

日本語の意味役割タグ体系を定義する試み

~FrameNet の視点から~

黒田 航

kuroda@nict.go.jp

独立行政法人 情報通信研究機構 (NICT)

04/24/2004

Kyoto Linguistics Colloquium

意味タグつきコーパス

- 既存のコーパスには限界が多い
 - その一つが意味情報の欠落
 - 品詞情報と統語情報の組み合わせからは意味情報は計算できない！
- それを補うコーパスには入念な設計が必要

よい意味タグ体系の条件

- 機械学習可能なくらい十分な体系性,
一貫性が備わっている
- 特定の知識領域や応用課題に限定され
ない十分な一般性, 網羅性をもつ
 - 工学者, 言語学者, 認知科学者の共同開発が必要
- 必要に応じて拡張可能

意味タグ体系に特有の問題

- 意味タグ空間は品詞タグ空間より広大
 - 最適なタグ列の決定過程で相互依存性大
- 意味タグづけの“単位”は今だに不明瞭
- 意味タグを定義するための“基準”が不明確
 - 意味要素の「自然な分類特徴」の発見は自明でない
- 最後の二つの問題は FrameNet の知見から克服できる(はず)

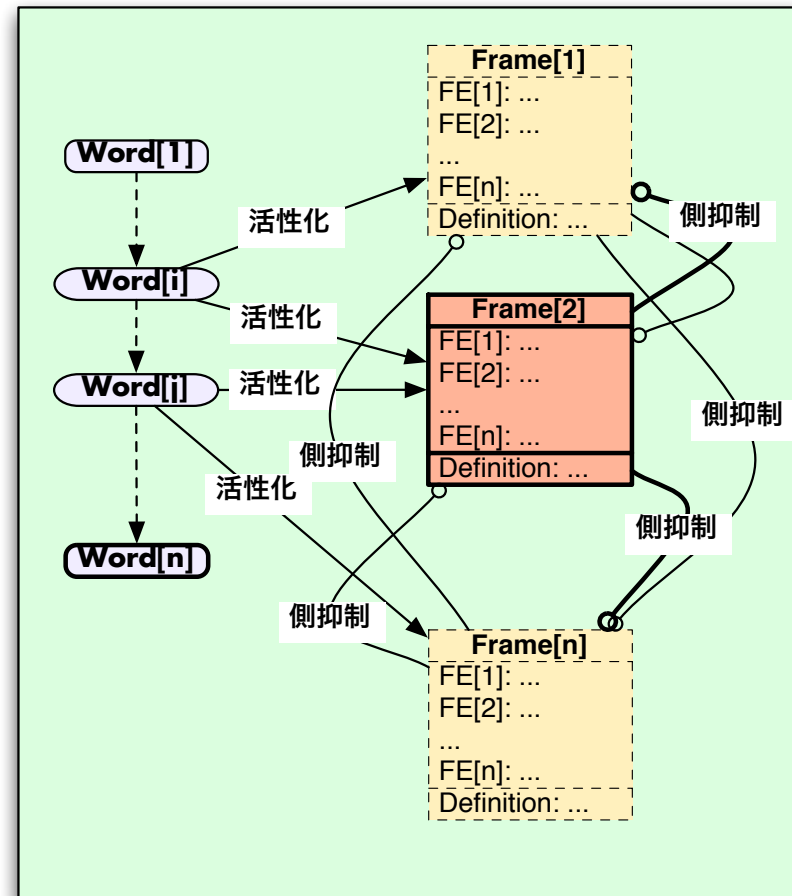
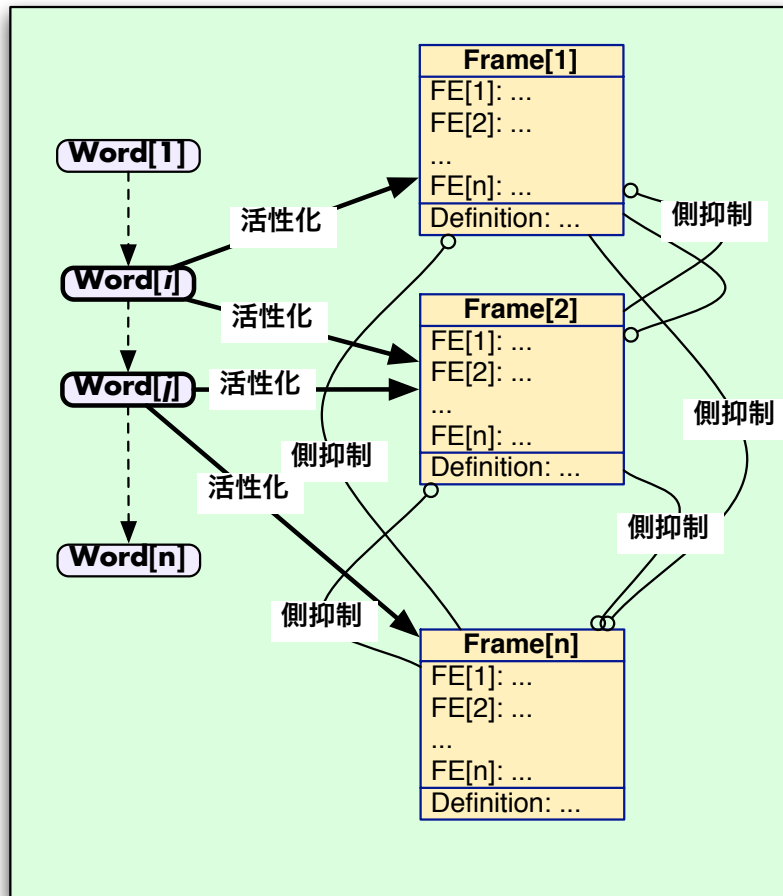
フレーム意味論の洞察

