

# 意味役割名と意味型名の区別に基づく 新しい概念分類の可能性

意味役割の一般理論はシソーラスを救う？

黒田 航      井佐原 均

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

電子情報通信学会 第168回自然言語処理研究会

7/23/2005 国民健康村 (沖縄)

# 本発表の主張

- ～ 概念体系は従来のシソーラス (e.g., 日本語語彙大系, EDR, WordNet) が基本としている自然分類の体系には納まり切らない
  - ～ 概念体系は意味型の体系が定義する自然分類の体系には還元不能
- ～ このため、十分に精緻で妥当な概念分類を与えるためには、自然分類の体系化の他に機能分類の体系化を考えなければならない
- ～ 意味フレームの一般理論 (黒田・井佐原 2004; 黒田・中本・野澤 2005) が意味役割の一般理論の基礎となり、機能分類の基礎となる
  - ～ 意味役割の体系が機能分類体系を定義する

# 発表の概要

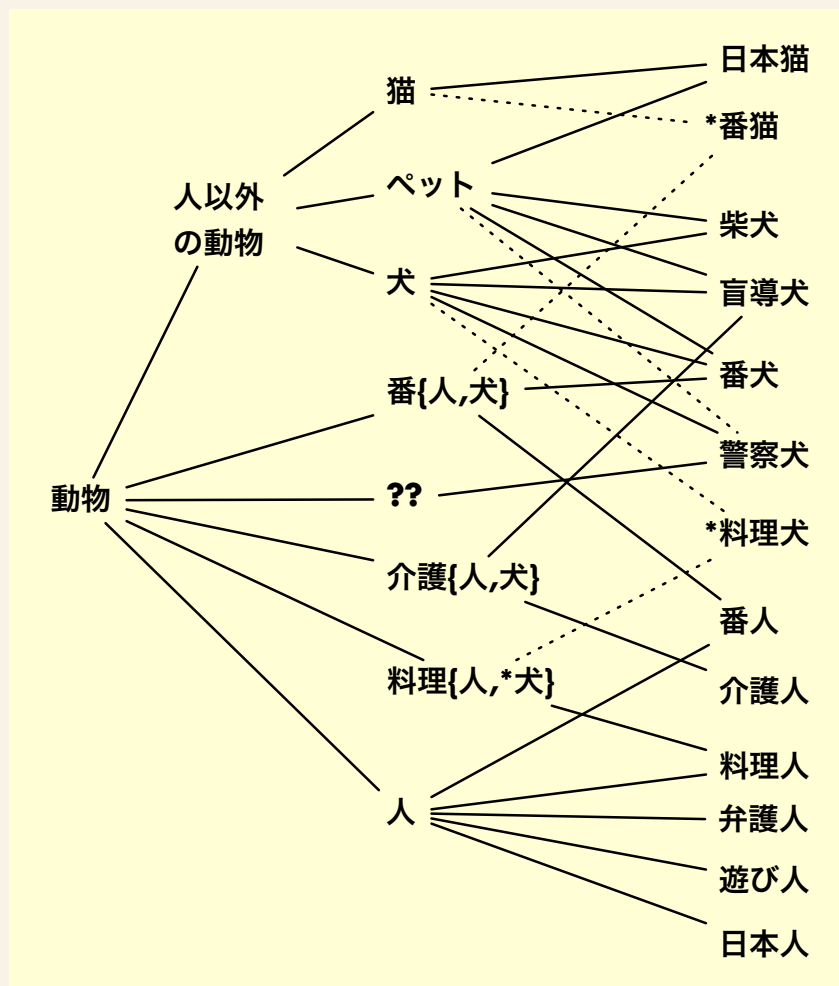
- ～ 概念分類の根本問題
  - ～ 柴犬と番犬の例を取り上げて、従来のシソーラスの限界を指摘
- ～ 意味型(名)と意味役割(名)の区別の提唱
  - ～ これから派生する対象名と機能名の区別の提案
- ～ 対象名と機能/意味役割名の区別の検証実験
- ～ まとめと今後の課題
- ～ 注意
  - ～ 意味役割の一般理論の詳細は時間の都合上、提示しない
  - ～ アフォーダンスの話はしない

