

MSFA を使った意味(役割)タグづけ

語彙意味論に基づく言語資源開発 [2]

黒田 航 // NICT

NLP14 Tutorial

2008/03/17

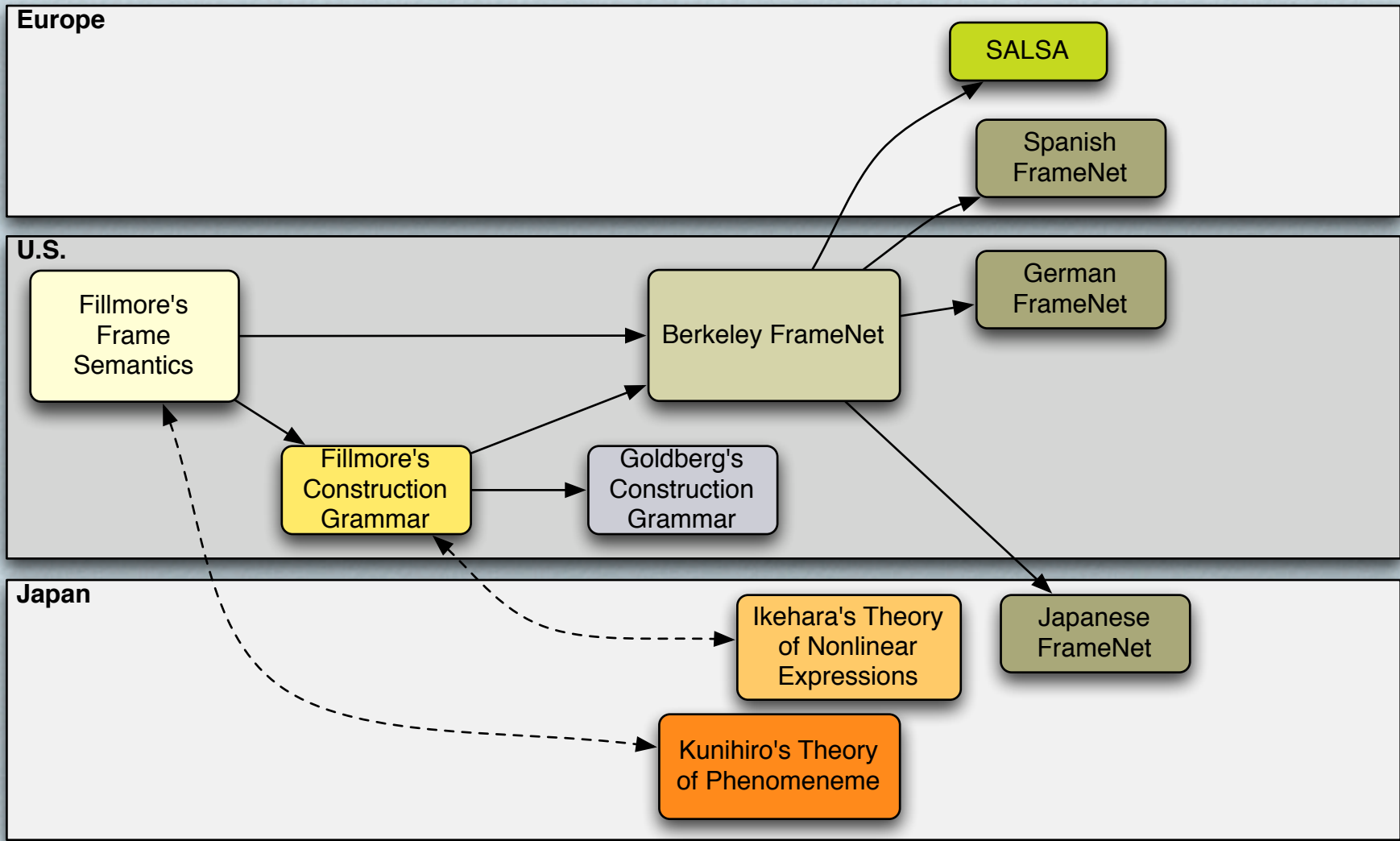
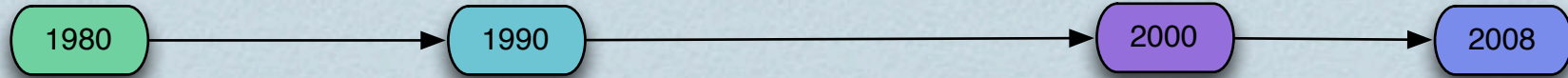
概要

- ❖ 背景
 - ❖ 自分の位置
 - ❖ Frame Semantics/FrameNet の予備知識
 - ❖ 辞書からタグつき用例集へ
- ❖ MSFA Lite を使ったタグづけの見本と解説
- ❖ 知見と議論
- ❖ これから

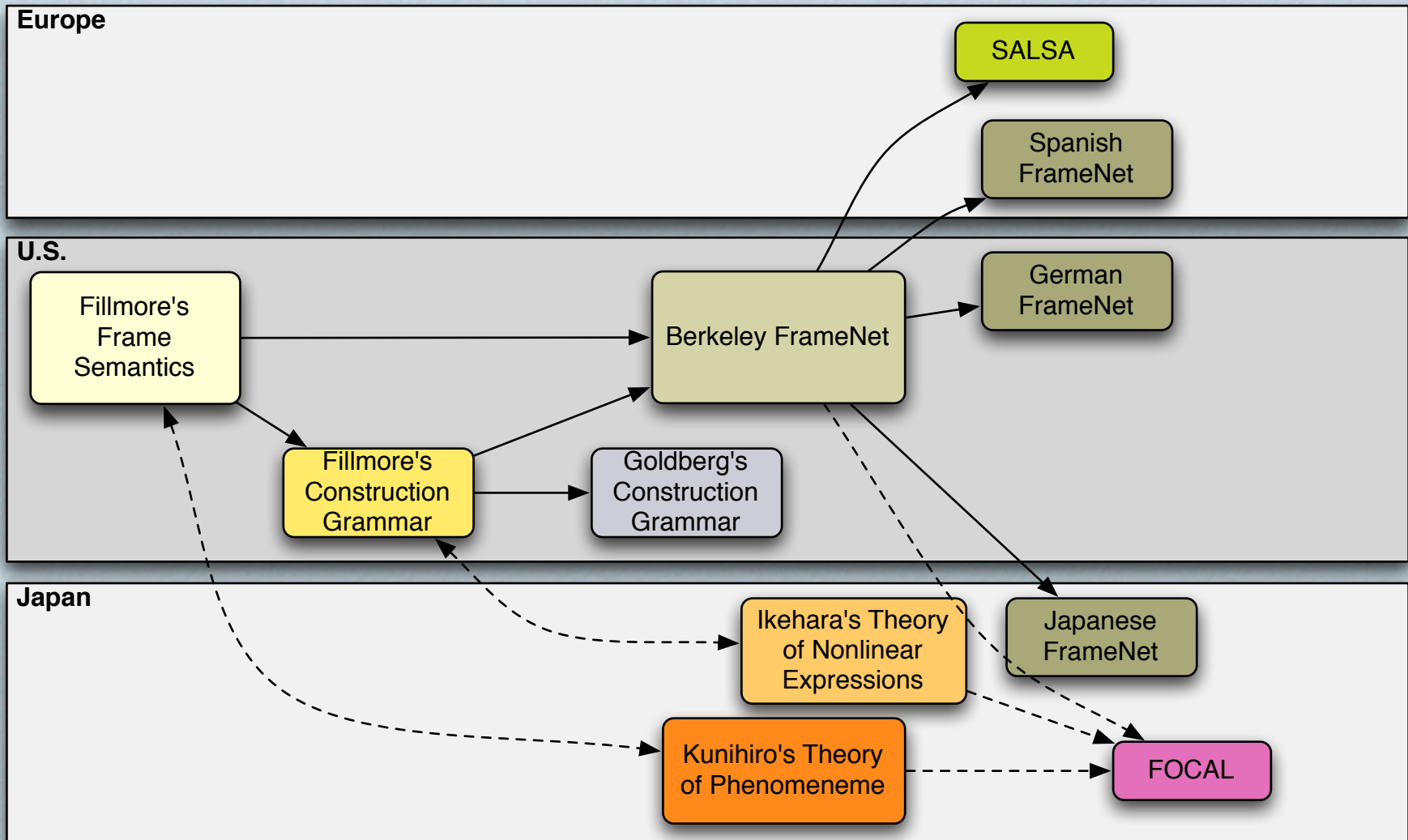
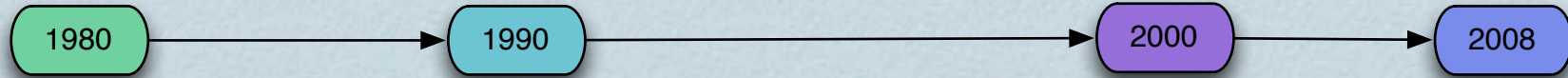


