

日本語の意味タグ体系を定義する試み

~FrameNet の視点から~

黒田 航 井佐原 均

{kuroda,isahara}@crl.go.jp

独立行政法人情報通信研究機構 (NiCT)

自然言語処理学会ポスター発表 (P1-9)

03/16/04
東京工業大学

よい意味タグ体系の条件

- 機械学習可能なぐらい十分な体系性, 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

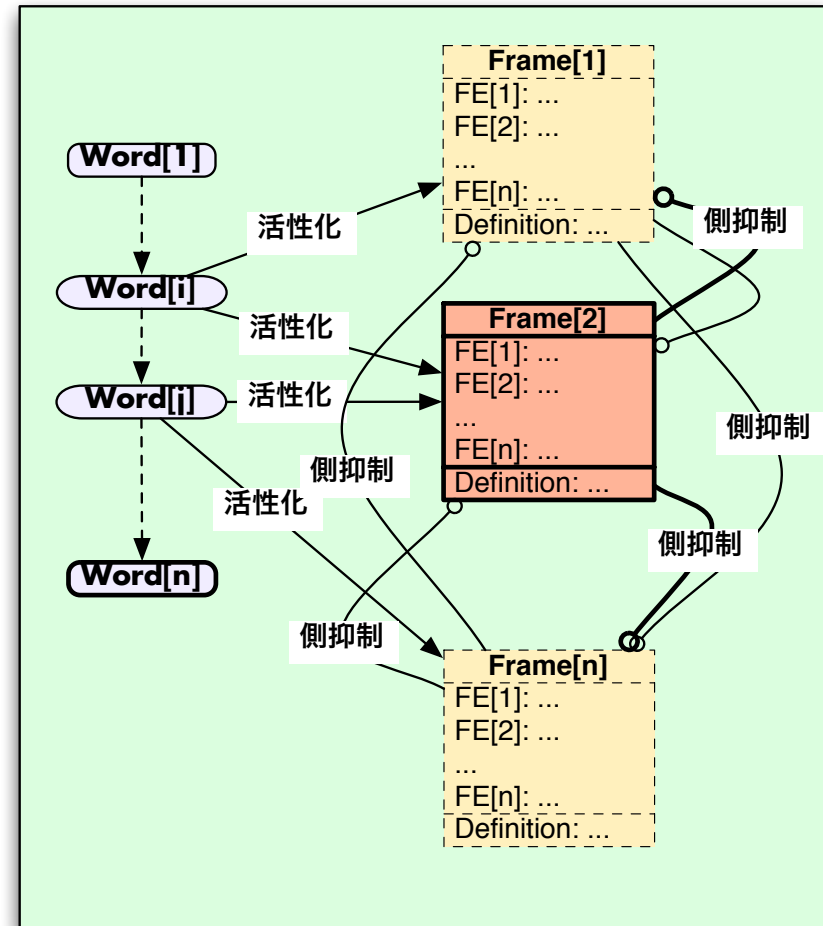
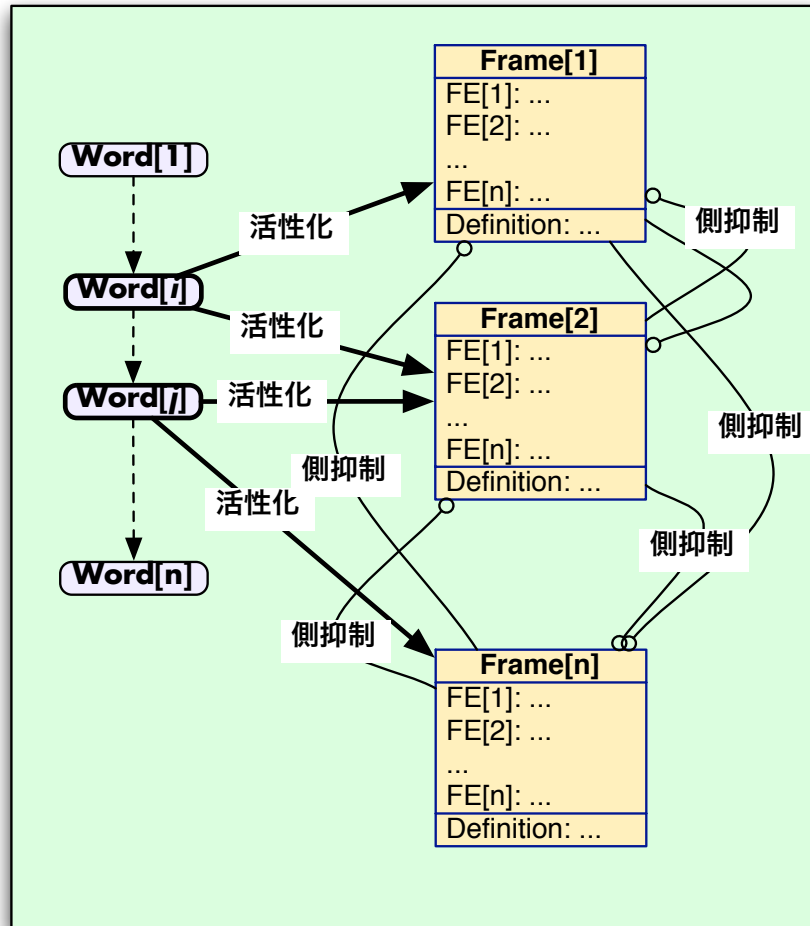
意味タグ体系に特有の問題