# 概念分類の根本問題

～二種類の犬の話～

# 「番犬」と「柴犬」ちがい

- ～ 従来のシソーラスの分類基準 (e.g., 日本語語彙大系) によると<番犬>と<柴犬>は共に<犬>の一種
- ～ この場合  $X = \text{番} + \text{犬} = a + Y$  で  $Y$  が意味的主要部
- ～ これは誤りではないが十分ではない。例えば次のクラスを記述できない
- ～ {番人, 番犬, \*番猫} をクラスとする分類基準は  $a = \langle \text{番} \sim \rangle$
- ～ {野良犬, 野良猫, \*野良馬, \*野良牛} をクラスとする基準は  $a = \langle \text{野良} \sim \rangle$



# 可能な概念/不可能な概念の区別が必要

- 〜 <番犬>と<柴犬>は明らかに異なる概念化の産物 (語彙大系の解説; 池原 2004; Lakoff 1987; Langacker 1987, 1991)
- 〜 問題は
  - 〜 可能な概念化 (e.g., 番{人, 犬}) と不可能な概念化 (e.g., \*番{猫, 鳥}, 料理{人, \*犬}) の区別を与える制約が必要だが, その体系化は自明ではない
  - 〜 それがないと意味記述に際限がなくなり「フレーム問題」が発生
- 〜 私たちの提案
  - 〜 意味役割の一般理論 (黒田・井佐原 2004, 2005; 黒田・中本・野澤 2005) がこの問題に有効な解決法を与える

# 概念の体系化に関する一般理論の必要性

## ～ 表面的問題

～ 1. 番{人, 犬}, 介護{人, 犬}, 野良{犬, 猫} が有意味な意味クラスであること

～ 2. \*番{猫, 鳥, 蛇}, \*料理{人, 犬}, \*野良人が無意味なクラスであること

～ 3. ???野良{鶏, 豚, 牛} が不可能ではないが奇妙なクラスであること

～ をどうやって記述するか

～ 従来のシソーラス (e.g., 語彙大系, EDR, WordNet) はそのような制約を表現していない

## ～ 根本的問題

～ では, <番~>や<野良~>のような意味クラスは何が定義するか?

意味型(名)と意味役割(名)の  
区別の提唱



# 意味型分類と意味役割分類

- 〜 「柴犬」「岩石」「鉱石」などは意味型名
  - 〜 意味型は自然界に存在する対象を特定するための概念化の産物
- 〜 「番犬」「野良犬」「宝石」などは意味役割名
  - 〜 意味役割は対象が果たす特定の機能や価値の概念化の産物
- 〜 重要な点
  - 〜 概念体系は意味型の体系であるという側面と意味役割の体系であるという側面を同時にもち、一方を他方に還元はできない
  - 〜 厄介な非排他性: 多くの名称は意味型名であると同時に意味役割名でもある (ただし単純な連続性があるわけではない)