- 「何が何のために何をどうした」という理解の単位が存在し、それを意味フレーム (SF) と呼ぶ
 - SFとはヒトが理解できる状況を定義する構造体
 - SFごとに設定される状況相対的な意味役割をフレーム要素 (FE) と呼ぶ
- 動詞と名詞群の組み合わせがSFを決める
 - 動詞は特にSFの特定に貢献するが一意化しない

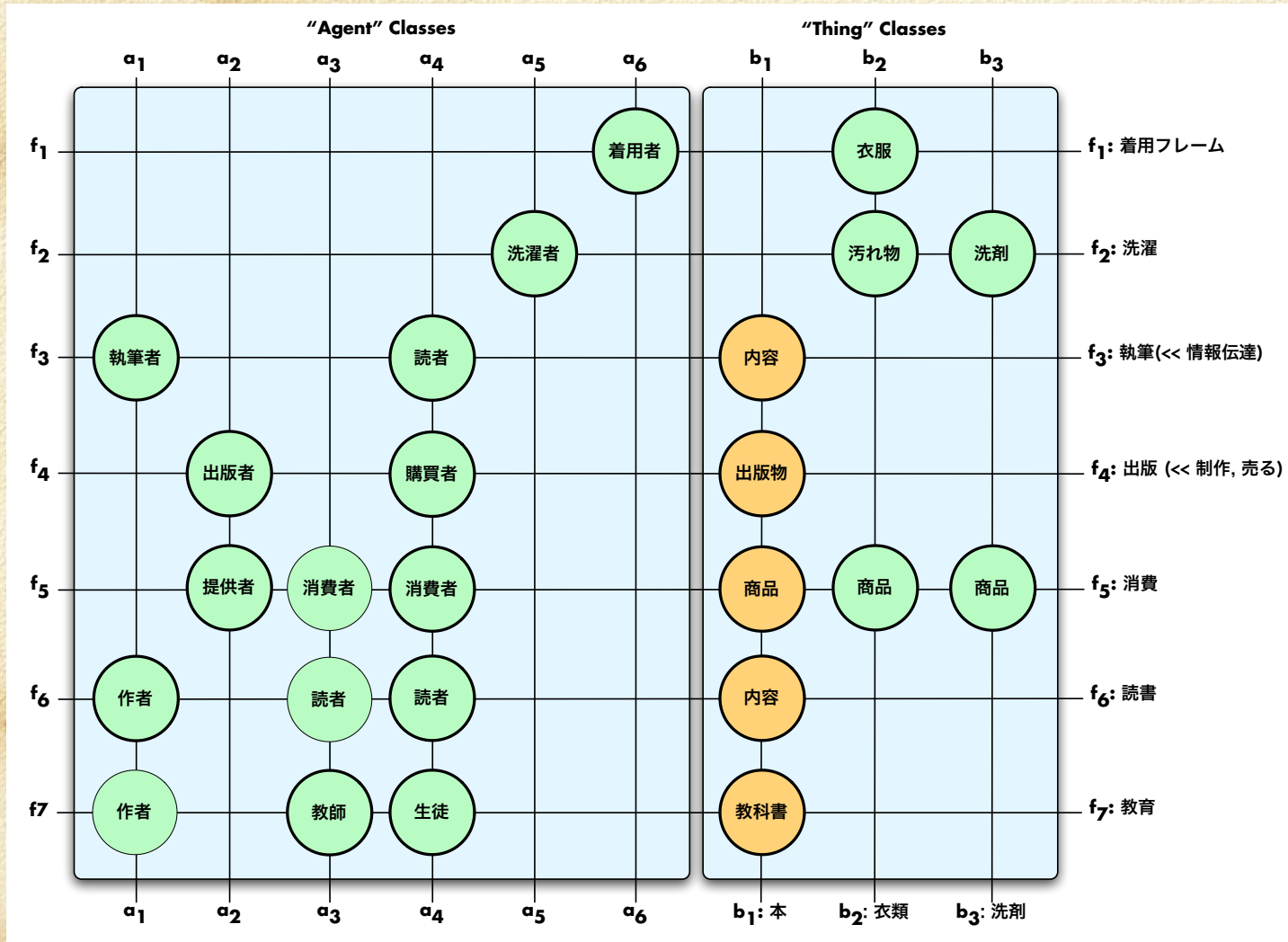
FNベースの意味タグづけ(案)

