

複層意味フレーム分析の紹介

領域を問わないオントロジー構築のための効果的な前処理として

Multilayered Semantic Frame Analysis Serves as a Preprocessing for Domain Ontology Construction

黒田 航*¹ 高梨 克也*² 竹内 和広*³ 井佐原 均*⁴
Kuroda, Kow Takanashi, Katsuya Takeuchi, Kazuhiro Isahara, Hitoshi

*^{1,3,4}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
National Institute of Information and Communications Technology

*²東京大学大学院 情報学環
Interfaculty Initiative in Information Studies, University of Tokyo

This paper introduces Multilayered Semantic Analysis (MSFA), which has been developed as a framework for semantic role tagging. We suggest that besides its usability for semantic annotation/analysis, MSFA would also serve as a useful heuristic procedure at the “preprocess” stage for domain ontology development due to its ability to reveal links from natural language to world knowledge.

1. はじめに

オントロジーが何であるかに関して相当に広い定義を考えても、私たちの研究は、オントロジーの理論、構築、利用のどれにも属さない。その位置づけは、おそらく「言語科学者が行う自然言語の分析とオントロジー研究の(意外な)接点」のようなものだろう。私たちはオントロジーの専門家ではないので、オントロジーの理論、開発の現状に関する無知に基づき、安易な提案、主張を行なう可能性がある。その点は大目に見て頂きたい。私たちの提案が仮にオントロジー研究ですすでに「自明なこと」に属している可能性はある。ただ、部外者の私たちに、[7, 8]のような成果紹介があるにしても、オントロジー開発の「内実」を知ることは困難だったし、これを相互に知ること自体、価値がないことではない。実際、言語の意味分析とオントロジー研究の接点は、少なくとも部外者である言語学者が前述のような基本文献から簡単に「見て取れる」内容でないのはまちがいない。以上の注意の下で、本論に入る。

2. 概念体系の明示化としてのオントロジーと言語の意味分析の関係づける

部外者的な立場から見る限り、AIの分野では(A)オントロジー構築は対象領域 D (工学, 生物学, 法律)を先に限定し、その実現手段の一つとして D に関連した自然言語の表現集合を収集し、それらの概念分析を行なうのが典型的な手法であるように思われる。これは目的志向的で、トップダウン的傾向の強いアプローチである。領域固有のオントロジーが存在するという前提が出発点にあり、それを効率的に抽出、表現するための技術としてオントロジー工学のようなものが存在する。

これはオントロジー開発の有効で正当な手法であることはまちがいないが、唯一無二の手法ではないだろう。(A)を補完するやり方として(B)オントロジー構築の対象領域を先に限定せず、言語の意味分析を広い意味でのオントロジー構築の作業工程の一つ(少なくともその前処理)であると見なすアプローチもありうるを私たちは考える。これは(A)に比べ、ボトムアップ的で汎用的である。

私たちは以下で(B)のアプローチの一例を示す。これが本当にうまく行く手法なのかは未知数だが、効率的なオントロジー開発には、おそらく(A)の一本槍よりは(A), (B)の両面作戦が必要であり、有効であると私たちは考える。

2.1 文章の複層意味フレーム分析

私たちの論点は以下の通りである: 私たちは将来的に「深い」意味処理を可能とする言語資源として、意味(役割)タグつきコーパスを構築するという目的[5]のために**複層意味フレーム分析(Multilayered Semantic Frame Analysis)**(以下MSFAと略す)と呼ぶ言語の意味の記述法を開発した[4]。MSFAは平均的な理解能力をもったヒトが文 s を読んだり、聞いたりしたときに理解する内容(の明示化可能な部分)を s に関与する複数の意味フレームが統合された状態として指定する手法である。私たちはMSFAを定義するにあたって、Berkeley FrameNet[2]の考えを参考にしながら、それを拡張した。

2.2 複層意味フレーム分析の解説

私たちは、ヒトが文 s を理解するとは、 s を聞いて(あるいは読んで)、何らかの「状況」を思い浮かべることであり、この意味での状況を構成する要素を表1に示した形で一つ一つ文章にタグづけすることを意味役割タグづけと呼んでいる。MSFAは意味役割タグづけのための一手法である。

2.2.1 MSFAの実例: 例えば(1)のMSFAは、表1に示す表である。

- (1) 空腹のライオンがインバラの群れを襲った。
- (2) スズメバチの大群が子供たちを襲った。
- (3) 大型の台風が九州(の人々)を襲った。
- (4) 覆面の男が都銀の支店を襲った。