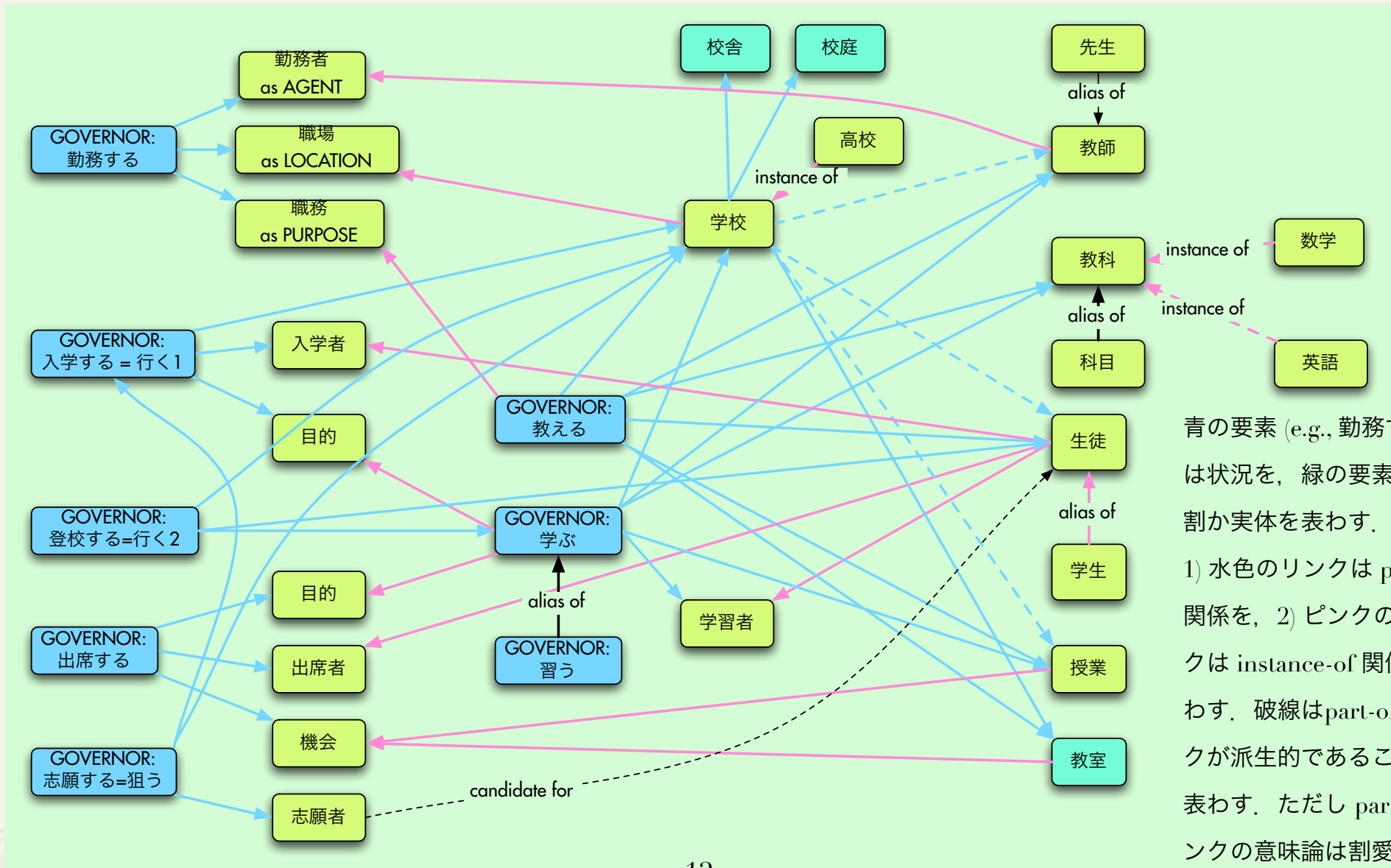
# 意味役割の概念化は意味フレームに基づく

- ～  $\langle (y \text{ の}) \text{番 } x \rangle$  という概念は次のフレームによって規定される:
  - ～  $\langle x \text{ が, } y \text{ (か } y \text{ の所有者 } z) \text{ のために, } y \text{ の}\{\text{番; 見張り}\} \text{ をする} \rangle$  とき,  $x$  は  $y$  の番である;
  - ～  $x$  がそうするのは,  $y$  が  $x$  (か  $x$  の  $\langle \text{主人} \rangle$  である  $z$ ) にとって何らかの意味で  $\langle \text{価値あるもの} \rangle$  で,  $\langle \text{泥棒} \rangle$  による  $\langle \text{盗難} \rangle$  や  $\langle \text{敵対者} \rangle$  によって意図的に, あるいは  $\langle \text{部外者} \rangle$  によって非意図的に  $y$  に加えられる  $\langle \text{損傷} \rangle$  を  $\langle \text{防止} \rangle$ ,  $\langle \text{予防} \rangle$  するため
- ～  $\langle \text{野良 } y \rangle$  という概念は  $\langle x \text{ が } y \text{ をペットとして飼う} \rangle$  というフレームに基づいて(正確にはその不成立によって)定義される

# 注意

- ～ 本稿で言う意味フレームはいわゆる深層格と同一ではない
- ～ EDR の意味フレーム, IPAL の述語フレームとも同一ではないようだ
- ～ 本稿で言う意味フレームとは状況レベルの概念化の単位のこと
- ～ Berkeley FrameNet (Baker, et al. 1998; Fillmore, et al. 2003) で意味フレームと呼ばれるもので, 様々な粒度の具体的な状況の心内モデル
- ～ 心理的妥当性は部分的に心理実験 (中本・黒田・野澤 2005) で検証されている (材料: “ $x$  が  $y$  を襲う”, “ $y$  が  $x$  に襲われる”)
- ～ 表層格フレーム (河原・黒橋 2002) の心理的実体
- ～ 定義の詳細は黒田・中本・野澤 (2005) を参照

