Combining Formal Concept Analysis and Translation to Assign Frames and Semantic Role Sets to French Verbs

Ingrid Falk, Claire Gardent

▶ **To cite this version:**

Ingrid Falk, Claire Gardent. Combining Formal Concept Analysis and Translation to Assign Frames and Semantic Role Sets to French Verbs. Annals of Mathematics and Artificial Intelligence, 2013, 10.1007/s10472-013-9377-3 . hal-00852861

HAL Id: hal-00852861

https://inria.hal.science/hal-00852861

Submitted on 21 Aug 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Noname manuscript No.
(will be inserted by the editor)

Combining Formal Concept Analysis and Translation to Assign Frames and Semantic Role Sets to French Verbs

Ingrid Falk · Claire Gardent

Received: date / Accepted: date

Abstract In Natural Language Processing, verb classifications have been shown to be useful both theoretically (to capture syntactic and semantic generalisations about verbs) and practically (to support factorisation and the supervised learning of shallow semantic parsers). Acquiring such classifications manually is both costly and errror prone however. In this paper, we present a novel approach for automatically acquiring verb classifications. The approach uses FCA to build a concept lattice from existing linguistic resources; and stability and separation indices to extract from this lattice those concepts that most closely capture verb classes. The approach is evaluated on an established benchmark and shown to differ from previous approaches and in particular, from clustering approaches, in two main ways. First, it supports polysemy (because a verb may belong to several classes). Second, it naturally provides a syntactic and semantic characterisation of the verb classes produced (by creating concepts which systematically associate verbs with their syntactic and semantic attributes).

Keywords Natural Language Processing · verb classification · concept selection indexes

Contents

1 Introduction 2
2 Problem Setting and General Overview of the FCA Approach to the Automatic Acquistion of Verb Classes 4

Ingrid Falk
LORIA UMR 7503, Campus Scientifique, F-54506 Vandoeuvre-lès-Nancy Cedex
E-mail: ifalk@unistra.fr

Claire Gardent
CNRS/LORIA UMR 7503 Campus Scientifique, F-54506 Vandoeuvre-lès-Nancy Cedex
Tel.: +33-38359-2039
E-mail: claire.gardent@loria.fr

3 Related Work 6
4 Formal Concept Analysis 9
5 Associating French Verbs with Syntactic Frames 10
6 Associating French Verbs with Semantic Grids. 17
7 Results and Discussion 19
8 Conclusion 27

1 Introduction

Natural language processing (NLP) aims to interpret and generate natural language. In particular, shallow semantic parsing (also referred to as Semantic Role Labelling or SRL) aims to detect the events described by a text together with their participants. For instance, given the sentence, *John eats bread*, a semantic role labeller will detect that an eating event takes place whose "agent" is John and whose "patient" is some bread.

"Agent" and "patient" are relations describing how the entities denoted by the syntactic arguments of a verb participate in the event. In the linguistic literature, these relations are referred to as semantic roles and the entities related by a semantic role to the event are said to be the semantic arguments of the verb denoting that event.

The syntactic arguments of the verb are constituents (e.g., noun phrases, sentences or prepositional phrases) which are related to the verb by a syntactic function capturing the distributional and morphosyntactic properties of each argument. For instance, the subject of a verb differs from other syntactic arguments in that its person and number must agree with those of the verb; and the object is defined as the argument of a verb which becomes its subject when the verb is used in the passive voice. Syntactic functions commonly used in NLP are subject, object, indirect object (cf. example 1f) and prepositional object i.e., an object introduced by a preposition (e.g., *with, on, at, in, above*).

As illustrated by the examples below, identifying the participants of an event and their semantic role requires both syntactic and semantic knowledge about a verb's arguments. Thus although the subject is often the agent of the event (examples 1a-e), some verbs accept a patient as a subject (cf. example 1d). Similarly, the object of a verb is often a patient (examples 1a,c) but not always (examples 1e,f). Finally, as shown in examples (1c,d), for the same verb and for the same syntactic function, the semantic role may vary.