Background

私たちの仕事の“位置”



私たちの仕事の“位置”



中心的な問題

- ❖ 表現の意味の非構成性=非線形性 (池原ら 2005) をどう扱うか?
 - ❖ 連語/ 慣用句/ 比喩表現などを全部合わせて
- ❖ 発想の転換:
 - ❖ 非構成的な場合が常態で “例外” 的なのは実は構成的に見える場合の方
 - ❖ ヒトの言語処理が Rule-based/Principle-based ではなく Example-based/Memory-based なら必然的な帰結

意味の非構成性

	Sは	Oを	捕まえた	逮捕した	捕獲した	呼び止めた
(1)	彼らは	犯人を	I	I	O	O*
(2)	彼らは	被害者を	0.5	O	O	O*
(3)	彼らは	ナマズを	I	O	I	O
(4)	彼らは	メダカを	I	O	0.5	O
(5)	彼らは	タクシーを	I	O	O	I
(6)	彼らは	バスを	0.5	O	O	0.5

状況=事態型のフレーム構造

- ❖ <逮捕>の状況のフレーム要素
 - ❖ <(行)逮捕者 x >, <逮捕の相手 y >, <逮捕の場所 l >, <逮捕の日付 t >, ...
 - ❖ x の代表値は {警察, 検察}, y の代表値は犯罪者
- ❖ <捕獲>の状況のフレーム要素
 - ❖ <捕獲者 x >, <捕獲の対象 y >, <捕獲の道具>, <捕獲の場所 l >, <捕獲の時間帯 t >, ...
 - ❖ y の代表値は x の獲物

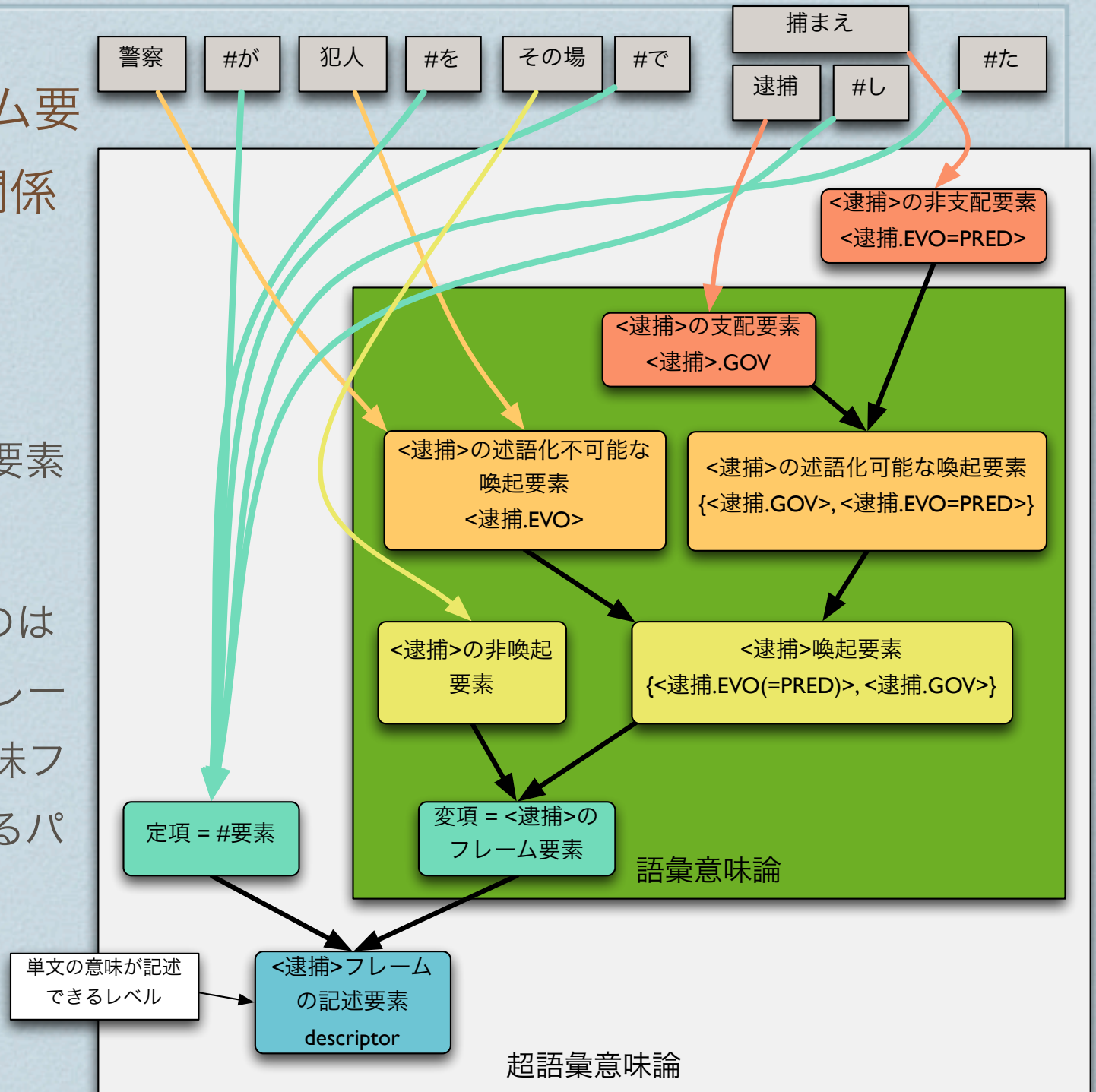
状況=事態型のフレーム構造

- ❖ <タクシーの呼止め>の状況のフレーム要素
 - ❖ <呼び止め手=潜在的な利用者 x >, <流しのタクシー y >, <手段>, <呼び止めの場所 l >, <呼び止めの時間帯 t >, ...
 - ❖ <手段>の代表値は合図 (電話も可?), <呼び止めの場所 l >の代表値は交通の多い通り