表1に示した意味フレーム群が(1)の理解内容を近似するのに十分に網羅的だと主張したいわけではないが、この表1から示唆されるのは、MSFAは意味フレームという記述形式を媒介にして世界知識と言語との結びつきを特定する効果がある、ということである。

例えば、F8:〈狩り〉という活動は次のようなフレームとして記述される:

連絡先: 第1,3,4著者は{kuroda,kazu, isahara}@nict.go.jp,
第2著者はtakanashi@iii.u-tokyo.ac.jp

Frame ID	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18
F-to-F Relations	elaborates F2; constitutes F3											elaborates F13						
Frame Name	感覚	感覚	欲求の満足	食物摂取	集団化	自衛	攻撃	狩り	捕食	選択	殺害	死亡	状態変化	使役	非生存	逃亡	生存[1]	生存[2]
	空腹(感)	感覚	欲求															脅威
	* GOVERNO																	
		GOVERNOR								目的								
			GOVERNOR	目的														
			手段	GOVERNOR														
の	MARKER																	
ライオン	空襲者	感覚者	欲求	食物摂取者		外敵	攻撃者	狩り手	捕食者	選択者	殺害者			使役者		外敵	脅威	生存者
が							MARKER											
																	逃亡者	生存者
インパラの					行動者	自衛者	標的	獲物	獲物の種類	除外項								
の					MARKER				候補									
					目的	GOVERNOR												
群れ					GOVERNO	手段			獲物の状態	:EVO								
				食物					獲物	選択項	犠牲者:EVO	死亡体	変化体	被役者	非生存者			
										GOVERNOR								
を							MARKER											
				手段				GOVERNOR	手段								起こされた事態	
								目的	GOVERNOR									逃れた災難
										副作用	GOVERNOR	結果状態					理由	理由
襲っ							GOVERNOR	手段										
た							EXTENDER											

図 1: (1) の MSFA

(5) {〈狩り手: ライオン(x)〉が, 〈獲物: インパラの群れ(y)〉を, 〈手段: F7: 〈攻撃〉し)て, 〈目的: F9: 〈捕食〉〉のために, ..., 〈GOVERNOR: 捕獲〉する}

この記述を構成する {〈狩り手〉, 〈獲物〉, ...} を意味役割と呼ぶ。意味役割は [8, p. 178] の言う「ロール」概念とよく似ている。私たちが意味フレームと呼ぶ概念構造は、このような意味役割の組織化の単位である。

この意味での意味役割の集合が所与のものとは私たちは考えず、実際の文章の解析を通じて特定し、記述し、再利用可能な形で体系化してゆこうとしている。意味役割タグづけは、そのような体系化の作業と同時並行的に行われる。

私たちは、状況とは {狩り, 逃亡, ...} のような活動や出来事の本質であると想定している。従って、私たちの言う(意味)フレームの多くは(AIで受け入れられているようなモノ)のモデル化ではなく)状況を構成する活動や出来事のモデル化である。状況を構成する個別的な活動や出来事の特徴づけが個別的な意味役割という形で特定されることになる。この意味での意味フレームは、上位オントロジーで成立するような一般的な推論規則から演繹できないような個別的な情報をパッケージ化する単位であると考えられる。

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13
Index	形態素	発生 [+inferred]	経路移動 [+inferred]	移動 [+inferred]	経路 [+inferred]	加害	被災 [+inferred]	経路 働きかけ	ヒトの生活 活動 [+indirect]	生物の棲息 [+inferred]	国土	収容	分割
1	大型	規模			規模?	規模?	規模?						
2	の	MARKER											
3	*	GOV											
4	台風	発生体	移動体EVO	移動体		加害体	原因	内容	影響源				
5	が				MARKER								
6	*	発生地	起点	起点				被影響体					外部
7	*		着点	着点									
8	*		通過点	通過点EVO							国境	収容 路EVO	境界
9	*												内部EVO
10	九州				場所	被災地	場所		生活地域	棲息地域	地 域EVO	収容物	
11	*								GOV				
12	(の人々)				被害者	被災者	経路者		生活者	棲息生物			
13	を				MARKER								
14	襲っ				GOV	EVO							
15	た												

図 2: (3) の MSFA

2.22 意味役割タグづけの効用: (1)で「インパラ」が例えば〈狩り〉の構成要素である〈獲物〉という意味役割の実現値であるという情報が明示されることの効果は、言語理解の際に心に構築される「理解内容」に(近似的)記述を与えるという目的からすると非常に大きい。このようなタグづけ例が大量にあれば、従来は不可能であった「深い」意味処理が可能になるだろうと期待されている。