- 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

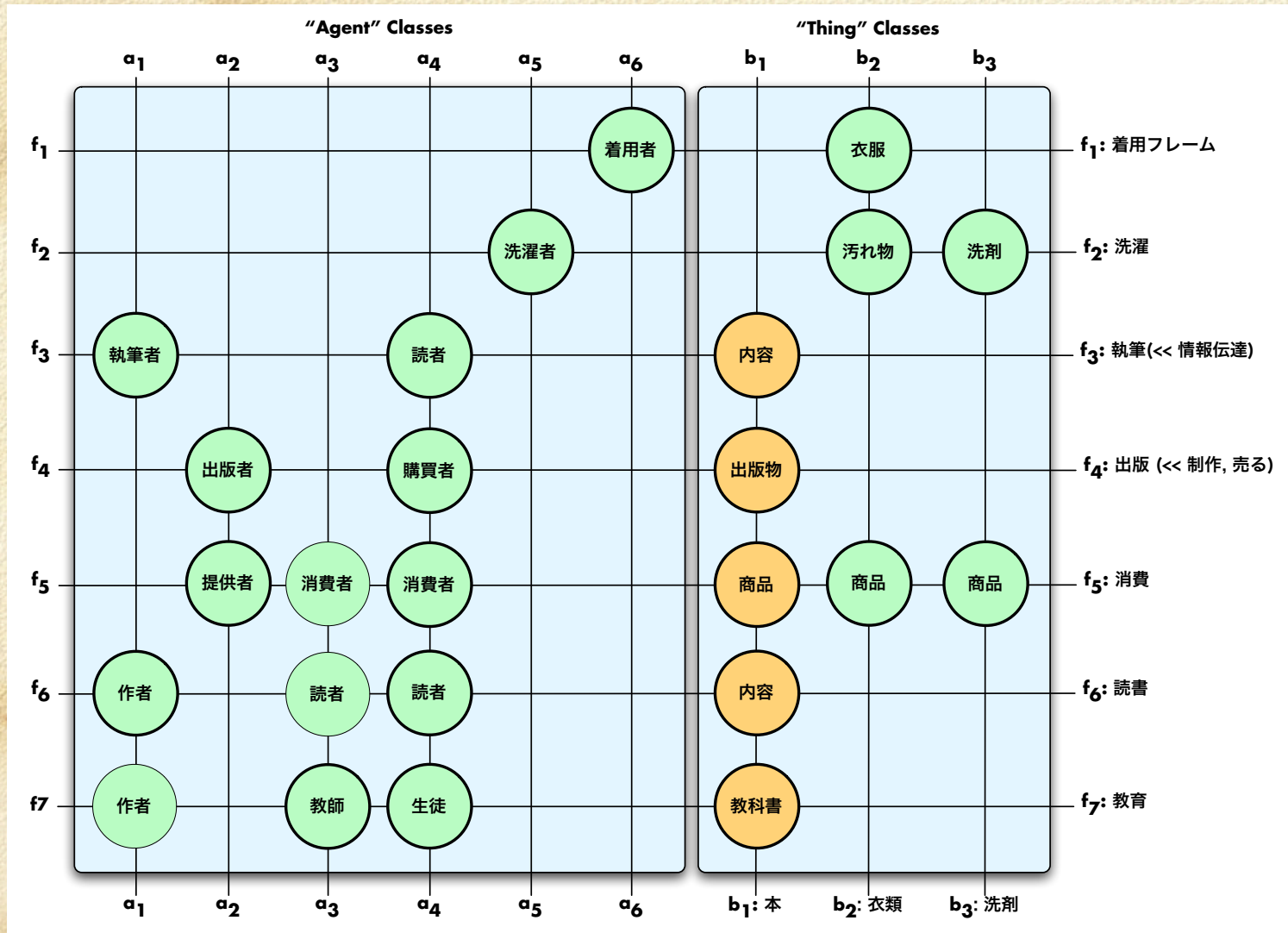
競合によるSFの特定



FNベースの意味タグづけ(案)

0 Segments	Frame:侮辱 [+explicit]	Frame:出版 [+explicit]	Frame:暴動の発生 [+explicit]	Frame:執筆 [-explicit]	Frame:表現 [-explicit]
1 西寧市での			FE:暴動の発生場所		
2 暴動は、			GOVERNOR		
3 イスラム教徒を	FE:侮辱の標的	FE:出版物の内 容	FE:暴動発生の原因	FE:執筆目的	FE:表現目的
4 侮辱する	GOVERNOR			FE:執筆内容	GOVERNOR
5 内容の				GOVERNOR	FE:表現手段
6 本が	FE:侮辱の手段	FE:出版物			
7 四川省で		FE:出版の場所			
8 刊行された		GOVERNOR			
9 ことが					
10 きっかけ					
11 NULL1	FE:侮辱者			FE:執筆者	FE:表現者
12 NULL2		FE:出版者			
13 。					