| Index | Segments | Frame:侮辱 [+explicit] | Frame:出版 [+explicit] | Frame:暴動の 発生 [+explicit] | Frame:(異常)事態 発生原因の特定 [+explicit] | Frame:執筆[- explicit,+gover norless] | Frame:表現[- explicit,+gover nor-less] | Frame:中国 [+implied] | |
|-------|----------|-------------------------|-------------------------|--------------------------------|----------------------------------------|-------------------------------------------|--------------------------------------------|------------------------|------------------|
| 1 | 西寧市 | | | FE:発生場所 | FE:発生場所 | | | FE:EVOKE: 都市名 | |
| 2 | で | | | MARKER | MARKER | | | | |
| 3 | の | | | LINKER | | | | | |
| 4 | 暴動 | | | GOVERNOR | FE:発生事態 | | | | |
| 5 | は、 | | | | LINKER | | | | |
| 6 | NULL1 | | FE:出版者 | FE:発生原因 | FE:発生原因 | | | | |
| 7 | NULL2 | FE:侮辱者 | FE:内容執筆者 | | | FE:執筆者 | FE:表現者 | | |
| 8 | イスラム教徒 | FE:標的 | FE:出版内容 | | | FE:執筆目的 | FE:表現目的 | | |
| 9 | を | MARKER | | | | | | | |
| 10 | 侮辱 | GOVERNOR | | | | | | | |
| 11 | する | EXTENDER | | | | | | | |
| 12 | 内容 | | EXTENDER | | | FE:EVOKE:執 筆内容 | FE:EVOKE:表 現内容 | | |
| 13 | の | | LINKER | | | LINKER | LINKER | | |
| 14 | 本 | FE:手段 | FE:出版 物:EVOKE | | | FE:EVOKE:執 筆対象 | FE:表現手段 | | |
| 15 | が | | MARKER | | | | | | |
| 16 | 四川省 | | FE:出版場所 | | | | | | FE:EVOKE: 地方名 |
| 17 | で | | MARKER | | | | | | |
| 18 | 刊行 | | GOVERNOR | HEAD | HEAD | | | | |
| 19 | された | | EXTENDER | EXTENDER1 | EXTENDER1 | | | | |
| 20 | こと | | | EXTENDER2 | EXTENDER2 | | | | |
| 21 | が | | | | MARKER | | | | |
| 22 | きっかけ | | | | GOVERNOR | | | | |
| 23 | NULL3 | | | | EXTENDER | | | | |
| 24 | 。 | | | | | | | | |

