

# Feedback from the ASFALDA French FrameNet project

Marie Candito, work in collaboration with  
Marianne Djemaa, Philippe Muller, Laure Vieu  
and also Pascal Amsili, Benoît Sagot, Lucie Barque, Richard Huyghe,  
Gaël de Chalendar, Farah Benamara, Yannick Matthieu

MLFN LREC Tutorial  
24 may 2016



# Outline

1. the ASFALDA - French FrameNet project
  - Methodology
  - Current status
  - Evaluation
2. Feedback: typical problems

# Motivation

- Objective = produce semantically annotated French data
- Why FrameNet ?
  - FrameNet more semantically oriented than e.g. PropBank
  - known to be quite portable across languages (Boas et al., 2009)

## Which strategy ?

- We could not target same coverage as Berkeley FrameNet
- → important to choose a development strategy
- **Frame-by-frame strategy**
  - (e.g. Berkeley FrameNet)
  - → **full lexical diversity** of a frame
- **Lemma-by-lemma strategy**
  - e.g. SALSA, Burchardt et al. 2006
  - → **coverage of all the senses of a lemma**
  - (in a given corpus)

## Which strategy ?

- Preliminary study:
  - Difficult to fully understand the exact semantic perimeter of a frame
  - Difficult to master very diverse semantic fields
- → we chose to work domain by domain
  - Objective: **full coverage of some chosen notional domains**
- Enforced coherence:
  - Close frames are either merged or their difference is made explicit
  - Missing frames for a given domain are created

## 4 annotated domains

- Commercial transactions
- Cognitive stances : belief, with various degrees of certainty, of a Cognizer for a given content
  - stative, with or without presupposition (to know, to think)
  - inchoative (to realize)
  - causal (to convince)
  - forecast (to predict) etc...
- Causality
  - various POS : because.c, to result.v, consequence.n, due to.prep ...
- Verbal communication (partially annotated only)

## Starting resources

- Berkeley FrameNet 1.5 release
- French lexicon obtained by projection from English
  - using bilingual dictionaries (Mouton et al., 2010)
  - projected using parallel corpora (Padò, 2007)
- Two syntactic treebanks (French Treebank and Sequoia Treebank)
  - corpus-oriented annotations: preserve natural probability distributions of senses and syntactic realizations of frame elements
  - syntactico-semantic lexicon can be extracted from annotations

# Development

- Selection of frames pertaining to the domains
- In **parallel**:
  - Adaptation of frames
  - Cleaning/extension of lexicon
- Annotation on corpus
  - Using the Salto tool (Burchardt et al., 2006)
  - Sometimes led to **further modification of frames and lexicon**



## Current status

- Release ... at the end of june 2016 (sorry)
- 98 frames with some annotations
- 872 LUs (= frame / lemma pairs)
- 12874 annotated frame instances
- plus 7116 occurrences marked as “out of domain”
- → can be used to train a framenet parser restricted to the 4 domains
- → syntactico-semantic lexicon re-extracted from the annotated data

## Evaluation : Inter-annotator agreement

For the lemma occurrences annotated by 2 independent annotators:

- Fscore for the frame selection
- Fscore for frame elements' exact match / partial match

	Nb of FEE	% of N	% of V	Inter-annotator Fscore		
				Frame	Exact FE	Partial FE
	17667	36	50	85.9	77.2	81.9
Break-down by notional domain						
Commercial	3307	60	40	92.0	73.4	80.4
Causality	7691	30	48	79.2	74.2	80.4
Cog. Stances	7886	28	62	90.6	81.1	86.0
Communic.	2221	23	76	89.6	82.3	87.5
Break-down by POS of the FEE						
V	8834	-	-	87.6	82.8	87.1
N	6234	-	-	86.8	68.3	72.5
other	2509	-	-	77.7	74.6	82.1

Rather high agreement

FE spans: **much easier for verbs** (cf. SALSA 2.0, Rehbein et al. 2012)

## Frame modifications

- 50 frames (only?) not modified from English frames
- 13 new frames: meant to complete a domain
- 37 frames are merges, splits, or slightly modified frames
  
- → more modification than expected
- (cf. Spanish FrameNet, SALSA reported few modifications)
  
- Merges resulting from difficulty to clarify frame differences
- Merges in order to limit polysemy
  - example Eventive\_cognizer\_affecting / Suasion

# Feedback

Main difficulties:

- Understanding the exact perimeter of a frame
- Coping with polysemy

Thank you

French FrameNet is coming soon..., check for announcement