Making Use of Existing Lexical Resources to Build a Verbnet like Classification of French Verbs

Ingrid Falk

École Doctorale IAEM
Spécialité Informatique

Soutenance de thèse 13/06/2012

Topic of the thesis

Explore ways of building a syntactic semantic classification of French verbs
where groups of verbs are associated with:

- syntactic information (subcategorisation frames)
- semantic information (thematic role sets)

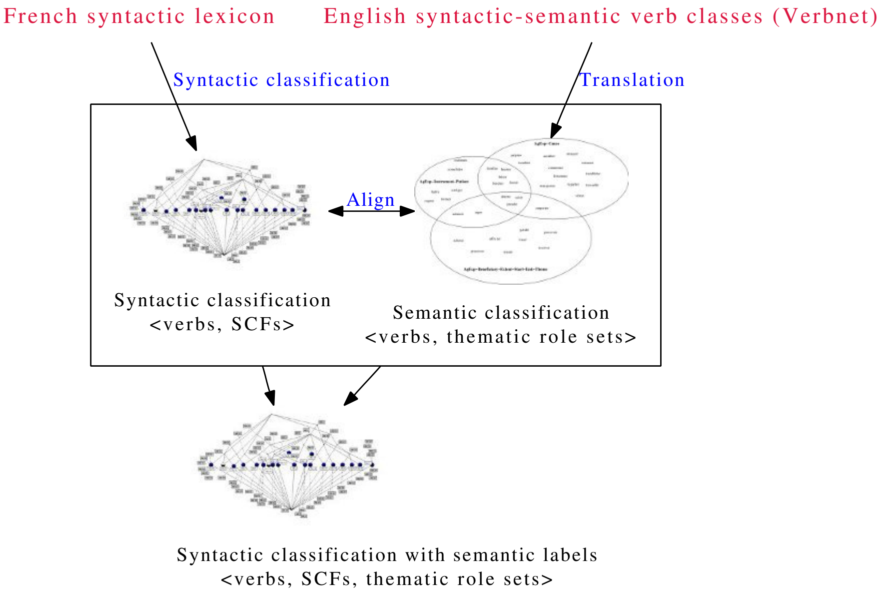
Using existing lexical resources for French and English.

More specifically

- we explore ways of building a syntactic classification
- using the classification methods
  - Formal Concept Analysis (FCA) – symbolic
  - Incremental Growing Neural Gas with Feature maximisation (IGNGF) – neural clustering
- two-fold evaluation
  1. on verb groups
  2. on associations of verbs with syntactic frames and thematic role sets

Contributions

- automatic acquisition of a syntactic-semantic classification
- two classification techniques not yet used for verb classification
- novel translation approach to build a semantic classification


French syntactic lexicon
English syntactic-semantic verb classes (Verbnet)
Syntactic classification
Translation
Align
Syntactic classification
<verbs, SCFs>
Semantic classification
<verbs, thematic role sets>
Syntactic classification with semantic labels
<verbs, SCFs, thematic role sets>

Outline

3. Lexical Resources
   - French Lexical Resources
   - English Lexical Resource

Lexical resources

French existing lexical resources: Dicovalence, Ladl tables, TreeLex

- merged into unique syntactic lexicon
- provide additional syntactic and semantic features
- both used for classification

English Verbnet classes

- translated to French
- provide associations with thematic role sets

Merged syntactic lexicon

- 5918 verbs, 345 subcategorisation frames, 20443 verb, frame pairs.

Verb: *expédier*

SCF	**Source info**
SUJ:NP,DUMMY:REFL	DV:41640,41650
SUJ:NP,OBJ:NP	DV:41640,41650;TL
SUJ:NP,OBJ:NP,AOBJ:PP	TL
SUJ:NP,OBJ:NP,POBJ:PP	DV:41640
SUJ:NP,OBJ:NP,POBJ:PP,POBJ:PP	LA:38L
SUJ:NP,OBJ:NP,POBJ:VPinf	LA:3
SUJ:NP,POBJ:PP,DUMMY:REFL	DV:41640

DV: Dicovalence, LA: LADL tables, TL: Treelex

Other features extracted from the lexicons

Mostly syntactic

Feature	Description	related VN class
ArgNbr	4 or more arguments	*get-13.5.1*, *send-11.1*
Event	arguments realised as clauses	*correspond-36.1*, *characterize-29.2*, *say-37.7*, ...
...		

