

並列パターン一致法の簡略版 (PNMAS)を用いた(共)項構造の指定法

黒田 航

(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

お話の概要

- 仕事の目的
- 並列パターン一致 / 統合法 (Parallel Pattern Matching / Intergration Analysis (Simplified)) 以下 **PMA(S)**) の概説
 - PMA(S) の定義
 - PMA(S) の実例
- まとめ

目的

- 動詞, 形容(動)詞のような**文法的項構造**をもつ要素の他に, 事態性抽象名詞 (e.g., 反撃, 反感) の**概念的項構造**や非事態性の一般名詞の概念的**共項構造** co-argument structure を特定
- 簡単に言うと NomBank の日本語版を作って照応解析などの特定の課題に役立てる
 - ゼロ代名詞の存在を仮定しないで照応を扱える

話の発端は

- 乾さんから
 - NL研究会の論文を読んで質問があるんだけど、名詞の項構造の一般理論のようなものはないの？
- という相談を受けたのが話の始まり。NLP12で小町くんの発表を聞いて質問の意図がわかった
 - NomBank の日本語版を作るのが目標なら「名詞の項構造の一般理論」ほど話は大袈裟にならない

要件

- ただ、少なくとも次のことは必要
 - 係り受け解析、句構造 / 木構造解析は (表現力が不足しているので) 前提にしない
 - これらに対して上位互換性をもつ記述法を用いる
- PMA (Kuroda 2000) は、この課題に則している
 - というより、PMA はこういう目的のために開発された統語解析の手法

PMASとは何か

- (Parallel, Multiple) Pattern Matching / Integration Analysis とは Kuroda (2000) の定義した統語解析法
 - 形態素ごとに項構造 argument structure, あるいは共項構造 co-argument structure を指定する手法
- PMAS とは今回の作業のための PMA の簡略版
 - PMA から格助詞や接続助詞の項解析をハズす

PMA(S)の実例

- [1] 太郎は花で玄関をきれいに飾った.
 - for PMA
- [2] K首相は迂闊な靖国神社訪問で中国政府 {の, に, から} 反感を買った.
 - for PMAS

[1] の PMA (初期状態)

s		太郎**	は**	花**	で**	玄関**	を**	きれい**	に**	飾っ**	た**	実現される統語パターン
p1	太郎	太郎*										realizes UNKNOWN
p2	は		は*									realizes UNKNOWN
p3	花			花*								realizes UNKNOWN
p4	で				で*							realizes UNKNOWN
p5	玄関					玄関*						realizes UNKNOWN
p6	を						を*					realizes UNKNOWN
p7	きれい							きれい*				realizes UNKNOWN
p8	に								に*			realizes UNKNOWN
p9	飾っ									飾っ*		realizes UNKNOWN
p10	た										た*	realizes UNKNOWN

[1] の PMA (完成状態)

s		太郎**	は**	花**	で**	玄関**	を**	きれい**	に**	飾っ**	た**	実現される統語パターン
p1	太郎	太郎*	P: ??							V	I(V)	realizes SV
p2	は	T	は*							{N, V}		realizes Topic Predicate
p3	花	S	P(S): (が)	花*	P(O): (を)					V[2]: (使っ)	I(V)	realizes SOV
p4	で	S	P(S): (が)	X	で*	O	P(O): (を)			V[2]: ??	I(V)	realizes SXOV (X=Means)
p5	玄関	S	P(S): (が)			玄関*	P(O): (を)			V[2]: ??	I(V)	realizes SOV
p6	を	S	P(S): (が)			O	を*			V[2]: ??	I(V)	realizes SOV
p7	きれい					S	P(S): (が)	きれい*	P(C): (に)	V[2]: (なっ)	I(V)	realizes SCV
p8	に					S	P(S): (が)	C	に*	V[2]: (なっ)	I(V)	realizes SCV
p9	飾っ	S	P(S): (が)	X	P(X): (で)	O	P(O): (を)			飾っ*	I(V)	realizes SXOV (X=Material)
p10	た	S	P: ??							V	た*	realizes SV

