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
けいはんな情報通信融合研究センター

NLP 11 [03/17/2005]

研究の背景と目的

- 黒田・井佐原 (2004) [NLP10] は Berkeley FrameNet (BFN) を拡張した意味役割タグつきコーパス開発を開始
- BFNと別に複層意味フレーム分析 (MSFA) と呼ぶ独自の意味タグづけ体系=手法を定義
- 今回の研究の目的: MSFA が技能として習得可能か確かめ，可能ならば文書化する

概要

- MSFA の目的，これまでの成果
- MSFA の理論と実例
- 現状と今後の方針

MSFA の目的

MSFA は何の記述か?

- 文 s の MSFA はヒト x が s を読んだり/聞いたりしたときに x が理解する内容 m(s), すなわち x による s の理解内容 F(x, m(s)) の可能な限り明示的で体系的な記述
- ただし最適でないし, 今は(まだ)完全でもない
  - m(s) ではなく F(x, m(s)) の最適性は x に依存する
  - これは“読み”の“深さ”, “観点”のパラメータ化

意味役割タグづけの動機

- 解析用の辞書とは別に文脈化された意味のデータベース化が必要
  - 意味(内容)分析は(当分)自動化できない
  - ヒトの意味直観は恐ろしく微妙，かつ正確である
- 言語学者だって“使える”(意味)分析を提供できることの証明
  - 役に立たない統語派生の分析や語の分類しかできない訳ではない

ちょいと大袈裟な話

- (自然言語)情報処理の十八番の「知識の自動獲得」もいいのだけれど
- イキナリ自動獲得に走る前に自然言語データから手動で, どんな知識が, どの程度まで獲得できるのかを見極めて/見積もっておく必要はないのでしょうか?
  - Text/Data Mining で「掘る」前に「どこ」に「何」があるか事前調査しなくていいのでしょうか?

これまでの成果

- 作業者四人による人手コーディングと結果の編集 (数値的評価はまだ)
- 京大コーパスから三記事文 (合計64文)
  - 950103083-001,018 (ラグビー)
  - 950101075-001,036 (将棋の名人戦)
  - 950107210-002,010 (マラドーナ逮捕)
  - フレームの延べ数にして500個ぐらい
- http://61.115.230.87/~mutiyama/cgi-bin/hiki/hiki.cgi?FrontPage で部分的に公開
  - 内山将夫 (NiCT) 氏の好意による

なぜ「評価」版?

- 記述の自由度の過剰を収拾するための情報収集
  - MSFA はキリがない = 終了条件が不明確
- 分野ごとの需要は尊重するが, 特定の利用目的に解析を特化させるつもりはない
  - 私たちの考える意味役割タグつきコーパスは NLP の特殊な用途の他にも言語学/認知科学/日本語教育を含め, 可能な限り広い研究分野への利用価値をもつもの
  - 言語学者の(無謀な)突っ走りへの「保険」

お断り

- MSFA は言語学者/言語の認知科学者の仕事なので，NLP での応用は視野に入っているとは言え，それ自体は目的ではない
- 「何のためにやっているの?」と「これは何の役にたつの?」という質問の答えは，別
  - 私が自信をもって答えられるのは前者のみ
  - 最も簡潔な後者への答えは「あなたの想像力次第」

MSFA の理論と実例

MSFA の設計思想

- 格フレーム辞書より細粒度の意味記述
- 読みのポテンシャル空間を特定する必要
  - 解釈は必ずしも“正しい読み”の特定ではない
- 読みの深さをパラメータ化する必要
  - 最適な読み(の深さ)は課題依存的である
- 意味解析の統語情報への依存性を軽減
  - 統語情報は意味解析に不可欠か? — No

MSFA の異例な特徴

- 理解内容の正確で詳細な記述に特化
- 格(助詞)パターンなどの意味の統語的実現状況は記述の対象外
  - 具現化のための最適な形式の選択の問題は捨象
  - この問題は, 語彙概念構造分析 (LCS), FrameNet, 格フレーム辞書にお任せします

意味役割タグづけとは?