Mostly semantic

Feature	Description	related VN class
Loc	location role	*put-9.1*, *remove-10.1*, ...
Nhum	concrete object, non human role	*hit-18.1* (eg. Instrument role), *other_cos-45.4*, ...
...		

English lexical resource – Verbnet

English Verbnet [Schuler, 2006]

- large scale syntactic semantic classification of English verbs
- verbs with similar syntactic and semantic behaviour manually grouped together

- Obtain associations of French verbs with Verbnet classes

English Verbnet

Verbnet example class *hit-18.1*:

Verbs	*batter, beat, bump, butt, drum, hammer, hit, jab, kick, knock, lash, pound, rap, slap, smack, smash, strike, tap*	
Thematic roles (semantics)	Agent, Instrument, Patient	
Frames (syntax)	SUJ:NP,P-OBJ:PP	Agent V Patient
	SUJ:NP,P-OBJ:PP,P-OBJ:PP	Agent V Patient Instrument
	SUJ:NP,OBJ:NP	Agent V Patient
		Instrument V Patient
	SUJ:NP,OBJ:NP,P-OBJ:PP	Agent V Patient Instrument

Translating English Verbnet classes

- using dictionaries
- noisy because of polysemy

Filter using two approaches:

1. Based on translation frequencies
   - Only keep most frequent translations
2. Machine Learning with Support Vector Machines
   - train classifier
   - for $\langle$French verb v_{fr}, English Verbnet class $C_{VN}\rangle$
   - has v_{fr} thematic roles of C_{VN}?

SVM classification performed best: Distribution of verbs

- most similar to English Verbnet
- most similar to FCA classification

Derived French lexical resources

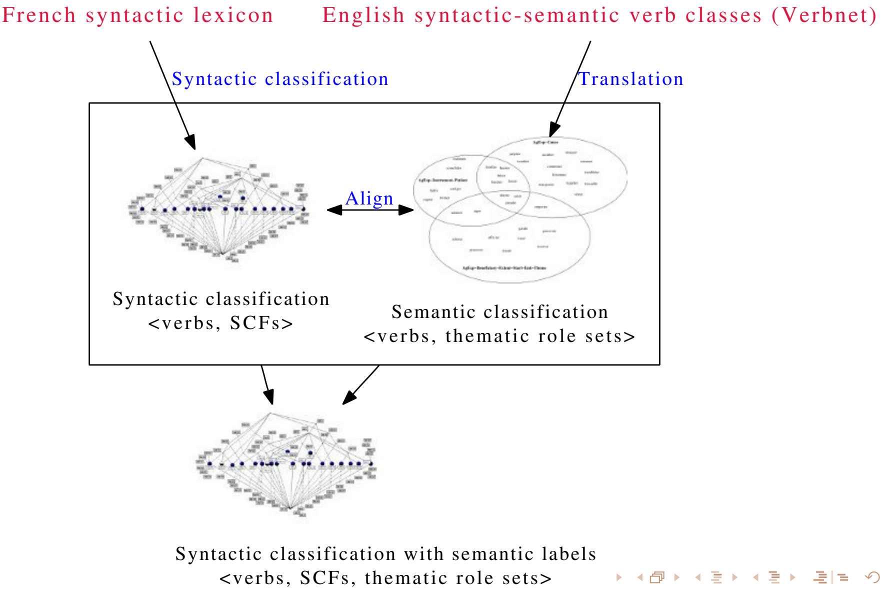
- merged syntactic lexicon – French
- syntactic and semantic features – French
- translated Verbnet classes – English

used to

1. extract features for classification
2. provide thematic role set to French verb classes

Extracted features

- from merged syntactic lexicon: subcategorisation frames
- from Dicovalence and Ladl resources: syntactic and semantic features other than subcategorisation frames
- from translated Verbnet classes: thematic role sets (grids)


French syntactic lexicon
English syntactic-semantic verb classes (Verbnet)
Syntactic classification
Translation
Align
Syntactic classification
<verbs, SCFs>
Semantic classification
<verbs, thematic role sets>
Syntactic classification with semantic labels
<verbs, SCFs, thematic role sets>

