

# Bootstrapping a Classification of French Verbs Using Formal Concept Analysis.

Ingrid Falk<sup>124</sup>   Claire Gardent<sup>34</sup>

<sup>1</sup>INRIA Nancy Grand Est

<sup>2</sup>Lorraine University, Nancy, France

<sup>3</sup>CNRS

<sup>4</sup>LORIA

Interdisciplinary Workshop on Verbs, November 4-5, 2010

# Summary

- 1 Overview
- 2 Lexical Resources
  - Dicovalence
  - The LADL Tables
  - VerbNet
- 3 Formal Concept Analysis
- 4 Building the Classification.
  - Concept Stability
- 5 Evaluation
  - Quantitative Evaluation
  - Qualitative Evaluation
- 6 Conclusion and future work.
- 7 References

Need French verb classes for semantic role annotation of a treebank.

⇒ We need classes with

- ▶ syntactic information,
- ▶ thematic role information.

This work:

We create a **syntactic-semantic classification of French verbs** ...

- ▶ starting from existing lexical resources:
  - ▶ a **French valency lexicon** and
  - ▶ a **coarse grained syntactic semantic classification** of French verbs
- ▶ and using Formal Concept Analysis.

We assess the adequacy of the obtained classification

- ▶ by a quantitative comparison with VerbNet,
- ▶ by a more detailed qualitative analysis of a sub-classification wrt. the thematic grid assignment.

- 1 Overview
- 2 Lexical Resources
  - Dicovalence
  - The LADL Tables
  - VerbNet
- 3 Formal Concept Analysis
- 4 Building the Classification.
  - Concept Stability
- 5 Evaluation
  - Quantitative Evaluation
  - Qualitative Evaluation
- 6 Conclusion and future work.
- 7 References

# Summary

- 1 Overview
- 2 Lexical Resources**
  - Dicovalence
  - The LADL Tables
  - VerbNet
- 3 Formal Concept Analysis
- 4 Building the Classification.
  - Concept Stability
- 5 Evaluation
  - Quantitative Evaluation
  - Qualitative Evaluation
- 6 Conclusion and future work.
- 7 References

# Dicovalence [van den Eynde and Mertens, 2003]

- ▶ valency lexicon of French verbs,
- ▶ 3936 verbs,
- ▶ created manually, based on “Pronominal Approach” [van den Eynde, Blanche-Benveniste et al. 1978-1989]

## Example entry

VAL\$	manifester: P0 P1 (P2)
VERB\$	MANIFESTER/manifester
EG\$	il ne manifeste jamais ses vrais sentiments <i>He never shows his true feelings (to sb.)</i>
TR_EN\$	show, display
P0\$	qui, je, nous, elle, il, ils, on, celui-ci, ceux-ci
P1\$	que, la, le, les, ça, celui-ci, ceux-ci
P2\$	0, qui, lui, leur, se réc., celui-ci, ceux-ci, l'un l'autre
RP\$	passif être, ?se passif
AUX\$	avoir

↪ *manifester* SUJ:NP, OBJ:NP, (A-OBJ:PP)

# The LADL Tables [Gross, 1975]

- ▶ for French,
- ▶ group verbs in tables according to syntactic and/or semantic criteria,
- ▶ developed manually over several years by a large team of linguists,
- ▶ too coarse grained for NLP.

## Example

Table 4: verbs where the object must be human and the subject is not restricted,

Table 8: verbs which allow for a sentential complement with *de*,

Table 32RA: transitive verbs which may be interpreted as *Make Adj<sub>v</sub>* where the adjective is morphologically related to the verb – eg. “*verdir*” (*to turn green*) is morphologically related to the adjective “*vert*” (*green*).

## VerbNet [Schuler, 2006] I

- ▶ for English,
- ▶ extension and refinement of Levin's classification [Levin, 1993],
- ▶ created manually,
- ▶ large coverage:  $\sim 5200$  verb senses and  $\sim 400$  classes,
- ▶ classes are defined by
  - ▶ a set of verb members,
  - ▶ a set of syntax frames,
  - ▶ a set of thematic roles.
- ▶ classes are
  - ▶ semantically and syntactically coherent,
  - ▶ organised in a hierarchy.