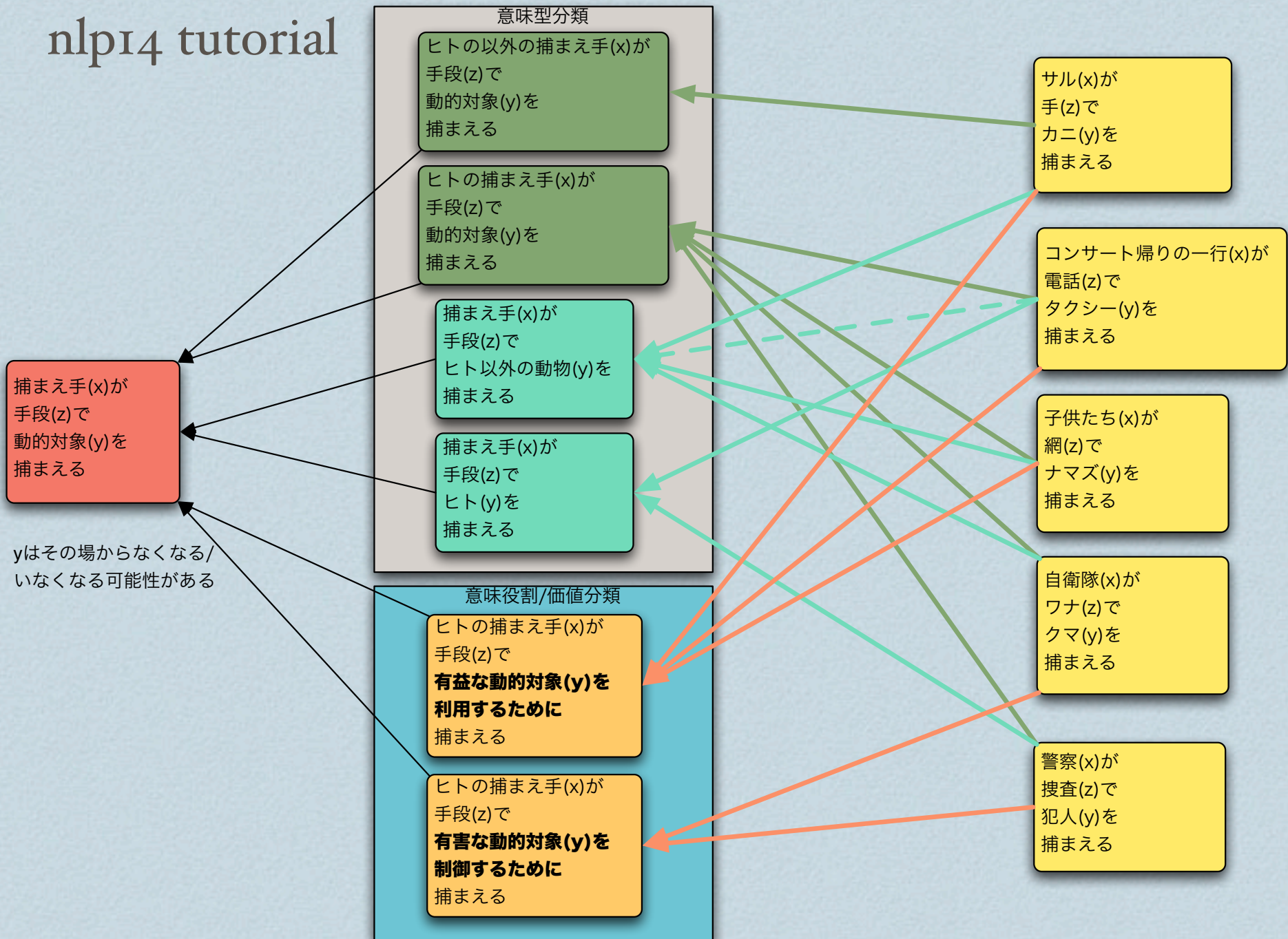
nlp14 tutorial

フレームとフレーム要素とパターンとの関係

- ❖ フレームの記述要素の階層
- ❖ #要素を含むものは正確には意味フレームではなく、意味フレームを実現するパターンの記述



nlp14 tutorial



<目的>の重要性

- ❖ “捕まえ手 x が手段 z で対象 y を捕まえる” の用例の x, y, z 名詞の意味タイプに基づく分類は単に効率が悪いだけでなく見当はずれな結果を生む
- ❖ “ x が z で y を目的 p のために捕まえる” の目的のレベルでの分類が効率が良く、有意義
- ❖ 「チャンスをつまめる」「いい男をつまめる」などのメタファー用法に拡張しているのは、<捕まえ手 x が、動的対象 y を利用するために捕まえる>の方だけ

“辞書”からタグつき用例集へ

- ❖ 辞書 (lexicon) = “文脈自由” な語の意味記述の難点
 - ❖ 低被覆率: 実例には辞書に載っていない例が山ほどある
 - ❖ 低精度: 実例には判別が難しい例が河のようにある
 - ❖ これらは表裏一体
- ❖ MSFA は低精度の問題に対する解決策の一つ
 - ❖ ただし被覆率は犠牲にしている

語の使用の創造性?

- ❖ 文脈依存的な意味記述が必要と言うと「それには際限がないのでは?」と言われる。理由は次:
 - ❖ 語 (の意味) の使われ方は“創造的”でしょ?
 - ❖ “creativity” in word use (Pustejovsky 1995)
- ❖ だが、これは額面通りに受取るべき主張ではない。
 - ❖ “創造的”な言い方で言われている“内容”には (少なくとも見かけほどは) 創造性も新規性もない
 - ❖ ヒトの理解する内容は本質的に紋切り型 (= “創造的”の反意)

nlp14 tutorial

何がしたいのか?

- ❖ Multidimensional Semantic Indexing

- ❖ ただし Indices は固定のリストではなく、一定の“文法”で生成されると考える

辞書と意味理論の関係 1/2

- ❖ 語義の分析・記述は辞書編纂の中で最も難しい部分であるが、その進歩がおそいのは、基になる現代日本語の意味研究が未発達であることが大きな原因であるように思われる。その理由として、学会におけるいままでの現代語研究の蔑視、意味研究よりも文法研究の重視ということが挙げられよう。(国広『理想の国語辞典』, p. 2)

辞書と意味理論の関係 2/2

- ❖ 現行国語辞書の語義記述の不完全な点は以下に示すところから明らかになると思うが、それならばと、多数の分析者と時間を投入すればすぐ成果が出るかということ、そう簡単にはことが運ばないところがこの問題の泣き所である。語義記述の進歩のためには、その理論的な基礎となる意味論が発達しなければならない。しかしこちらの進歩は遅々としている。ともかく、意味論の発達と辞典の発達は雁行しているのである。(国広『理想の国語辞典』, p. 3)

言語資源開発に先立つもの

- ❖ それは
 - ❖ 日本語の意味記述ができる意味分析の専門家の育成
 - ❖ これがMSFA の実践を通して本当に実現したいこと
- ❖ 専門家の育成が意味を扱う言語資源の開発に先決するべきなのは当然なのに、それはNLPで一般に理解されていない
- ❖ 専門家の育成をせず、適当な業者への外注で済ませる風潮は、長い目で見ると良くない影響を生んでいる
 - ❖ 仕様書を書く人に応用から独立した“理論”がないと致命的では?

人手の無用化は可能か? 1/2

- ❖ 大規模データの統計処理の産物は、将来的に人手解析を完全に無用にするような精度に至るか?
 - ❖ 多くのNLP研究者の夢
- ❖ それが原理的に不可能だと考える理由はないが、それを“教師なし”の手法で実現するのは不可能かも知れない
- ❖ 100年後にはどうなっているかわかりませんが、自分が生きている間にそんな夢の時代が来るとは信じ難い

人手の無用化は可能か? 2/2

- ❖ 教師アリの手法が不可欠なら、ちゃんとした教師データがあることが自動化のための必須条件
- ❖ だが、誰がちゃんとした教師データを作成するのか?
 - ❖ 業者に外注して作成した産物にそれが期待できるのか?
- ❖ 目的ごとにちゃんと教師データを作成できる人材を育成しないとまずいのでは?
 - ❖ MSFA は言語分析者の訓練、育成のための枠組み



MSFA Lite Step by Step

発表後の補足

- ❖ “語義”の扱いに関して
 - ❖ MSFA は (分類ラベルとしての) 語義を想定しない
 - ❖ 語義に想定するのは
 - ❖ 動詞の場合は状況のクラスター
 - ❖ 名詞の場合はおのおのの状況を構成する意味役割の配列

MSFA Lite

- ❖ 2007年夏に導入された新しい仕様で、次が主な特徴
 - ❖ 効率化のためにタグづけ対象となるフレーム範囲の制限
 - ❖ 当初は“述語一つについてフレーム一つ”と考えていたが実は ...
 - ❖ NLPの利用条件を考え、フレームごとにフレーム要素の MARKER の値を明示化
 - ❖ 代表的な述語を選び、その表層格パターンを明示化
 - ❖ 状況 (のフレーム) は本来、述語とは別に存在する抽象的な実体
 - ❖ MSFA のパターン辞書化の土台

注意

- ❖ MSFA の目的は辞書の構築ではないが、意味タグづけの副産物として次のものが構築される
 - A. フレームのデータベース (Frame List) (10000+)
 - B. フレーム要素のデータベース (FrameElement List) (70000+)
 - C. フレーム要素の実現値になっている部分文字列のデータベース (Word/Phrase List) (40000+)
 - ❖ A, B の階層化の産物として C が複合表現のシソーラスになる
 - D. 意味に対応したパターン/構文のデータベース (未実装)