(1) a. [John (Subject,Agent)] eats [the bread (Object,Patient)]
b. [John (Subject,Agent)] eats
c. [John (Subject,Agent)] broke [the window (Object,Patient)]
d. [The window (Subject,Patient)] broke
e. [John (Subject,Agent)] throws [the ball (Object,Theme)]
f. [John (Subject,Agent)] throws [Mary (Indirect Object,Destination)] [the ball (Object,Theme)]

To support the development of semantic role labellers and in particular, to annotate the corpora necessary for training such systems, verb classifications

need to be defined which specify for each verb in a given language, its syntactic frame i.e., the function and the syntactic type (Noun Phrase, Prepositional Phrase, Sentential argument) of its arguments; and its semantic grid i.e., the semantic roles contributed by each of its syntactic arguments to the event denoted by the verb.

For English several such resources exist which capture different intuitions about semantic roles. The Propbank corpus [26] relies on a classification where the syntactic arguments of a verb are labeled as numbered arguments (Arg0, Arg1, ... , Arg5) and where verb modifiers are assigned functional labels such as manner (MNR), locative (LOC), temporal (TMP) etc. Framenet [1] is centered around the more abstract notion of frames which generalises descriptions accross similar verbs (e.g., "describe" and "characterise") as well as nouns and other words (e.g., "description"). As a result, the semantic roles used by Framenet are more specific describing the participation of the verb arguments to the frame denoted by that verb. For instance, in Framenet, a cooking event involves up to 4 participants namely, a Cook (the person doing the cooking), the food that is to be cooked (Food), something to hold the food while cooking (Container) and a source of heat (Heating_instrument). Finally, Verbnet [16] takes a middle path between both approaches and associates verbs with "linguistic" semantic roles i.e., roles such as "Agent" and "Patient" which have been identified in the linguistic literature as universal across languages [11].

As we shall see in Section 3, fewer work has been done on producing verb classifications for French. Moreover, existing verb classifications are either too restricted in scope (Volem, [30]) or not sufficiently structured (the LADL tables, [14]) to be directly useful for NLP.

In this paper we show how Formal Concept Analysis (FCA) can be used to automatically acquire a Verbnet style classification of French verbs based on information contained in existing linguistic resources. Using verbs as objects and verb properties extracted from these linguistic resources as attributes, we build a formal context and the corresponding concept lattice. To derive a verb classification from this lattice, we then explore methods for selecting concepts which were proposed in the literature to identify those concepts in the lattice that are most relevant to a given application. We show in particular that the stability and separation indices proposed in [20] and [17] are relevant for identifying those concepts that most closely capture verb classes. The results produced are compared against the gold standard independently developed by [35] and both a qualitative and a quantitative evaluation of the results obtained is provided.

The paper is structured as follows. Section 3 reviews existing verb classifications for French as well as work on automatically acquiring verb classifications and previous applications of FCA to natural language. Following a brief introduction to Formal Concept Analysis (FCA) in Section 4, Section 2 gives an overview of how we use FCA to derive a verb classification for French similar to the English Verbnet classification. The following sections present the details of the approach with Section 5 showing how FCA is used to associate verb sets with sets of syntactic frames and Section 6 presenting the approach

used to enrich these syntactic classes with semantic role information. Section 7 describes the results obtained and Section 8 concludes.

2 Problem Setting and General Overview of the FCA Approach to the Automatic Acquistion of Verb Classes

Our goal is to automatically create a Verbnet style classification of French verbs, where groups of verbs are associated with both their syntactic frames and their semantic grids. In this section, we start by summarising the linguistic terminology used throughout the article. We then briefly describe the Verbnet classification. Finally we sketch the overall procedure we developed for constructing a Verbnet-like classification of French verbs using FCA and a set of available linguistic resources.

2.1 Linguistic Terminology

In what follows, we list the main linguistic concepts used in this paper. For each concept, the first line (**D**) gives an informal definition of that concept while the second (**X**) shows some illustrating examples.

Syntactic Category
D Label defining the distributional behaviour of a word or a constituent.
X "the" is a determiner (Det); "cat" is a noun (N); "the cat" is a noun phrase (NP).

Constituent
D Word or Group of words which linguistically functions as a unit.
X "sleeps", "has slept", "eats an apple" are all Verb Phrase (VP) constituents; "the cat", "a small cat", "John" are Noun Phrase (NP) constituents; "with a cat", "near John" are Prepositional Phrase (PP) constituents.

