FrameNet translation using bilingual dictionaries with evaluation on the English-French pair

Claire.Mouton@gmail.com
Gael.de-Chalendar@cea.fr
Benoit.Richert@student.ecp.fr

Agenda

- Introduction
- Proposed approach
- Evaluation
- Resource enrichment
- Conclusions

- **FrameNet : a resource for Semantic Role Labeling**
  - Semantic Role Labeling (SRL)
    - Detect and identify **predicate** of a given situation
    - Detect and identify **roles** of a given situation
    - Aims at helping Textual entailment, Question-Answering systems...
  - FrameNet
    - Language: English
    - Structure: Frame = set of triggering predicates + set of specific roles
    - Number of predicate-frame pairs: more than 10,000
    - Number of roles: 250 (specific subset for each frame)
  - Example

Attempt_suasion [Advise, beg, discourage,encourage, exhort, press,urge (...)]

[A number of embassies]$_{SPEAKER}$ are warning [their citizens]$_{ADRESSEE}$ [against traveling to Thailand's capital]$_{CONTENT}$.

- **Real need for other languages than English**
  - Case of French
    - Volem [Fernandez et al., 02]
      - Semantic resource for French, Spanish and Catalan
      - 1,500 verbs
      - ~20 generic semantic roles
      - Comparison to FrameNet
        - Much lower coverage
        - Less specific roles
        - Only verbs, no other part-of-speech
        - Entries are verbs (and not sets of predicates grouped by "senses" as in FrameNet)
    - FrameNet transposition to French [Pado and Pitel, 07]
      - ~7000 predicate-frame pairs
      - Precision 77%

Agenda

- Introduction
- **Proposed approach**
- Evaluation
- Resource enrichment
- Conclusions

Overview of the proposed method

- **For each frame and each predicate in this frame**
  - Extraction of translation pairs from bilingual dictionaries
  - Base score representing the confidence we have in the translation of the given predicate in the given frame
  - 5 variations of this score based on different heuristics
- **Linear combination of the scores**
- **Filtering with a parameter threshold**
- **Run with different parameters and weights on a development set to find the best settings**

Extraction of translation pairs

- **Bilingual dictionaries we use in our experiments**
  - Wiktionary
    - Creative Commons license
    - 27,109 French-English translation pairs in January 2009 version
    - Distinction of senses for some of the translations
  - EuRADic
    - Distributed by ELDA
    - 243,539 entries
- **Extraction of translation pairs**
  - English Lexical Unit (LU) present in predicates of a frame
    → French Lexical Unit(s) (LU)
  - 2 different resources by dividing EuRADic and Wiktionary results

- **Score S1: redundancy of translations**
  - If many English LU of the same frame translate to the same French LU → confidence for the translation to be correct is high.
  - French LU-Frame score=Nb of translation pairs for the LU in the given frame
  - If a translation pair is found in several sense distinctions in the Wiktionary, they are all summed up.
  - Example:
    - **Ingestion**

      …
      remettre.v {put back.v:1} 1
      boire.v {quaff.v:1, drink.v:2} 3
      alimenter.v {feed.v:1} 1
      déjeuner.v {lunch.v:1, dine.v:1, feed.v:1, eat.v:1} 4
      …

Wiktionary
consume liquid through the mouth
drink.v → boire.v
consume alcoholic beverages
drink.v → boire.v

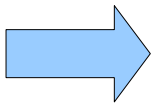
- **Structural score S2: polysemy of source LU**
  - Hypothesis
    - Polysemous source LU (present in more than one frame)
      → higher risk that translation is erroneous
    - S2 = confidence score S1 lowered depending on the number of frames containing the source LU
  - Example
    - *rise* appears in 9 different frames

      Getting_up

      get up → se lever

      rise → augmenter

      → se lever

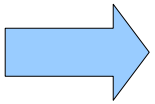


Se lever : S1 = 2 S2 = $2/10^{\alpha}$

Augmenter : S1 = 1 S2 = $1/9^{\alpha}$

Structural Scores II

- **Structural score S3: number of English LUs in the frame**
  - Hypothesis
    - Source frame contains lots of LUs
      → higher risk that redundant translations appear
    - S3 = confidence score S1 lowered depending on the number of source LUs in the given frame
  - Example
    - **Containers** has 116 English LUs
      bac.n is the French translation of 15 of the English LUs
      (WRONG) nigaud.n (← mug) is the French translation of 1 English LU
    - **Operational_testing** has 8 English LUs
      tester.v is the French translation of 1 of the English LUs



bac_Containers : S1 = 23 S3 =15/116$^{\alpha}$
nigaud_Containers : S1 = 1 S3 = 1/116$^{\alpha}$
tester_Operational_testing : S1 = 1 S3 = 1/8$^{\alpha}$

