

自動構築された格フレーム辞書の精度向上を見こんだ人手評価

黒田 航 李在鎬 渋谷 良方 河原 大輔 井佐原 均

(独) 情報通信研究機構 知識創成コミュニケーション研究センター

{kuroda, jhlee, yshibuya, dk, isahara}@nict.go.jp

1 はじめに

1.1 目的

ウェブに存在する文書から自動構築された大規模格フレーム辞書 [5, 6] には大きな期待が寄せられている。その精度向上を見こんで、言語学者が述語 (= 用言) の用例クラスターの結合精度を評価し、それに基づいて用例結合の優先順位の設定法に関する改良案の提案を行う。実装された案については精度向上の効果があつたかどうかを評価する実験を行い、その結果を報告する(ただし評価実験は執筆時点で未完)と共に、実装されてはいるが精度向上に繋がらなそうな、現行の格フレーム辞書の問題点も幾つか指摘する。

1.2 方法

従来の評価は、適当な課題 (e.g., 格解析) を設定し、格フレーム辞書の利用した条件で得られた結果と使用しない条件で得られた結果 (ベースライン) を比べ、ベースラインよりも得られた結果が向上したことをもって、格フレームの利用が有用だとするもので、辞書の記述の精度を直接評価したものではない。これは評価方法としては間接的であり、どんな変更が、どう辞書の品質の向上に貢献したかを見るのは難しい。格フレーム辞書の記述精度をもっと直接的に評価するため、私たちは次の二つを行った:

- (1) 用例結合法の見直しとして、§2.2 で後述の定性的手法と定量的手法で、述語ごとに共起する格の優先順位を決め、この順序に従って構築された用法クラスターを使った評価実験の精度を、元々の直前格を優先する手法で構築された用法クラスターの精度と比較し、期待された精度向上の効果が認められるか判断した。
- (2) 言語学者が、一定の基準で用法クラスターを幾つかの等級にわけ、その割合と誤りの傾向を調べた。

(1) の評価は [5] が評価に使った 20 個の述語¹⁾ に「襲う、逃げる、逃れる、避ける (「さける」「よける」の混合)、防ぐ、守る」を加えた計 26 個の述語に対して、(2) の評価は 16 個の述語²⁾ に対して行った。(1) の結果については §2 節で、(2) の結果については §3 節で詳細を述べる。

1.3 お断り

(1) では、提案された結合順位の変更によって格フレーム辞書構築の精度の向上の効果があつたかを判断する評価実験が

必要だが、被覆率が低さが問題にならないような評価方法を考える必要がある。このために、今回評価を済ませた 16 個の述語を含む文からなる評価用のデータセットを用意し、これに対する独立の実験 (e.g., 格解析) の精度の変化を見て、それを判定することにしたが、この実験は残念ながら論文の完成には間に合わなかった。本番までに間に合わせたい。

2 格要素の結合優先順位の変更の効果

2.1 背景

述語の用例のクラスター化では、どの格要素から順番に結合させるかを決める必要がある。[5] では、表層で述語の直前に現れている格 (これは直前格と呼ばれる) が最優先される用例結合を行って格フレーム辞書を構築しているが、これは唯一のクラスター化の結合優先順位の設定法とは言えず、他の決定法 (例えば統計量) を使ったクラスター化の結果と比較しない限り、妥当性を正しく評価できない。少なくとも頻度効果を考慮に入れた結合を考える余地は十分にある。例えば現行の手法では、用例結合の是非を決める閾値は一定である。これには工夫の余地があると思われるし、その必要は実際、§3 で説明する解析結果からも示唆されている。

これとは別に、(NLP 業界では必ずしも信頼の高くない) 言語学者の直観では、述語の項 (= 必須格) の間には内在的な重要度の違いがあり、それを反映するような用例の結合ができれば、もっとよいクラスター化が結果するように思える。

2.2 提案の内容

以上の理由を踏まえて結合法の改良のために提案された方法は、次の定性的、定量的手法の二つに大別できる:

