

日本語フレームネット 拡張版開発企画の紹介

黒田 航

[kuroda@\(khn.\)nict.go.jp](mailto:kuroda@(khn.)nict.go.jp)

05/10/2004

NiCT内部向け発表

概要

- Japanese FrameNet eXtended (JFNX)の中間目標と最終目標
- FrameNet (FN)の特徴
- JFNXのロードマップ
- JFNXの作業工程
- 既存の解析データとJFNXの関係

JFNX の最終・中間目標

- **FrameNet に緩やかに準拠した日本語のための意味(役割)タグつきコーパスの開発**
- **日本語のための意味(役割)タグ体系の定義**
- **日本語の意味(役割)タグの母体となる意味フレームのデータベース (JFN)の開発**
- **意味(役割)タグづけ支援ツールの開発**

FN の特徴

- **既存の語彙解析, 統語解析に依存しない**
 - 特定の統語解析(e.g., 構文木, 係り受け)を仮定しない意味構造の直接記述
- **一般性, 実用性への指向**
 - 任意の言語の任意の構文の意味構造の記述
 - 意味フレームの概念に基づく(真理条件ではなく知識構造の表現を指向した)「理解のモデル化」

(意味) フレームとは？

- 「何が何を何のためにどうした」を表現する状況理解の単位
- 理解の「トップダウン効果」の素になる
- 状況相対的意味役割の構造体
 - 格フレームの発展形
 - 継承などによりネットワークを形成
- 語の意義 sense を決定するのが意味フレーム
 - 語義曖昧性解消問題の解決に特に有効

JFNに“X”が付く意味

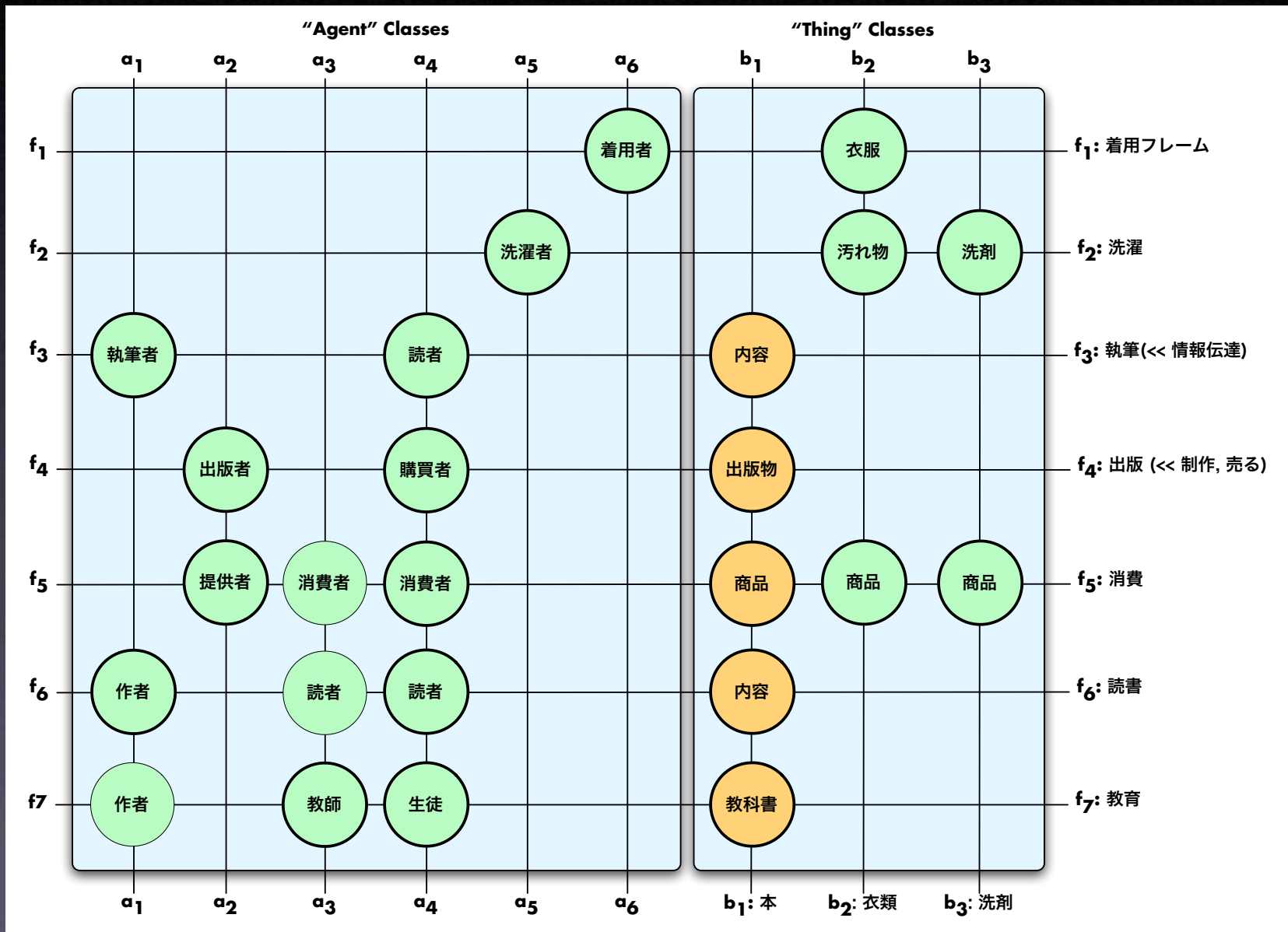
- 以下の理由により BFN と慶応大学(小原代表)の「本家」JFN との互換性は現時点では保証されていない
- 多重コーディングが動機となつて
 - BFN に存在しない独自概念 (EVOKER, MARKER, LINKER, EXTENDER) の導入
 - FE の言語実現基準で無視できない違いが存在
 - BFN に存在する概念 (supporters) の放棄

