中立性は意味フレームの同定という目的には好ましい。以上のことから, (5) FN は自然言語の意味構造の体系的・網羅的記述の第一歩を可能とし (6) 長期的にそのような記述が充実すれば, それは意味研究に重要な研究資源を提供すると言える。

参考文献

- [1] C. J. Fillmore, C. Wooters, and C. F. Baker. Building a large lexical databank which provides deep semantics. In *Proceedings of the Pacific Asian Conference on Language, Information and Computation*. 2001.
- [2] 中本敬子, 野澤元, 黒田航. 動詞「襲う」の多義性: カード分類課題と意味素性評定課題による検討. 認知心理学会第二回大会口頭発表論文集, p. 38, 2004. [<http://clsl.hi.h.kyoto-u.ac.jp/~kkuroda/papers/Nakamoto-et-al-CogPsy20%04-Original.pdf>].
- [3] 黒田航, 井佐原均. 日本語の意味タグ体系を定義する試み: FrameNet の視点から. 言語処理学会第 10 回年次大会発表論文集, pp. 148–151. 言語処理学会, 2004. [増補改訂版: <http://clsl.hi.h.kyoto-u.ac.jp/~kkuroda/papers/jfn-nlp10-rev4.pdf>].
- [4] 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.
- [5] 河原大輔, 黒橋禎夫. 用言と直前の格要素の組を単位とする格フレームの自動獲得. 自然言語処理, Vol. 9, No. 1, pp. 1–16, 2002.