- (3) 言語学者が基本格、準基本格を認定し、それらを順位づけによって結合の優先順位を決定 [定性的]
- (4) 基本格、準基本格のような区別は設定せず、統計量に基づく単純な判断基準で結合の優先順位を決定 [定量的]

(4) に関しては、「足切り」の設定法に関して、次のような幾つかのバリエーションがある:

- (5) 格要素 C の最大の延べ頻度を $N(C)_{\max}$ としたとき、
 - a. 延べ頻度が大きい順に適当な数 M 個の格を選び (それ以外のものは「足切り」し)、 M 個について延べ頻度の小さい順に結合する
 - b. $N(C)_{\max}$ の何割かに満たない格を切り捨て、残りを延べ頻度の小さい順に結合する
 - c. 適当な θ 個以下の格を切り捨て、残りを延べ頻度の小さい順に結合する(5b), (5c) については何個の候補ができるかは不明。

(4) は手法としては (3) とは別だが、(3) の作業過程で有効性が気づかれたものである。

いずれにも共通しているのは「格要素の結合の優先順位は、 θ 個以下の格を切り捨て、残りを延べ頻度の小さい順に結合する」という想定である。これはあらかじめ

¹⁾ 1. 著しい (A), 2. 好評だ (B), 3. 積極的だ (B), 4. 大半だ (B), 5. 問題だ (B), 6. 賛成だ (B), 7. 積む (V), 8. 解く (V), 9. 折る (V), 10. 冷やす (V), 11. 味わう (V), 12. 裏切る (V), 13. 復元する (V), 14. 躍進する (V), 15. 恐れる (V), 16. 飾る (V), 17. 寄与する (V), 18. 心掛ける (V), 19. 散る (V)

²⁾ 1. 固まる (V), 2. 散る (V), 3. 賛成する (V), 4. 逃げる (V), 5. 逃れる (V), 6. 躍進する (V), 7. 寄与する (V), 8. 解く (V), 9. 飾る (V), 10. 恐れる (V), 11. 心掛ける (V), 12. 裏切る (V), 13. 襲う (V), 14. 積む (V), 15. 折る (V), 16. 冷やす (V).

表1 人手評価で決めた優先順位: n=項=基本格; qa=準項=準基本格; na=非項=非基本格; (k) は優先順位が k であることを意味する. 優先順位指定のつかなかった格 (ニトツテ, 二続イテ, ヨリ, 〈無格〉, 〈外の関係〉, 〈時間〉) は割愛した

	固まる	散る	賛成	逃げる	躍進	寄与	解く	飾る	恐れる	心掛ける	裏切る	襲う	積む	祈る	冷やす
ガ	a (4)	a (4)	a (1)	a (1)	a (1)	a (1)	a (1)	a (1)	a (1)	a (1)	a (1)	a (1)	a (1)	a (1)	a (1)
ガ2	a (4)	a (4)	na	na	na	na	na	na	na	na	na	na	na	na	na
ヲ	na	na	na	a (3)	na	na	a (2)	a (3)	a (2)	a (4)	a (2)	a (3)	a (7)	a (2)	a (3)
ヲ通ジテ	na	na	na	na	na	na	qa (5)	na	na	na	na	na	na	na	na
ニ	a (3)	a (2)	a (2)	a (5)	a (3)	a (3)	qa (3)	a (2)	na	a (2)	na	na* (2)	a (3)	qa (4)	a (2)
ニ対シテ	na	na	a (2)	na	na	na	na	na	na	na	na	na	na	na	na
ニツイテ	na	na	a (2)	na	na	na	na	na	na	a (2)	na	na	na	na	na
二關スル	na	na	a (2)	na	na	na	na	na	na	na	na	na	na	na	na
ニヨツテ	qa (2)	na	na	na	na	na	na	na	na	na	na	na	na	na	na
へ	na	a (1)	na	a (4)	a (2)	a (2)	na	a (2)	na	na	na	na	a (2)	qa (4)	na
カラ	qa (1)	na	qa (3)	a (2)	qa (5)	na	na	na	na	na	na (4)	na (4)	na	na	qa (4)
マデ	qa (1)	na	na	a (5)	qa	na	na	na	na	na	na	na	qa (4)	na	na
デ	qa (2)	qa (3)	na	qa	qa (4)	qa (4)	qa (4)	qa (4)	na	na	na	qa (5)	qa (5)	qa (5)	qa (5)
ト	na	na	na	na	qa	na	na	na	na	na	a (3)	na	na	qa (3)	na
トシテ	na	na	qa (3)	na	na	na	na	na	na	na	na	na	na	na	na
トイウ	na	na	qa (4)	na	na	na	na	na	na	na	na	na	na	na	na
ノ	na	qa	na	na	na	na	na (5)	na	qa	na	qa	qa (6)	qa (6)	qa (6)	qa (6)