MSFA Lite の見本

- ❖ 高知県の橋本大二郎知事は三十一日、都道府県で初めて一般事務職の採用資格から国籍条項を撤廃する方針を明らかにした。[京大コーパス: 950101009-001]

Step 0: 形態素への分割

- ❖ JUMANの解析結果を人手修正
- ❖ [高知, 県, の, 橋本, 大二郎, 知事, は, 三十一, 日, 、, , 都道府県, で, 初め~, て, 一般, 事務, 職[員], の, 採用, 資格, から, 国籍, 条項, を, 撤廃, する, 方針, を, 明らか~, に, し~, た, 。]
- ❖ [...]は補った要素
- ❖ “X~Y” は JUMAN より細かい分割

nlp14 tutorial

Step I

- ❖ 動詞の活用形とサ変名詞を選んで、フレーム f の喚起要素 $f.EVO$ か支配要素 $f.GOV$ の認定と暫定的な命名
- ❖ Setting の列に背景情報の明示化 [任意]

	A	B	C	D	E	F
1	Frame-ID	F0	F1	F2	F3	F4
2	Frame-to-Frame Relations					
3	Frame Name	Setting	初回性の指定	採用	撤廃	明示化
4	%% Predicate for GOV					
5	高知	X[instance-of X[日本の1都1道 2府43県]]				
6	県					
7	の					
8	橋本					
9	大二郎					
10	知事	X[高知県知事]				
11	は					
12	三十一	X[Y年M月31日]				
13	日					
14	,					
15	都道府県	X[日本の1都1道 2府43県]				
16	で					
17	初め~		初回性の指 定.EVO			
18	て					
19	一般					
20	事務					
21	職					
22	[員]					
23	の					
24	採用			採用.GOV		
25	資格					
26	から					
27	国籍	X[日本国籍]				
28	条項					
29	を					
30	撤廃				撤廃.GOV	
31	する					
32	方針					
33	を					
34	明らか~					明示 化.EVO=PRED
35	に					
36	し~					
37	た					
38	。					

nlp14 tutorial
Step I

9	大二郎				
10	知事	X[高知県知事]			
11	は				
12	日	X[Y年M月31日]			
13	日				
14					
15	都道府県	X[日本の1都1道 2府43県]			
16	で				
17	初め~		初回性の指 定.EVO		
18	て				
19	一般				
20	事務				
21	職				
22	[員]				
23	の				
24	採用			採用.GOV	
25	資格				
26	から				
27	国籍	X[日本国籍]			
28	条項				
29	を				
30	撤廃				撤廃.GOV
31	する				
32	方針				
33	を				
34	明らか~				明示 化.EVO=PRED
35	に				
36	し~				
37	た				
38	。				

nlp14 tutorial

Step 2

- ❖ Step I で特定されたフレーム f_1, \dots, f_n のフレーム要素の特定
- ❖ f 者, f 体, f の対象, f の手段, f の時期, f の内容, ..
- ❖ フレーム要素名の指定には [Attribute: Value] (e.g., f .EVO: f 者) 表記を許している
- ❖ [γ ; (f .EVO:) f . X] という表記では γ が語彙化された要素 (e.g., 獲物)

	A	B	C	D	E	F
1	Frame-ID	F0	F1	F2	F3	F4
2	Frame-to-Frame Relations					
3	Frame Name	Setting	初回性の指定[事態の]	採用	撤廃	明示化
4	%% Predicate for GOV				止	明示化し
5	*	X[記者][著者]	初回性の指定[事態の]者			
6	**		#によると			
7	*	X[読者]	相手[初回性の指定[事態の]の]			
8	**		#NONE			
9	高知	X[instance-of X[日本の1都1道2府43県]]	対象[初回性の指定[事態の]の]<1,2>: 事態<1,2>	採用者 [+metonymic]	撤廃者 [+metonymic]	対象,内容[明示化の]<1,2>
10	県					
11	の		#が,で	#が,で	#が,で	#が,で
12	橋本					明示化者
13	大二郎					
14	知事	X[高知県知事]				
15	は					#が
16	三十一	X[Y年M月31日]				時期[明示化の]
17	日					
18	**					#(に)
19	,					
20	都道府県	X[日本の1都1道2府43県]	範囲[初回性の指定[事態の]の]		様態,条件[撤廃の]	様態,条件[明示化の]
21	で		#で			
22	初め~		初回性の指定[事態の].EVO			
23	て				#て	#て
24	一般		対象[初回性の指定[事態の]の]<2,2>: 事態<2,2>	対象[採用の]	場所[撤廃の]	対象,内容[明示化の]<2,2>
25	事務					
26	職					
27	[員]					
28	の			#を		
29	採用			採用.GOV		
30	資格			条件[採用の]		
31	から			#に基づいて	#から	
32	国籍	X[日本国籍]			対象[撤廃の]	
33	条項					
34	を				#を	
35	撤廃				撤廃.GOV	
36	する				撤廃.GOV.SUP+EXT	
37	方針					
38	を					#を
39	明らか~					明示化.EVO=PRED
40	に					
41	し~					
42	た					明示化.EVO.EXT
43	。					

4	for GOV					
5	*	X[記者][著者]	初回性の指定[事 態の]者			
6	**		#によると			
7	*	X[読者]	相手[初回性の指 定[事態の]の]			
8	**		#NONE			
9	高知	X[instance-of X[日本の1都1道 2府43県]]	対象[初回性の指 定[事態の] の]<1,2>: 事態 <1,2>	採用者 [+metonymic]	撤廃者 [+metonymic]	対象,内容[明示 化の]<1,2>
10	県					
11	の		#が,で	#が,で	#が,で	#が,で
12	橋本					明示化者
13	大二郎					
14	知事	X[高知県知事]				
15	は					#が
16	三十一	X[Y年M月31日]				時期[明示化の]
17	日					
18	**					#(に)
19	,					
20	都道府県	X[日本の1都1道 2府43県]	範囲[初回性の指 定[事態の]の]		様態,条件[撤廃 の]	様態,条件[明示 化の]
21	で		#で			
22	初め~		初回性の指定[事 態の].EVO			
23	て				#て	#て
	一般		対象[初回性の指 定[事態の] の]<2,2>: 事態	対象[採用の]	場所[撤廃の]	対象,内容[明示 化の]<2,2>

alpha4 tutorial

Step 2

Step 3

1	Frame-ID	F0	F1	F2	F3	F4	F5	F6	F7
2	Frame-to-Frame								
3	Frame Name	Setting	報告[-v_evoked]	初回性の指定[事態の]	採用	条件の指定[行為の][-v_evoked]	撤廃	予定[-v_evoked]	明示化
4	%% Predicate for GOV		報告している		採用する	求める		予定{がある, ている}	
5		V[記者]者	報告[-v_evoked]者	初回性の指定[事態の]者					
6			#が	#によると					
7	*	X[読者]	相手[報告[-v_evoked]の]	相手[初回性の指定[事態の]の]					
8	**		#に	#NONE					
9	高知	X[instance-of X[日本の1都1道2府43県]]	内容[報告[-v_evoked]の]	対象[初回性の指定[事態の]の]<1,2>: 事態<1,2>	採用者 [+metonymic]	条件の指定[行為の][-v_evoked]者 [+metonymic]	撤廃者 [+metonymic]	予定[-v_evoked]者 [+metonymic]	対象,内容[明示化の]<1,2>
10	県								
11	の			#が,で	#が,で	#が,で	#が,で	#が	#が,で
12	橋本								明示化者
13	大二郎								
14	知事	X[高知県知事]							明示化者.EXT
15	は								#が
16	三十一	X[Y年M月31日]							時期[明示化の]
17	日								
18	**								#(に)
19	,								
20	都道府県	X[日本の1都1道2府43県]		範囲[初回性の指定[事態の]の]			様態[撤廃の].ARG	様態[予定[-v_evoked]の].ARG	様態[明示化の].ARG
21	で			#で					
22	初め~			初回性の指定[事態の].EVO			様態[撤廃の]	様態[予定[-v_evoked]の]	様態[明示化の]
23	て						#で	#で	#で
24	一般			対象[初回性の指定[事態の]の]<2,2>: 事態<2,2>	対象[採用の].MOD	対象[条件の指定[行為の][-v_evoked]の]	元[撤廃の].MOD	内容[予定[-v_evoked]の]	対象,内容[明示化の]<2,2>
25	事務								
26	職				対象[採用の]				
27	[員]								
28	の				#を				
29	採用				採用.GOV				
30	**					#に			
	資格				条件[採用の]	条件の指定[行為の][-v_evoked].EVO	元[撤廃の]		