Target Scores I

- **Target score S4: number of translation pairs**
  - Hypothesis
    - High number of translation pairs
      → higher risk that redundant translations appear
    - S4 = confidence score S1 lowered depending on the number of translation pairs for the given frame
  - Example
    - Same idea as previous score

Target Scores II

- **Target score S5: number of LUs in the target frame**
  - Hypothesis
    - Target frame contains lots of LUs
      → Some LUs may carry slightly different meanings
    - S5 = confidence score S1 lowered depending on the number of target LUs in the given frame

- **Target score S6: polysemy of the target LU**
  - Hypothesis
    - Polysemous target LU (present in more than one frame)
      → LU less informative in the given frame
    - S6 = confidence score S1 lowered depending on the number of frames containing the target LU
  - Example
    - *Prendre* appears in 83 frames and *Porter* appears in 75 frames

Agenda

- Introduction
- Proposed approach
- **Evaluation**
- Resource enrichment
- Conclusions

Experimental setup

- **Evaluation criteria**
  - Precision, Recall, $F_{0.5}$-measure
  - Computed on each frame and averaged
- **Two FrameNet subsets**
  - Obtained from the union of FrameNet.FR [Pado and Pitel,07], unfiltered translations with EuRADic and with Wiktionary
  - Subset 1: Development set
    - Sample of 10 frames: Nb of LUs representative of the global distribution (quantiles)
    - Manually corrected
  - Subset 2: Test set
    - Sample of 10 frames: the ones used by [Pado and Pitel, 07]
    - Manually corrected
- **Scores combination and parameter settings**
  - Normalization and linear combination
  - Maximization of recall at $P_{0.95}$ and maximization of $F_{0.5}$-measure

Results

Resource	Linear combination	All frames		Test Set
		#LU-Frame	#Frames	P
Berkeley FrameNet		11,171	796	
FrameNet.FR [Pado and Pitel, 07]		6,659	480	77%
Wi_nofilter		19,912	781	70%
Wi_P_{095}	¼S2+ ¼ S5 + ½ S6	2,889	686	94%
$Wi\_F_{0.5}max$	**¼ S1 + ½ S4 + ¼ S6**	**15,720**	**781**	**74%**
Eu_nofilter		57,787	795	58%
Eu_P_{095}	¾ S2 + ¼ S6	616	210	100%
$Eu\_F_{0.5}max$	**¼ S2 + ¾ S6**	**24,885**	**767**	**74%**
$Wi\_F_{0.5}max$ U $Eu\_F_{0.5}max$		**34,121**	**793**	**70%**
Wiktionary /\ EuRADic		**12,211**	**773**	**82%**
$Wi\_F_{0.5}max$ /\ $Eu\_F_{0.5}max$		**6,484**	**724**	**95%**

Agenda

- Introduction
- Proposed approach
- Evaluation
- **Resource enrichment**
- Conclusions

Enrichment by similarity

- **Resources used to perform the enrichment**
  - Semantic spaces computed with MI on syntactical co-occurrences
  - Cosine similarity
- **Classification of nouns**
  - Classes ↔ frames
  - Learning data ↔ set of triggering Lus of each frame
  - K-NN classifier on multi-represented data [Kriegel et al, 05]
  - In every semantic space, weights the confidence on the neighbors by taking into account density of neighbors belonging to the same class
- **Variation of parameters**
  - K: 10, 25, 50
  - Filter thresholds
  - Selection of semantic spaces
  - Use of the size of the classes in confidence vector
  - Use of the translation score S1 into the learning process

Enrichment Results

- **Setting parameters**
  - Optimizing precision / coverage against union of three resources:
    - FrameNet.FR [Pado and Pitel, 07]
    - Translation using Wiktionary
    - Translation using EuRADic
- **Results**