め予見されていたことではなく, [3, 2] が得た結果. すなわち「[x が y を襲う] と [y が x がに襲われる] では x の方が y よりも用例の体系への説明力が高い」という事実に基づいた用例分析の過程で洞察されたものである. 「頻度の高いものほど多義的である」⇔「頻度の低いものほど一義的である」という一般的な性質があるので, 格要素になる名詞に関して, 多義性の影響がなるべく少なくなるように結合すれば, 全体の精度が向上するのではないかとというのが洞察の内容だった. これが正しい予想であるとすると, 述語が複数の項をもつ時, それらに許される曖昧性の幅には違いがあり, それは述語のタイプに対応しているらしいということが予測が得られる. この妥当性を検討することは理論的にも経験的にも興味深い.

2.3 結合法の改良のための提案

以上の背景から, 用例結合法の改良案として (7) を提案し, それを従来の手法 (6) と比較した.

- (6) 項, 準項, 非項の区別は想定せず, 特に頻度効果も考慮に入れず, 直前格から結合する
- (7) 標準化された格要素 $[C_1, C_2, \dots, C_n]^3$ について, C_i の要素にある名詞 $w_1/C_i, \dots, w_k/C_i$ ごとの頻度 $f(w_1)/C_i, \dots, f(w_k)/C_i$ の総和を格の延べ頻度 $N(C_i)$ とし,
 - a. 項, 準項, 非項の区別を想定し, 項と準項を延べ頻度 $N(C)$ の小さい順に結合する [定量的で定性的]
 - b. 項, 準項, 非項の区別は想定せず, 延べ頻度 $N(C)$ の小さいものから結合する [定量的で非定性的]
 - c. 項, 準項, 非項の区別は想定せず, 単純に一定一定頻度に満たないものを切り捨てた上で, 延べ頻度 $N(C)$ の小さいものから結合する [定量的で非定性的]

(7a) の結果は表 1 の通りである. 〈無格〉, 〈外の関係〉, 〈時間〉はいずれも na で優先順位は設けられなかった. 述語ごとに最大のべ頻度をもつ格とその頻度を表 2 に示した. これから述語のだいたいの頻度が推定できる.

項, 準項, 非項の定義は簡単ではないが, 概略, 項=必須格, 非項=任意格, 準項=必須格のような挙動をする任意格 (ノ格やデ格) と考えてよい. 定量的手法は (期待通りの効果を持つ限り) この問題を回避できる分だけ信頼できると言える⁴⁾.

³⁾ 標準化された格要素 $[C_1, C_2, \dots, C_n]$ は [ガ格, ガ2格, ヲ格, ヲ通ジテ格, 二格, 二対シテ格, ニツイテ格, 二關スル格, ニヨツテ格, ニトツテ格, 二続イテ格, へ格, カラ格, マデ格, デ格, ヨリ格, ト格, トシテ格, トイウ格, ノ格, 〈無〉格, 〈外の関係〉, 〈時間〉] である.

