

「逃れる」の階層的意味フレーム分析とその意義

「言語学・心理学からの理論的、実証的裏づけ」のある言語資源開発の可能性

中本 敬子*

黒田 航†

1 はじめに

本研究では、フレーム指向概念分析 (Frame Oriented Concept Analysis of Language: FOCAL) [11, 4] に基づく“y が x { を; から } 逃れる”の意味フレーム分析の結果を報告する。本研究の提案は下記の通り: (a) NLP 技術の成果 (表層格フレーム辞書) を利用すれば言語学者による人手解析の省力化が可能, (b) 人手解析とその妥当性確認のための心理実験を行うことで小規模ながら質の高い資料を構築することが可能, さらに, (c) 表層格フレーム辞書に対し直接 (b) の手法を適用することで意味資源に変換できる可能性がある。

1.1 階層意味フレーム分析の概要

階層意味フレーム分析 (Hierarchical Frame Network Analysis; HFN(A)) は黒田らによって考案されたフレーム指向概念分析 (FOCAL) の一手法である。基本的には、次の分析手続きからなる。

(a) 動詞を中心にコーパス事例を偏りなく収集し, (b) 状況レベルの意味フレーム (e.g., 〈〈捕食者〉が, 〈獲物〉を, …, 〈襲う〉〉) として, その動詞および共起する語が実現している文脈に最適化された意味を手作業でコード化する。このコーディングの基本精神は次の意味で「発見」的なものである: 「辞書」的な意味を, 実際に使われている語句の意味に (強引に) 当てはめるのではなく, 逆に辞書的な意味記述を仮定せずに文脈に特殊化した意味を同定し, これを通じて得られた記述を帰納的に一般化することで辞書的な意味を得ようとする。この際, 意味型 (semantic types) の指定と意味役割 (semantic roles) の指定を区別する (この区別の詳細は [10] を参照のこと)。具体的には, 共起語 (主に主語, 目的語などの名詞句) の「自然類」的な意味型 (の階層) (e.g., [インパラ] IS-A [草食哺乳動物] IS-A [哺乳動物] ...) と, 状況相対的に定義される意味役割 (の階層) (〈インパラ〉 IS-A 〈獲物〉 IS-A 〈被害体〉 ...) をおのおの同定する。意味フレームの区別と意味役割の同定は, 動詞の項となっている名詞句の意味型の共変関係を元に特定されると言ってもよい。

更に, (c) 意味フレーム同士を“具体化: [上位フレーム] → [下位フレーム] OR [具体例]”, “抽象化: [下位フレーム] OR [具体例] → [上位フレーム]” の関係を指定し, ラティス構造を得ることによって関係づける。これによって, 意味フレームを単なる用例の分類ではなく, 互いに関係づけられた, 異なる粒度での意味記述を可能にする資料として体系づけることができる。また, 並行して, 意味フレームの内実を, フレーム間の弁別を可能にする意味素性の集合として記述する。

FOCAL では, 更に (d) HFN(A) が基本的に少数の言語学者の直観に基づいていることから, 結果の妥当性を確認するた

めに一般話者の直観との整合性を確認する心理実験を行っている。

1.2 格フレーム辞書の利用

これらから明らかな通り, 先行研究での HFN(A) には次のような難点があった。

一つ目は, 非常に労力と時間を要するという点である。偏りのないサンプルに対し解析を行うにはある程度数の (数百程度) の文を扱う必要がある。しかも, サンプル収集の対象となるコーパスにどのようなフレームがどのような分布で含まれているか不明なため, サンプルの抽出は全数, あるいは適当な数を指定しての無作為抽出となる。しかし, このように得たサンプルには類似の文が多数含まれており, しかもそれは人手で解析するまで分からない。これは作業の非効率性につながっている。

二つ目は, 逆に数百のサンプルに全てのフレームを具現化する文が含まれているとは限らないことである。サンプルの規模を大きくすることは理屈では可能だが, 実際の労力や分析者の注意力, 記憶力の限界を考えると, それは現実的ではない。この点については, 分析者の直観によりサンプルには含まれていないがあり得るフレームの存在を予測し, 作例や Web 検索によりその存在を確認するという手段が取られてきた。だが, 直観は常に揺れを含むので, 不確実な直観による予測誤差を減らすことができれば, それに越したことはない。