Step 3 の要点

- ❖ 事態喚起名詞 (e.g., X の資格, X する方針, X の姿勢) の認定
 - ❖ “ X の資格を{もつ, 有する}”, “ X の方針を{取る, 進める}”, “ X の姿勢を{取る, 見せる}” のように支援動詞構文をなす
 - ❖ “もつ”, “取る” “見せる” が喚起する事態は, 連語が喚起する事態と一致しない
- ❖ <予定>フレームの認定
 - ❖ 「*予定する」とは言えないが, 元の文で「方針」と「予定」を交換しても文意が保存される

nlp14 tutorial

Step 4

任意だが重要

- ❖ フレーム間関係の特定
- ❖ 語義の階層関係, ズレの指定によって意味の多重性の明示化

	A	B	C	D	E	F	G	H	I	J	K	L
1	Frame-ID	F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
2	Frame-to-Frame Relations		presupposes F10	characterizes F5		presupposes F3	elaborates_&_targets F7; presupposes	targets F5; ?targets F7	presupposes F4		targets F10	
3	Frame Name	Setting	報告[-v_evoked]	初回性の指定[事態の]	採用	条件の指定[行為の][-v_evoked]	撤廃	除去 [+metaphoric]	廃止	予定[-v_evoked]	明示化	公表
4	% Predicate for GOV		報告している		採用する	求めている				予定{がある, ている}		
5	*	X[記者][著者]	報告[-v_evoked]者	初回性の指定[事態の]者								
6	**		#が	#によると								
7	*	X[読者]	相手[報告[-v_evoked]の]	相手[初回性の指定[事態の]の]								
8	**		#に	#NONE								
9	高知	X[instance-of X[日本の1都1道2府43県]]	内容[報告[-v_evoked]の]	対象[初回性の指定[事態の]の]<1,2>; 事態<1,2>	採用者 [+metonymic]	条件の指定[行為の][-v_evoked]者 [+metonymic]	撤廃者 [+metonymic]	除去 [+metaphoric]者 [+metonymic]	廃止者 [+metonymic]	予定[-v_evoked]者 [+metonymic]	対象,内容[明示化]の<1,2>	対象,内容[公表]の<1,2>
10	県											
11	の				#が,で	#が,で	#が,で	#が,で	#が,で	#が,で	#が	#が,で
12	橋本										明示化者	公表者
13	大二郎											
14	知事	X[高知県知事]										
15	は										#が	#が
16	三十一	X[Y年M月31日]									時期[明示化]の	時期[公表]の
17	日											
18	**										#(に)	#(に)
19	,											
20	都道府県	X[日本の1都1道2府43県]		範囲[初回性の指定[事態の]の]			様態[撤廃]の.ARG	様態[除去 [+metaphoric]の].ARG	様態[廃止]の.ARG	様態[予定[-v_evoked]の].ARG	様態[明示化]の.ARG	様態[公表]の.ARG
21	で			#で								
22	初め~			初回性の指定[事態の].EVO			様態[撤廃]の	様態[除去 [+metaphoric]の].MOD	様態[廃止]の	様態[予定[-v_evoked]の]	様態[明示化]の	様態[公表]の
23	て						#で	#で	#で	#で	#で	#で
24	一般			対象[初回性の指定[事態の]の]<2,2>; 事態<2,2>	対象[採用]の.MOD	対象[条件の指定[行為の][-v_evoked]の].ARG	元[撤廃]の.MOD	元[除去 [+metaphoric]の].MOD	対象[廃止]の.MOD.MOD	内容[予定[-v_evoked]の]	対象,内容[明示化]の<2,2>	対象,内容[公表]の<2,2>
25	事務											
26	職				対象[採用]の							
27	[員]				#を	#の						
28	の				採用.GOV	対象[条件の指定[行為の][-v_evoked]の]						
29	採用					#に						
30	**											
31	資格				条件[採用]の	条件の指定[行為の][-v_evoked].EVO; 条件; 対象[指定]の]	元[撤廃]の	元[除去 [+metaphoric]の]	対象[廃止]の.MOD.MOD			
32	から				#に基づいて	#を	#から	#から	#にある,の			
33	国籍	X[日本国籍]					対象[撤廃]の.MOD	対象[除去 [+metaphoric]の].MOD	対象[廃止]の.MOD			
34	条項						対象[撤廃]の	対象[除去]の	対象[廃止]の			
35	を						#を	#を	#を			
36	撤廃						撤廃.GOV	除去 [+metaphoric].GOV	廃止.GOV			
37	する						撤廃.GOV.SUP+EXT	除去 [+metaphoric].GOV.SUP+EXT	廃止.GOV.SUP+EXT			
38	**									#(という)		
39	方針									予定[-v_evoked].EVO =PRED		
40	を										#を	#を
41	明らか~										明示化.EVO	公表.EVO=PRED
42	に											
43	し~										明示化.EVO.SUP	公表.EVO.SUP
44	た										明示化.EVO.SUP+EXT	公表.EVO.SUP+EXT
45	**			#と,ことを								
46	-											

Step 4

2	Frame-to-Frame Relations		presupposes F10	characterizes F5		presupposes F3	elaborates_& targets F7; presupposes	targets F5; ?targets F7	presupposes F4		targets F10	
3	Frame Name	Setting	報告[-v_evoked]	初回性の指定[事態の]	採用	条件の指定[行為の][-v_evoked]	撤廃	除去 [+metaphoric]	廃止	予定[-v_evoked]	明示化	公表
4	%% Predicate for GOV		報告している		採用する	求めている				予定{がある,ている}		
5	*	X[記者][著者]	報告[-v_evoked]者	初回性の指定[事態の]者								
6	**		#が	#によると								
7	*	X[読者]	相手[報告[-v_evoked]の]	相手[初回性の指定[事態の]の]								
8	**		#に	#NONE								
9	高知	X[instance-of X[日本の1都1道2府43県]]	内容[報告[-v_evoked]の]	対象[初回性の指定[事態の]の]<1,2>: 事態<1,2>	採用者 [+metonymic]	条件の指定[行為の][-v_evoked]者 [+metonymic]	撤廃者 [+metonymic]	除去 [+metaphoric]者 [+metonymic]	廃止者 [+metonymic]	予定[-v_evoked]者 [+metonymic]	対象,内容[明示化の]<1,2>	対象,内容[公表の]<1,2>
10	県											
11	の			#が,で	#が,で	#が,で	#が,で	#が,で	#が,で	#が	#が,で	#が,で
12	橋本										明示化者	公表者
13	大二郎											
14	知事	X[高知県知事]										
15	は									#が	#が	
16	三十一	X[Y年M月31日]									時期[明示化の]	時期[公表の]
17	日											
18	**										#(に)	#(に)
19	,											
20	都道府県	X[日本の1都1道2府43県]		範囲[初回性の指定[事態の]の]			様態[撤廃の].ARG	様態[除去 [+metaphoric]の].ARG	様態[廃止の].ARG	様態[予定[-v_evoked]の].ARG	様態[明示化の].ARG	様態[公表の].ARG
21	で			#で								
22	初め~			初回性の指定[事態の].EVO			様態[撤廃の]	様態[除去 [+metaphoric]の].ARG	様態[廃止の]	様態[予定[-v_evoked]の]	様態[明示化の]	様態[公表の]
23	て						#で	#で	#で	#で	#で	#で
24	一般			対象[初回性の指定[事態の]の]<2,2>: 事態<2,2>	対象[採用の].MOD	対象[条件の指定[行為の][-v_evoked]の].ARG	元[撤廃の].MOD	元[除去 [+metaphoric]の].MOD	対象[廃止の].MOD.MOD.MOD	内容[予定[-v_evoked]の]	対象,内容[明示化の]<2,2>	対象,内容[公表の]<2,2>
25	事務											
26	職				対象[採用の]							
27	[員]											
28	の				#を	#の						
29	採用				採用.GOV	対象[条件の指定[行為の][-v_evoked]の]						
30	**					#に						
31	資格				条件[採用の]	条件の指定[行為の][-v_evoked].EVO: 条件; 対象[指定の]	元[撤廃の]	元[除去 [+metaphoric]の]	対象[廃止の].MOD.MOD			
32	から				#に基づいて	# 3 5	#から	#から	#にある,の			