競合によるSFの特定



FE Grid



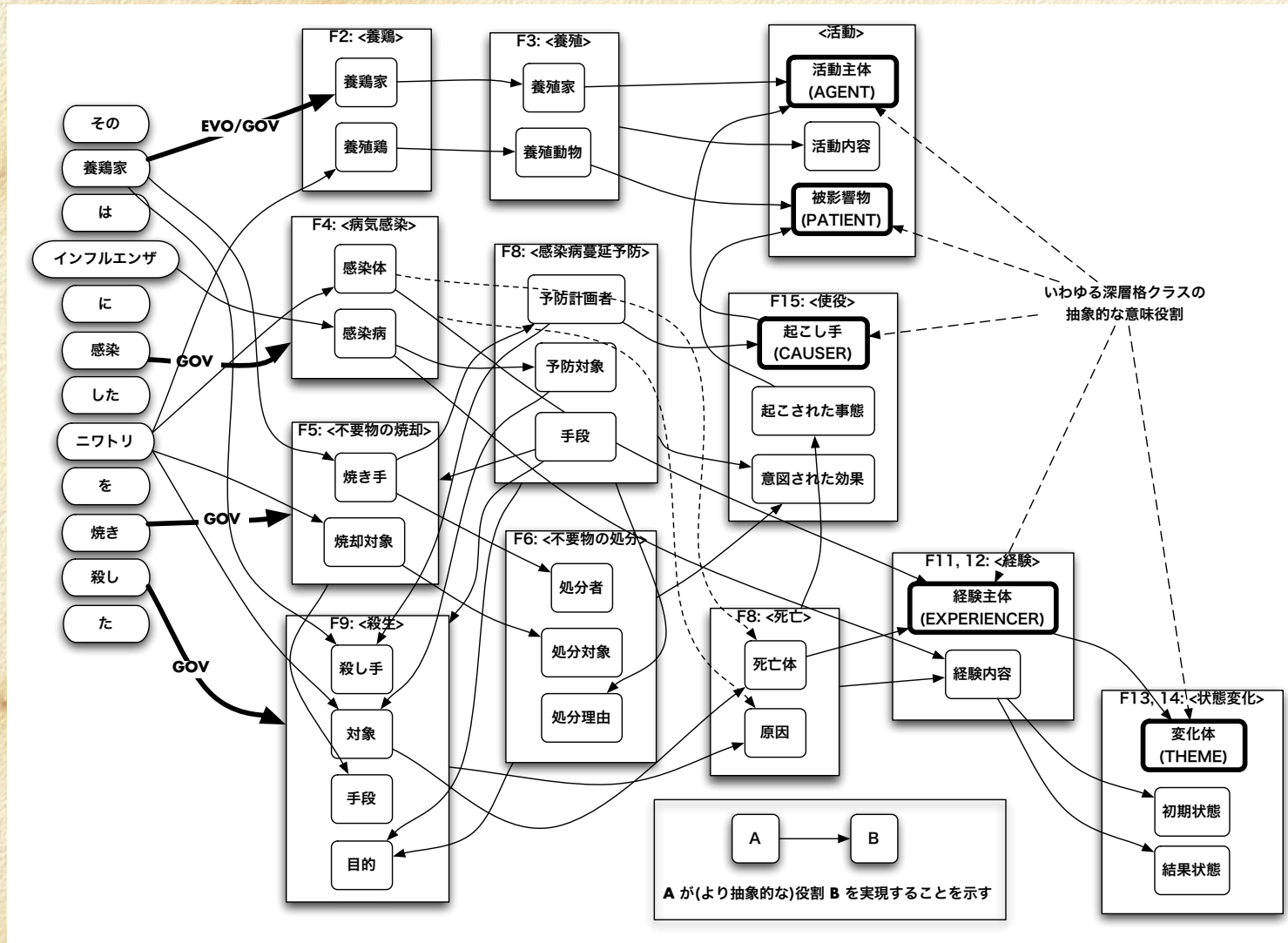
深層構造の再解釈

(1) その養鶏家はインフルエンザに感染した鶏を焼き殺した

| | Frame Index | F1 | F2 | F3 | F4 | F5 | F6 | F8 | F7 | F9 | F10 | F11 | F12 | F13 | F14 | F15 |
|-------|-------------|----------------------|----------|---------------------|---------------|--------------------------|----------------------|------------------------|------------------|--------|------------------|---------------------|---------------------|-----------------------|-----------------------|-------------------|
| Index | Frame Name | 文脈内での関係づけ | 養鶏 | 動物の養殖 [+implied] | 病気感染 | 不要物の焼却 | 不要物の処分 [+implied] | 疫病流行の予防 [+inferred] | 燃焼 [+implied] | 殺傷 | 死亡 [+implied] | 経験 [1][+implied] | 経験 [2][+implied] | 状態変化 [1][+implied] | 状態変化 [2][+implied] | 使役 [+implied?] |
| 0 | NULL | 先行状況 | | | | | | | | | | | | | | |
| 1 | その | GOVERNOR[+anaphoric] | | | | 焼却.主体 | 処分.主体 | 予防計画者 | | 殺し手 | | | | | | 起こし手 |
| 2 | 養鶏 | 指示 | GOVERNOR | EVOKER | | | | | | | | | | | | |
| 3 | 家 | | 主体指標 | 主体指標 | | | | | | | | | | | | |
| 4 | は | | | | | | | | | | | | | | | |
| 5 | インフルエンザ | | | | 感染病 | 焼却.対象 | 不要となった原因 | 予防する病気 | | 殺傷.対象 | | 経験内容1(病気) | | EVOKER:結果状態 | | |
| 6 | に | | | | | | | | | | | | | | | |
| 7 | 感染 | | | | GOVERNOR | | | EVOKER | | | | EVOKER | | | | |
| 8 | した | | | | | | | | | | | | | | | |
| 9 | ニワトリ | (食料としての)鶏 | 養殖動物 | 感染主体 | 焼却.対象 HEAD | 不要物 | 感染体 | 燃焼物[+間接] | 殺傷.対象 HEAD | 経験主体 | 経験主体 | 経験主体 | 変化体 | 変化体 | 起こされた事態 | |
| 10 | を | | | | | | | | | | | | | | | |
| 11 | 焼き | | | | GOVERNOR | GOVERNOR [+composite] | 手段 | EVOKER | 手段 | | | | | | | |
| 12 | 殺し | | | | | | [+composite] | [+composite] | GOVERNOR | EVOKER | | EVOKER:経験内容2(死) | | EVOKER:結果状態 | | |
| 13 | た | | | | | | | | | | | | | | | |
| 14 | . | | | | | | | | | | | | | | | |