これらの問題点を克服するために, “逃れる”の HFN(A) では, コーパスからの事例の分析に加え, 表層格フレーム辞書 [12] を資料として参照した。

格フレーム辞書を参照することには以下のような利点がある。まず, 格フレーム辞書は動詞の直前格要素の意味属性の類似性によって用例パターンを分類することで作成されている。意味属性の決定は, NTT 日本語語彙大系 [6] に基づいてなされている。語彙大系は一種のソーラスであり, これは格要素の意味型の類似性により用例を分類する試みと見なしうる。つまり, 格フレーム辞書は特定の格要素の意味型に基づいて用例を分類している。この特性が, HFN(A) に格フレーム辞書を利用する第一の効用につながる: つまり, 格フレーム辞書は HFN(A) で問題となる類似用例の重複に対し, (大まかではあるが) ある格要素の意味型での一致による分類という点で一つの解決を与えている。意味型と意味役割が排他的でない点に関しては [10] を参照されたい。

HFN(A) での対象用例数の問題に関しては, 表層格フレーム辞書は計算機による自動化に基づいて構築されていることから, 大量のデータを扱っていることは間違いない。そのため, これを参照することは, 人手解析で扱える標本数の限界を間接的に解消すると考えられる。

* 京都大学教育学研究科

† (独) 情報通信研究機構 けいはんな情報通信融合研究センター

表1 “逃れる”の理解に利用される意味フレーム群

フレーム	例文
F01a 発見の回避	黒ヒョウが密猟者の捜索を逃れた
F01b 追跡からの逃げ切り	インパラの群れが追ってくるライオンを逃れた
F01c 捕獲のかわし	キタキツネが猟師のしかけた罠を逃れた
F02a 発覚の回避	飲酒運転の車が深夜の検問を逃れた
F02b 追跡からの逃げ切り	新宿のカジノの経営者が一斉摘発を逃れた
F02c 逮捕の回避	潜伏中のテロリストが踏み込んできた公安を逃れた
F03a 発見の回避	ステルス戦闘機が相手基地のレーダーを逃れた
F03b 追跡からの逃げ切り	撤退中の戦車部隊が敵兵の追撃を逃れた
F03c 攻撃の回避	戦闘中の兵士がライフル兵の狙撃を逃れた
F04 非戦闘員が敵の攻撃を逃れる	アフガンの人々が米軍の空爆を逃れた
F05a 義務の発生の回避	借金持ちの親戚を持つ女性が連帯保証を逃れた
F05b 責任負担からの逃げ切り	不祥事を起こした会社の社長が追求を逃れた
F05c 責任取りの回避	大学に進学した若者が兵役を逃れた
F06 脱獄	懲役中の凶悪犯が独房を逃れた
F07 脱走	強制入院中の患者が隔離病棟を逃れた
F08 逃亡	14歳の少女がすさんだ家庭を逃れた
F09a 人為災害からの避難	農夫達が長引く故郷での内乱を逃れた
F09b 自然災害からの避難	山村の人々が突然の山火事を逃れた
F10a 自然災害からの避難	サル群れが火山の噴火を逃れた
F10b 人為災害からの避難	中国西部のカモンカが乱獲を逃れた
F11 脱却	自立した女性が古くからの慣習を逃れた

2 “yがxを逃れる”の階層意味フレーム分析

HFN(A)は、まず、日英対訳コーパス [8] から収集した動詞“逃れる”を含む全ての文 86 例を対象として始めた。まず、従来の方法と同じく、文内 (e.g., で〈逃れるモノ〉、〈避ける対象〉、〈逃げ元になる場所〉、〈逃れ先になる場所〉を表す文字列を取り出し、それらの意味役割を特定した。例を上げれば、“都会の児童が戦火を逃れた”は次の形の特徴づけを与えるられる:

- (1) 〈脱出: 〈Agent: 脱出者: 児童〉が、〈Purpose: 危険: 戦火〉を避けるため、〈Source: 危険な場所: 都会〉から、〈Goal: 安全な場所: NULL〉に、〈Gov: 逃れる: 脱出する〉〉