Step 4 の要点

- ❖ 「(条項の) 撤廃」から「(条項の) 廃止」へ語義の一般化すると「廃止」がカラ格を項に取らないことがわかる
- ❖ <廃止者>が, (<条件>{にある, の, *から}) <条項>を, 廃止する
- ❖ つまりこの文脈での<撤廃>の意味は<廃止>には帰着できないことがわかる
- ❖ <撤廃>の上位クラスにもう一つ, カラを要求する<除去 [+metaphoric]>を認定して解決

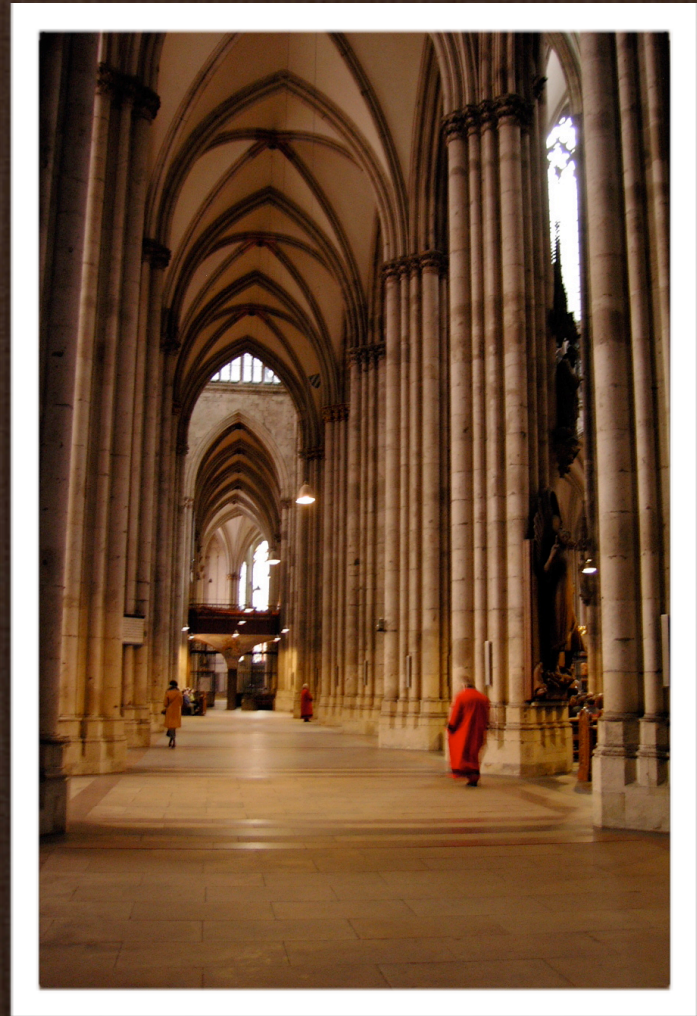
Steps 3, 4 の重要性

- ❖ 次は Step 4 (か Step 3) から判るが Step 2 から判る保証なし
 - ❖ 「撤廃」は<廃止>の特殊な場合を意味するが、その格パターン“動作主(x)が 場所(l)から モノ(y)を V”が<除去>のメタファーに由来
 - ❖ 「明らかにした」の字義通りの意味<明示化>がこの文脈で<公表>になる
 - ❖ これは「<団体>が (<部外者>に対して) <行動>の方針を明らかにする」という連語の効果

精度と被覆率のジレンマ

- ❖ フレームが決まっても述語は一つには決まらない
 - ❖ フレームと述語の対応は一对多
- ❖ だが、述語を特定しないと#要素 (定項)の値が決まらない
 - ❖ 述語と表層格パターンの対応は一对他
- ❖ フレーム記述の一般性=被覆率の低下 => 精度の向上

Managing Annotation



精度保証のためのチェック

- ❖ 文生成プログラム (fbsgen) の出力を見てタグづけの妥当性を評価
 - ❖ 定項と変項の実現値をフレームごとに自動連結したものが自然な日本語文にならないなら，記述は妥当性を欠く
 - ❖ フレーム要素への区切り方と#要素の値の妥当性は検査できるが，フレーム要素名の妥当性は検査できない

理論

- ❖ MSFA Lite のフレーム記述を非線型表現/パターン (池原ら 2005, 2006) と同一視するなら, MSFA Lite を使った文 s の分析は, s の複数の非線型表現/パターン p_1, p_2, \dots, p_n への分解だと見なせる
- ❖ 元の s は p_1, p_2, \dots, p_n の行ごとの“重ね合わせ”で得られる
- ❖ これが正しいなら, p_i がおのこの非線型な単文を記述しているはずなので, p_i の妥当性評価が MSFA の間接的な妥当性評価になる

nlp14 tutorial fbsgen の動作

- ❖ #で始まる要素 (e.g., #が) を定項, #で始まらない要素 (e.g, 採用者) を変項
- ❖ 列を縦にスキャンし, 範囲が定項ならその値を, 変項ならその実現値を一系列の形態素列に取りに行く. $f.GOV, f.X=PRED$ の場合だけ, フレーム名か Predicate for GOVの値を参照

Frame Name	Setting	報告[-v_evoked]	初回性の指定[事態の]	採用	条件の指定[行為の][-v_evoked]	撤廃
%% Predicate for GOV		報告している		7 採用	求めている	
*	X[記者][著者]	報告[-v_evoked]者	初回性の指定[事態の]者			
**		#が	#によると	1		
*	X[読者]	相手[報告[-v_evoked]の]	相手[初回性の指定[事態の]の]			
**		#に	#NONE			
2	X[instance-of X[日本の1都1道2府43県]]	内容[報告[-v_evoked]の]	対象[初回性の指定[事態の]の]<1,2>: 事態<1,2>	採用者 [+metonymic]	条件の指定[行為の][-v_evoked]者 [+metonymic]	撤廃者 [+metonymic]
	の		#が,で	#が,で	#が,で	#が,で
	橋本 大二郎					
	知事 X[高知県知事]					
	は					
	三十一 X[Y年M月31日]					
	日					
	**					
	,					
	都道府県 X[日本の1都1道2府43県]		範囲[初回性の指定[事態の]の]			様態[撤廃の].ARG
	で		#で			
	初め~		初回性の指定[事態の].EVO			様態[撤廃の]
	て					#で
4	一般		対象[初回性の指定[事態の]の]<2,2>: 事態<2,2>	対象[採用の].MOD	対象[条件の指定[行為の][-v_evoked]の].ARG	元[撤廃の].MOD
	事務			対象[採用の]		
5	職 [員]					
	の			6 #を	#の	
	採用			採用.GOV	対象[条件の指定[行為の][-v_evoked]の]	
	**			8	#に	
9	資格			条件[採用の]	条件の指定[行為の][-v_evoked].EVO: 条件; 対象[指定の]	元[撤廃の]
	から			10 基づいて	#を	#から
	国籍 X[日本国籍]					対象[撤廃の].MOD
	条項					対象[撤廃の]
	を					#を
	撤廃					撤廃.GOV