⇒ **VerbNet** is a model for the classification we would like to achieve for French.



# VerbNet [Schuler, 2006] II

## Example

### amuse-31.1

verbs (242): abash, affect, afflict, amuse, annoy, ...

$\theta$ -grid: EXPERIENCER [+animate]  
CAUSE

frames (6):	NP V NP	EXPERIENCER V CAUSE
	NP V ADV-Middle	EXPERIENCER V Adv
	NP V NP-PRO-ARB	CAUSE V
	NP V NP PP.oblique	CAUSE V EXPERIENCER <i>with</i> OBLIQUE
	NP.cause V NP	CAUSE('s) OBLIQUE V EXPERIENCER
	NP V NP ADJ	CAUSE V EXPERIENCER Adj

# Summary

- 1 Overview
- 2 Lexical Resources
  - Dicovalence
  - The LADL Tables
  - VerbNet
- 3 Formal Concept Analysis**
- 4 Building the Classification.
  - Concept Stability
- 5 Evaluation
  - Quantitative Evaluation
  - Qualitative Evaluation
- 6 Conclusion and future work.
- 7 References

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

- ▶ method for deriving conceptual structures – **concepts** out of data,
- ▶ a concept is defined by a **set of objects**  $\mathcal{O}$  and a **set of properties**  $\mathcal{P}$  shared by every object in  $\mathcal{O}$ ,
- ▶ FCA organises concepts into a **hierarchy** – **concept lattice**.

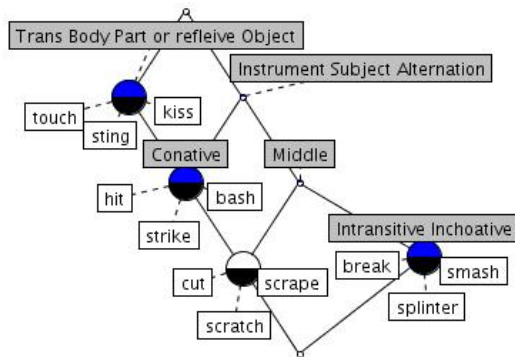
# FCA Example – the Formal Context

Given a **formal context** ...

	Instrument Subject	Middle	Trans. Body-part or refl.	Intransitive Inchoative	Conative
break	x	x		x	
splinter	x	x		x	
smash	x	x		x	
cut	x	x	x		x
scratch	x	x	x		x
scrape	x	x	x		x
hit	x		x		x
strike	x		x		x
bash	x		x		x
touch			x		
sting			x		
kiss			x		

# FCA Example – Formal Concepts and Concept Lattice

FCA produces **formal concepts** ordered in a **concept lattice**:



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

- ▶ method for deriving conceptual structures – **concepts** out of data,
- ▶ a concept is defined by a **set of objects**  $\mathcal{O}$  and a **set of properties**  $\mathcal{P}$  shared by every object in  $\mathcal{O}$ ,
- ▶ FCA organises concepts into a **hierarchy** – **concept lattice**.

VerbNet classification and FCA concept lattices are similar

- ▶ **VerbNet** classes are defined by **verb members** and **syntax frames**.
- ▶ **FCA** concepts are defined by subsets of **objects** and **properties**.
- ▶ sub/superclass relation and sub/superconcept relation are defined similarly.