- 事実上, Grammatical Constructions (GC) = <Frame, Form> Pairs
- この際, Frames の一部は変形生成文法の「深層構造」と同一の役割を果たす
- ただし, 列の相対的順番には意味がない
 - 異なるフレーム要素の束縛があるのみで「派生」は存在しない

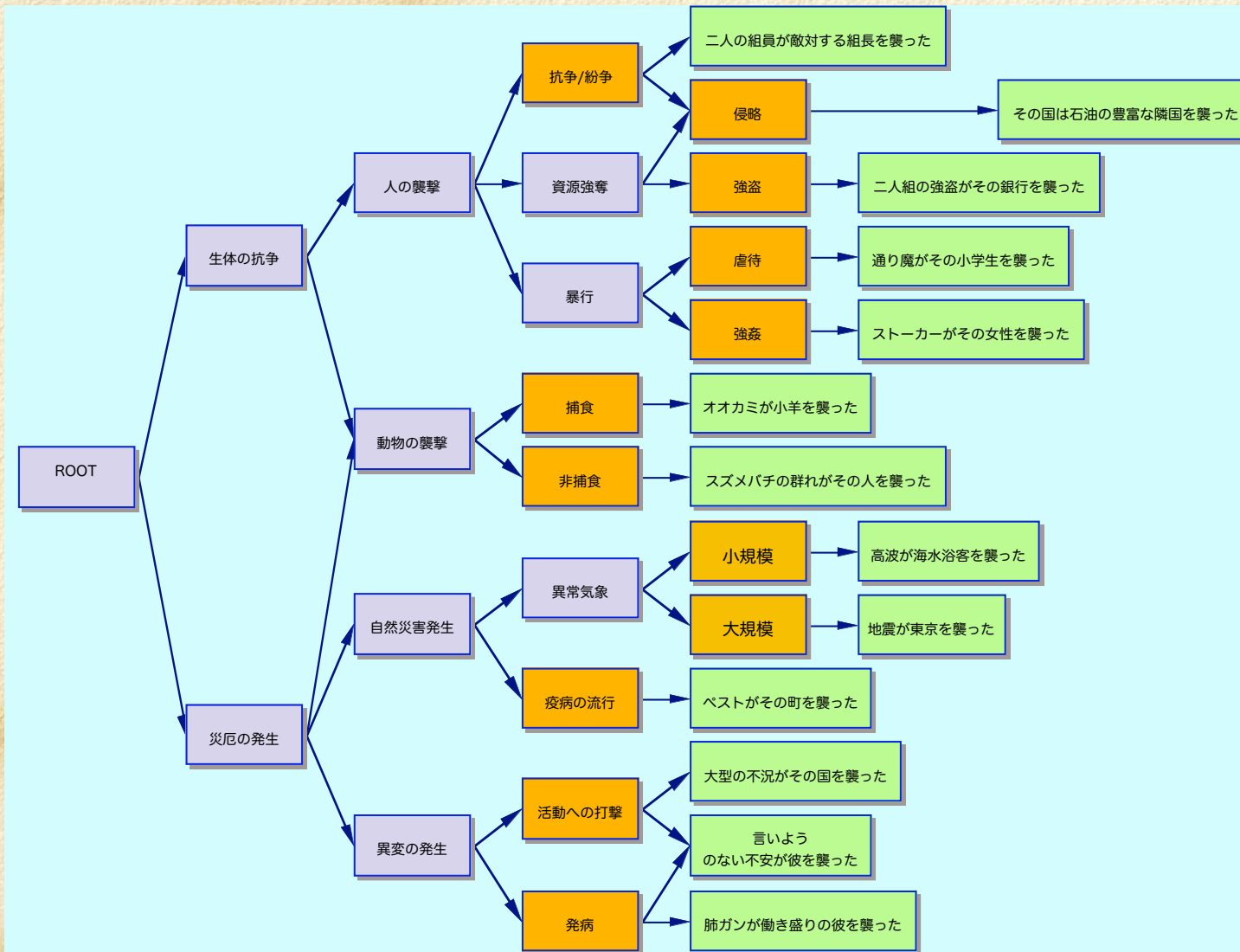
「養鶏」文の FrameNet



意味フレームは実在する

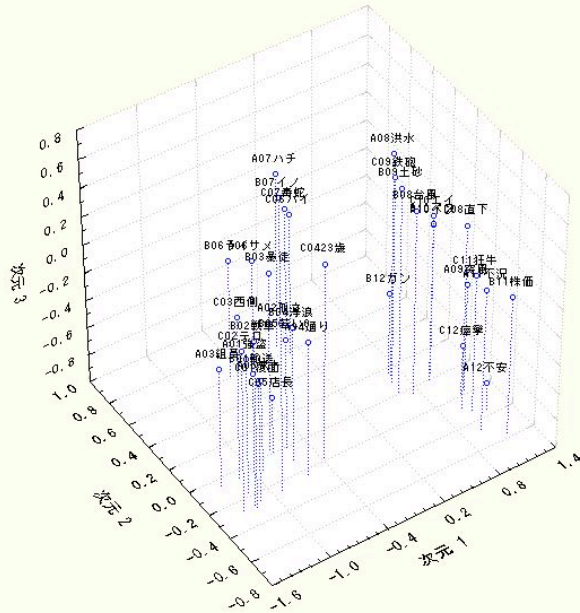
- Cluster01 = <資源強奪>: {
 - Cluster01.1 = <営利目的の強奪>: {F01a, b, c; F03a},
 - Cluster01.2 = <グループ間抗争>: {F02a, b, c; F03c}}
 - Cluster02 = <虐待>: {F04a, b, c}
 - Cluster03 = <暴動>: {F03b}
 - Cluster04 = <強姦>: {F05a, b, c}
 - Cluster05 = <動物の襲撃 (捕食)>: {F06a, b}
 - Cluster06 = <動物の襲撃 (非捕食)>: {F06c; F07a, b, c}
 - Cluster07 = <自然災害: 異常事態の発生 (超個人規模 1)>: {F08a, b; F09b, c}
 - Cluster08 = <活動への打撃: 異常事態の発生 (超個人規模 2)>: {F09a*, F11a, b, c}
 - Cluster09 = <疫病の流行: 異常事態の発生 (超個人規模 3)>: {F10a, b, c; F08c*}
 - Cluster10 = <発症: 異常事態の発生 (個人規模 1)>: {F12a, c}
 - Cluster11 = <発病: 異常事態の発生 (個人規模 2)>: {F12b}
- x^* は x が予想外の位置にあることを表す