Source MSFA	Column	Frame	Source Frame Name	Generated Sentence With Annotation	Generated Sentence Without Annotation
				高知/ 県/ の/ 橋本/ 大二郎/ 知事/ は/ 三十一/ 日/ **=/、/ 都道府県/ で/ 初め~/ て/ 一般/ 事務/ 職/ [員]/ の/ 採用/ **=/ 資格/ から/ 国籍/ 条項/ を/ 撤廃/ する/ **=/ 方針/ を/ 明らか~/ に/ し~/ た#と、ことを=MARKER// 報告している =PRED[+added]//	// *=X[記者][著者]#が// *=X[読者]#に// 高知/ 県/ の/ 橋本/ 大二郎/ 知事/ は/ 三十一/ 日/ **=/、/ 都道府県/ で/ 初め~/ て/ 一般/ 事務/ 職/ [員]/ の/ 採用/ **=/ 資格/ から/ 国籍/ 条項/ を/ 撤廃/ する/ **=/ 方針/ を/ 明らか~/ に/ し~/ た#と、ことを// 報告している//
950101009-0	3	F1	報告[-v_evoked]	// *=X[記者][著者]#によると=MARKER// *=X[読者]#NONE=MARKER// 高知/ 県#が、で=MARKER// 都道府県#で=MARKER// 初め~/ て/ 一般/ 事務/ 職/ [員]/ の/ 採用/ **=/ 資格/ から/ 国籍/ 条項/ を/ 撤廃/ する/ NULL=PRED[+added]//	// *=X[記者][著者]#によると// *=X[読者]#NONE// 高知/ 県#が、で// 都道府県#で// 初め~/ て/ 一般/ 事務/ 職/ [員]/ の/ 採用/ **=/ 資格/ から/ 国籍/ 条項/ を/ 撤廃/ する/ NULL//
950101009-0	4	F2	初回性の指定[事態の]	// 高知/ 県#が、で=MARKER// 一般=MOD/ 事務/ 職/ [員]#を=MARKER// 採用する=PRED/ 資格#で、に基づいて=MARKER//	// 高知/ 県#が、で// 一般/ 事務/ 職/ [員]#を// 採用する/ 資格#で、に基づいて//
950101009-0	5	F3	採用	// 高知/ 県#が、で=MARKER// 一般=ARG/ 事務/ 職/ [員]#の=MARKER// 採用#に=MARKER// 資格#を=MARKER// 求めている=PRED[+added]//	// 高知/ 県#が、で// 一般/ 事務/ 職/ [員]#の// 採用#に// 資格#を// 求めている//
950101009-0	6	F4	条件の指定[行為の] [-v_evoked]	// 高知/ 県#が、で=MARKER// 都道府県=ARG/ で/ 初め~/ #て=MARKER// 一般=MOD/ 事務/ 職/ [員]/ の/ 採用/ **=/ 資格#から=MARKER// 国籍=MOD/ 条項#を=MARKER// 撤廃=PRED/ する=SUP/	// 高知/ 県#が、で// 都道府県/ で/ 初め~/ #て// 一般/ 事務/ 職/ [員]/ の/ 採用/ **=/ 資格#から// 国籍/ 条項#を// 撤廃/ する/
950101009-0	7	F5	撤廃	// 高知/ 県#が、で=MARKER// 都道府県=ARG/ で/ 初め~/ #て=MARKER// 一般=MOD/ 事務/ 職/ [員]/ の/ 採用/ **=/ 資格#から=MARKER// 国籍=MOD/ 条項#を=MARKER// 除去 [+metaphoric]=PRED/ する=SUP/	// 高知/ 県#が、で// 都道府県/ で/ 初め~/ #て// 一般/ 事務/ 職/ [員]/ の/ 採用/ **=/ 資格#から// 国籍/ 条項#を// 除去 [+metaphoric]/ する/
950101009-0	8	F6	除去[+metaphoric]	// 高知/ 県#が、で=MARKER// 都道府県=ARG/ で/ 初め~/ #て=MARKER// 一般=MOD/ 事務/ 職/ [員]/ の/ 採用/ **=/ 資格=MOD#にある,の=MARKER// 国籍=MOD/ 条項#を=MARKER// 廃止=PRED/ する=SUP/	// 高知/ 県#が、で// 都道府県/ で/ 初め~/ #て// 一般/ 事務/ 職/ [員]/ の/ 採用/ **=/ 資格#にある,の// 国籍/ 条項#を// 廃止/ する/
950101009-0	9	F7	廃止	// 高知/ 県#が=MARKER// 都道府県=ARG/ で/ 初め~/ #て=MARKER// 一般/ 事務/ 職/ [員]/ の/ 採用/ **=/ 資格/ から/ 国籍/ 条項/ を/ 撤廃/ する#(という)=MARKER// 予定{がある,でいる}=PRED/	// 高知/ 県#が// 都道府県/ で/ 初め~/ #て// 一般/ 事務/ 職/ [員]/ の/ 採用/ **=/ 資格/ から/ 国籍/ 条項/ を/ 撤廃/ する#(という)// 予定{がある,でいる}/
950101009-0	10	F8	予定[-v_evoked]	// 高知/ 県#が、で=MARKER// 橋本/ 大二郎/ 知事#が=MARKER// 三十一/ 日#(に)=MARKER// 都道府県=ARG/ で/ 初め~/ #て=MARKER// 一般/ 事務/ 職/ [員]/ の/ 採用/ **=/ 資格/ から/ 国籍/ 条項/ を/ 撤廃/ する/ **=/ 方針#を=MARKER// 明らか~/ に/ し~/ た#と、ことを=MARKER// 報告している =PRED[+added]//	// 高知/ 県#が、で// 橋本/ 大二郎/ 知事#が// 三十一/ 日#(に)// 都道府県/ で/ 初め~/ #て// 一般/ 事務/ 職/ [員]/ の/ 採用/ **=/ 資格/ から/ 国籍/ 条項/ を/ 撤廃/ する/ **=/ 方針#を// 明らか~/ に/ し~/ た/
950101009-0	11	F9	明示化	// 高知/ 県#が、で=MARKER// 橋本/ 大二郎/ 知事#が=MARKER// 三十一/ 日#(に)=MARKER// 都道府県=ARG/ で/ 初め~/ #て=MARKER// 一般/ 事務/ 職/ [員]/ の/ 採用/ **=/ 資格/ から/ 国籍/ 条項/ を/ 撤廃/ する/ **=/ 方針#を=MARKER// 公表=PRED/ し~/ た#と、ことを=MARKER// 報告している =PRED[+added]//	// 高知/ 県#が、で// 橋本/ 大二郎/ 知事#が// 三十一/ 日#(に)// 都道府県/ で/ 初め~/ #て// 一般/ 事務/ 職/ [員]/ の/ 採用/ **=/ 資格/ から/ 国籍/ 条項/ を/ 撤廃/ する/ **=/ 方針#を// 公表/ し~/ た/
950101009-0	12	F10	公表		

MOD/ 条項#を=MARKER// 撤廃=PREDF	// 高知/ 県#が, 都道府県/ で/ 初め~#て// 一般/ 事務/ 職/ [員]/ の/ 採用/ **=/ 資格#から// 国籍/ 条項#を// 撤廃/ する/
MARKER// 都道府県=ARG/ で/ 初め~#て // 事務/ 職/ [員]/ の/ 採用/ **=/ 資格#	// 高知/ 県#が, 都道府県/ で/ 初め~#て// 一般/ 事務/ 職/ [員]/ の/ 採用/ **=/ 資格#から// 国籍/ 条項#を// 除去 [+metaphoric]/ する/
MOD/ 条項#を=MARKER// 除去 する=SUP/	
MARKER// 都道府県=ARG/ で/ 初め~#て // 事務/ 職/ [員]/ の/ 採用/ **=/ 資格	// 高知/ 県#が, 都道府県/ で/ 初め~#て// 一般/ 事務/ 職/ [員]/ の/ 採用/ **=/ 資格#にある, の// 国籍/ 条項#を// 廃止/ する/
MARKER// 国籍=MOD/ 条項#を=MARKER//	
MARKER// 都道府県=ARG/ で/ 初め~#て // 事務/ 職/ [員]/ の/ 採用/ **=/ 資格/ から/	// 高知/ 県#が// 都道府県/ で/ 初め~#て// 一般/ 事務/ 職/ [員]/ の/ 採用/ **=/ 資格/ から/ 国籍/ 条項/ を/ 撤廃/ する #(という)// 予定{がある, ている}/
MARKER// 橋本/ 大二郎/ 知事#が #(に)=MARKER// 都道府県=ARG/ で/ 一般/ 事務/ 職/ [員]/ の/ 採用/ **=/ 資 格/ から/ 撤廃/ する/ **=/ 方針#を=MARKER// 公表/ し~ / た=EXT/	// 高知/ 県#が, 橋本/ 大二郎/ 知事#が// 三十一/ 日 #(に)// 都道府県/ で/ 初め~#て// 一般/ 事務/ 職/ [員]/ の/ 採 用/ **=/ 資格/ から/ 国籍/ 条項/ を/ 撤廃/ する/ **=/ 方針#を // 明らか~/ に/ し~/ た/
MARKER// 橋本/ 大二郎/ 知事#が #(に)=MARKER// 都道府県=ARG/ で/ 一般/ 事務/ 職/ [員]/ の/ 採用/ **=/ 資 格/ から/ 撤廃/ する/ **=/ 方針#を=MARKER// 公表/ し~ / た=EXT/	// 高知/ 県#が, 橋本/ 大二郎/ 知事#が// 三十一/ 日 #(に)// 都道府県/ で/ 初め~#て// 一般/ 事務/ 職/ [員]/ の/ 採 用/ **=/ 資格/ から/ 国籍/ 条項/ を/ 撤廃/ する/ **=/ 方針#を // 公表/ し~/ た/