Aligning French verb groups with translated Verbnet classes

- using F-measure between recall (R) and precision (P)
- verb cluster C_{cluster}, translated Verbnet class C_{VN}

$$
\begin{aligned}
\mathrm{R}(\mathrm{C}_{\text{cluster}}, C_{\text{VN}}) &= \frac{|\text{verbs} \in C_{\text{VN}} \cap C_{\text{cluster}}|}{|\text{verbs} \in C_{\text{VN}}|} \\
\mathrm{P}(\mathrm{C}_{\text{cluster}}, C_{\text{VN}}) &= \frac{|\text{verbs} \in C_{\text{VN}} \cap C_{\text{cluster}}|}{|\text{verbs} \in C_{\text{cluster}}|} \\
\mathrm{F}(\mathrm{C}_{\text{cluster}}, C_{\text{VN}}) &= \frac{2RP}{R+P}
\end{aligned}
$$

Associating French verb groups with thematic role sets

- C_{cluster} aligned with translated class C_{VN}
- C_{cluster} is assigned thematic role set of C_{VN}

- Verbnet classes identified with their thematic role set
- Verbnet roles grouped:

  AgExp: Agent, Experiencer
  Start: Source, Material
  End: Product, Destination, Recipient

Outline

4. Clustering Methods
   - Formal Concept Analysis (FCA)
   - Incremental Growing Neural Gas with Feature Maximisation (IGNGF)

Formal Concept Analysis (FCA) [Ganter and Wille, 1999]

- symbolic method for deriving conceptual structures – concepts – out of data
- FCA organises concepts into a hierarchy – concept lattice
- Concepts determined by:
  - extent: set of objects shared by attributes in intent
  - intent: set of attributes shared by objects in extent

The data

Objects: 2091 verbs

Attributes:

- 238 frames from merged syntactic lexicon
- additional syntactic and semantic features from Dicovalence and Ladl

Example

	frames					
	SUJ:NP,OBJ:NP,AOBJ:PP	SUJ:NP,OBJ:NP,DEOBJ:PP	Sym	ArgNbr	Loc	Nhum
expédier	X		X		X	X

The concept lattice

12 802 concepts

- need to filter

How to select the most relevant concepts?

Concept selection indices

- introduced in [Klimushkin *et al.*, 2010]
- select relevant concepts
- in concept lattices built on noisy data

Stability
- How much does a concept depend on individual members in extent/intent?

Separation
- How well does a concept sort out verb and frames it covers from other verb and frames.

Probability
- What is the probability of a concept intent/extent to be a concept intent/extent by chance?

Which indices to select the best classes?

Method:
Using fixed combination of indices

- select $N, (N \in \{1500, 1000, 500\})$ concepts from concept lattice with highest index combination
- align classes translated from Verbnet with these concepts
- select FCA concepts with associated Verbnet class
- compare obtained ⟨verb, Verbnet class⟩ associations with a reference

Best combination of indices:

- ⟨verb, VN class⟩ associations are closest to reference
- concepts associated to VN classes cover large proportion of verbs

Best combination of concept selection indices

stability + separation

- $F_2 = 25.16$
- close to upper bound (no selection)
- coverage 98.04%

Final classification method

1) use FCA to build classes grouping French verbs and SCFs
2) select 1500 concepts where *stability* + *separation* is highest
3) align translated Verbnet classes with selected concepts
4) keep FCA concepts aligned with a translated Verbnet class
5) associate these FCA concepts with the Verbnet class thematic role sets

Effectively we obtain a classification associating:

- groups of French verbs
- groups of subcategorisation frames
- sets of thematic roles

Resulting classification: sample concept

Concept 5312 – verbs of movement

verbs: *bouger, déplacer, emporter, passer, promener, envoyer, expédier, jeter, porter, transmettre, transporter*

syntactic frames: SUJ:NP,OBJ:NP,POBJ:PP,POBJ:PP

thematic roles: AgExp (Agent or Experiencer), Theme, Start, End

Incremental Growing Neural Gas with Feature Maximisation, [Lamirel *et al.*, 2011b]

Growing neural gas clustering method

- based on Hebbian learning
- incremental
- *winning* clusters determined through distance function

IGNGF

- uses feature maximisation to determine *winning* cluster
- supports cluster labeling with distinguishing features

Incremental Growing Neural Gas with Feature Maximisation, [Lamirel *et al.*, 2011b]

- crisp, non-overlapping
- flat, non-hierarchical structure
- features can be weighted:
  weight of feature f for verb $v \longmapsto W_v^f$
- choose number of classes

Feature maximisation

Used for

- guiding the clustering
- cluster labeling
  i.e. associating relevant features to clusters

Feature f maximal for cluster c: $FF_c(f)$ higher for c than other cluster.