(1)の理解は(5)のレベルの記述が与えられたときによく達成される。これは(1)の理解内容と(3)や(4)の理解内容を比較すれば明らかになることである。記述にどれも同じく「襲う」という語が使われているけれど、(1, 3, 4)で記述されているのは、まったく異なる状況である。この点は、例えば、表1の(1)のMSFAと表2の(3)のMSFAの比較から十分に明らかであると思う。

2.3 MSFA とオントロジー研究との接点

表1のF7:〈攻撃〉は「xがyを襲う」の辞書的意味、あるいは語義(word sense)の一つだと見なせる。実際、語彙大系[3]にある「襲う」の三つの定義の一つ*1)はF7と概ね一致する。

問題は、このような語義が与えられただけでは、(1)の状況レベルの理解内容は特定されたことにならない、という点にある。理解内容の特定が課題である場合、語義の脱曖昧化は解決のための手段というより、むしろ結果である。

「xがyを襲う」の基本的意味が例えば〈xがyに(特に予期, 予防できない仕方)で危害を加える〉のように抽象的に与え、(1, 2, 3, 4)で達成される具体的理解との差分を「推論」で補えばよいか? そのようなアプローチは過去に幾つも提案されたようだが、それによって一般的な形で解答を与えるのは不可能である、というのが過去の研究の結論であるようだ。これは「推論は基本的に事例ベースだ」と言う見解がAIの一部に定着している点[1]にも伺えるのではないかなと思う。

実際、(3, 4)の場合とは異なり、F7は(1)ではF8と結びつかないと意味をなさないが、F7, F8の結びつきは無条件ではない。(2)ではF7の目的はF8ではなく、〈自己防御〉である

これが示唆しているのは、F8:〈xがyを狩る〉は{〈yがxから逃げる〉, 〈yがxから身を守る〉, 〈yがx(の攻撃)を警戒す

*1) 襲う(おそう)(2)23 身体動作(動作): N1がN2を襲う; N1 attack N2 [N1(3 主体 535 動物 2416 病気) N2(2 主体)]].

る),...}と共に,自然界で〈捕食者〉と〈被捕食者〉が置かれている(非友好的な相互作用)の概念化が,理解を支える前提知識となっている,ということである.このような知識は—オントロジーと呼ばれるべきだろうとそうでなかろうと—深い意味処理を実現するためには明示的に記述されている必要がある.対象領域の限定は必ずしも専門知識の特定を保障しない.

2.31 アドホックな疑似オントロジーの発見支援: 問題は, {〈xがyを狩る〉, 〈yがxから逃げる〉, 〈yがxから身を守る〉, 〈yがx(の攻撃)を警戒する〉,...}の相互作用の集合は広義のオントロジーの一部をなすと考えられるが,それは目標指向的,トップダウンなアプローチでは発見困難な知識だという点である.それが発見困難である理由は,この種の知識は専門知識と民間知識の境界にある知識で,アドホックな性質をもつ「疑似」オントロジーだからである.

疑似オントロジーがAIのモデル化の対象となるオントロジーなのかどうかは怪しい.だが,この種の知識構造がヒトの言語理解に不可欠であるのは明白である.実際,疑似オントロジーと真正のオントロジーを区別しない点がMSFAの長所であり,また短所であるだろう.文の意味が結びつけられているのが,真性オントロジーなのか疑似オントロジーかという問題は,ヒトが何を理解しているかを特定する手段としての意味(役割)タグづけの目的には差し当たって重要な問題ではない.

{襲う,守る,逃げる,...}のような概念化のパターンは,ドメイン,あるいはタスクを先に定めるトップダウン的アプローチでうまく発見できる保証がないように思われる.この点が仮に正しいとすれば,MSFAはアドホックな疑似オントロジーの発見を支援すると言える.これがいいことなのか悪いことなのかは,目的次第だろう.視点を変えると,MSFA自体はオントロジーを提供することはないけれど,オントロジー構築の対象となる知識の「部品」を特定するという目的に有用である. MSFAは対象分野を選ばず,記述内容がオントロジー化するに値するしないに関係なく知識の特定するのに有効で,本格的なオントロジー構築のための「前処理」として使える.

2.32 MSFAはそれ自体で完結していない: 以上のような利点,あるいは不利点があるとはいえ,MSFA単独では完結した知識表現を達成しないのは明らかである.

文(e.g., (1))の理解に関与する意味フレームには(i)文に現われている語彙に直接反映しているものと,(ii)推論で結びついているものとの二種類がある.後者について(iia)事例ごとの推論と,(iib)事例によらない一般的な推論とに区別できるが,MSFAは(iia, b)を明示的に区別できない.