同様の分析を数多く行なうことにより、〈避ける対象〉は〈危険(なモノか場所)〉と〈不利(なモノか場所)〉に大別できることが明らかになった。その結果が図3にあるHFNである。

この分析を元に、格フレーム辞書を参照し、分析を進めた。具体的には辞書を参照し、格フレームの一覧から高頻度で特徴的な用例を中心にガ格、ヲ格、カラ格名詞句を拾い出した¹⁾。それを元にして作例により具体例を補いつつ、フレームの分化と統合を繰り返す作業を行った。格フレーム辞書の参照は、当初の86文には現れなかった意味フレームの存在を確認するだけでなく、辞書無しでの作例に比べ、分析者のバイアスを低減する働きがあると考えられる。また、実際にコーパスから得る生のデータに比べ、分析に必要な要素だけを取り出す手間が少なく、省力化につながるとされる。

これらの手順を経て特定されたフレームの一覧を、表1に示す。それらの元になったHFNを図3に示す。

基本的には11個のフレームが同定されたが、幾つかのフレームでは更に下位に分化する可能性があったため、合計で21個のフレームを検証の対象とした。

¹⁾ ただし、特徴性の判断の明示的な基準はなかった。この明示化は今後の検討課題である。

3 心理実験

3.1 言語材料

表1のフレームに対応する文を、“yがxを逃れた”形式で各フレーム2文ずつ作成した。また、フレーム弁別の意味素性を評定項目化した (e.g., 「y(逃れた側)はx(xが〈逃した側〉) (逃れた対象) によって命を失う可能性があった」)。

変数 x, y は “x₁ が y₁ を襲った” [4]。“y₂ が x₂ から逃げた” [3] の x_{1,2}, y_{1,2} にある程度合わせている (例えば, “ライオンがインパラの群れを襲った” ↔ “インパラの群れがライオンを逃れた”)。ただし, x_i と x_j, y_i と y_j との間には相関があるが、完全に一致するわけではない (例えば “逃げる” の〈具体物の流出〉(暖まった空気がドアの隙間から逃げた) に対応するフレームは “襲う” にも “逃れる” にもない)。更に、同じ表層形が使えるとしても、意味型の強要 [7] によって意味型が変更されている可能性もある。X = {x₁, x₂, x₃} , Y = {y₁, y₂, y₃} の包含関係を調べ、X, Y の間の相関を見ることも、“襲う”, “逃げる”, “逃れる” の分析で変数を揃えたことの目的の一つである。ただし、結果はまだ得られていない。

3.2 カード分類課題

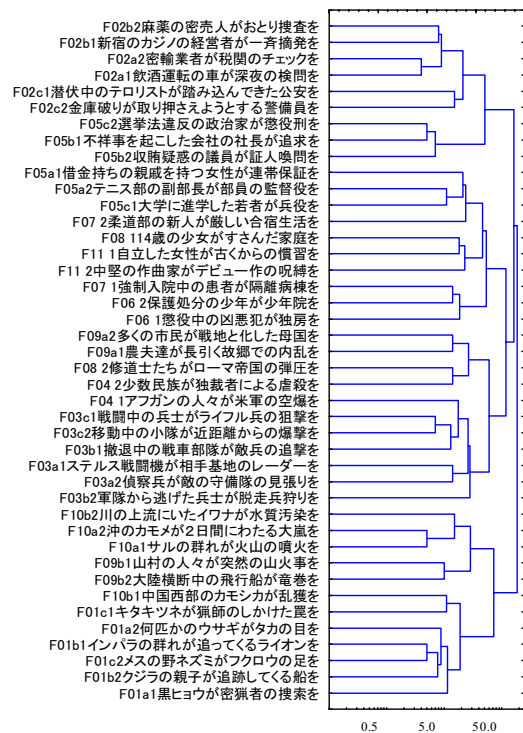


図1 “逃れる”のカード分類のクラスター

方法 大学生 37 名に材料文を 1 文ずつ印刷したカード 42 枚を 1 組ずつ配布し、文意の類似性に基づいてカードを任意の数のグループに纏めるよう指示した。

結果と考察 “逃げる”で表される状況の分類パターンを見るため、2つの文が同じカテゴリーに配置された頻度を類似性の指標と見なし、Ward法による階層クラスター分析を行った。結果として得られた樹形図を図1に示す。図1に示される通り、一般の日本語話者の直観的な分類は、HFN(A)とよく一致している。