$FF_c(f)$ Feature F-measure for cluster c

verbs in c having f vs.
all verbs having f

$$FR_c(f) = \frac{\sum_{v \in c} W_v^f}{\sum_{c' \in C} \sum_{v \in c'} W_v^f}$$

(f, verb) combinations in c vs.
all (feature, verb) combinations in c

$$FP_c(f) = \frac{\sum_{v \in c} W_v^f}{\sum_{f' \in F_c, v \in c} W_v^{f'}}$$

IGNGF vs. FCA

Differences

- crisp, non-overlapping, no hierarchical structure
- features can be weighted (not only binary):

$$\text{weight of feature } f \text{ for verb } v \longmapsto W_v^f \in [0,1]$$

Analogy

[Lamirel, 2010]: A cluster c where for all maximal features f:

$$FP_c(f) = 1 \text{ and } FR_c(f) = 1$$

$\Longrightarrow$ c is formal concept:

- extent: verbs in c
- intent: maximal features for c

IGNGF classification method

Objects:
- verbs

Features:
- same as for FCA
- + grid (thematic role set) feature

- IGNGF produces verb clusters
- label clusters with
  - syntactic frames
  - thematic role sets

Associations with syntactic frames and semantic grids I

Syntactic frames

- Fmax: cluster maximising features
- Fpos: feature f-measure is above a global threshold

Thematic role sets

- θ features: feature f-measure is above a global threshold
- θ trans: assigned by alignment with translated classes

Best configuration

best performance in task based evaluation (simplified SRL)

- syntactic frames: feature f-measure above global threshold – Fpos
- thematic role sets: alignment with translated Verbnet classes – θ trans

Example IGNGF Cluster

C6- 14(14) [197(197)]
———-

Prevalent Label — = AgExp-Cause

0.341100 G-AgExp-Cause
0.274864 C-SUJ:Ssub,OBJ:NP
0.061313 C-SUJ:Ssub
0.042544 C-SUJ:NP,DEOBJ:Ssub
****
****
0.017787 C-SUJ:NP,DEOBJ:VPinf
0.008108 C-SUJ:VPinf,AOBJ:PP
. . .
[**déprimer 0.934345 4(0)] [affliger 0.879122 3(0)] [éblouir 0.879122 3(0)] [choquer 0.879122 3(0)] [décevoir 0.879122 3(0)] [décontenancer 0.879122 3(0)] [décontracter 0.879122 3(0)] [désillusionner 0.879122 3(0)] [**ennuyer 0.879122 3(0)] [fasciner 0.879122 3(0)] [**heurter 0.879122 3(0)] . . .

Outline

5 Evaluation and Comparison
- Evaluating Semantic Verb Classes wrt. Existing Reference
- Evaluating Syntactic-Semantic Verb Classes wrt. Corpus Annotations
- Summary

Evaluation

Goal: evaluate both FCA and IGNGF wrt.

- groups of verbs
- associations with syntactic frames – ⟨verb, frame⟩ pairs
- associations with thematic grids – ⟨verb, thematic role set ⟩ pairs
- associations with both syntactic frames and thematic grids – ⟨verb, syntactic frame, thematic role set⟩ triples

Other question:

- Which features work best for what classification technique?

Resources for evaluation

V-gold by [Sun *et al.*, 2010]

- groups ≈160 verbs in 16 Levin classes

VN class **role set**	**French translations in gold**
amalgamate-22.2 AgExp, PatientSym	incorporer; associer; réunir; mélanger; mêler; unir; assembler; combiner; lier; fusionner
amuse-31.1 Cause, AgExp	abattre; accabler; briser; déprimer; consterner; anéantir; épuiser; exténuer; écraser; ennuyer; éreinter; inonder
. . .	