「襲う」の意味フレーム階層

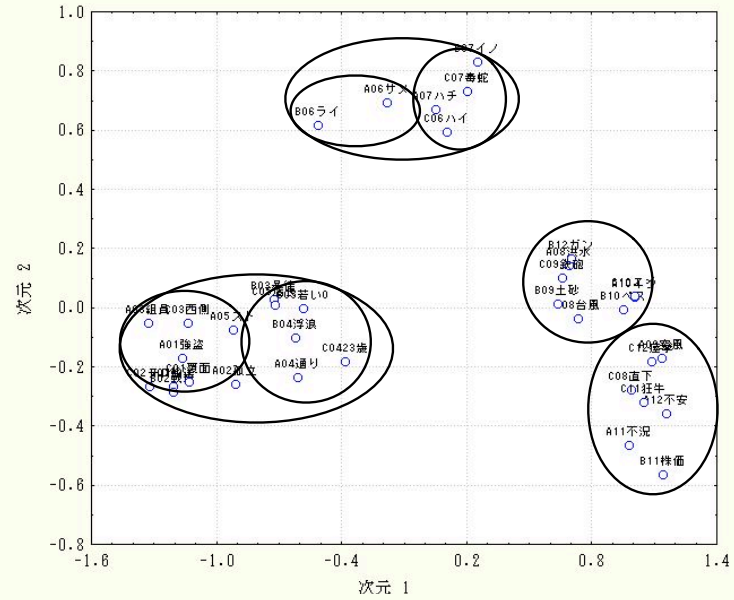


- **F01a:** 二人組の強盗がその銀行を襲った; **F01b:** 外国人のグループが現金輸送車を襲った; **F01c:** 覆面を被った男が銀座の宝石店を襲った
- **F02a:** 政治的に孤立した国が隣国を襲った, **F02b:** ドイツの戦車部隊がパリを襲った; **F02c:** テロリストの集団がアメリカ軍基地を襲った
- **F03a:** 二人の組員が敵対する組長を襲った; **F03b:** 暴徒と化した民衆が警察隊を襲った; **F03c:** 森の西側の部族が北側の部族を襲った
- **F04a:** 通り魔がその小学生を襲った; **F04b:** 5,6人の少年たちが公園にいた浮浪者を襲った; **F04c:** 23歳の男性が通行人を次々と襲った
- **F05a:** ストーカーがその女性を襲った; **F05b:** 無職の男が一人暮らしの若いOLを襲った; **F05c:** 店長がアルバイトの女子店員を襲った
- **F06a:** サメが傷ついたイルカを襲った; **F06b:** ライオンがインパラの群れを襲った, **F06c:** ハイエナの群れが国立公園の監視員を襲った