例えば,〈捕食者〉としての「ライオン」が〈生存〉に必要な〈食料〉を得るために〈狩り〉を行なうことは,確かに(1)の理解を構成するが,これは「ライオン」のような語彙的にコードされているというよりは,〈肉食動物〉という上位レベルの概念で特定される内容であるように思われる.このレベルの知識は状況ごとに記述するより,(上位)オントロジーとして一括して記述した方がよいレベルの知識となっている.このことは図3に示した意味フレームの階層化を伴った統合に明らかであろう.このように,MSFAという形で特定された記述(例えば表1)の一部—特に上位階層—は明確にオントロジーとして解釈されないと,十分に有効な意味記述にはなっていない.

図3に示した統合的状态は表1の記述にオントロジー的解釈を行なったものとして得られる.ここに言語学者が行なう意味分析と知識工学者が行なう概念化の記述体系としてのオントロジーとの接点があるように思われる.

2.33 オントロジーの「内」と「外」を前提にしない知識構造の中立的記述の必要性: オントロジーの定義はまちまちであ

る.仮にそれが「概念化の規定」であるとしても,その概念の有効な範囲はどこまでか?という問題が常につきまとう.

実際,特にオントロジー構築を意識せず,ありふれた言語に現われる概念構造の分析をしていると,次のようなことが見えてくる:特に概念体系の下位の部分では,(i)概念化のパターンは集団によって異なりうるし,実際に異なる.(ii)概念化のパターンは個人によって異なりうるし,実際に異なる.(iii)同一の個人の中でも,複数の概念化の体系 C_1, \dots, C_n が存在し,それらのあいだで一貫性が保たれている保証はない.実際,個人が複数の領域ごと,課題ごとに異なるオントロジーを「使いわけ」している可能性は十分にありうることである.その場合,(iv)それらのうちのただ一つが正しい概念化であるとする十分な理由がないことが多い.

概念化でないものがオントロジーであるはずはないが,概念化が単純にオントロジーと同一視できるとは限らない.オントロジーとは言えないレベルでも概念化の体系が存在することは経験的に明らかであると思われる. MSFAのようなボトムアップ式のアプローチは,このようなオントロジー未満の概念化パターンをあぶり出すという目的には正しいアプローチであると思われる.

知識はオントロジー化に値するような共約性の高いものと,そうでない共約の低いものがあるように思われる.とすれば,どこかでオントロジー化が有意義であるレベルとそうでないレベルのあいだに線引きが行われなければならない.だとすれば,なおのこと,共約可能性の保証された体系化を前提としないで,うまい記述単位(例えばフレーム)でオントロジー化の候補となる知識の断片を取り出す,結果に対して中立的な前処理がある方が好ましい. MSFAはそのような手法として機能するように思われる.

2.4 まとめ

MSFAは自然言語とある種のオントロジーの記述との間で二重の役割を演じる.第一に,既存の(ある種の)オントロジーと自然言語文を結びつけるインターフェースとして機能しうる.第二に,MSFAは不特定の文書からオントロジー化の対象となる概念化を探り出す手段ともなる.つまり,オントロジー化の候補を見つける発見的手順として役立つ.これがこの論文で私たちが示唆したいと思ったことである.知識工学者は概念体系の上から,言語学者は概念体系の下から,おのおの探索を始め,概念体系のだいたい中間地点で出会う運命になっているように思われる.

3. 終わりに

呼び名はともかく,高品質の領域オントロジーを構築することは,今も昔も人工知能研究の重要な基礎作業の一つである.その品質は,構想のよしあし,利用アルゴリズムのよしあしと並んで,成果物のよしあしを大きく左右する.だが,私たちが見る限り,高品質の領域オントロジー構築のための汎用的で効果的な解析の手法は確立していないように思われる.特に洗いだしの部分,つまりオントロジー開発の「前処理」の段階では,相当に行き当たりばつたりの作業を強いられているのではないかと想像する.

これがないために,オントロジー開発者は研究の表に出てこない裏の部分では今でも新しい領域に挑戦する度に専門文書の意味解析で四苦八苦しているように思う.知識の自動獲得の技術の精度や性質を考えると,初期データになる人手解析の高度化,省力化,一貫性の向上を可能にする技法は,多くの人々が恩恵に預かれる,有用な技術だと思われる.