FE Grid

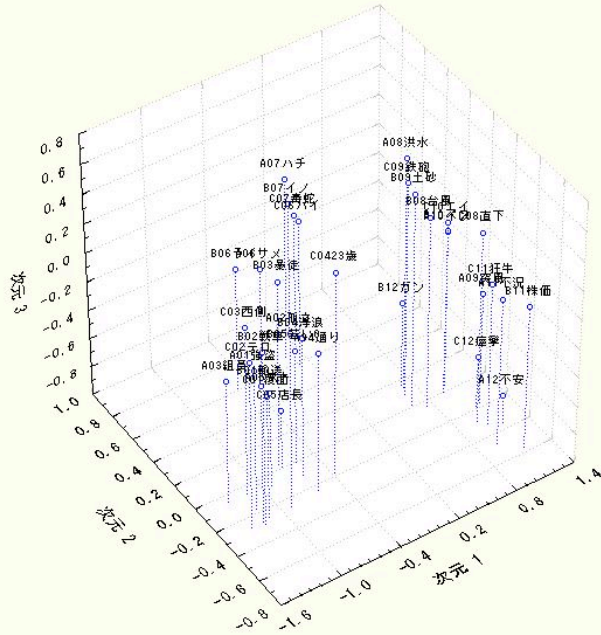


意味フレームは実在する

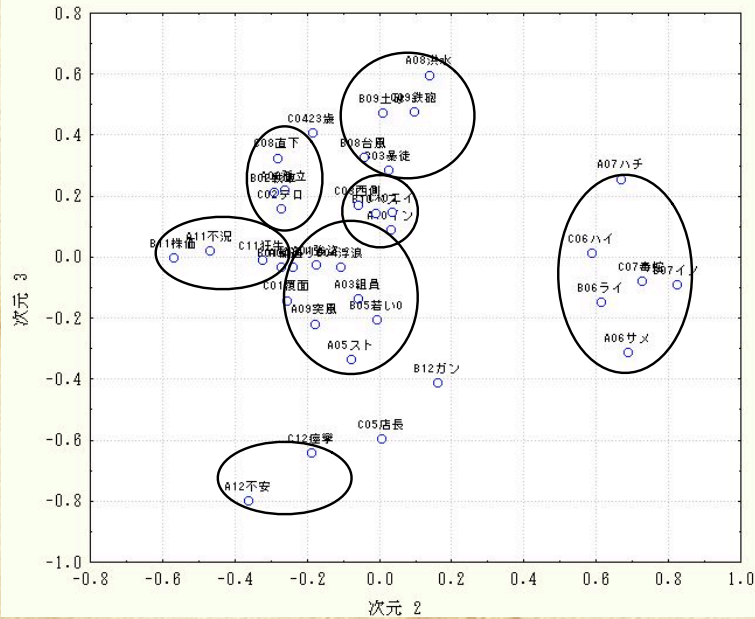
- Cluster01 = <資源強奪>: {
 - Cluster01.1 = <営利目的の強奪>: {F01a, b, c; F03a},
 - Cluster01.2 = <グループ間抗争>: {F02a, b, c; F03c}}
- Cluster02 = <虐待>: {F04a, b, c}
- Cluster03 = <暴動>: {F03b}
- Cluster04 = <強姦>: {F05a, b, c}
- Cluster05 = <動物の襲撃 (捕食)>: {F06a, b}
- Cluster06 = <動物の襲撃 (非捕食)>: {F06c; F07a, b, c}
- 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

- **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歳の男性が歩行者を次々と襲った
- **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

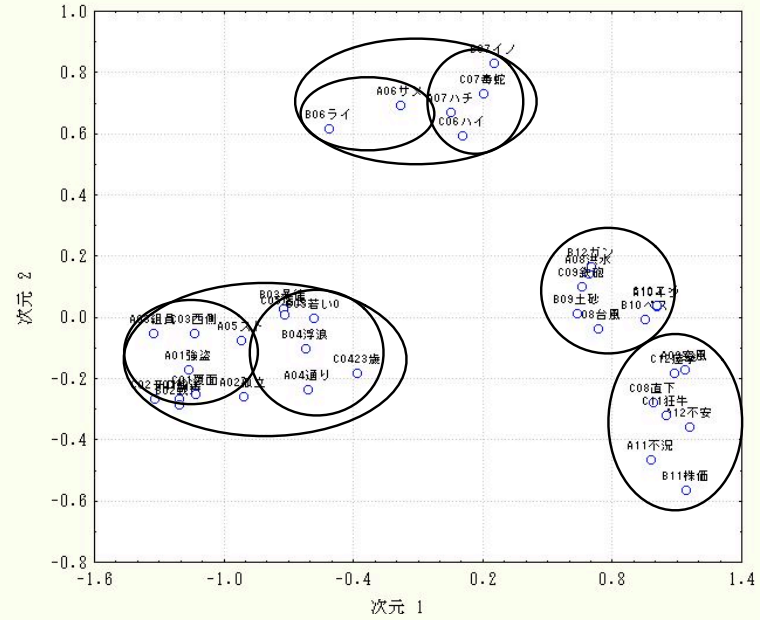
素性評定平均値（5項目削除）によるMDS（3次元解）



素性評定平均値（5項目削除）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1 vs 次元3

