

# 複層意味フレーム分析を用いた 意味役割タグつきコーパス評価版の公開

黒田 航 井佐原 均  
(独) 情報通信研究機構

{kuroda, isahara}@nict.go.jp

## 1 はじめに

私たちは、[7] で予告した通り、Berkely FrameNet (BFN) のアプローチ [1, 2] を参考にしながら、日本語のための意味役割タグ体系  $S$  を、BFN やその日本語版に相当する日本語フレームネット (JFN) [3] とは独立に定義し、(BFN/JFN の意味(役割)タグ体系ではなく)  $S$  を用いて、日本語の意味役割タグつきコーパス  $C_S$  を開発する研究を進めている<sup>1)</sup>。

これまでの研究の結果 [4, 6]、**複層意味フレーム分析法** (Multilayered Semantic Frame Analysis: MSFA) をタグづけの枠組みに採用することが決定しているが、タグづけの仕様の細部は固定されていない。今回の評価版の公開は、未決定仕様を決定するための手がかりを得るためのものである。

今回、評価版として公開するデータは、京大コーパスの三記事分 (950103083-001,018; 950101075-001,036; 950107210-002,010, 合計 64 文) に §3 で述べる内容で作業者に依頼した意味タグづけ結果を修正したものである (この作業は本稿執筆時点で継続中)。この結果は決定的ではないが、これに対する NLP の研究者の率直な意見を求め、タグづけ仕様を最終調節の参考にしたいと考えた<sup>2)</sup>。

### 1.1 なぜ「評価版」か？

本格的な開発を始める前に小規模な「評価版」を公開し、NLP 業界の反応を見ることにした理由は、次の三点である。

#### 1.1.1 理由 1: 「市場調査」の必要性

私たちは NLP に限らず、言語学、認知科学を含めた多くの分野の研究者にとって有益な言語資源の開発を望んでいるが、**どのような情報が、どの研究分野で (例えば NLP の分野で) 強く望まれている情報なのか**は、初めからわかっているわけではない。それは市場調査しない限り見えてこない需要だと判断し、評価版の公開に対する反応を見てそれを推測することがもっとも効果的な評価方法だと考えた。

特に、NLP にはすでに (表層) **格フレーム辞書** [8, 5] というすぐれた研究成果が存在する。実際、格フレーム辞書程度の粒度をもたない「粗い」意味記述の体系は (少なくとも)

NLP の分野で無条件に有益な資源とは見なし得ないだろう。私たちの目標の一つは、言語学者による人手解析の結果が格フレーム辞書以上の精度で意味解析ができることを示すことだったが、評価版がそれに見合った結果を出しているかは、専門家に率直な意見を求めるのが適切だと考えた。

#### 1.1.2 理由 2: 「保険」の必要性

私たちが見る限り、言語学者主導の言語資源開発の最大の懸念は、次の点にある: MSFA に限らず**言語学者が考案した言語分析の手法に開発者が思いこんでいる以上の有用性が伴っている保証はない**。従って、有用性の見こみに「保険」をかけずに (ただ「結果が出る」という理由だけで) 本格的な開発に手を出すのは、使い道のない資源を手間も暇も金もかけて開発するような愚行に繋がる可能性がある。私たちはそのような事態は避けたい。評価版の公開は、その種の保険の一つであり、開発予定の資源に最低限の有用性の見こみがあるかを NLP 関係者に判断してもらうための機会の一つとした。

#### 1.1.3 理由 3: 優先順位の特定し自由度の過剰を抑さえこむ

最後は技術的な問題になるが、私たちは以下のような形で「過剰な自由度の制御」の問題を抱えている: MSFA では課題に依存して任意に意味分析の精度を上げ下げできる、つまり、課題に依存して「浅い」分析も「深い」分析も可能であり、意味分析の最適な「深さ」(意味の粒度) は、課題が決まらない限り決まらない。これは理論的には望ましいことなのであるが、実践的には自由度が過剰で、好ましくない状態でもある。実際、**どの種のフレームを優先的に認定し、定義を与えるかは、何らかの「外的な基準」がないと決まらないが、この基準をどう与えるかは、自明な問題ではない**。言語学者、認知科学者の観点から優先順位を設定することも可能なのだが、それでは多くの研究分野の需要を反映した領域横断的な利用価値をもつ研究資源を提供するという目標に反する。