⁴⁾ (7b) と (7c) の有効性は, (7a) を定義する過程で気づかれたものである. 実際, この意味では (7b) と (7c) は (7a) の近似と見なせる.

表 2 述語と最大頻度をもつ格

述語	頻度最大の格	頻度	品定め
開く	ヲ格	261933	
守る	ヲ格	260402	
防ぐ	ヲ格	112890	
避ける	ヲ格	77161	
問題だ	ガ格	55233	
飾る	ヲ格	52790	
味わう	ヲ格	49046	
解く	ヲ格	40472	1
積む	ヲ格	39288	1
襲う	ヲ格	30516	1
恐れる	ヲ格	26676	1
寄与する	二格	23576	1
裏切る	ヲ格	20434	1
祈る	ヲ格	12675	1
逃げる	二格	10169	1
冷やす	ヲ格	9646	1
大半だ	ガ格	8402	
心掛ける	ヲ格	7933	1
著しい	ガ格	7915	
固まる	ガ格	7915	1
逃れる	ヲ格	7882	1
賛成する	二格	7646	1
好評だ	ガ格	6569	
散る	ガ格	5639	1
復元する	ヲ格	4063	
積極的だ	二格	2259	
躍進する	二格	387	1

2.4 結果と優先順位の決定作業を通じて得られた知見

(7) で得られた結果は <http://nlp.kuee.kyoto-u.ac.jp/nl-resource/caseframe/NLP2007/> に公開した. 定量的評価は現在も進行中である.

これに先立つ人手による 26 個の述語の格要素の優先順位づけの過程で次の知見が得られた:

- (8) a. ガ格は他動詞と他述語で同じ挙動をしない. 自動詞のガ格の一部は他動詞のヲ格に似た振る舞いをする [非対格動詞の存在から予想可能]
- b. 形容動詞, 形容詞のガ格と他動詞のガ格は同じ挙動をせず, 自動詞のガ格の一部と (従って他動詞のヲ格と) よく似た振る舞いをする
- (9) 「x が y の z に寄与する」のように疑似補文のような項を取る動詞ではノ格を項扱いする必要がある.

ノ格は他の多くの述語では同じ基準で結合すると精度を下げる可能性が高いので, 見極めが必要である. この意味でも述語のクラスの別に応じた結合法を工夫する必要がある.

3 言語学者による用法クラスターの品定め

3.1 用法クラスターの区別

従来の研究を見ると, 格フレーム辞書が表現すべき情報とは何か明確でないように思われる. これが決まらないうちに辞書の精度の結果を直接評価する方法を定義できない.

[1] の示唆に従うならば、「うまく行った」述語の用例クラスターとは次のようなものである：

- (10) クラスターが「うまく行った」格フレーム辞書項目とは、 $[C_1, \dots, C_n]$ を総合的に判断して、[1] の言う意味での特定の状況を表わすと判断できるクラスターである。

この基準の下で用法クラスターは次の四つに大別される：

- (11) A. 「ゴミ」の混じっていない理想に近いクラスター
 B. 少数の「ゴミ」が混じっているが、それらを取り除けば A になるクラスター
 C. 異なる種類の「ゴミ」が数多く混じっており、適切に純度を保つことの難しいクラスター
 D. 意味不明なクラスター

「ゴミ」とは用例クラスター C の格要素のどれかに、C が表わすのが S なら入っているべきではない要素のことである。

例えば「焼く」の用例クラスターに C_1 と C_2 と C_3 の三つがあり、それぞれのヲ格要素の一覧が (12) にある通りだとすると、 C_1 と C_2 はゴミのない用例クラスターだが、 C_3 は「世話」がゴミになり、 C_1 になり損なった格要素集合である⁵⁾。

- (12) a. C_1 のヲ格: {パン: a, 魚: b, ギョウザ: c}
 b. C_2 のヲ格: {世話: d}
 c. C_3 のヲ格: {パン: e, 魚: f, ギョウザ: h, 世話: h}
 ただし a, b, ..., h はおのおの頻度を表わす