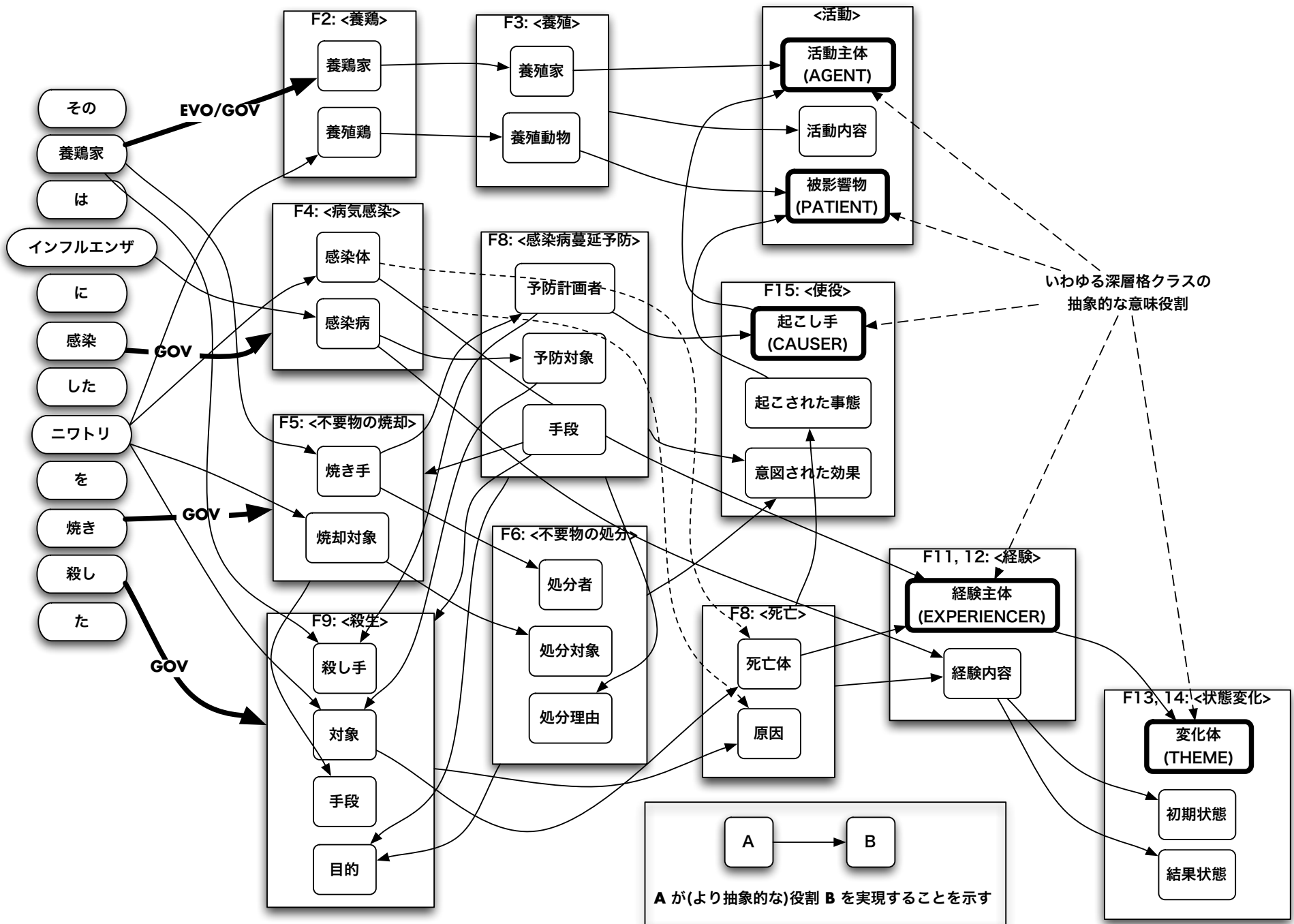
JFNX符号化の実例

# S-ID:950113180-014 KNP:98/08/19								
0	Frame	設置	寄付	選挙[+implied]	選挙上の支援 [+implied]	団体活動 [+implied]	資金確保	事務作業 [+implied]
1	議員ら	設置主	受取手	当選者:EVOKER	支援相手	活動主	働き手	働き手
2	が	MARKER				MARKER	MARKER	
3	事務局	設備	受付			窓口:EVOKER		EVOKER
4	を	MARKER						
5	つくり	GOVERNOR						
6	,							
7	市民		寄付主	投票者	支援主	支援者	資金源	
8	の		LINKER					
9	カンパ		寄付金:EVOKER		支援法:EVOKER	資金源:EVOKER		
10	で						MARKER	
11	供託金			供託金:EVOKER	支援内容		資金	作業内容
12	や							
13	運動					活動内容		
14	費用					活動費用:EVOKER		
15	を						MARKER	
16	ねん出						GOVERNOR	HEAD
17	。							

- 意味役割を意味タグに使用(列の順序, 色には意味なし)
- フレームは名称空間 ==> 意味役割名はローカル
- 各種の省略に対応

意味役割の格子





JFNX 開発 Roadmap

- 部分コーパスに基づいて評価版の制作
 - 当面は試験期間
 - 来年 NLP11 までの公開を目差す
- 評価者の反応に応じて軌道修正し，規模の拡大，注釈作業の本格化