	All frames			Test set	
Resource	**#LU-frame**	**#New attributions**	**#Frames**	P	R
Berkeley FrameNet	**11,171**		**796**		
FN.1 *precision*	9,536	7,581 (79%)	295	82%	7%
FN.2 *coverage*	27,371	24,997 (91%)	359	61%	10%
TFN+EFN.1	**15,132**	8,648 (57%)	727	**86%**	18%

- **Comments**
  - TFN + EFN.1 = (Wi_$F_{0.5}$max ∩ Eu_$F_{0.5}$max) ∪ FN.1
  - Combined resource: 15,132 pairs with an estimated precision of 86%

Agenda

- Introduction
- Proposed approach
- Evaluation
- Resource enrichment
- **Conclusions**

Conclusions and future work

- **New approach to transfer FrameNet into another language**
  - Validated for French
- **Resources resulting from translation**
  - A robust one: 95% estimated precision - 58% of BerkeleyFN size
  - A balanced one: 70% estimated precision – 3 times BerkeleyFN size
- **Enrichment**
  - Performed on nouns
  - Significant results incite to go further with verbs and adjectives
- **Future work**
  - Try to apply the translation method to the heads of the phrases filling the different roles in order to build learning data for a SRL system.

Questions

- **Approaches with bilingual corpora**
  - German: [Pado and Lapata, 05]
  - French: [Pado and Pitel, 07]
  - Italian: [Tonelli and Pianta, 08], [Basili et al.09]

- **Approaches with bilingual dictionaries and filtering**
  - Chinese: [Fung and Chen, 04]
  -

Parameter tuning

Resource	α	P	R	$F_{0.5}$
Wiktio		63%	40%	53%
Wiktio+S1 $F_{0.5}max$		63%	40%	53%
Wiktio+S2 $F_{0.5}max$	1	65%	**40%**	54%
Wiktio+S3 $F_{0.5}max$	1	63%	40%	53%
Wiktio+S4 $F_{0.5}max$	0.5	66%	38%	53%
Wiktio+S5 $F_{0.5}max$	0.75	66%	38%	53%
Wiktio+S6 $F_{0.5}max$	1	70%	36%	**55%**
EuRADic		51%	93%	56%
EuRA+S1 $F_{0.5}max$		74%	34%	58%
EuRA+S2 $F_{0.5}max$	0.75	59%	**75%**	60%
EuRA+S3 $F_{0.5}max$	0.25	69%	51%	59%
EuRA+S4 $F_{0.5}max$	0.1	71%	46%	60%
EuRA+S5 $F_{0.5}max$	0.25	71%	46%	60%
EuRA+S6 $F_{0.5}max$	0.25	68%	55%	**64%**

Results

	Linear combination	All frames		Test set			
Resource		#LU-frames	#Frames	P	R	F_{05}	F
Berkeley FrameNet		**11,171**	**796**				
Wi_nofilter		19,912	781	70%	33%	57%	44%
Wi_P095	$\frac{1}{4}.S2 + \frac{1}{4}.S5 + \frac{1}{2}.S6$	2,889	686	94%	11%	33%	18%
Wi_F05max	$\frac{1}{4}.S1 + \frac{1}{2}.S4 + \frac{1}{4}.S6$	**15,720**	**781**	**74%**	**30%**	**56%**	**42%**
EuRADic_nofilter		57,787	795	58%	84%	61%	67%
EuRADic_P095	$\frac{3}{4}.S2 + \frac{1}{4}.S6$	616	210	100%	2%	10%	4%
EuRADic_F05max	$\frac{1}{4}.S2 + \frac{3}{4}.S6$	**24,885**	**767**	**74%**	**44%**	**63%**	**53%**
FrameNet.fr_nofilter		**6,659**	**480**	**77%**	**23%**	**43%**	**31%**
Union							
$Wi \cup EuRADic$		65,488	796	57%	92%	61%	69%
$W\_P_{0.95} \cup E\_P_{0.95}$		3,256	695	94%	12%	35%	20%
$W\_F_{0.5max} \cup E\_F_{0.5max}$		**34,121**	**793**	**70%**	**59%**	**67%**	**63%**
Intersection							
$Wi \cap EuRADic$		**12,211**	**773**	**82%**	**25%**	**56%**	**38%**
$W\_F_{0.5max} \cap E\_F_{0.5max}$		**6,484**	**724**	**95%**	**15%**	**43%**	**25%**
$Wi\_F_{0.5}max \cap Eu\_F_{0.5}max \cap FN.fr$		7,814	742	95%	18%	49%	29 %