長期的には、なるべく多くの粒度で意味記述を充実させる必要がある。とはいえ、今回の評価版は、そのための足がかりということになる。その次に来る中期的な目標は、小規模ながらも高品質な教師データという位置づけである。

データの公開は 3/17 日の本発表時に行ないたい。

## 1.2 以下の議論の構成

全体の構成は以下の通り、§2 では意味役割タグづけの仕様を述べ、§3 では作業者による意味役割タグづけの現状を報告し、§4 では問題点と展望を述べる。

<sup>1)</sup>  $S$  と BFN/JFN との互換性はなるべく確保したいが、JFN の研究成果が未公開である以上、現時点では何とも言えない。

<sup>2)</sup> なお、これ以外の参考データも Wiki ページ <http://61.115.230.87/~mutiyama/cgi-bin/hiki/hiki.cgi?FrontPage> で公開している。このページの構築は内山将夫氏 (NICT) のご好意による。

## 2 MSFA を用いた意味役割タグづけの方針と仕様

### 2.1 MSFA の解析方針

MSFA の仕様に関しては、次の点に注意を促しておきたい:

- (1) 非排他性(あるいは「寛容」)の原理: 多くの言語解析の手法と異なり、**MSFA は唯一の「正解」を指定するものではない**: それは、特定の意味解釈を与える理解者の読みの「深さ」と「幅」を指定するものである。

MSFA は実際、この寛容性の原理により個人の解釈の食い違いを「吸収」する。ただ、これが前述の「自由度の過剰」が現われる原因でもある。どうしてそうなるかは、以下の説明と実例から明らかになると考える。

### 2.2 MSFA の作業仮説

詳細は [4, 6] や準備中の作業マニュアルに譲るが、MSFA は概略、次のような作業仮説を設ける:

- (2) 一つの文に(“喚起”の効果によって)結びつけられる意味フレームの数には理論的上限はない。矛盾が生じない限り、フレームは一文に幾つ対応づけられてもよい
- (3) 仮に文  $s = m_1 \cdots m_n$  に  $\mathcal{F} = \{f_1, \dots, f_N\}$  のフレーム群が結びつけられている ( $n, N$  の間に因果関係はない) とすると、 $s$  のおのおのの形態素  $m_i$  について、(i)  $m_i$  は、 $\mathcal{F}$  のうち少なくとも一つのフレーム  $f_j$  の意味役割の一つ  $f_{j.r_k}$  を実現する必要がある(ただし、 $f_j$  について、その中で実現する意味役割の個数は一つに限られなくてもよい。つまり  $k$  は単一でなくてもよい)。 (ii) それと同時に、 $m_i$  は  $\mathcal{F}$  のうち可能な限り数多くのフレームの意味役割を実現するべきである。つまり、 $m_i$  が実現する  $f_{j.r}$  の数は  $j$  について多いほど好ましい

また、以下の仮定がなされている点も注意が必要であろう:

- (4) 同一の形態素  $m$  が  $f$  ごとに異なる意味役割をもつことが、**文脈ごとの語の意味の記述**—すなわち語の意味の曖昧性の多次元的解消の効果の記述—と、**語の意味の文意への統合の記述**の核心である。
- (5) 意味役割(の極く極く一部)は“深層格”と同一視可能だが、格文法の「単文異格」の原則は文(あるいは述語)についてではなく、意味フレームごとに成立する。
- (6) 意味フレームの間にはしばしば概念的な依存関係が存在するが、それらは統語派生に由来する効果ではなく、例えば統語移動などを使って表示する必要はない。

意味フレームの要素としての意味役割の体系は深層格の体系よりケタ違いに豊かであり、その豊かさに一貫した方法で対処する必要がある。私たちが BFN を参考にしたのは、この問題に妥当な解決を与えられると判断したためである。

### 2.3 MSFA の具体例

(7) [=京大コーパス: 950103083-006] の MSFA は表 1:

(7) 京産大のゲームメーカー広瀬に圧力をかけ続け、集団パワーで京産大の攻め手を封じた。

ただし、形態素列 (A4, A5) に現われている \* は意味役割が形態素によって実現されないことを示す補助記号である。表 1 のフレーム群はフレーム間関係を基にして**階層ネットワーク解析 (HFNA)**として再構築される。詳細は [6] を参照。