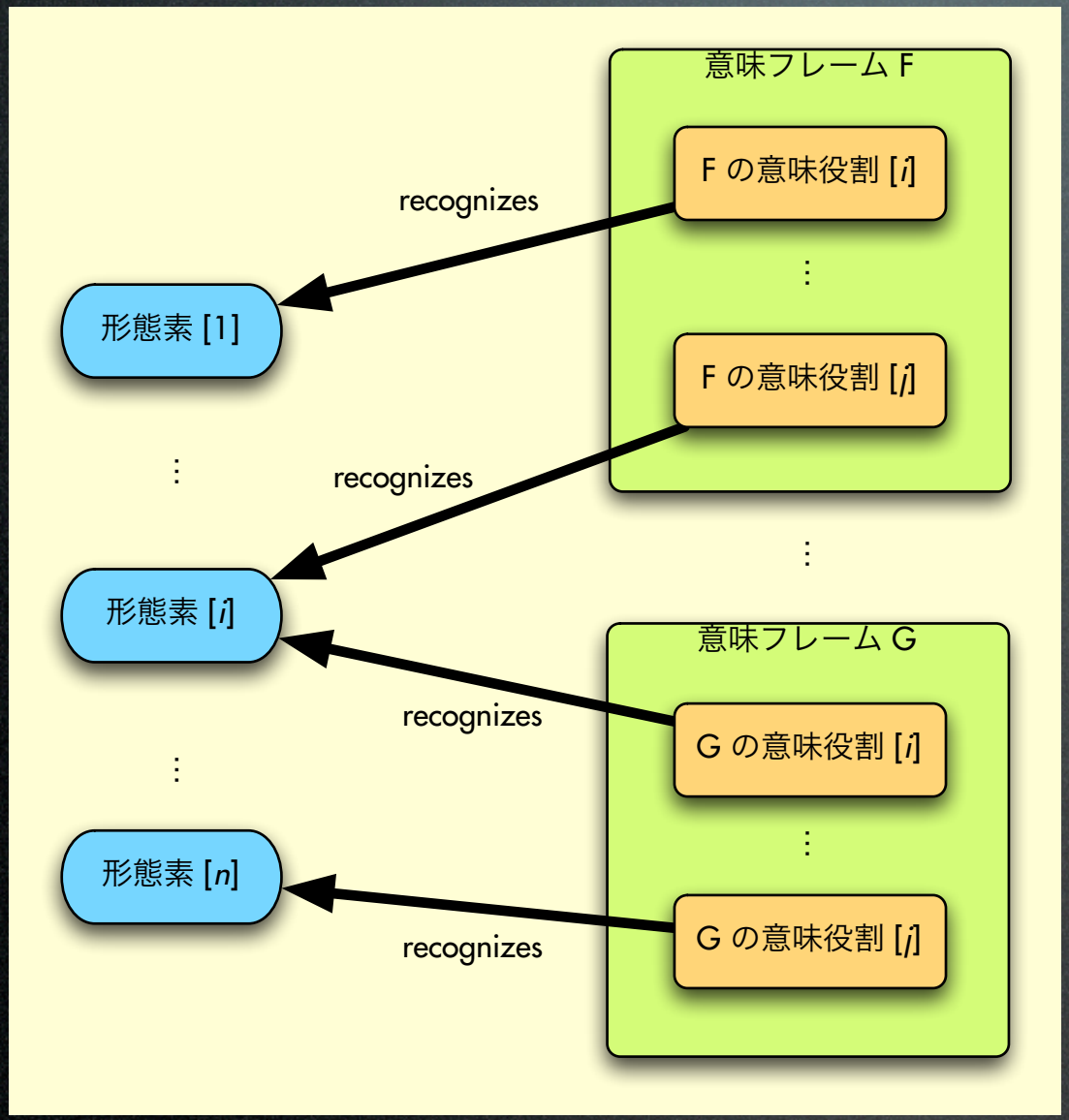
- 状況に依存しない意味型 (e.g., ヒト, 地震) の特定ではなく
- 状況内の意味役割 (e.g., 加害体, 被害者, 犠牲者) の特定
- MSFA は形態素列と複数の意味フレーム列 (= 意味役割のグループ化の列) との対応づけシステム

望ましい意味分析の条件

1. 機械学習可能なぐらい十分な一貫性
2. 文書分野を選ばない網羅性
3. (半)自動化可能な解析 (e.g., 格フレーム辞書) の精度を越える具体性，特定性
4. 用途 (e.g., 機械翻訳) を限定しない汎用性
5. 文脈内の語の意味の多元性を記述する柔軟性