Allows semantic evaluation:

- verb groups
- association with thematic role sets – ⟨verb, thematic role set⟩ pairs.

Evaluating verb groups – metrics

Modified Purity: How well can the clustering be embedded into gold?

Cluster C $\rightarrow \text{prev}(C) \in$ gold classification with maximal $|\text{prev}(C) \cap C|$

$$mPUR = \frac{\sum_{C \in \text{Clustering}, |\text{prev}(C)|>1} |\text{prev}(C) \cap C|}{\sum_{C \in \text{Gold}} \text{Verbs}_{\text{Clustering} \cap \text{C}}},$$

Weighted Class Accuracy: How well can the gold be embedded into the clustering?

gold class C $\rightarrow \text{dom}(C) \in$ clustering with maximal $|\text{dom}(C) \cap C|$

$$ACC = \frac{\sum_{C \in \text{Gold}} |\text{dom}(C) \cap C|}{\sum_{C \in \text{Gold}} \text{Verbs}_{\text{C}}}$$

Evaluating verb groups – the classifications

- Verbs
  - in Verbnet classes from V-gold translated to French
  - 2100 verbs
- Features
  - scf: subcategorisation frames
  - sem/synt: additional syntactic and/or semantic features
  - grid: translated classes a verb is a member of (IGNGF only)

Evaluating verb groups – results

Classifying 2100 verbs:

	Purity	Accuracy	F-measure
FCA	32.30	95.61	48.29
IGNGF	86.00	59.00	70.00
[Sun *et al.*, 2010], corpus based features, slightly different gold			55-65.4

Discussion

- IGNGF outperforms FCA wrt. F-measure
- IGNGF: better results than related work by [Sun *et al.*, 2010]
- IGNGF: higher purity, verb groupings more similar to gold
- FCA: higher accuracy, gold groups can be embedded in FCA groupings more easily.

Evaluating association with thematic role sets I

FCA and IGNGF

provide associations of clusters with thematic role sets.

Compare resulting ⟨verb, thematic role set⟩ pairs with those given by gold

using Recall (R), Precision (P) and their F-measure (F):

$$R = \frac{|\text{pairs in gold} \cap \text{pairs in classes}|}{|\text{pairs in gold}|}$$

$$P = \frac{|\text{pairs in gold} \cap \text{pairs in classes}|}{|\text{pairs in classes}|}$$

Evaluating association with thematic role sets II

Results

	Precision	Recall	F
FCA	24.09	75.00	36.47
IGNGF	27.16	26.67	27.16

Discussion

- FCA outperforms IGNGF wrt. ⟨verb, thematic role set⟩ associations.
- FCA better represents polysemy – overlapping classification

What are the best features?

FCA - ⟨verb, thematic role set⟩ evaluation

Features	cov.	prec	rec	f
scf & sem.	96.17	24.09	75.00	**36.47**
scf & synt. & sem.	96.05	23.95	75.00	36.31
scf (frames only)	95.37	23.48	73.80	35.63
scf & synt.	96.34	21.51	74.40	33.38

IGNGF - Evaluating groups of verbs

Features	mPUR	ACC	F
grid & scf & sem	86.00	59.00	**70.00**
grid & scf & sem & synt	99.00	52.00	69.00
grid & scf	94.00	54.00	68.00
scf & sem	83.00	55.00	66.00
scf	93.00	48.00	64.00
grid & scf & synt	87.00	50.00	63.00
scf & synt	91.00	45.00	61.00
scf & sem& synt	89.00	47.00	61.00

For both IGNGF and FCA

- semantic features improve classification
- syntactic features degrade classification

Possible reason for *syntactic feature* behaviour:

- information missing from lexicons

Evaluating syntactic-semantic verb classes

- Goal: evaluate associations
  - $\langle$verb, syntactic frame$\rangle$
  - $\langle$verb, syntactic frame, thematic role set$\rangle$
- V-gold does not provide associations with French syntactic frames
- Create SRL-gold reference
  providing $\langle$verb, syntactic frame, thematic role set$\rangle$ associations.
- Evaluate
  - recall for $\langle$verb, syntactic frame$\rangle$, $\langle$verb, thematic role set$\rangle$
  - task based: simplified Semantic Role Labeling

The SRL-gold reference

- sentences from Paris 7 Dependency Treebank [Candito *et al.*, 2009]
- annotate ⟨verb, syntactic argument⟩ instances with Verbnet thematic roles.