理想的な格フレーム辞書とは A タイプの用法クラスターのみからなるクラスターの集合(か、これらの体系化)であるが、どの述語を見ても現状はこの理想からはほど遠い。現実的な対処法は、予備調査から比較的数量の多いとわかっている B タイプのクラスターを A タイプのクラスターに変換するための対策を取ることである。そのための実態調査を行うのが今回の評価の動機であった。

3.2 評価の結果と個別の傾向の分析

3.2.1 評価基準と結果

述語の用例クラスターは格要素集合ごとに (13) に示した通りに得点を与え、その合計点から、次のように分類した: 極上品 (>150), 美品 (150~125), 美良品 (125~100), 良品 (100~70), 並品 (70~50), 難あり (70~25), 評価不可 (25~0)。

- (13) 大で特良=85, 中で特良=80, 小で特良=75, 極小で特良=60, 大で良=70, 中で良=65, 小で良=55, 極小で良=50, 大で並=45, 中で並=40, 小で並=35, 極小で並=30, 大で不良=25, 中で不良=20, 小で不良=15, 極小で不良=10, 小で評価不可=5, 極小で評価不可=0。

16 個の述語の精度評価の結果をを 図 1 と 図 2 に個数グラフと割合グラフにして示す (個数グラフは述語の使用頻度を表わすわけではないことに注意)。

3.2.2 二格, デ格の脱曖昧化の向上の必要

印象的なのは述語ごとの精度のばらつきが大きいことである。だが、細かく見てゆくと次のようなことがわかった:

- (14) 述語の頻度のクラスター精度への影響は顕著ではない (表 2 参照)
 (15) ヲ格が強い述語はうまく用例がクラスター化されているが、ヲ格を取らない述語のクラスター化精度が低い

(14) を考慮に入れ (15) の結果を見ると、クラスター化精度はおそらく (格) 助詞の意味の広がり反比例していると言え

⁵⁾ C_3 は C_1 と C_2 の結合から得られるが、これは意味フレームの理論から見れば過剰般化に結果する結合である。

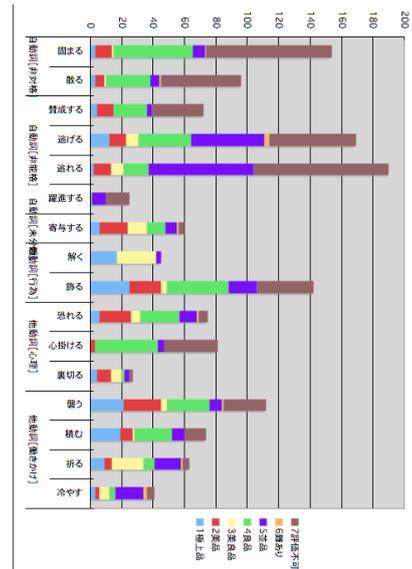


図 1 横軸=述語ごとのクラスターの個数 (品質別)

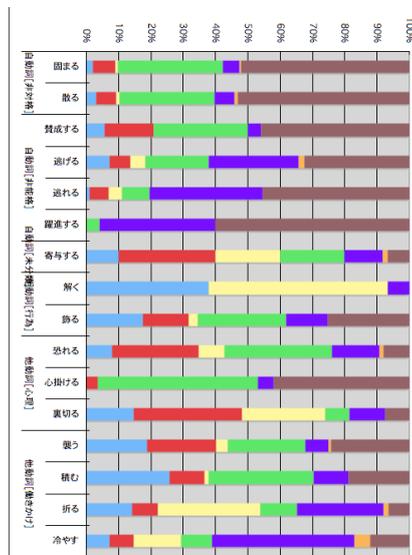


図 2 横軸=述語ごとのクラスターの相対度数 (品質別)

る。用例クラスターを実際に考察すると、曖昧性が著しいのは二格, デ格, ノ格であり、比較的曖昧性が低いのがヲ格, ガ格である。ガ格は曖昧性は低いが、省略率と八への吸収率が高く獲得される実例数が少ないため、精度の向上に貢献していないと思われる。