## 3 タグづけ作業の現状

この節では第三者による意味役割タグづけ作業の意義と目的、その結果の現時点での先取的評価を述べる。

### 3.1 第三者による意味役割タグづけ作業の意義

MSFA は  $C_5$  開発の要となる技法であるが、[6] で言及したように、第一著者が行なった MSFA が第三者にも実行可能かどうか、それが可能だとしても、どれほど再現性があるのかには少なからず疑問の余地があった。今回の第三者への依頼による意味タグづけ作業はその懸念を晴らすと期待された。

今回の第三者による意味役割タグづけの目的は特に:

- (8) MSFA による意味タグづけ作業が、第一著者以外の第三者に実行可能かどうかを確かめる
- (9) 十分に実行可能な場合、第一著者の行なった MSFA に第三者による再現性があるか、また再現性があるとなれば、どの程度かを確かめる
- (10) 再現性が十分に認められる場合、どれぐらいの粒度のフレームを公式にタグづけの対象として選定するのが現実的かを判断するために必要な情報を特定する
- (11) 本格化するタグづけ作業に不可欠な作業の手引きを整備するための下準備的研究とする

詳細を論じる前に、簡単に結果の総合評価を述べる。ただし、数値的な評価はまだ行っていない。

### 3.2 作業結果の総合評価

- (12) (8) に関しては、十分に肯定的に評価できる結果が得られた。速度は別にして、MSFA に基づく意味タグづけ作業は第三者にも十分に可能であった。この評価には §3.4 で後述する作業モデルの変更も反映している。
- (13) (9) に関しては、まだ最終的な判断は下せる段階ではないが、現時点ではそれなりに肯定的に評価できる結果が出ているように思われる。特に、作業速度は別にして、第三者による意味タグづけの内容は、十分に道理の通ったものであった。それに加えて、作業内容の慣れによる技能の伸びが顕著であった。
- (14) (10) に関しては、現時点では判断できない。作業モデルの修正を迫られ、先送りになったからである。この問題には §3.4 で立ち戻る。
- (15) (11) に関しては、有益な結果が得られている。これに関しては、準備中の作業マニュアルに反映される予定

### 3.3 作業内容

作業者に依頼した作業の内容は、以下の通りであった:

1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
2	Frame ID	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14
	F-to-F relations	targets F2	targets F3	elaborates F11; constitutes F4; presupposes F5	elaborates F13				constitutes F3	constitutes F3	targets F11	elaborates F12			presupposes F13
3	Frame Identifier	加圧 [+metaphoric]	押さえつけ [+metaphoric]	防衛の態の相手の審判	動作の継続	行動の先導	<部の接続>	行動の連続	集団行動?	攻撃	封印 [+metaphoric]	防衛者による攻撃の無力化	大学ラグビーの試合	ラグビーの試合戦術	観戦内容の報告
4	*														
5	*[+anaphoric; target=前大]	加圧体	押さえつけ 攻撃	審判者	継続者		共通項	行動体	集団行動体	攻撃対象	封印者	防衛者	チーム1	観戦者	報告者
6	京産大	加圧対象 特徴	押さえつけ る相手 特徴	審判相手 特徴	動作	全体	節1	行動1		攻撃体	手段1	手段1	チーム2		
7	の					MARKER 先導体 特徴									
8	ゲーム														
9	メーカー														
10	広瀬	加圧対象	押さえつけ る相手	審判相手		先導体									
11	に	MARKER	MARKER	MARKER											
12	圧力	GOV[+composite]	GOV[+composite]	GOV[+composite]											
13	を														
14	かけ														
15	続け			GOV											
16	.						EVOKER	EVOKER							
17	集団	目的	目的	目的	目的				GOV		手段2	手段2			
18	パワー								目的						
19	で									MARKER 攻撃体	MARKER 攻撃者	MARKER 攻撃者			
20	京産大												チーム2		
21	の									GOV 攻撃球	MARKER 攻撃球	MARKER 攻撃球			
22	攻撃									手段: 攻撃球		GOV			
23	手														
24	を									MARKER GOV					
25	対じ														
26	た						EXT	EXT					EXT		
27	.														