- MSFA は [3,4,5] を最重要視
- シソーラス型 (e.g., EDR, IPAL, WN) は [5] で失格

形態素と意味役割の対応



- 意味解釈は並列パターン認識の一種
- 文中での形態素 (=語) と意味役割 (=フレーム要素) の対応は一対多
- 矛盾のない限り，形態素は幾つの意味役割を実現しても構わない

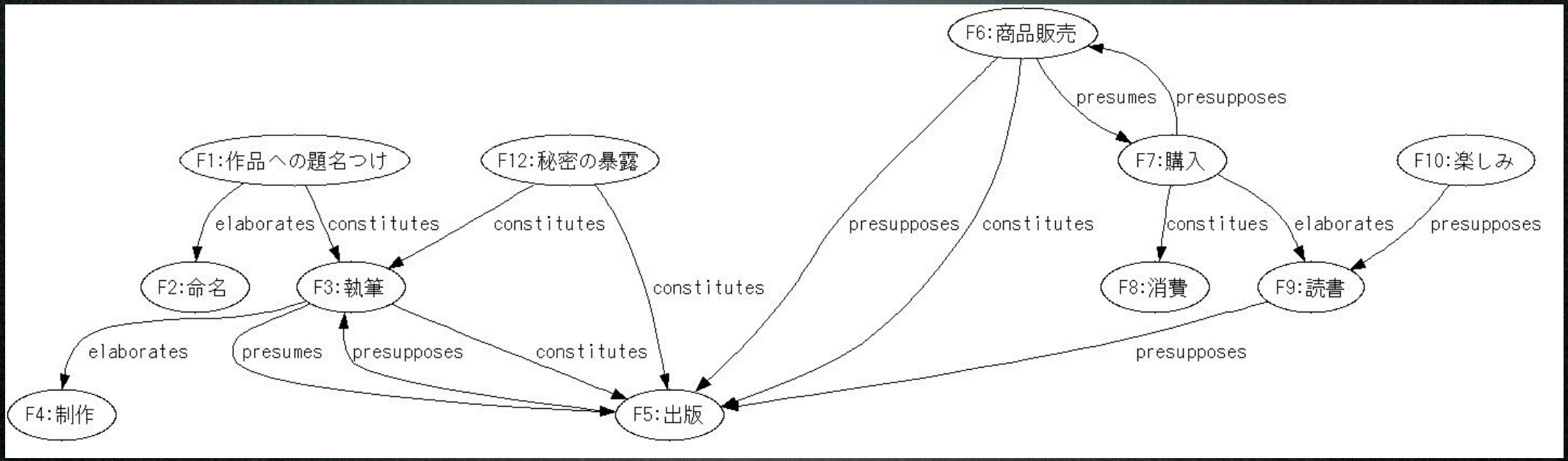
意味フレーム分析の要点

- 「先日二人組の男に襲われたXY銀行の支店は，その先日にも不渡りに見舞われたばかりだった」
  - [二人組の男, XY銀行]: --linked to--> <銀行強盗>フレームの<強盗, 銀行>役割
  - [不渡り, XY銀行]: --linked to--> <打撃発生>フレームの<打撃, (経済)活動体>役割

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Frame ID	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14
2	F-to-F relations	targets F2	targets F3	elaborates F11; constitutes F4; presupposes F5	elaborates F13				constitutes F3	consitututes F3	targets F11	elaborates F12			presupposes F13
3	Frame Identifier	加圧 [+metaphoric]	押さえつけ [+metaphoric]	防御の際の相手の牽制	動作の継続	行動の先導	<節の接続>	行動の連続	集団行動?	攻撃	封印 [+metaphoric]	防御者による攻撃の無力化	大学ラグビーの試合	ラグビーの試合観戦	観戦内容の報告
4	*													観戦者	報告者
5	*[+anaphoric: target=明大]	加圧体	押えつける者	牽制者	継続者		共通項	行動体	集団行動体	攻撃対象	封印者	防御者	チーム1	試合内容	報告内容
6	京産大	加圧対象.特徴	押えつける相手.特徴	牽制相手.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									
8	ゲーム					先導体.特徴							チーム2の主力選手		
9	メーカー														
10	広瀬	加圧対象	押えつける相手	牽制相手		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		節2	行動2	GOV		手段2	手段2			
18	パワー								目的						
19	で										MARKER	MARKER			
20	京産大									攻撃体	攻撃者	攻撃者	チーム2		
21	の										MARKER	MARKER			
22	攻め									GOV	攻撃技	攻撃技			
23	手									手段: 攻撃技		GOV			
24	を										MARKER				
25	封じ										GOV				
26	た						EXT	EXT			EXT	EXT			
27	。														