ヲ格を取る述語の精度が高いという傾向の例外は「逃れる」と「心掛ける」の二つだが、これは、それぞれの格支配が A. [x ガ, y カラ, z 二, 逃れる] か B[x ガ, y ヲ, (z 二), 逃れる], A. [x ガ, z 二, 心掛ける] か B[x ガ, y ヲ, 心掛ける] で B の用法が十分に強くないためであると思われる。

二格が関係する述語の精度が低い理由は二つある。一つには、用法によって二格が項に相当する要素だったり、しなかったりする (例えば「躍進する」も「x が y (e.g., n 位, z の地位, 一流企業) に躍進する」という常套句で二格を項に取る) こと、もう一つは二格要素として収集されているのは二で終わっている文節で二の項になっている名詞であり、必ずしも

二格名詞のみではない(これには「上手に」のような形容動詞連用形も含まれている)ということである。いずれの場合でも慣用句の影響が大きい。二番目に関しては品詞の別は本質的な問題ではないだろう。本質的に重要な問題は、二で終わっている文節の下位分類が効果的ではないということに帰着される。形態素解析の一層の精度向上が望まれる。

ヲ格を項に取らず、二格を項に取る述語の中では「寄与する」が例外的に精度がよい。これは例外的にこの動詞と共起する二格の曖昧性が低いからであると解った。程度や様態を表わす副詞類が共起しないことに原因があるようだ。

二格要素と同じことがデ格要素の集合に関しても言えるが、デ格はそれでもまだ二格よりは曖昧性が少なく、また、二格と違って(準)項をマークする率も低いので、デで終わっている文節の下位分類が効果的ではないことは、深刻な精度の低下には繋がっていないように思える。

例外的にヲ格を取るのに精度が低い述語に挙げた「心掛ける」「逃れる」に関しては、提案の三つの結合法(7)のいずれでも従来の結合法に較べて精度が向上している。特に「心掛ける」に関しては向上は顕著である。とはいえ、精度向上の原因がつかみ切れているとは言えない。

3.2.3 述語のクラスとクラスター化精度の関係

結合のための閾値が一律に設定されているが、述語の意味クラスごとに別の閾値を最適化したり、統計量を反映させることには精度向上の効果があるのではないかと思う。

本来であれば述語ごとに最適な閾値を発見すべきであるが、どれくらい労力に見合うかわからないので、述語の頻度の違いや、意味クラスの別(e.g. 動詞か非動詞か、自動詞か他動詞か、非対格動詞か非能格動詞か)によって最適な閾値が異なっている可能性が高い。図1の結果を見る限り、ヲ格をもつ述語ともたない述語に区別するだけ、あるいは他動詞とその他の述語を区別するだけでも、それなりの効果があるのではないかとと思われる。

3.2.4 用例収集・結合の活用形ごとの層別化

大きな効果は期待できないだろうが、[4]の結果を考えると、述語をタ/タイトル/テイル形ごとに採集し、活用形ごとに用例クラスター化した後に一つに統合することで、より良い結果が得られる可能性もある。平均した効果は些細かも知れないが、一部の述語にはこれで大幅にクラスター精度が向上することがあるかも知れない。

4 議論: 現行の格フレーム辞書構築の主な問題点

この節では結論に代えて、現行の辞書構築の主な問題点を、評価に当たった言語学者三人が気づいた限り指摘し、格フレーム辞書の品質改善のための提案を幾つか行う。

4.1 同表記異語の扱いが未熟

「避ける」には「さける」と「よける」が混在していた。読みの推定と語義の推定は表裏一体なので、問題の解決は単純でないように思われる。語義の推定の手法が進歩しない限り、この種の問題を根本的に解決するのは難しいだろう。ただ、「避ける」のように語義の異なる述語が一つの表記に集約する場合は数が少ないので、目立った精度の低下には繋がっていないかも知れない。