⇒ **FCA naturally models this classification task.**

# Summary

- 1 Overview
- 2 Lexical Resources
  - Dicovalence
  - The LADL Tables
  - VerbNet
- 3 Formal Concept Analysis
- 4 Building the Classification.**
  - Concept Stability
- 5 Evaluation
  - Quantitative Evaluation
  - Qualitative Evaluation
- 6 Conclusion and future work.
- 7 References

# The Method

- ▶ build a formal context from Dicovalence and LADL lexical resources,
- ▶ compute the concept lattice using **Galicia software**<sup>1</sup>,
- ▶ filter using concept stability – [Kuznetsov, 2007],

---

<sup>1</sup><http://www.iro.umontreal.ca/galicia/>



# The Formal Context

**Objects:** verbs that are the intersection of Dicovallence and LADL,

- Attributes:**
- ▶ frames assigned to verb by Dicovallence (**syntactic features**),
  - ▶ LADL table identifier if verb is in that LADL table (**semantic features**).

## Example

	frames		table ids		
	SUJ:NP, OBJ:NP, A-OBJ:NP	SUJ:NP, OBJ:NP, DE-OBJ:NP	31H	5	4
<i>manifeste</i>	X		X	X	

⇒ Context of:

- ▶ 3536 objects (verbs) and
- ▶ 186 attributes (frames and table identifiers).

# The Concept Lattice

## Filtering

The obtained concept lattice (classification) has **36065** concepts (classes)  
Most are not interesting:

- ▶ only 1 or 2 verbs,
- ▶ too few attributes (frames, tables).

⇒ only keep concepts with:

- ▶ at least 1 frame,
- ▶ at least 1 table identifier,
- ▶ high **concept stability**.

## Concept stability

- ▶ a measure from FCA,
- ▶ helps discriminating interesting patterns from irrelevant information.

### Definition ([Kuznetsov, 2007])

Let  $(V, F)$  be a formal concept and  $'$  the derivation operator. It's **intensional stability** is defined as:

$$\sigma_i((V, F)) := \frac{|\{A \subseteq V \mid A' = F\}|}{2^{|V|}}$$

- ▶ the proportion of the subsets of the verb set ( $V$ ) which have the same attributes ( $F$ ).
- ▶ a more stable concept is less dependant on individual members in the object (verb) set.

**Problem:** keeping only the most stable concepts affects coverage.

# Resulting Classification I

- ▶ 430 classes with highest stability ( $> 0.9995$ ), at least 1 frame and 1 table identifier,
- ▶ coverage of
  - ▶ 98.41% of verbs,
  - ▶ 25% of frames,
  - ▶ 83.17% of verb, frame pairs,
- ▶ 18 – 498 verbs,
- ▶ 1 – 7 valency frames.

# Resulting Classification II

## Example

Verbs blanchir (to whiten), bleuir (to turn blue), blêmir (to turn pale), pâlir (to turn white), rajeunir (to become younger), rosir (to turn pink), rougir (to blush), verdir (to turn green), viellir (to age).

Tables                    32RA    Make *Adj<sub>v</sub>*  
                                       8 verbs with sentential complement in *de*.

Frames    SUJ:NP; SUJ:NP,OBJ:NP

Class groups together verbs which

- ▶ indicate a change of state (colour or age) – table 32RA,
- ▶ can be used with or without direct object – frames,
- ▶ may be used with a sentential *de*-object – table 8.

# Summary

- 1 Overview
- 2 Lexical Resources
  - Dicovalence
  - The LADL Tables
  - VerbNet
- 3 Formal Concept Analysis
- 4 Building the Classification.
  - Concept Stability
- 5 Evaluation**
  - Quantitative Evaluation
  - Qualitative Evaluation
- 6 Conclusion and future work.
- 7 References

# Evaluation

**Quantitative:** by comparing with VerbNet,

**Qualitative:** do the classes group verbs with identical thematic grids?

- ▶ case of psychological verbs (LADL Table 4).

# Quantitative Comparison with VerbNet

	<i>Class430</i>	VerbNet
Nb. of classes	430	411
Max. verbs	498	383
Max. frames	7	25
Average class size (verbs/frames/harm. mean)	70.13/2.71/7.02	14.96/4.02/4.67
Verbs covered (%)	98.41% (of 3536)	3626
Total number of frames	116	117

- ▶ numbers suggest good generalisation and factorisation power,
- ▶ differences in frame inventories:

**VerbNet and Dicovalence**  $\sim$  number of frames,

**Dicovalence:** high number of low frequency frames  $\Rightarrow$  only 29 frames in classification,

**VerbNet** frames are finer grained: NP and PP specialised with thematic roles, differentiated sentential arguments.

$\Rightarrow$  VerbNet classes have more frames and less verbs (avg. and max.).