# 状況を媒介にした概念の組織化 (例)



青の要素 (e.g., 勤務する) は状況を、緑の要素は役割か実体を表わす。

1) 水色のリンクは part-of 関係を、2) ピンクのリンクは instance-of 関係を表わす。破線は part-of リンクが派生的であることを表わす。ただし part-of リンクの意味論は割愛

# 基本的な考え方

- ～ 意味フレームは幾つかの意味役割の組織化
  - ～ 幾つかの概念 (e.g., <裁判官>, <被告>, <検察>, <容疑>, <弁護士>, <弁護士>, <判決>) が一つの概念化 (e.g., <裁判>フレームとそれを構成する部分フレーム{<検挙>, <弁護>, ...}) に帰属
  - ～ 意味フレーム自体にも階層的埋めこみがある
  - ～ 従来のシソーラスの弱みは Part-of 関係が明示されていないこと
- ～ 課題
  - ～ 基本となる概念化の明示化, データベース化
  - ～ 意味役割名の同定は意味フレームのデータベースに不可欠

# 疑問の解消の必要性

- ～ 理論は(どうでも)いい. <柴犬>と<番犬>の概念化は本当に異なっているのか?
- ～ 直観的には明らかにそう思えるが, 直観に訴えるだけでは不十分
- ～ (たかが) 言語学者の直観はそんなに信用できるのか?
  - ～ どんなすぐれた研究者の直観でも常に実験的に検証する必要がある
- ～ この疑問を解消する目的で, 仮定されている意味型と意味役割の区別を実験的に検証した

# 対象名と意味役割/機能名 の区別の検証実験

# 実験手法

- 〜  $X=a+Y$  (e.g.,  $X=柴+犬$ ) を満足する17個の名称対  $X, X^*$  (次ページ) について次の環境 1-5 の生起可能性を三段階 (0, 0.5, 1.0) で評定 (n=7)
  - 〜 (1) この  $Y$  は  $X$  に向いて {いる; いない}
  - 〜 (2) 彼は, ある  $Y$  を  $X$  にした
  - 〜 (3) その  $Y$  は, この場では  $X$  だ
  - 〜 (4) この  $Y$  は  $X$  になる {恐れがある; 心配} がある
  - 〜 (5) この  $Y$  は  $X$  になる {見こみ; 見通し} がある
- 〜  $X$  の機能性の指標  $F(X) = [(1)+(2)+(3) + \text{Max}((4), (5))] / (\text{有効試験数}=4)$
- 〜 値域:  $0 \leq F(X) \leq 1$



語彙大系	Y	X	X*	特徴/素性
816 布	布	リンネル	フィルター	+concrete, animate, human
815 糸	糸	絹糸	横糸	+concrete, animate, human
775 石材	石	隕石	敷石	+concrete, animate, human
775 石材	石	石灰石	墓石	+concrete, animate, human
773 板	板	ベニヤ板	床板	+concrete, animate, human
770 紙	紙	和紙	型紙	+concrete, animate, human
549 昆虫	虫	カメムシ	害虫	+concrete, +animate, human
538 鳥	鳩	山鳩	伝書鳩	+concrete, +animate, human
537 獣	犬	柴犬	番犬	+concrete, +animate, human
537 獣	馬	白馬	名馬	+concrete, +animate, human
537 獣	猫	シャム猫	野良猫	+concrete, +animate, human
537 獣	犬	柴犬	猛犬	+concrete, +animate, human
049 女	女性	美人	保母	+concrete, +animate, +human
481 洞穴	洞穴 (ほら穴)	洞窟	抜け穴	?concrete, animate, human
461 土地	土地	砂地	聖地	?concrete, animate, human
2372 風	風	潮風	追い風	?concrete, animate, human
2346 光	光	月光	照明	?concrete, animate, human