4.2 形態素の過分析/過分割

二格の生起例として「嗟」が挙がっていた。これが二格に挙がっているのは、「咄嗟に」が二格として扱われ「咄」と「嗟」への過分割の結果である。前者は解決困難なので、後者のため形態素解析の辞書を充実する必要が示唆される。

過分割の発生頻度は高く、影響は無視できない。「襲われる」のガ格に過分割の「者」が現れ、元の形である「被害者」「犠牲者」が現れないのは、重要な意味情報の損失に繋がって

いる。実際、人手評価の段階で、格要素の一覧を見ただけでは用例クラスターの精度を判断できない用例クラスターの数はそれなりに多く、それを復元する手間は無視できないものだった。これはかなり顕著な過分割の悪影響である。「者」のような接尾辞、「上」のような接尾辞化した形式名詞が一見かけの被覆率を向上させつつ—精度の低下させている可能性は高いので、扱いの見直しが必要だと思われる。

4.3 複合動詞の扱いが未熟

接尾辞の扱いに関連したことだが、V(e.g. 「襲う」「解く」)がV+W(e.g. 「襲い+かかる」「解き+放つ」)という形で複合動詞をなす場合に、Wの格要素となっている形がVの格要素として抽出されている。「xがyを襲う」と「xがyに襲いかかる」とでは、同一のyについてヲ格支配と二格支配である。だが、後者で二格を要求しているのは「かかる」である。同様に、「xがy(へ)のzを解く」と「xがzからyを解き放つ」とでは、同一のzについてヲ格支配と、カラ格支配がある。だが、後者でカラ格を要求しているのは「放つ」である。「襲う」の格要素として二格が、「解く」の格要素としてカラ格が拾われている。これは過剰般化に繋がりが、見かけの被覆率の向上と裏腹に精度の低下を招くのは確実である。

4.4 共起情報の損失

現行の格フレーム辞書では名詞(句)同士の共起情報が利用されていない。このため「xがyを襲った」から二格名詞としてyが採られた時のガ格名詞が何であったかが明示されない。格フレームの心理的実体が「〈加害者(x)〉が、〈被害者(y)〉を、襲う」のような項間の共変動情報を表現した意味フレームで[3]、それを復元することが「良い」格フレームを作ることなら、これは重要な情報の損失であるように思える⁶⁾。

4.5 八に吸収されている格の扱いの改善

係助詞八の要素は収集から外されている。これは八に化けて現れやすい格要素の事例が十分に収拾できないという結果を生んでいる。評価の結果を見ると、影響を受けやすいのはガ格である。ガ格は一般に省略されやすいと考えられているが、八に吸収されているために取りこぼし率も高いと思われる。ガ格の被覆率を上げるには八に隠された格の推定が有効ではないかと思われる(この推定は特定の述語から見た格で十分である)。これは述語と共起する名詞要素を組みで収拾したデータが十分にあれば、解決可能な問題かと思われる。

参考文献

- [1] 中本敬子, 黒田航. 「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. 言語処理学会第12回大会発表論文集, pp. 592-595, 2006.
- [2] 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.
- [3] 黒田航, 中本敬子, 野澤元. 状況理解の単位としての意味フレームの実在性に関する研究. 日本認知科学会第21回大会発表論文集, pp. 190-191, 2004.
- [4] 李在鎬, 永田由香, 鈴木幸平, 黒田航, 井佐原均. 語形と意味: 「xガ流レ{ル, タ, テイル, テイタ}」の意味の多変量解析を用いた比較を通じて. 第13回言語処理学会発表論文集, 2007.
- [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.
- [6] 河原大輔, 黒橋禎夫. 格フレーム辞書の漸次的自動構築. 自然言語処理, Vol. 12, No. 2, pp. 109-131, 2005.

⁶⁾ 同一文内共起を結合の条件にすると得られるデータが稀薄になる懸念があるなら、それを用例結合の閾値を下げる要因とすれば良い。