

GETTING DEEPER SEMANTICS THAN BERKELEY FRAMENET WITH MSFA

KOW KURODA MASAO UTIYAMA HITOSHI ISAHARA

NATIONAL INSTITUTE OF INFORMATION AND COMMUNICATIONS TECHNOLOGY (NICT), JAPAN

Multilayered Semantic Frame Analysis (MSFA) (Partial) of

アルゼンチンの元サッカー選手、ディエゴ・マラドーナ氏が六日、同国の検察当局に身柄を拘束された。 [Kyodai Corpus S-ID 950107210-002]

On January 6, Diego Maradona, a renowned former soccer player of Argentina, was temporarily taken into custody by the local prosecutor's office in his country.

Other 62 MSFA's (also from Kyodai Corpus) are available for evaluation (You need an account to access the data, however).

Structure of MSFA

1. **Columns** encode frames.
2. **Rows** encode realizations of semantic roles by linguistic units at the first column.
3. **Values of cells** identify the semantic roles for the frames columnwise.

The first column encodes the segmentation of a target sentence into semantic units — often different from the traditional morphological analysis. This is motivated by the need to deal with mismatches between semantic and morphosyntactic units.

This part of a row encodes a **role array**, a list of as many semantic roles as assigned to a substring (e.g. "Diego Maradona") in a text.

Frame-to-Frame Relations	A	B	C	D	E	F	G	H	I	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO
Frame Name	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]
Frame Map	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]	捜査[捜査の]

realizes one role of the role array

This cell encodes the realization relation between the semantic role <Criminal> (= <対象[捜査の, 犯罪の]>) of the frame <Investigation[Criminal]> (= <捜査[犯罪の]>) and the substring "Diego Maradona" (= "ディエゴ・マラドーナ(氏)").

This column encodes a frame <Investigation[Criminal]> (= <捜査[犯罪の]>) consists, for example, of such semantic roles as: Investigator (IS-A Agent), Criminal [+potential] (IS-A Patient), Charge (IS-A Reason)

Based on the Frame-to-Frame Relations (e.g., "Frame F subordinates Frame G") specified on this row, a FrameMap is automatically generated. Note: The list of F-to-F relations is still open.

Background

Adopting the (exciting) idea of Fillmore's Frame Semantics/FrameNet, we assume that a semantic frame F is a collection of semantic roles (= frame elements (FEs)) $R(F) = \{F.R_1, F.R_2, \dots, F.R_n\}$. MSFA specifies the set of "realization" relations and "evocation" effects that hold between all morphemes and major semantic roles for all frames evoked in a target sentence.

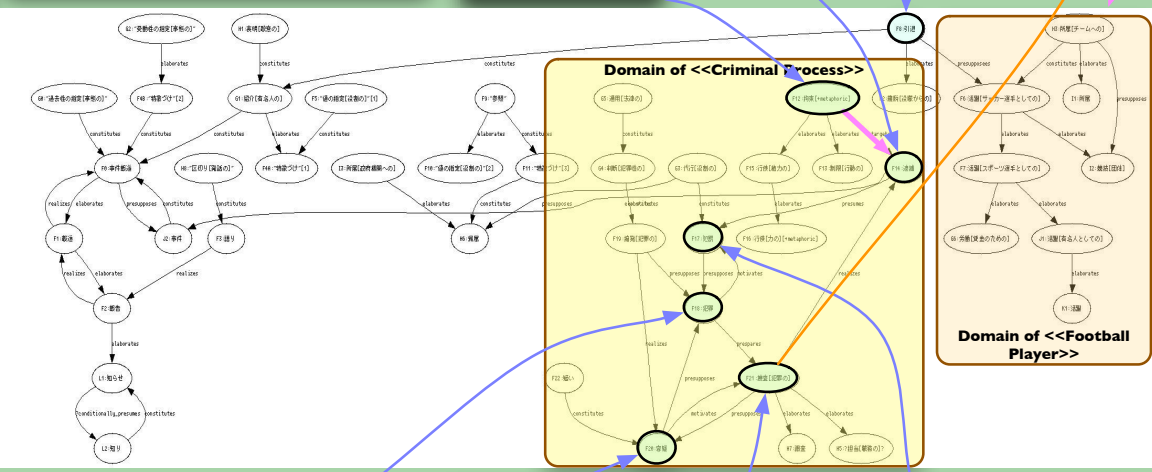
Motivation: Dealing with multiple, distributed frame-evocations

Frames are evoked by certain frame-evoking elements, or **frame-evokers**. Two types of evokers are distinguished: **frame-governors** (encoded as GOVs in MSFA) and **mere evokers** (encoded as EVOs in MSFA). A governor "names" a situation or activity, whereas a mere evoker doesn't. But many frames within a sentence are evoked by mere evokers, rather than by governors. MSFA is designed for specifying and encoding this effect. Very often, frame-evocation is achieved in a "distributed" fashion. This cries for explicit specification.

Goals (and Excuses)

We are NOT trying to provide a database of semantic frames, at least for now. This is the job FrameNet is dedicated for. Our purpose is to define a workable framework for serious annotation/analysis of real texts for semantic information in terms of semantic roles. This requires the estimation of (1) the number of frames needed to specify the understood content of a sentence at a reasonable level, and (2) the way all the evoked frames are integrated within a sentence — these are NOT trivial problems at all. Specifically, we DEFINITELY need to explore into what kinds of frame-to-frame relations are required, and estimate how many of them will exist.

MSFA is exploratory in that through MSFA, frames and the frame-to-frame relations among them are identified and defined inductively.



Acknowledgements to Jea-Ho Lee (NICT), Keiko Nakamoto (Bunkyo University), Hajime Nozawa (NICT), Yoshikata Shibuya (NICT), Daisuke Yokomori (Kyoto University)