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	F-ID	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13
2	F-to-F relations	elaborates F2; constitutes F3		constitutes F5; presumes F5; elaborates F4		presupposes F3	presupposes F5; constitutes F5; presumes F7	presupposes F6; constitues F8; elaborates F9		presupposes F5	presupposes F9		constitutes F3,F5	
3	Frame idenfitier	作品への題名つけ	命名	執筆	制作	出版	商品販売	購入	消費	読書	楽しみ	米国での大統領による為政	秘密の暴露	報告
4	*													報告者
5	*					目的		GOVERNOR	GOVERNOR	手段				報告内容 [start1,end]
6	*							目的	手段	GOVERNOR	手段			
7	*								目的	目的	GOVERNOR			
8	*					小売り業者	販売者	販売者	提供者 [ternary]					
9	*					客	客	購入者	消費者	読者	楽しむ者			
10	*	題名付与者 [secondary]	命名者[+aux]			出版社			提供者 [secondary]					
11	*	題名付与者 [primary]	命名者	執筆者	作者	協力者?			提供者 [primary]				暴露者	
12	*			目的1									GOVERNOR	
13	「	MARKER[1,2]	MARKER[1,2]	特徴		出版物.特徴	商品.特徴	商品.特徴	消費物.特徴	読み物.特徴	楽しみ.特徴			報告内容 [start2,end]
14	ホワイトハウス	題名	名称									大統領官邸	標的	
15	の												MARKER	
16	内側												秘密: EVOKED	
17	」	MARKER[2,2]	MARKER[2,2]											
18	と	EVOKER1	EVOKER1											
19	題	GOVERNOR	GOVERNOR	EVOKER1										
20	する	EXTENDER	EXTENDER	EVOKER2										
21	本	作品	命名対象	書物: EVOKER3	作品	出版物	商品	商品	消費物	読み物	楽しみ.源			
22	が					MARKER	MARKER							
23	十四					日時:日付	日時:日付							
24	日													
25	、													
26	米国					場所	場所							
27	で					MARKER	MARKER							
28	発売			目的2		GOVERNOR	GOVERNOR						手段	
29	され					EXTENDER1	EXTENDER1							
30	る					EXTENDER2	EXTENDER2							
31	。													

MSFA を基にした意味フレームの組織化の記述



- “「ホワイトハウスの内側」と題する本が十四日，米国で発売される” (S-ID: 950112062-001) の MSFA を基に FOCAL Wiki ページ (http://61.115.230.87/~mutiyama/cgi-bin/hiki/hiki.cgi?FrontPage) で自動生成 (Graphviz 使用)

文中の語の意味の多元性

- 任意の文 s 中で任意の形態素 m/s に割り当てられる解釈役割 (= 意味役割) r(m/s) はただ一つだとは限らない
  - 意味役割の文脈内相対化による語義の曖昧性解消への非一義化的アプローチ
- (BFN と異なり)形態素に付与される意味役割の数に理論上の上限なし
  - 理論言語学の“常識”からは意識的に逸脱

メタファーやメトニミー

- 生成辞書理論 (Pustejovsky 1995) と同じぐらいはうまくメトニミーを扱える
  - GL と異なり，Qualia 構造の記述はフレーム動詞の相互作用の結果として動的に与えられる
- メタファー (“圧力をかける”, “手を封じる”) がちゃんと扱える
  - メタファーは動的に生成されると考えるので，概念メタファー理論 (Lakoff & Johnson 1980, 1999) とは違い，メタファーの閉じた目録があるとは考えない

現状と今後の方針

現状

- 基礎が固まる前に開発してもしょうがない
  - 評価版を段階的に開発，公開し，軌道修正しながら小規模だが良質の(教師)データの提供
- 難しいのはタグづけ仕様の決定だけでなくタグづけ作業者の確保 (育成を含めて)
  - MSFA は教授可能，習得可能な技能だと判明したが，習得/体得には集中的な訓練が必要
  - 適当な作業者を必要に応じて雇うより，適切な大学の言語学研究室と連携したほうが得なはず

究極の未来?