3.3 意味素性評定課題

方法 意味素性を翻案し 33 個の評定項目について、各文が表す状況にこれらの項目が当てはまる程度を 5 段階 (1. 全くそう思わない — 5. 強くそう思う) で評定するよう求めた。80 名の被験者が実験に参加した。各被験者は 21 文に対し、16 個もしくは 17 個の素性評定項目に解答した。

結果と考察 例文ごとに、各評定項目の被験者間平均を算出した。文間のばらつきの小さい 4 項目を除き、市販ソフト Viscovery SOMine 4.0 を用いて自己組織化マップにより文の類似性を視覚化した。マップの大きさは 31 × 25 とした。結果を図 2 に示す (歪み測度 0.0113; 量子化誤差 0)。図から明らかな通り、素性評定からも上手く意味フレームの意味的布置が得られている。

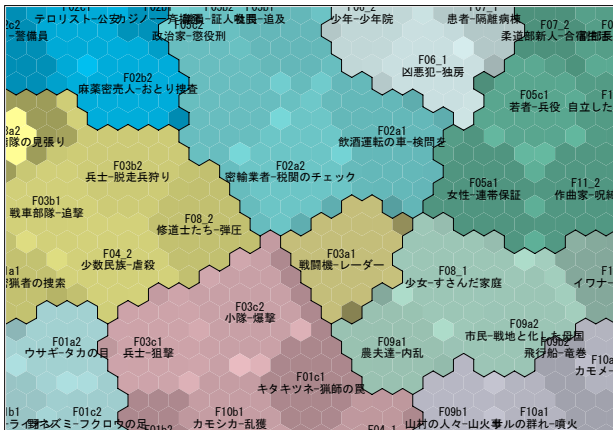


図 2 素性評定値に基づく“逃れる”の自己組織化マップ

4 総合考察

以上の結果は次のことを示している:

- (2) a. 従来の HFN(A) に格フレーム辞書の参照の利用を加えたことで分析が効率化され、
- b. 得られた HFN は従来の解析結果と同様に心理実験の結果と適合している。

この結果を元に、以下で簡単な議論を加える。

4.1 格フレーム辞書利用についての評価

本研究での HFN(A) 構築に利用した生の事例は、先行研究“襲う” [4] 約 450 例，“逃げる” [3] 約 200 例に比べると、86 例と圧倒的に少ない。にも関わらず、特定されたフレームは 11 個 (さらに下位のレベルならば 21 個) と数多く、それらの妥当性を心理実験によって裏付けることができた。このことは、表層格フレーム辞書 [12] の用例格フレームを適切に解釈することができれば、分析を効率化し、かつ分析の精度を損なわなくともよいことを示唆する。さらに言えば、表層格フレーム辞書を直接の分析対象にできれば、生コーパスからの事例収集や意味役割を実現している文字列の切り出しといった HFN(A) の“下処理”の過程を完全に省略できるかも知れないことを示唆している。逆の観点から見れば、これは表層格フレーム辞書を言語学者が適切に解釈することで、意味フレームのデータベースという形の意味資源に変換できる可能性があるということである。

4.2 HFN(A) のこれまでの成果: 先行研究との比較

本研究により、コーパスから帰納的に発見されたフレーム分析と心理実験による検証を行った HFN は、“襲う” [4]，“逃げる” [3] に続き 3 つとなった。これまでの成果について簡単にまとめてみる。

まず第一に、これら 3 つの分析はいずれも Berkeley FrameNet (BFN) [1] を参照せず、独自に構成された。これには二つの理由がある。第一に、BFN の意味記述の粒度が私たちが意味記述において目指しているものほど細くないこと (詳細な比較は [2] で行われた)。第二に、BFN の日本語版である日本語フレームネット [9] が利用可能なデータを提供していないことによる。