- **F07a:** スズメバチの大群が子供たちを襲った; **F07b:** イノシシがキノコ狩りに来ていた男性を襲った; **F07c:** 毒蛇が近づいてきた登山客を襲った
- **F08a:** 大洪水が東海地方を襲った; **F08b:** 大型台風が日本列島を襲った; **F08c:** 直下型地震が神戸の町を襲った
- **F09a:** 突風がテレビのリポーターを襲った; **F09b:** 土砂崩れが民家を襲った; **F09c:** 鉄砲水が避難する住民を襲った
- **F10a:** 悪性のインフルエンザがわが国を襲った; **F10b:** ペストがその町を襲った; **F10c:** エイズがアフリカの国々を襲った
- **F11a:** 大型の不況がその国を襲った; **F11b:** 株価の暴落が市場を襲った; **F11c:** 狂牛病問題が畜産業界を襲った
- **F12a:** 言いようのない不安が彼を襲った; **F12b:** 悪性のガンが働き盛りの彼を襲った; **F12c:** 突然の痙攣が少女の全身を襲った

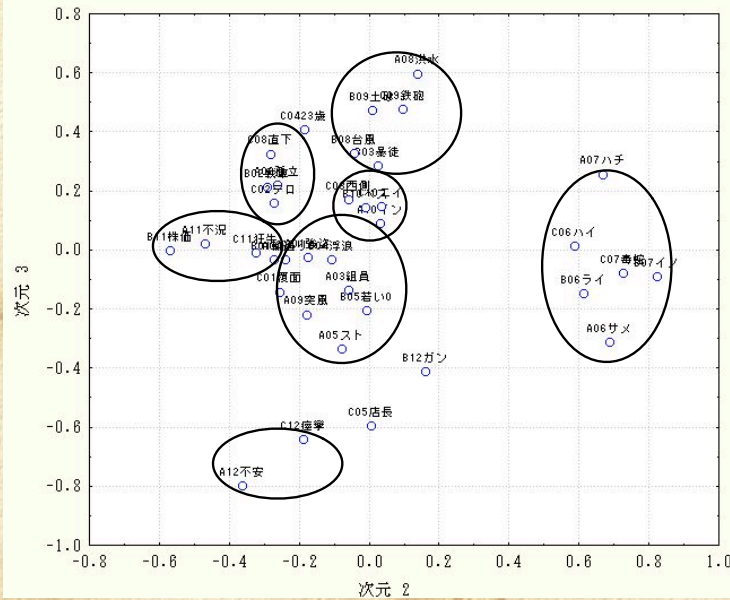
素性評定平均値（5項目削除）によるMDS（3次元解）



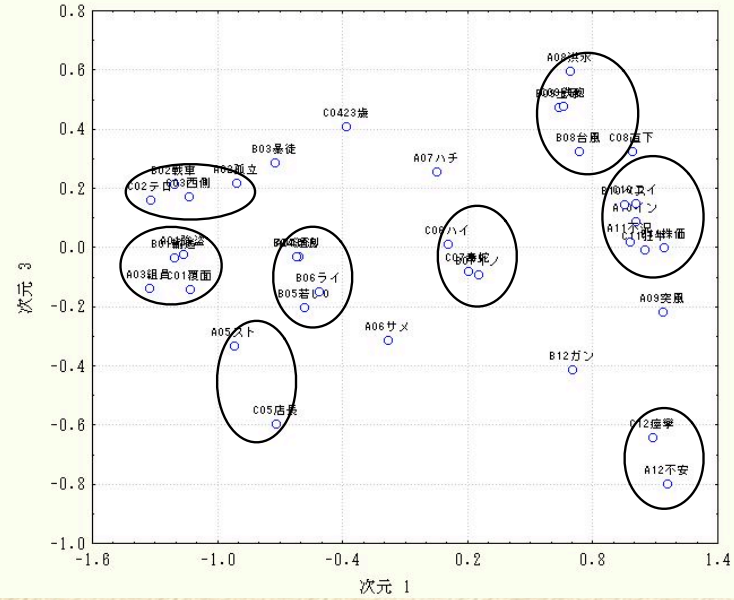
素性評定平均値（5項目削除後）によるMDS（3次元解）
ユークリッド距離，次元1 vs 次元2