図1 (7)のMSFA

- (16) 京大コーパスから適当な文章(一記事ごと)を選び、その一文ごとにMSFA [4, 6]を用いた意味役割タグづけを上記の作業員四人に依頼した。
- (17) 作業指示内容としては、作業の簡単な解説 [4]を示すと共に、第一著者が作成した数十文程度の見本を示し、それを見習って解析するように指示した。なお、見本自体が試行錯誤的なもので、標準化されたものではない。
- (18) 作業は、一週間につき8時間、4週間で1ラウンド(1記事)とし、全部で3ラウンド(3記事)あった。2週間ごとに報告会を行ない、時間の許す限り、タグづけ内容の修正、タグづけの秘訣に関して指導を行なった。その詳細の報告は、目下、別の文書に用意中である。

以上の内容で作業を進め、概算で一文につき平均10個程度の有効フレーム<sup>3)</sup>を付加したデータが得られた(詳細は執筆時点で未定)。だが、途中で作業モデルの修正を迫られた。

### 3.4 二段階作業モデルの必要性

今回の作業では、将来的に最適な作業枠を確立するため、実験的に二つの作業条件を比較してみた。

作業員四名は、二つの作業グループにわけた、第一のグループの二人は常に二人一組になり、相談しながら**二人組コーディング**(Pair Coding: PC)する。第二のグループの二人は常に単独で、相談なしで**単独コーディング**(Single Coding: SC)する。このような比較を行なったのは、当初、MSFAでのコーディングが未経験者には困難と思われたため、**二人組プログラミング**の効果と同じような相乗効果を期待してのことであった。見こみでは、PCの方がSCよりもよい結果を出すと思われていたが、期待と予想に反し、PCは目的に

好ましくないことが判明した。理由はおそらく次の二つであると思われる: (i) 意味役割タグづけが、予想よりも困難度が低かった。(ii) 困難な部分は確かに存在するが、それはPCで支援される部分ではなかった。

結果の評価を客観的に示す指標はなかったが、合計4回の定例報告会の結果を見る限り次のことは明らかであった: 二人で相談することで一致度は増し、分析結果の見かけの信頼性は深まったが、それと同時に、探索すべき可能性が意見の不一致により抑制され、“摘み取られる”傾向が強かった<sup>4)</sup>。

これはおそらく、分析者が未経験者レベルでは、二人の合意が得られないで放棄される直観に重要な可能性が含まれている率が高いことを意味する。つまり、熟練者が二人組になった場合にはPPと同じような相乗効果、相互保証効果が得られるかも知れないが、未経験者レベルでは反対に、抑制効果の方が勝って出るように思われる。

これは予想に反しており、PCを作業モデルに見込んでいたので、モデルの修正が必要となった<sup>5)</sup>。これにより、作業モデルを二段階モデルに改訂する必要が生じた。改訂により、モデルは次のようになる:

- (19) 意味タグづけ作業を、**探索的段階**と探索結果の**認証段階**の二段階に分ける。
- (20) 探索的段階は作業者が担当し意味フレームを最大限に探索的に記述する。この際、精度には重きを置かない。
- (21) 認証は、独立に存在する**認証委員会**が担当する。この際、作業者の探索的分析の結果から選択的に値する意味フレームを選定する(正式な定義の候補となるのは、認証委員会によって選定されたフレームのみ)。

<sup>3)</sup> 非公式の観察によると、新聞の文の意味フレームの「濃度」は高い。ただ、これは十分な数値的なデータに基づく解釈ではない。

<sup>4)</sup> 慎重を期するならば、これが作業者の個性によるものでないかは確かめる必要があるだろう。

<sup>5)</sup> PCが熟練者による場合、モデル修正は不要だった可能性はある。

ただし、委員会制の認証がうまく機能するための条件として、

- (22) 委員会は熟練者によって構成される必要がある。
- (23) 意見が食い違った場合、有効な合意形成の方法を有している必要がある。おそらく、特殊権限をもった人間が一人いて、判断が収束しない場合に特例的解決をもたらせる用意を作っておくことが必要だろう。

このモデルの下では、タグづけ作業は最大限に探索能力を発揮するべきであり、PCよりもSCの方がその目的に適している。PCでは相談することで、一致が見られなかったフレームはその場で摘み取られ、認証の候補に残らないからである。このような摘み取りの判断を下すのが熟練者である場合には、全体の作業効率を上げる効果が期待できるが、作業者に熟練が不足している場合、見落としの可能性の方が増える恐れがある。このような判断に基づいて、最初の二ラウンドでPCは取りやめ、全員SC体制に移行した。

#### 4 今後の課題と展望

作業モデルの改訂で問題が解消されたわけではない。現実には、次のような課題が残っている:

- (24) 委員会方式は、本当に思惑通り機能するか。それ以前に、委員会の構成に十分な数(少なくとも3名程度)の熟練者が確保できるか。
- (25) 委員会方式が機能するとしても、作業者が発見した非常に多数のフレームのうちからどれかを有効なフレームとして採用するための判定基準を明示化できるか。
- (26) 作業結果に不一致が生じた場合にはどう解消するか。