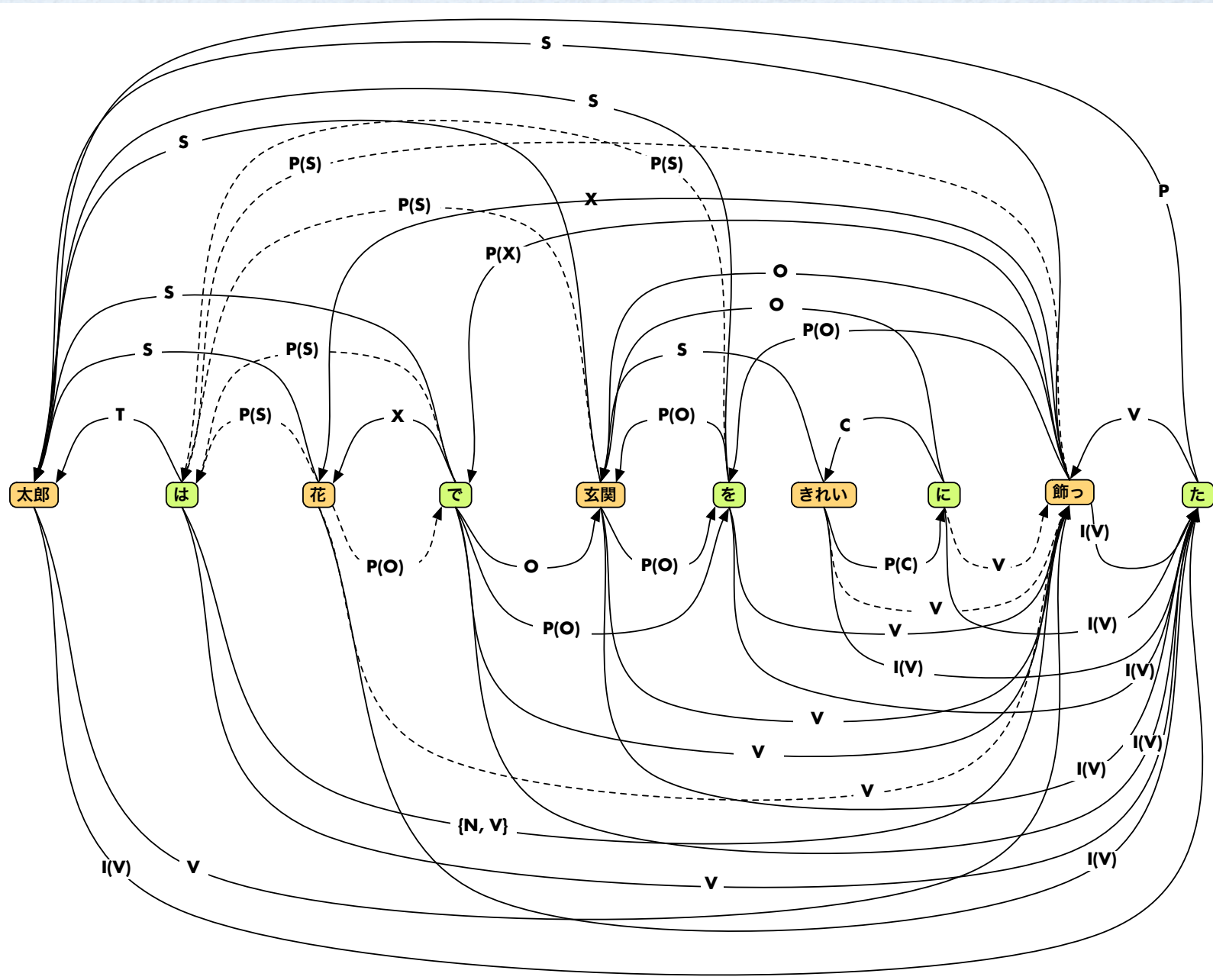
解説

- PMA は $s = w_1 w_2 \dots w_n$ の統語構造を $n \times n$ の共起制約行列として表現
 - 部分パターン p_i は s 中の w_i の(共)項構造を指定
 - w は語でも形態素でもよく, $p_{i,j}$ のような規約を用いれば, 句の内部構造 (埋めこみの効果) の記述も可能
 - 部分パターンの列ごとの重ね合わせが全体を規定
 - 厳密には「深層構造」からの「派生」は考えないが, おの