Sentences chosen as follows:

- for 116 verbs in V-gold and P7
- randomly choose upto 25 sentences containing verb

Results in:

- 1600 verb instances associated with thematic grid,
- 3605 ⟨verb, syntactic argument⟩ instances associated with thematic roles.

Larger classifications

Verbs

- all verbs in syntactic lexicon – 4200

Features/Attributes

- scf: subcategorisation frames
- sem: additional semantic features
- grid: derived from translated classes (IGNGF only)

Associations with frames and thematic role sets

SCFs (types)	SRL gold	SRL gold & classif	Recall
IGNGF	316	163	**59.59**
FCA	316	243	**88.69**

Grids (types)	SRL gold	SRL gold & classif	Recall
IGNGF	318	153	**48.11**
FCA	318	280	**88.05**

FCA better reflects associations with frames and grids

Linking

- How good are the induced ⟨verb, synt. arg., sem. role⟩ associations?
  - Adapt SRL method by [Swier and Stevenson, 2004]
- [Swier and Stevenson, 2004]:
  - Associate ⟨verb, syntactic argument⟩ instances in English corpus with Verbnet thematic roles
  - By aligning syntactic frames from corpus parses with Verbnet thematic grids
- Our adaptation:
  - Associate ⟨verb, syntactic argument⟩ instances in French P7 corpus with Verbnet thematic roles
  - By aligning syntactic frames from classification with Verbnet thematic grids

Semantic role labeling example

⟨voler, SUJ:NP,OBJ:NP,DEOBJ:PP⟩

FCA class role set	theta-grids for *voler*	syntactic construction			%θ	%SCF	Score
		SUJ:NP	OBJ:NP	DEOBJ:PP			
6583	Agent-Theme	Agent	Theme		100	67	167
Agent, Benef	**Agent-Theme-Start**	**Agent**	**Theme**	**Start**	100	100	**200**
Start, Theme	**Agent-Theme-Benef**	**Agent**	**Theme**	**Benef**	100	100	**200**
(*steal-10.5*)	Agent-Theme-Start-Benef	Agent	Theme	Start/Benef	75	100	175

FCA concept *6583*:

Verbs:	*acheter, assurer, attendre, . . . ,* **voler**
Thematic roles	Agent, Beneficiary, Start, Theme
Frames	SUJ:NP
	SUJ:NP,OBJ:NP
	SUJ:NP,OBJ:NP,AOBJ:PP
	SUJ:NP,OBJ:NP,DEOBJ:PP

Semantic role labeling example

⟨voler, SUJ:NP,OBJ:NP,DEOBJ:PP⟩

FCA class role set	theta-grids for *voler*	syntactic construction			%θ	%SCF	Score
		SUJ:NP	OBJ:NP	DEOBJ:PP			
6583	Agent-Theme	Agent	Theme		100	67	167
Agent, Benef	**Agent-Theme-Start**	**Agent**	**Theme**	**Start**	100	100	**200**
Start, Theme	**Agent-Theme-Benef**	**Agent**	**Theme**	**Benef**	100	100	**200**
(*steal-10.5*)	Agent-Theme-Start-Benef	Agent	Theme	Start/Benef	75	100	175

FCA concept *6583*:

Verbs:	*acheter, assurer, attendre, . . . ,* **voler**
Thematic roles	Agent, Beneficiary, Start, Theme
Frames	SUJ:NP
	SUJ:NP,OBJ:NP
	SUJ:NP,OBJ:NP,AOBJ:PP
	SUJ:NP,OBJ:NP,DEOBJ:PP

Semantic role labeling example

⟨voler, SUJ:NP,OBJ:NP,DEOBJ:PP⟩

FCA class role set	theta-grids for *voler*	syntactic construction			%θ	%SCF	Score
		SUJ:NP	OBJ:NP	DEOBJ:PP			
6583	Agent-Theme	Agent	Theme		100	67	167
Agent, Benef	**Agent-Theme-Start**	**Agent**	**Theme**	**Start**	100	100	**200**
Start, Theme	**Agent-Theme-Benef**	**Agent**	**Theme**	**Benef**	100	100	**200**
(*steal-10.5*)	Agent-Theme-Start-Benef	Agent	Theme	Start/Benef	75	100	175