これらに関して自明な解はなく、今後も実験的調査を継続によって妥当な方向性を模索する予定である。

ただ、(26)の問題には慎重に対処する必要がある。というのは、作業者が意味フレームに与える名称、意味役割名には違いがあることは表面的な現象だからである。問題は同一の意味フレームが同じ要素に対して認定されるかどうかであり、名称が同一かどうかではない。従って、(26)への正しい対処法は、**同一の意味フレームが正しく作業者に認定されているかを、誰かが正しく判定できるかどうか**に依存する。これは前述の二段階制を必然的なものとする。これは回り道のようにだが、結局は近道だと考える。

以上の実験的タグづけから一つ重要なことが明らかになったように思われる。それは、MSFAが習得可能な「技能」であるという見こみの下で、その熟練者の数を増やすことは、企画を成功させるための必要条件であり、そのためには、**意味役割タグ作業の熟練者を育成システムを確立することが、本質的に重要だ**ということである。これは**言語学系の研究室との密接な連携なしには達成しえない**と思われるが、企画の成功がこの可否にかかっていると云っても過言ではないだろう。

だが、この点に関しては、次のような否定的評価も十分にありうる: これは要するに、ある程度の解析能力をもつ人物が(独断による恣意性を避けるために)複数人で協力して開発

にあたらないと、本当に有益な言語資源は開発できない(し、人件費もかかるのでサッサと自動化してしまおう)という、NLPでは「あたり前」の教訓の再現でしかない。

だが、私たちが強調したいのは、将来的には(半)自動化に委ねられる見通しがあるとは言え、経費節約のために初期段階で高品質の人手解析を導入する手順をなしにする開発はしばしば結果の頭打ちを招き、結局のところ「高くつく」という点と、(1)で示した非排他性の原理により**MSFAがそのような共同開発に適しており、分析精度の高品質化には期待も**てるという点である。MSFAは「言語学者に見える言語」のことしか考えないで開発された意味記述のモデルではない。

NLPが本音では(経費の割りに信頼性の低い結果しか与えない)言語学者を無用にしたい気持ちはわかる。だが、相手が有益な情報を提供する限りでは、完全に手を切るの賢くない選択であろう。文脈に置かれた語の意味の実体化がようやく始まったばかりであり、それが全自動獲得できるようになるのは、楽観的に見積もっても数十年も先の話だと考えられる現状では、なおさらそうだと私たちは思う。

#### 参考文献

- [1] C. R. Johnson and C. J. Fillmore. The FrameNet tagset for frame-semantic and syntactic coding of predicate-argument structure. In *Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics (ANLP-NAACL 2000)*, pp. 56–62, 2000.
- [2] J. B. Lowe, C. F. Baker, and C. J. Fillmore. A frame-semantic approach to semantic annotation. In *Proceedings of the SIGLEX Workshop on Tagging Text with Lexical Semantics: Why, What, and How? 1997*.
- [3] K. H. Ohara, S. Fujii, T. Ohori, R. Suzuki, H. Saito, and S. Ishizaki. The Japanese FrameNet project: An introduction. In *Proceedings of LREC-04 Satellite Workshop "Building Lexical Resources from Semantically Annotated Corpora" (LREC 2004)*, pp. 9–11, 2004.
- [4] 中本敬子, 黒田航, 野澤元, 金丸敏幸, 龍岡昌弘. FOCAL/PDS 入門: フレーム指向概念分析/並列分散意味論の具体的紹介. [未発表論文: <http://cls1.hi.h.kyoto-u.ac.jp/~kkuroda/papers/introduction-to-focal.pdf>], 2004.
- [5] 荻野孝野, 小林正博, 井佐原均. 日本語動詞の結合価. 東京: 三省堂, 2003.
- [6] 黒田航, 井佐原均. 意味フレームを用いた知識構造の言語への効果的な結びつけ. 電子情報通信学会技術研究報告, 第104(416)巻, pp. 65–70. 電子情報通信学会, 2004. [増補改訂版: <http://cls1.hi.h.kyoto-u.ac.jp/~kkuroda/papers/linking-1-to-k-v3.pdf>].
- [7] 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://cls1.hi.h.kyoto-u.ac.jp/~kkuroda/papers/jfn-nlp10-rev4.pdf>].
- [8] 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.