## 評定の例

～ 例 1:  $F(\text{柴犬}) = [0 + 0 + 0.5 + 0] / 4 = 0.125$

～ (1) \*この犬は柴犬に向いて {いる; いない} → 評定値 = 0

～ (2) \*彼は, ある犬を柴犬にした → 評定値 = 0

～ (3) ???その犬は, この場では柴犬だ → 評定値 = 0.5

～ (4) \*この犬は柴犬になる {恐れがある; 心配} がある → 評定値 = 0

～ (5) \*この犬は柴犬になる {見こみ; 見通し} がある → 評定値 = 0

～ 例 2:  $F(\text{番犬}) = [1 + 1 + 1 + 1] / 4 = 1.0$

～ (1) この犬は番犬に向いて {いる; いない} → 評定値 = 1

～ (2) 彼は, ある犬を番犬にした → 評定値 = 1

～ (3) その犬は, この場では番犬だ → 評定値 = 1

～ (4) ?\*この犬は番犬になる {恐れがある; 心配} がある → 評定値 = 0.5

～ (5) この犬は番犬になる {見こみ; 見通し} がある → 評定値 = 1

語彙大系の X, X* の 上位概念	Y	X	F(X)	X*	F(X*)	F(X*)-F(X)	
537 獣	犬	柴犬	0.09	番犬	0.96	0.87	G1: F(X)<0.2; 0.7<F(X*)
538 鳥	鳩	山鳩	0.05	伝書鳩	0.86	0.81	
2346 光	光	月光	0.16	照明	0.82	0.66	
775 石材	石	隕石	0.09	敷石	0.80	0.71	
773 板	板	ベニヤ板	0.16	床板	0.80	0.64	
775 石材	石	石灰石	0.05	墓石	0.77	0.72	
537 獣	馬	白馬	0.18	名馬	0.75	0.57	
816 布	布	リンネル	0.13	フィルター	0.71	0.58	
815 糸	糸	絹糸	0.13	横糸	0.71	0.58	
クラス1内平均			0.12		0.80	0.68	
537 獣	猫	シャム猫	0.07	野良猫	0.57	0.50	G2: F(X)<0.2; 0.5<=F(X*)<=0.7)
549 昆虫	虫	カメムシ	0.05	害虫	0.54	0.49	
481 洞穴	洞穴(ほら穴)	洞窟	0.14	抜け穴	0.54	0.40	
537 獣	犬	柴犬	0.09	猛犬	0.50	0.41	
クラス2内平均			0.09		0.54	0.45	
クラス1,2内平均			0.11		0.72	0.61	
2372 風	風	潮風	0.13	追い風	0.46	0.33	G3: F(X)<0.2; F(X*)<0.5
クラス3内平均			0.13		0.46	0.59	
049 女	女性	美人	0.64	保母	0.82	0.18	G4: 0.6<F(X)<F(X*)
770 紙	紙	和紙	0.21	型紙	0.79	0.58	G1*: 0.2<F(X)<0.3; 0.7<F(X*)
461 土地	土地	砂地	0.39	聖地	0.66	0.27	G5: F(X)<0.2; 0.5<F(X*)<0.7
クラス4,1*,5内平均			0.52		0.76	0.34	
全体の平均			0.15		0.71	0.55	

# 結果

## ～ 重要な点

～  $F(X) < F(X^*)$  は常に成立

～  $d = F(X^*)$   $F(X)$  の最大値は 0.87 ( $F(\text{番犬})$   $F(\text{柴犬})$  の場合), 最小値は 0.18 ( $F(\text{保母})$   $F(\text{美人})$  の場合)

～ あらかじめ  $X$  は対象名として  $X^*$  は意味役割名として選ばれていたの  
で, この結果は発見的ではなく確証的

## ～ 参考値

～ クラス1内の  $F(X)$  の平均 0.12;  $F(X^*)$  の平均 0.80

～ クラス2内の  $F(X)$  の平均 0.09;  $F(X^*)$  の平均 0.54

～ クラス1, 2 内の  $F(X)$  の平均 0.11;  $F(X^*)$  の平均 0.72

# まとめと今後の課題

# まとめと今後の課題

- ～ 従来のシソーラスの限界を指摘し，それを解決するために自然分類の記述に加えて機能/役割分類の記述を補う必要性を指摘
- ～ その実装のために必要な意味役割の体系化に意味フレームの理論が使えることを主張し，その妥当性を部分的に実証
- ～ 今後の課題
  - ～ 記述の範囲を広げて，単なるオモチャ問題を扱っているわけではないことを証明する必要
  - ～ 機能性から独立に対象性を測るテストの開発が必要