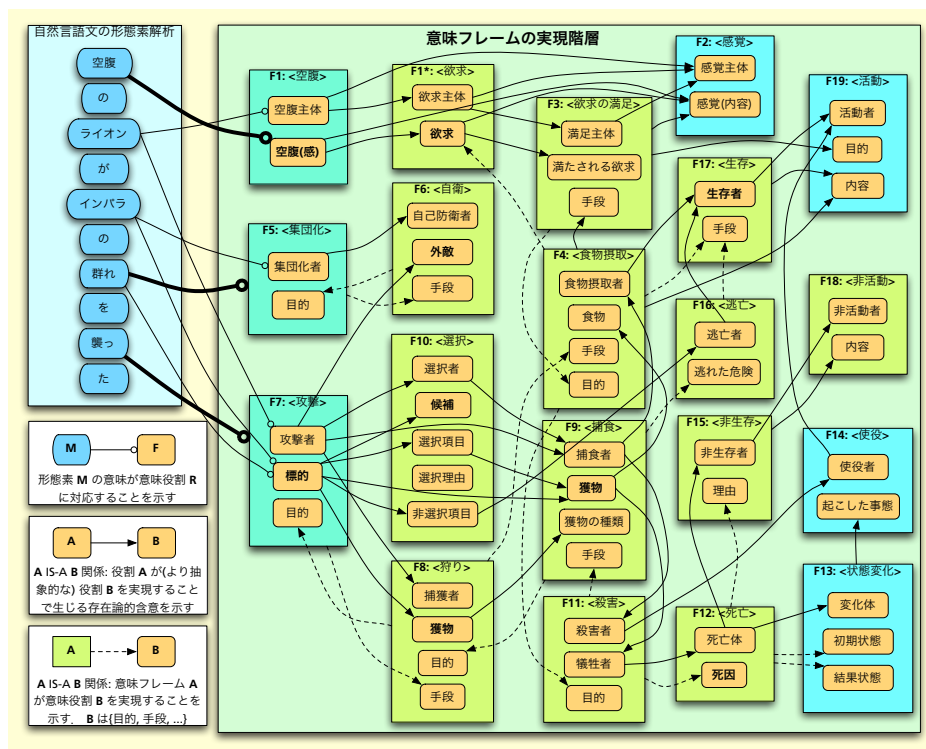


図 3: 意味フレームの階層化を伴った統合

私たちは、自分たちが BFN の研究成果 [2] を参考にしながら開発した MSFA という分析法がその目的のために有効である可能性を示唆した。MSFA は [4] で意味役割タグづけの技術として考案されたものだが、それは言語と世界知識の結びつき「切り出す」という目的のために一般的に使える技法であると考えられる。

オントロジー開発という研究の性質上、試行錯誤な部分を完全に消滅させることは不可能だが、データマイニングの段階で MSFA の手法を適当に利用することで、重要な基本情報の見落としや、誤った体系化の危険や、人手分析の際の時間的負荷、能力的選択性を減らすことが可能だと私たちは考える。実際、MSFA は超人的な能力をもたない平凡な作業者に妥当な知識記述が行えるように作業を支援する環境の一部として機能するように考案されている。この点は [6] に詳しい。

これが正しい示唆であるならば、MSFA は領域オントロジー開発の初期状態に存在する、試行錯誤的で、名人芸の性質の強い解析作業を効率化し、体系化し、おそらく省力化、高品質化する基礎技術の一つとなりうる可能性があるように思われる。

重要なのは、[8] にも明言されているように、オントロジー開発には、ある種の熟練が必要とされるという点である。(A)、(B) のいずれの方略をとるにせよ、熟練性の重要性は明らかである。熟練工の育成のためのノウハウが「暁の上の水練」であるのが現状であるような気がする。そのような熟練を支援する道具は、幾らあってもありすぎるということはないのではないだろうか？

参考文献

[1] 人工知能学会誌: 特集「事例ベース推論」, Vol. 7, No. 14. 日本人工知能学会, 1992.

[2] T. Fontenelle, editor. *International Journal of Lexicography — Special Issue: FrameNet and Frame Semantics*, Vol. 16 (5). Oxford University Press, 2003.

[3] NTT コミュニケーション科学研究所 (監修). 日本語語彙大系. 東京: 岩波書店, 1997.

[4] 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>].

[5] 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>].

[6] 黒田航, 井佐原均. 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の評価版の公開. 自然言語処理学会第 11 回大会発表論文集, pp. A5–3. 自然言語処理学会, 2005. [<http://cls1.hi.h.kyoto-u.ac.jp/~kkuroda/papers/sr-tagging-nlp11.pdf>].

[7] 溝口理一郎. 特集「開発されたオントロジー」. 人工知能学会誌: 特集「開発されたオントロジー」, Vol. 19, No. 2, pp. 135–193, 2004.

[8] 溝口理一郎. オントロジー工学. オーム社, 2005.