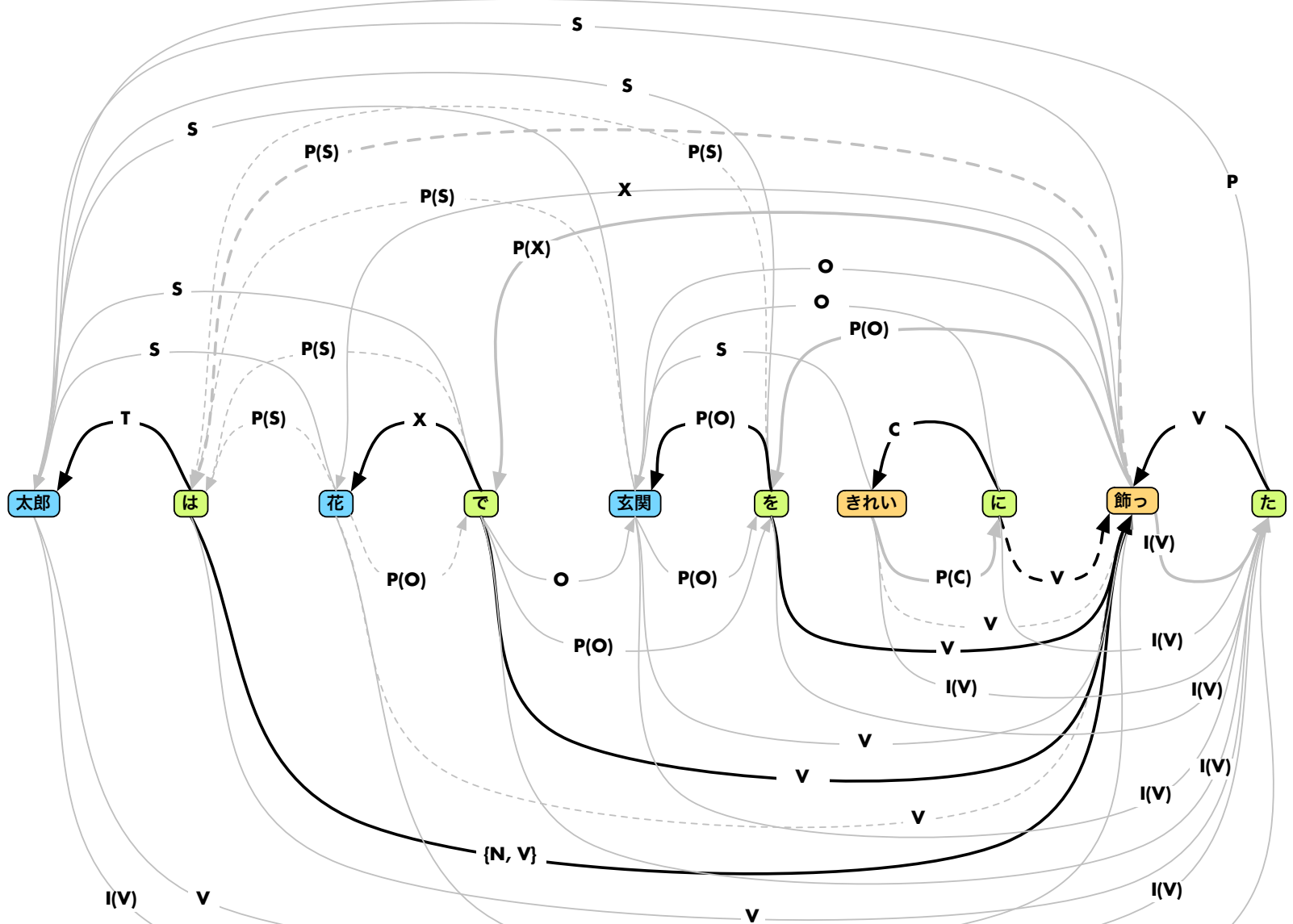
PMAの表現力

- 依存構造解析に対して上位互換の関係にある
- 以下, この点を実例を用いて示す

PMIA が表現する [1] の依存関係ネット



係り受け解析の表わす部分ネット



[2] の PMAS (結果状態)

s		K首相**	が**	迂闊**	な**	靖国神社**	訪問**	で**	中国**	政府**	{に**, から **, の**}	反感**	を**	買った**
p1	K首相	K首相*												
p2	が		が*											
p2	迂闊	Arg0		迂闊* as A[1]	(だ)									
p3	な			Arg1	な*	Arg0[1,2]	!Arg0[2,2]							
p4	靖国神社					靖国神社*								
p5	訪問	Arg0				Arg1	訪問* as V[2]							
p6	で							で*						
p6	中国								中国*					
p7	政府									政府*				
p8	{に, から, の}	Arg0							Arg2[1,2]	Arg2[2,2]	に*	Arg1	P: (を)	V: もたれた
p9	反感	Arg1[1,3]	P: (の)			Arg1[2,3]	Arg1[2,3]		Arg0[1,2]	Arg0[2,2]		反感* as V[2]	P: (を)	V: ({もつ, 抱く})
p10	を												を*	
p11	買った	Arg0		Reason[1,4]	Reason[2,4]	Reason[3,4]	Reason[4,4]		Arg2[1,2]	Arg2[2,2]	P: (から)	Arg1		買った* as V[3]

名詞の(共)項構造解析

- 助詞を除いた要素の(共)項構造解析
- $X[n]$ は n -ary の述語 (e.g., $V[n]$, $A[n]$) を表わす
 - n の数は語義ベースに、経験的に決める
 - LCS データベースとの整合性も気にしたい
 - $Arg0, Arg1, \dots$ は PropBank に従った
 - 作業は { $Arg1$ =主語, $Arg2$ =直接目的語} の方が楽は楽
- $R[i, i]$ は 全体で i 個ある部分の i 番目の要素

現状と展望

- 京大コーパス 9501ED の 0000 から 0041 までの合計 42 文に黒田がほぼ網羅的な (共) 項構造解析を行なった
- これを元に作業内容を明確にし, 京大の院生に (ボランティアの) 作業を依頼する予定
 - 黒田の解析は全体に overspec 気味なので (苦), 飯田くんと協力して作業課題に落としこむ予定

まとめ

- PMA(S) を用いれば,
 - 動詞や形容(動)詞のように文法的項をもつ要素の他に, 名詞の(概念的)項を指定することができる
 - その指定は動詞や形容動詞の項構造の指定に干渉しない(し, 逆にそれを補う)
- (一部の) 言語学者には「使える」記述モデルを開発する能力があることを示せる

謝辞

飯田 龍 (奈良先端技術大学大学院)

加藤 鉦三 (信州大学)

今回の企画/成果は、以上の方々との有益な議論の
賜物です