先述の通り、HFN(A) は数名の言語学者による人手解析だが、少なくとも一般話者の直観と一致するという意味では心理実験によって妥当性が保証された。これは、HFN(A) に基づいた概念分析が単なる一個人の主観ではなく、一般性を持つことを示している。逆に言えば、HFN(A) を行える分析者を確保できれば、より多くの語に対して同様の解析を行い、状況レベルの意味フレームとその実現例を意味資料として蓄積できる可能性があることになる。HFN(A) の具体的な手順は簡単な手引き [5] が存在するのみだが、コーディング済みの事例の公開といった支援も行われており、今後、これらの資料を拡充し、HFN 分析者の訓練・養成法を開発することは不可能ではないかも知れない。

第二の点として、意味フレームの内容を記述するために心理学的手法を利用することへの示唆が得られてきたと言える。言語表現から人が何を理解しているかを記述する試みは、これまでのところ決して多いとは言えない。本研究および先行研究で利用したカード分類と意味素性評定という課題自体は、心理学的手法としては極めて頻繁に用いられるが、意味の問題そのものに適用した例は (特に日本語では) さほど多いとは言えない。しかし、一連の研究は、そのような心理学的手法の応用が意味記述に貢献しうる可能性を示していると言える。さらに、本研究では、文の読み手が平均的に理解する内容の記述に特化するのであれば (つまり、個人差などを検討しないのであれば)、課題を多くの被験者に分割して課しても、全てを同一被験者に行わせた場合と、同じような結果が得られるであろうことを示している。このことは、例えば、被験者一人あたりの負担を減らせることを意味し、一般に向けて Web を通じた評定の実施するといった新たな心理データ取得方法の可能性を示唆する。

第三に言えることは、HFN(A) が従来のシソーラスを越えたレベルで自然な分類を見出しているという点である。格フレーム辞書は日本語語彙体系の概念体系 C を利用している。従って、今回の HFN は間接的に C を利用していることになるが、得られた結果は必ずしも、 C から予測できるものではない。特に HFN は次のような意味役割によって“ y が x を逃れる”の意味分類を行なっている点に注意が必要である:

- (3) a. F01: 〈脱走者 y 〉: が 〈搜索者: x 〉 (の 〈搜索: x^* 〉) を逃れる
- b. :
- c. F09,10: 〈受難者 y 〉: が 〈災難: x 〉 (の 〈影響: x^* 〉) を逃れる