# 今後の見通し

- ～ NLP 関係者の関心は「既存のシソーラスとの違いと主張はわかった。では、次のどの方向で進むのか?」だろう:
  - ～ I. 既存のシソーラスを修正する方向に行くのか
  - ～ II. 新しくシソーラス (ぽい) ものを構築していくのか
  - ～ III. あくまでも理論研究であり、当面は分析の蓄積でいくのか
- ～ 答え: 資金的, 人材的制約から当面は III と II の組み合わせ, 条件が整えば I を目指したいが, そのためには共同研究者が必要

# 池原先生からのメール (07/21づけ)

- ご承知のように「日本語語彙大系」は、日英機械翻訳における単文の訳文品質（特に動詞の訳語選択）向上を目的に開発したものです。開発の当初、辞書的な語義の定義があってもそれだけでは表現の意味が解析できませんので、用法に関する知識が必須であると考え、意味的用法の体系化を試みました。
- 論文でご指摘されているように、日本語単語の意味的用法として十分な体系化ができているとは言えませんし、またどうすればそのような体系が完成できるか明らかになったとも言えません。しかし、言語処理のための言語知識ベースを構築するには、意味的用法の体系化は避けて通れない課題だと思います。今後の意味処理技術の発展のためには、より精密な体系化が必要だ思っていたところ、理論家の黒田さんによってこの問題が正面から採り上げられ、体系化が試みられていることを知り、大変嬉しく思いました。大いに期待しております。 [...]
- 私は、工学的な立場から言語処理をやっており、現実のシステム開発に使える研究成果に期待しています。黒田さんの論文は、正に現実に直面している問題に正面から答えられようとしており、嬉しく思いましたので、少し長くなりましたが、感じたことなどを書かせて頂きました。今後、どうか宜しくお願い致します。



# 謝辞

- ～ 金丸 敏幸 (京都大学 教育学研究科)
- ～ 竹内 和広 (情報通信研究機構)
- ～ 中本 敬子 (京都大学 人間・環境学研究科)
- ～ 李 在鎬 (情報通信研究機構)

# 資料

# 意味役割の一般理論 (素描)

- ～ 意味役割の大別
  - ～ A. 状況的役割 [+abstract, +temporal]
  - ～ B. 社会的役割 [+abstract, -temporal]
  - ～ C. 構造的役割 [-abstract, ?temporal]
  - ～ D. NONE [-abstract, -temporal]
- ～ 以下の例では A, B の区別が重要
  - ～ 運転者 (A のみ), 運転主 (A, B)
  - ～ 講演者 (A のみ), 講師 (A, B), 教師 (B のみ) と

# 応用

- ～ 意味役割名の同定は意味フレームのデータベースに不可欠
- ～ 意味フレームのデータベースがあれば、**複層意味フレーム分析** (MSFA) (黒田・井佐原 2004, 2005; Kuroda and Isahara 2005) の標準化が可能となり、その有効性が向上する

# “空腹のライオンがインパラの群れを襲った”

Frame ID	F1	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 F2; constitutes F3											elaborates F13						
Frame Name	感覚	欲求	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	空腹(感)	欲求	感覚	欲求															脅威
*	GOV																		
*			GOV																
*				GOV	目的					目的									
*				手段	GOV														
の	MARKER																		
ライオン	空腹者	欲求主体	感覚者	欲主	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											
*											除外項						逃亡者	生存者	
インパラ						行動者	自衛者	標的	獲物	獲物の種類	候補								
の						MARKER													
*						目的	GOV												
群れ						GOV	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			
*											GOV								
を								MARKER											
*																	GOV	理由	
*					手段				GOV	手段					起こされた事態				
*									目的	GOV		目的?					逃れた災難		理由
*												副作用	GOV	結果状態		理由			
襲った								GOV	手段		目的		死因						
た								EXT											

自然言語文の形態素解析

空腹  
の  
ライオン  
が  
インパラ  
の  
群れ  
を  
襲っ  
た

意味フレームの実現階層