# Do the classes group together verbs with identical thematic grid?

## Thematic Grids and Linking for Psych Verbs

⇒ Case study on **LADL table 4**

► Show graph

- 616 emotional (psychological) verbs,
- transitive construction,
- object is human, subject not restricted
- roles for emotion verbs: EXPERIENCER and CAUSE
- subject may be phrasal ⇒ EXPERIENCER is object (not subject)
- subject accepts both CAUSE (non-agentive) and AGENT (volitional) reading

⇒

Thematic grid for sub-classification is:

[ (CAUSE or AGENT), EXPERIENCER ]

### Next 2 slides:

Classes obtained by FCA approach help analyse phenomena of

- ▶ polysemy,
- ▶ linking (mapping between syntactic arguments and thematic roles)

for the sub-classification of Table 4 verbs.

## Table 4: Polysemy

[► Show graph](#)

≡ classes with additional table identifier in attribute set

### Example

Class 4562 has additional table 32C (transitive verbs with concrete object)

⇒ verbs have two readings:

1. psychological reading (table 4): *toucher le public*  
*touch (move) the audience*
2. concrete object reading (table 32C): *toucher le mur*  
*touch the wall*

⇒ 2 thematic grids and linkings:

Table 32C	Table 4
NP.AGENT NP.PATIENT	NP.CAUSE/AGT NP.EXP.

## Table 4: Linking I

Mapping between syntactic arguments and thematic roles

► Show graph

Determine linking for classes:

semantically homogeneous: Table 4 only

syntactically varied: at least 3 frames

## Table 4: Linking II

Mapping between syntactic arguments and thematic roles

⇒ 20 classes ⇒ linking worked well for almost all

Example: Class 14650 (28 verbs)

SUJ:NP,OBJ:NP

NP.CAUSE OR AGENT NP.EXPERIENCER

*Jean irrite Marie/Jean irritates Mary*

SUJ:NP

NP.CAUSE OR AGENT

*Jean irrite/Jean irritates.*

SUJ:NP,DUMMY:REFL

NP.EXPERIENCER, reflexive

*Marie s'irrite/Marie irritates herself*

SUJ:NP,DUMMY:REFL,POBJ:PP

NP.EXPERIENCER, PP.CAUSE OR AGENT, reflexive

*Marie s'irrite contre Jean/  
Marie irritates herself against Jean.*

## Table 4: Linking III

Mapping between syntactic arguments and thematic roles

Linking is problematic: Class 25647 (37 verbs)

SUJ:NP,OBJ:NP	NP.CAUSE NP.EXPERIENCER <i>Ceci exaspère/anime Marie.</i> <i>This exasperates/animates Marie.</i>
SUJ:NP,DUMMY:REFL	NP.EXPERIENCER, reflexive <i>Marie s'exaspère/s'anime.</i> <i>Marie exasperates/animates herself.</i>
SUJ:NP,OBJ:NP,POBJ:PP	NP.CAUSE/AGENT, NP1, P-OBJ:PP <i>Elle exaspère [en moi]<sub>P-OBJ:EXPERIENCER</sub> ce désir.</i> <i>She exasperates [in me]<sub>P-OBJ:EXPERIENCER</sub> this desire.</i> <i>Paul anime [les jeunes]<sub>NP1:EXPERIENCER</sub> contre moi.</i> <i>Paul animates [the young]<sub>NP1:EXPERIENCER</sub> against me.</i>

⇒ Need more fine-grained representation of prepositional objects!