このような形で、HFNA は外延の集合で定義された格フレーム辞書を意味役割という内包で再定義し、その有意義な

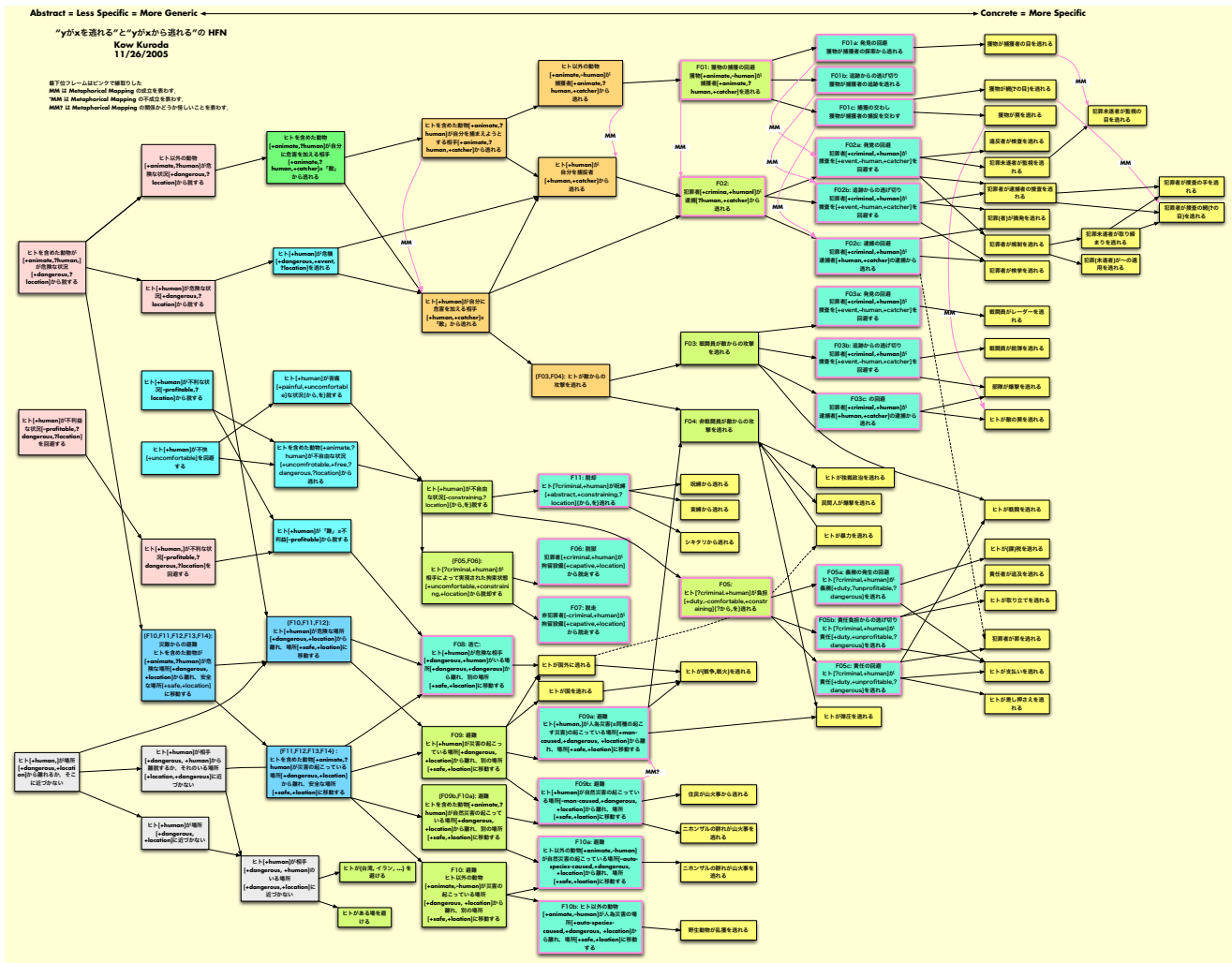


図3 “逃れる”のHFN (実験試料の作成のための版)

「解釈」を提供することになる。これは NLP 自体にとっては特に嬉しいことではないかも知れないが、言語学、言語心理学を含めた言語の認知科学への寄与は大きい。このような再解釈のプロセスを通じて格フレーム辞書の有用性が理解され、それを継起にした連携が確立することで、今後 NLP と関連領域との意見交換が活発化することは、決してありえないことではない。

参考文献

[1] FILLMORE, C. J., JOHNSON, C. R. and PETRUCK, M. R. L. Background to FrameNet, *International Journal of Lexicography*, 16, 3 (2003), 235–250.
 [2] KANAMARU, T., MURATA, M., KURODA, K. and HITOSHI, I. Obtaining Japanese Lexical Units for Semantic Frames from Berkeley FrameNet using a Bilingual Corpus, *Proceedings of the 6th International Workshop on Linguistically Interpreted Corpora (LINC-05)* (2005), 11–20.
 [3] 中本敬子, 黒田航 「y が x から逃げる」の理解内容の階層的意味フレーム分析: コーパスの人手解析と心理実験を通して (2005).
 [4] 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, 3, 1 (2005), 65–81.

[5] 中本敬子, 黒田航, 野澤元, 龍岡昌弘, 金丸敏幸 FOCAL/PDS 入門: フレーム指向語彙概念分析/並列分散意味論の具体的な紹介. [未刊]
 [6] NTT コミュニケーション科学研究所 日本語語彙大系, 岩波書店, 東京 (1997).
 [7] PUSTEJOVSKY, J. *The Generative Lexicon*, MIT Press (1995).
 [8] 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*, 10 (2003), 201–220.
 [9] 小原京子, ほか 日本語フレームネット: 意味タグつきコーパスの試み, *言語処理学会第 11 回大会発表論文集* (2005).
 [10] 黒田航, 井佐原均 意味役割名と意味型名の区別による新しい概念分類の可能性: 意味役割の一般理論はシソーラスを救う?, *信学技報*, 105 (204) (2005), 47–54.
 [11] 黒田航, 中本敬子, 野澤元 意味フレームに基づく概念分析の理論と実践, *認知言語学論考第 4 巻* (山梨正明ほか (編)), ひつじ書房 (133–269).
 [12] 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*, 9, 1 (2002), 1–16.