 - データベースの導入
 - 注釈ツールの開発

JFNXの作業行程

- NICT コーパスから数十～数百程度の文 $\{S[1], S[2], \dots\}$ を採取し部分コーパス T を作成
 - 選定基準は検討中
- $S[i]$ ごとに意味フレームを網羅的に特定し、それを人手でコーディング
 - JFNX は Berkely FN にはない多層コーディング法を採用
- T に現れたすべてのフレームを規格化，データベース化し，全体作業量の見積もりを得る

当面の課題

- 評価版コーディングのための部分コーパス T の選定基準の明確化
- JFNX 解析法の明示化, 確立
 - 黒田が個人的に行っているタグづけ作業の明示化からマニュアル化が不可欠だが, これは容易ではない
 - 黒田以外の人物がコーディングできるのか, 実験する必要アリ
 - 可能だとして, どれほどの作業量なのか見積もる必要アリ

既存データとの関係 1

- 当面, ほかの解析には依存しない独立のレイヤーに意味タグつき部分コーパスを「参考データ」として提供
- 形態素解析, 係り受け解析, 格フレーム辞書の「信頼性」の指標を提供
 - 意味フレームの心理学的妥当性は部分的に検証されている
- **ただし「本気で」開発すれば, JFNX 流の意味解析は統語解析を無用にする可能性もある?**

既存データとの関係 2

形態素解析	手動解析可能	互換性は期待しうる
係り受け解析	仮定せず	上位互換性は期待しうる
格フレーム辞書	前提とせず	互換性は期待しうる

- JFNX の解析結果は、従来の諸解析結果に対し、独立のレイヤーをなし
- おそらく「監督」的役割を果たす