- 「英辞郎」のように有志の (Web 上での) Open Development に委ねるのがイチバン
  - 何しろ開発費はタダ
  - 作業者(言語学者)も納期に縛られないで嬉しい
- そのために必要なのは
  - MSFA の体系化，作業の明示化
  - 解析結果の品質管理 (FOCAL Wiki で部分的に実現)

今後の方針と課題

- タグづけ結果から有意味な意味フレームを選定し明示的な定義を与える
  - 当面は一部の比較的優先順位の高いフレームのみに定義を与える予定
  - 評価版公開の目的は選定の条件の特定
- 次の段階で意味フレームのデータベース化
- タグづけ仕様の文書化と公開

お断りとお願い

- 仮に意味役割タグづけの企画が正しいゴールをめざしているとしても，先はまだまだ長い
- 従って
  - 過度の期待はもたないでください
  - フィードバックがあれば，その分だけ皆さんの期待に答えられる可能性が高くなります
  - “こういう文はどうするの?!”のような挑戦も歓迎

謝辞

内山将夫 (NICT)
金丸 敏幸, 中本 敬子, 野澤 元, 龍岡昌弘
(FOCAL 研究グループ)
黒宮 公彦 (大阪学院大学)
竹内孔一 (岡山大学)
石山 昌代, 大谷 直輝, 鬼頭 修, 横森大輔
(京大山梨研究室)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Frame ID	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13
2	F-to-F relations	elaborates F2; constitutes F3		constitutes F5; presumes F5; elaborates F4		presupposes F3	presupposes F4; constitutes F5; presumes F7	presupposes F6; elaborates F9		presupposes F5	presupposes F9		constittues F3,F5	
3	Frame	Title Giving	Name Giving	Writing	Authoring	Publishing	Selling	Purchasing	Consuming	Reading	Having Fun	Presidential Government in the U.S.	Disclosure	Reporting
4	*													Reporter
5	*					Purpose		**GOVERNOR**	**GOVERNOR**	Means				Report[start 1,end]
6	*							Purpose	Means	**GOVERNOR**	Means			
7	*								Purpose	Purpose	**GOVERNOR**			
8	*					Retailer	Seller	Seller	Provider3					
9	*					Customer	Customer	Purchaser	Consumer	Reader	Enjoyer			
10	*	Title Giver[secondary]	Name Giver[2]		Supporter	Publisher	Provider		Provider2					
11	*	Title Giver[primary]	Name Giver[1]	Writer	Author	Supporter?			Provider1				Revealer	
12	*			Purpose1							Domain=Topic		**GOVERNOR**	
13	A	Work	Object	Book	Work[+Piece]	Publication	Goods	Goods	Commodity	Book	Fun Source			Report[start 2,end]
14	book													
15	titled	**GOVERNOR**	**GOVERNOR**	Book.attribute	Work.attribute	Publication.attribute	Goods.attributes	Goods.attributes	Commodity.attribtute	Book.attribute	Fun Source.attribute			
16	"	MARKER[1,2]	MARKER[1,2]											
17	The	Title	Name										Secrets: **EVOKER**	
18	Inside													
19	White											Presidential Office: **EVOKER**	Target	
20	House													
21	"	MARKER[2,2]	MARKER[2,2]											
22	will					EXTENDER2	EXTENDER2							
23	go					EXTENDER1	EXTENDER1							
24	on			Purpose2		**GOVERNOR[+composite]**	**GOVERNOR[+composite]**						Means	
25	sale													
26	in					MARKER	MARKER							
27	the					Place	Place							
28	U.S.													
29	on					MARKER	MARKER							
30	January					Time: Date	Time: Date							
31	14													
32	.													

Frame	ラインアウトの場面	集合	突破	前進	時間関係	ラックの場面	文関係（並列）	集中	ラグビー
明大	闘争体1	活動体	活動体	活動体		闘争体1		活動体	闘争体1
は		MARKER	MARKER	MARKER				MARKER	
ラインアウト	EVOKER	場所					項目 1	集中するもの1	ラインアウト：EVOKER
で		LINKER							
の									
密		様態							
集		GOVERNOR							
、							GOVERNER		
*	闘争体2		相手	相手		闘争体2			闘争体2
突			GOVERNOR	様態	前の出来事		項目 2	集中するもの 2	
進				GOVERNOR					
後					GOVERNOR				
の									
ラック					後の出来事	EVOKER			ラック：EVOKER
に								MARKER	
力								力	
を								MARKER	
集中								GOVERNOR	
した								EXTENDER	
。									