粒度コントロール

- ❖ フレーム要素が認定される粒度は自明でも一定でもないの
で、管理が必要
- ❖ 統語解析との対応づけの良し悪しは、この分析の粒度に強
く依存する

nlp14 tutorial

粒度の異なり

- ❖ 左に行く粗い
- ❖ 右に行くと細かい
- ❖ 細くなるほど、作業者への要求大
- ❖ L₂ ぐらいが最適?
- ❖ MSFA 粒度の統一の予定はないが、シートに粒度指標を明示化する予定

	A	B	C	D	E	F	G
1	F-ID	F1 at Level 0	F2 at Level 1	F3 at Level 2	F4 at Level 2+		
2	Frame-to-Frame						
3	Frame Name	廃止	廃止	廃止	廃止		
4	%% Predicate for GOV						
5	高知	廃止者	廃止者	廃止者	廃止者		
6	県						
7	が	#が	#が	#が	#が		
8	全国	条件,様態[廃止の]	条件,様態[廃止の].MOD: 基準	条件,様態[廃止の].ARG: 基準	条件,様態[廃止の].ARG: 基準		
9	で		#で	#で	#で		
10	最初		条件,様態[廃止の]	条件,様態[廃止の]	条件,様態[廃止の]		
11	に	#に	#に	#に	#に		
12	一般	対象[廃止の]	対象[廃止の].MOD: 全体	対象[廃止の].MOD.ARG: ?? .ARG	対象[廃止の].MOD: 全体.ARG: ?? .ARG.MOD: 種別		
13	事務						
14	**				#(のための)		
15	職				対象[廃止の].MOD: 全体.ARG: ?? .ARG		
16	[員]						
17	の			#の	#の		
18	採用			対象[廃止の].MOD.ARG: ??	対象[廃止の].MOD: 全体.ARG: ??		
19	**			#のための	#のための		
20	資格			対象[廃止の].MOD	対象[廃止の].MOD: 全体		
21	の		#にある,の	#にある,の	#にある,の		
22	国籍		対象[廃止の]	対象[廃止の]	対象[廃止の]		
23	条項						
24	を	#を	#を	#を	#を		
25	停止	廃止.GOV	廃止.EVO=PRED	廃止.EVO=PRED	廃止.EVO=PRED		
26	す		廃止.EVO=PRED.S UP	廃止.EVO=PRED.S UP	廃止.EVO=PRED.S UP		
27	る			廃止.EVO=PRED.S UP.EXT	廃止.EVO=PRED.S UP.EXT		
28	。						

Kow Kuroda:
 <廃止者>が, <目的>のために,
 <決まり>を廃止する

Kow Kuroda:
 関係名詞「最初」の項であることを記述

Kow Kuroda:
 「国籍条項」の修飾語句であることを記述

Kow Kuroda:
 関係名詞「採用」の項であることを記述

Kow Kuroda:
 名詞「職員」の修飾語句であることを記述

Kow Kuroda:
 関係名詞「採用」の項であることを記述

Kow Kuroda:
 関係名詞「資格」の項であることを記述

Granularity Levels

- ❖ L₀ ではフレーム要素 (*f.GOV* を含む) と MARKER の記述
 - ❖ Berkeley FrameNet のアノテーションは L₀ と同等で意外とスカスカ
- ❖ L₁ では L₀ にフレームの喚起要素 (= *f.EVO*) と *e.SUP* と埋めこみなしの *e.MOD* の記述を追加し, “=PRED” を許容
- ❖ L₂ では L₁ に埋めこみありの *e.MOD*, *e.EXT* の記述を追加
- ❖ L₃ では L₂ に *e.ARG* の記述を追加し, 強制置換 “=>” を許容



Findings & Open Issues

知見 I

- ❖ 名詞句がフレームごとに異なる意味役割を担うのは常態 (これはメトニミーという効果を生む)
 - ❖ 国広 (1997) の多面的多義性 (cf. 排他的多義)
 - ❖ Pustejovsky (1995) の (特質構造に由来する) 論理的多義
- ❖ 名詞の意味記述として文脈不変な型 (e.g, human, place) のみを指定するのは不十分

知見 2

- ❖ “一つの述語に一つの意味” というのは多くの場合, 不成立
 - ❖ MSFA では “ F targets G ” で文字通りの意味と理解されるべき (超語彙的) 意味の関係を記述
 - ❖ F のみの記述が不十分なのは当然だが, G だけでもダメ
 - ❖ 動詞も多面的多義をもつ
 - ❖ 意味タグづけ作業者に課せられる “語義の択一選択” が実は実行可能な課題として成立しない可能性を示唆

知見 3

- ❖ 動詞の語義の overload を回避する必要性
 - ❖ 文の意味を動詞以外の共起要素に“分担”させる
- ❖ 多層化をしないと文意が動詞の語義レベルで決まっているのか、共合成の結果として決まっているのかわからない

知見 4

- ❖ (国研の“短単位”を使った場合でも) 文中で喚起される全フレーム記述に共通な形態素境界は保証されない
- ❖ 「日本国内外」のような重複形態では{日本国内, 日本国外, 日本国, 日本内, 日本外, 国内外, 内外, 日本, 国, 内, 外}の全部が認定される必要がある
 - ❖ [日本国 = 日本+国内外] なら“日本国”が落ち, [日本国 = 日本国+内外] なら“国内外”が落ちる
- ❖ 最小単位をフレームごとにチャンキングするのが最良?

知見 5

- ❖ 修飾/非修飾の関係はフレームごとに違ってもよい
 - ❖ 就任後初めて地元の大分県へ里帰りしていた村山富市首相は三十一日夕[方]、三泊四日の日程を終え ... [S-ID: 950101014-001]
- ❖ “里帰り”で喚起される<帰郷>フレーム内では，“首相”は“村山富市”という個人の属性を記述する修飾要素，“日程を終え”で喚起される<用件の完了>フレーム内では，“村山富市”は“(日本の)首相”という役割の値を記述する修飾要素

知見 6

- ❖ 名詞句だけからなる“述語を含まない文”は統語解析では無用なので京大コーパスから除外されているが、このために後続文の意味が決められない場合は少なくない。

知見 7

- ❖ 長い文の解析は大変だが短かいほど良いわけではない
 - ❖ 短すぎる文は不定性が高くなり，解釈の決定が難しい



Things to Do

問題点 & 課題

- ❖ MSFA は文の“意味”自体は表わしていない
 - ❖ MSFA は“意味”へのリンク集
- ❖ 問題
 - ❖ リンク先にある“状況”の定義がまだない
 - ❖ それらに定義を与えるなら、オントロジーの専門家に相談して、ちゃんとしたものにしたい
 - ❖ だが、“フレーム”の実体は何? 単なるオントロジー?

最後に

❖ 理論も重要!!

- ❖ デザインが具体的アプリケーションから要請であっても別にいいのですが、**そればかりではダメ**だと思います
- ❖ 工学系ではアプリケーションなしのデザインを云々することは (机上の空論として) 煙たがられますが
- ❖ (完全に教師なしの手法で構築できるものでない限り) “理論の支援なし” の言語資源を開発しても使えるようになるとは限らないと (反論を覚悟で) 主張しておきます

nlp14 tutorial

謝辭

- ❖ 李在鎬 (NICT)
- ❖ 渋谷良方 (京都外国語大学/NICT)
- ❖ 野澤元 (NICT)

Thank you