Thematic role set **Agent, Beneficiary, Start, Theme**: English Verbnet class *steal-10.5*:

Verbs:	*abduct, annex, cabbage, capture,. . . , steal, . . .*	
Thematic roles	Agent, Beneficiary, Start, Theme	
Frames	SUJ:NP,OBJ:NP	**Agent V Theme**
	SUJ:NP,OBJ:NP,P-OBJ:PP	**Agent V Theme Start**
		Agent V Theme Benef
	SUJ:NP,OBJ:NP,P-OBJ:PP,P-OBJ:PP	**Agent V Theme Start Benef**

Semantic role labeling example

⟨voler, SUJ:NP,OBJ:NP,DEOBJ:PP⟩

FCA class role set	theta-grids for *voler*	syntactic construction			$\%\theta$	%SCF	Score
		SUJ:NP	OBJ:NP	DEOBJ:PP			
6583	Agent-Theme	Agent	Theme		100	67	167
Agent, Benef	**Agent-Theme-Start**	**Agent**	**Theme**	**Start**	100	100	**200**
Start, Theme	**Agent-Theme-Benef**	**Agent**	**Theme**	**Benef**	100	100	**200**
(*steal-10.5*)	Agent-Theme-Start-Benef	Agent	Theme	Start/Benef	75	100	175

	SUJ:NP	Agent
voler	OBJ:NP	Theme
	DEOBJ:PP	Beneficiary, Start

resulting labeling: non-ambiguous associations

- ⟨voler, SUJ:NP⟩ $\rightarrow$ Agent
- ⟨voler, OBJ:NP⟩ $\rightarrow$ Theme
- ⟨voler, DEOBJ:PP⟩ no label

Results

Comparison with SRL gold:

	%total (R)	%labeled (P)	F	%not labeled
baseline (default associations)	65.21	65.21	65.21	0.00
FCA	30.87	70.40	42.92	56.14
IGNGF	47.43	71.91	**57.39**	34.79
S&S (English, baseline 74.00)			76.00	38.00

- IGNGF outperforms FCA
- IGNGF & FCA lower than baseline
- precision better than baseline

Evaluation Summary

	Reference	FCA	IGNGF	Related work
verb groups	V-gold (PUR/ACC F)	48.29	**70.00**	Sun et al. 55-65
⟨verb, thematic role set⟩	V-gold (F)	**36.47**	27.16	
	SRL-gold (R)	**88.05**	48.11	
⟨verb, scf⟩	SRL-gold (R)	**88.69**	59.59	
⟨verb, synt. arg, θ role⟩	SRL-gold (F)	42.92	**57.39**	S&S 76

semantic and *syntactic* features

- similar effect on FCA and IGNGF classification

Major issues

Associations with syntactic frames:

- FCA: too general → classes associated to high frequency frames
- IGNGF: too specific → classes associated to low frequency frames

Associations with thematic role sets:

- Large heterogeneous classes aligned
  to small, very specific Verbnet classes
- How to better align translated classes with clusters/concepts?

Outline

Conclusion

Large scale syntactic-semantic classification of French verbs

- based on existing French and English lexical resources
- using the FCA and IGNGF clustering methods

Classification methods

- useful verb classes associated with syntactic frames and thematic role sets
- complementary
  - FCA: better associations with frames and thematic role sets
  - IGNGF: better support in SRL task.
- main shortcoming: association with syntactic frames

- lexicon: `http://talc.loria.fr/tl_dv2_ladl-a-subcategorisation.html`
- classifications: `http://talc.loria.fr/-Classifications-.html`

Future Work

Improve classifications

- Better associations with syntactic frames:
  - FCA
    - attribute (scf) based selection indices
    - exploit hierarchical structure
  - IGNGF
    - cluster labeling depending on individual frames
    - towards creating overlapping classifications
- Better associations with thematic grids:
  - better methods of aligning clusters and translated Verbnet classes
  - explore other methods of associating verbs/frames with thematic role sets.
- Better evaluation method:
  - How significant is comparison with $< 10\%$ reference data?
  - Use unsupervised evaluation measures (eg. cumulated micro precision [Lamirel *et al.*, 2011a]).