素性評定平均値（5項目削除後）によるMDS（3次元解）
ユークリッド距離，次元2 vs 次元3



素性評定平均値（5項目削除後）によるMDS（3次元解）
ユークリッド距離，次元1 vs 次元3

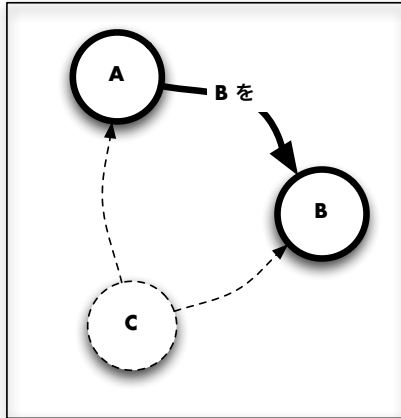


{襲う, 守る, 防ぐ, ...} フレームクラスター

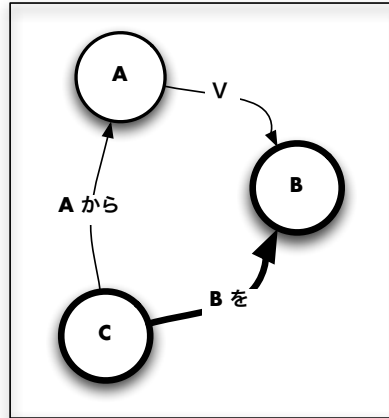
{襲う, 守る, 防ぐ} と危険との関係
Kow Kuroda 04/21/2004

太線は profiling に基づく項選択, および関係選択を表す
{守る, 防ぐ} の違いは, 視点 perspective の違いである
襲うは

状況 1: A が B を襲う(という危険性)



状況 2: C が (A の V) から B を守る



A: ATTACKER (襲い手)

B: VICTIM (被害者)

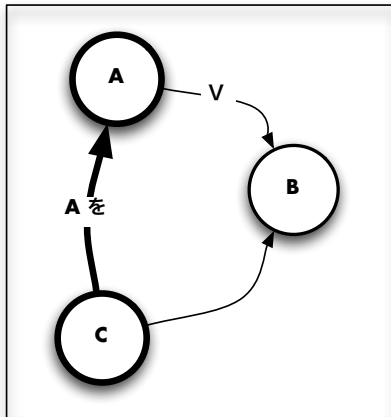
C: PROTECTOR (保護者), **DEFENDER** (衛手)

状況 1 は B (と C) にとって<危険: danger>となる

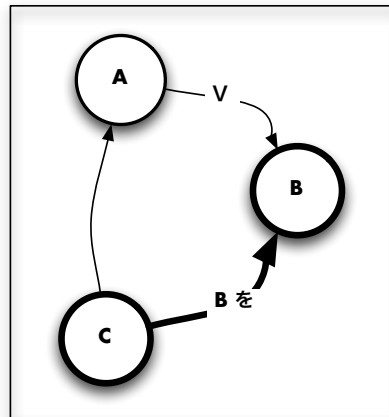
状況 2, 4 は<防御: defense>というより<保護: protection>

状況 3 は<防御>

状況 3: C が A (の V) を防ぐ (cf. 妨害)



状況 4: C が (結果として生じる) B を防ぐ



- a. A国が隣のB国を襲った
- b1. C国が(A国(の侵略)から)B国を守った
- b2. B国が(A国(の侵略)から)自国を守った
- c1. C国がA国(の(B国への)侵略)を防いだ [C/=B]
- c2. B国がA国(の(自国への)侵略)を防いだ [C==B]
- d1. C国がB国の(A国による)征服を防いだ [C/=B]
- d2. B国が自国の(A国による)征服を防いだ [C==B]

- e1. *C国が(A国(の侵略)から)B国を防いだ
- e2. *B国が(A国(の侵略)から)自国を防いだ
- f1. *C国がA国(の(B国への)侵略)を守った [C/=B]
- f2. *B国がA国(の(自国への)侵略)を守った [C==B]
- g1. *C国がB国の(A国による)征服を守った [C/=B]
- g2. *B国が自国の(A国による)征服を守った [C==B]

謝辞

- 中本敬子 (京大教育学部)
- 野澤 元 (京大人間なかぐる環境学)
- 龍岡昌弘 (Freelance?)
- ほか京大山梨研究室の面々