# Summary

- 1 Overview
- 2 Lexical Resources
  - Dicovalence
  - The LADL Tables
  - VerbNet
- 3 Formal Concept Analysis
- 4 Building the Classification.
  - Concept Stability
- 5 Evaluation
  - Quantitative Evaluation
  - Qualitative Evaluation
- 6 Conclusion and future work.**
- 7 References

# Conclusion

Results suggest that FCA is an appropriate framework for bootstrapping a verb classification for French from existing lexical resources:

- ▶ concepts obtained by FCA naturally model the association btw. objects (verbs) and attributes (syntactic and/or semantic features),
- ▶ FCA permits soft clustering – a data element may belong to several classes,
- ▶ stable concepts have good generalisation and factorisation power and linguistically sound empirical content,
- ▶ (preliminary) evaluation shows classes adequately describe association btw. verb sets, syntactic frames and thematic grids.



## Future Work

- ▶ Improve the classification
  - ▶ use more and more detailed syntactic and semantic features,
  - ▶ model and make use of relations and dependencies between attributes.
- ▶ Use classification to assign  $\theta$ -grids for semantic role labeling.

Thank you!

# Summary

- 1 Overview
- 2 Lexical Resources
  - Dicovalence
  - The LADL Tables
  - VerbNet
- 3 Formal Concept Analysis
- 4 Building the Classification.
  - Concept Stability
- 5 Evaluation
  - Quantitative Evaluation
  - Qualitative Evaluation
- 6 Conclusion and future work.
- 7 **References**

# Lexical Resources. I



Baker, C. F., Fillmore, C. J., and Lowe, J. B. (1998).

The berkeley FrameNet project.

International Conference on Computational Linguistics, Montreal, Quebec, Canada.



Fellbaum, C., editor (1998).

*WordNet: An Electronic Lexical Database.*

MIT Press, Cambridge, MA.



Gross, M. (1975).

*Méthodes en syntaxe.*

Hermann, Paris.

## Lexical Resources. II



Levin, B. (1993).

*English Verb Classes and Alternations: a preliminary investigation.*  
University of Chicago Press, Chicago and London.



Saint-Dizier, P. (1999).

Alternation and verb semantic classes for french: Analysis and class formation.

In *Predicative forms in natural language and in lexical knowledge bases*. Kluwer Academic Publishers.



Schuler, K. K. (2006).

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

## Lexical Resources. III



van den Eynde, K. and Mertens, P. (2003).

La valence: l'approche pronominale et son application au lexique verbal.

*Journal of French Language Studies*

## FCA



Bernhard Ganter and Rudolph Wille.

1999.

*Formal concept analysis: Mathematical foundations.*

Springer, Berlin-Heidelberg.



Kuznetsov, S. O. (2007).

On stability of a formal concept.

*Annals of Mathematics and Artificial Intelligence*, 49(1-4):101–115.

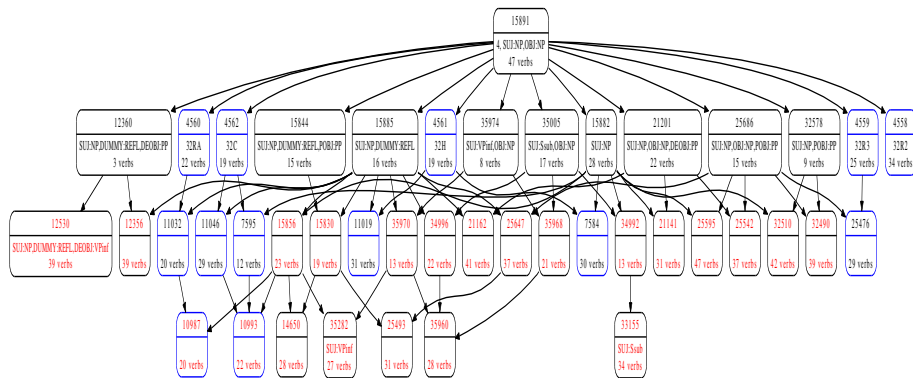


Figure: Blue contours: classes with more table identifiers. Red font: classes with more than 3 frames.