Future Work

Polysemy

- How to adequately represent it?
- How to evaluate?

Explore fully unsupervised approach

- using distributional data – eg. LexSchem

Publications

Ingrid Falk, Claire Gardent, and Jean-Charles Lamirel.
Classifying French Verbs Using French and English Lexical Resources.
In *Proceedings of the 50th annual meeting of the ACL*, July 2012.

Ingrid Falk and Claire Gardent.
Combining Formal Concept Analysis and Translation to Assign Frames and Thematic Grids to French Verbs.
In *Concept Lattices and their Applications*, October 2011.

Ingrid Falk and Claire Gardent.
Bootstrapping a Classification of French Verbs Using Formal Concept Analysis.
In *Interdisciplinary Workshop on Verbs*, November 2010.

Ingrid Falk, Claire Gardent, and Alejandra Lorenzo.
Using Formal Concept Analysis to Acquire Knowledge about Verbs.
In *Concept Lattices and Their Applications*, October 2010.

Associations with frames and thematic role sets (more detailed)

⟨verb, frame⟩ pairs in corpus: recall 59.59 for IGNGF, 88.69 for FCA.

SCFs (types)	SRL gold	classif	SRL gold & classif	SRL gold & lex ¬ classif	SRL gold ¬ lex	Recall	Recall w/o missing in lex
IGNGF	316	1149	163	111	42	51.58	**59.59**
FCA	316	16542	243	31	42	76.90	**88.69**

⟨verb, thematic grid⟩ pairs in corpus: recall 48.11 for IGNGF, 88.05 for FCA.

Grids	gold	gold & classif	R
IGNGF	318	153	**48.11**
FCA	318	280	**88.05**

FCA better reflects associations with frames and grids in SRL gold.

IGNGF vs. FCA

Differences

- crisp, non-overlapping, no hierarchical structure
- features can be weighted (not only binary):

$$\text{weight of feature } f \text{ for verb } v \longmapsto W_v^f \in [0,1]$$

Analogy

[Lamirel, 2010]: A cluster c where for all maximal features f:

$$FP_c(f) = 1 \text{ and } FR_c(f) = 1$$

$\Longrightarrow$ c is formal concept:

- extent: verbs in c
- intent: maximal features for c

M.-H. Candito, B. Crabbé, and M. Falco.
Dépendances syntaxiques de surface pour le français.
Technical report, Université de Paris 7, 2009.

Bernhard Ganter and Rudolph Wille.
Formal concept analysis: Mathematical foundations.
Springer, Berlin-Heidelberg, 1999.

Mikhail Klimushkin, Sergei Obiedkov, and Camille Roth.
Approaches to the selection of relevant concepts in the case of noisy data.
In Léonard Kwuida and Baris Sertkaya, editors, *Formal Concept Analysis*, volume 5986 of *Lecture Notes in Computer Science*, chapter 18, pages 255–266. Springer Berlin / Heidelberg, Berlin, Heidelberg, 2010.

J. C. Lamirel, P. Cuxac, and R. Mall.

A new efficient and unbiased approach for clustering quality evaluation.
In *QIMIE'11, PaKDD*, Shenzen, China, 2011.

J.-C. Lamirel, R. Mall, P. Cuxac, and G. Safi.
Variations to incremental growing neural gas algorithm based on label maximization.
In *Neural Networks (IJCNN), The 2011 International Joint Conference on*, pages 956 –965, 2011.

Jean-Charles Lamirel.
A new multi-viewpoint and multi-level clustering paradigm for efficient data mining tasks.
In Kimito Funatsu, editor, *New Fundamental Technologies in Data Mining*, INTECH E-Book Series, pages chapitre 15, pp. 283–304. INTECH Open Access Publisher, 2010.

Karin Kipper Schuler.

VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon.
PhD thesis, University of Pennsylvania, 2006.

Lin Sun, Anna Korhonen, Thierry Poibeau, and Cédric Messiant.
Investigating the cross-linguistic potential of VerbNet-style classification.
In *Proceedings of the 23rd International Conference on Computational Linguistics*, COLING '10, pages 1056–1064, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics.

Robert S. Swier and Suzanne Stevenson.
Unsupervised semantic role labellin.
In *EMNLP*, pages 95–102, 2004.