Syntactic Function
D Captures the distributional behaviour of a word or a constituent with respect to other words or constituents.
X In the sentence "the man put an apple on the table", "the man" is the Subject of "put"; "an apple" is its Object; and "on the table" is its Prepositional Object. Subject (Subj), Object (Obj) and Prepositional Object (PObj) are syntactic functions.

Syntactic Functor
D A word or constituent which requires the presence of other words or constituents for the sentence to be grammatical.
X Verbs are syntactic functors because they require a subject and in some cases, an object and/or a prepositional object.

Syntactic Argument
D A word or constituent which satisfies the requirement of a syntactic functor.
X In the sentence "the man eats an apple", "the man" and "an apple" are syntactic arguments of "eats".

Semantic Role
D Describes the semantic relation between a verb and its argument.
X In the sentence "the man eats an apple", "the man" is the AGENT of "eats" and "an apple" is its PATIENT. AGENT and PATIENT are semantic roles.

Syntactic frame
D The set of arguments required by a verb, each argument being specified by its syntactic category and its syntactic function.
X The syntactic frame of "sleep" is { NP:Subj }. The syntactic frame of "put" is { NP:Subj, NP:Obj, PP:PObj }.

Semantic Grid
D The set of semantic roles associated with the arguments of a syntactic functor.
X The semantic grid of "eat" is { AGENT, PATIENT }.

Syntactic Lexicon
D A lexicon associating each verb with its syntactic frame and the set of required arguments, each argument being specified by its syntactic category and its syntactic function.

2.2 Verbnet

In Verbnet, a *syntactic frame* characterises the number and the semantic role of the syntactic arguments expected by a verb. Further, the *semantic grid* is taken to be a set of semantic roles describing how the entities denoted by the syntactic arguments of a verb participate in the event described by that verb.

Figure 1 shows an excerpt of the *throw-17.1*[1] Verbnet class, with its verbs, semantic roles and syntactic frames.

[1]The naming conventions for Verbnet classes refer partly to the class names used in Levin's manual classification of English verbs and partly to the hierarchy structure. Verb classes are numbered according to shared semantics and syntax, and classes which share a top-level number (9-109) have corresponding semantic relationships. For instance, verb classes related to putting, such as put-9.1, put_spatial-9.2, funnel-9.3, etc. are all assigned to the class number 9 and related to moving an entity to a location. Classes that share a top class can also be divided into subclasses, such as wipe verbs in wipe_manner (10.4.1) and wipe_inst (10.4.2) which specify the manner and instrument of wipe verbs in the "Verbs of Removing" group of classes (class number 10). Class numbers 1-57 are drawn directly from [21]'s classification. Class numbers 58-109 were developed later in the work of [18].

Verbs (32):	kick, launch, throw, tip, toss, ...
Sem. grid:	AGENT, THEME, SOURCE, DESTINATION
Frames (8):	Subj:Agent V Obj:Theme
	John throws a ball
	Subj:Agent V Obj:Theme PObj:Destination
	John throws a ball to Mary
	Subj:Agent V Obj:Destination Obj:Theme
	John throws Mary a ball
	...

Fig. 1: Simplified Verbnet class *throw-17.1*.

2.3 Overview of the FCA Approach to the Acquisition of a Verbnet like Classification of French Verb

Figure 2 gives a graphical description of the procedure we use to derive a Verbnet like classification from a set of existing linguistic resources. The left hand side of the picture shows how the associations of verb groups and syntactic frames are created. We start from a syntactic lexicon of French verbs and build a *syntactic* classification using Formal Concept Analysis, where groups of verbs are associated with syntactic frames. However, to obtain a Verbnet like classification these groups of verbs also need to be associated with semantic grids. Our approach to build this association is depicted on the right hand side of the picture. We start from the English Verbnet classes, translate them to French and obtain a *semantic* classification associating groups of French verbs with Verbnet classes and so indirectly with the semantic role sets (semantic grid) of these classes. The *syntactic* and *semantic* classifications are then aligned (based on the member verbs of the classes), resulting in a syntactic-semantic classification which associates groups of verbs with both syntactic frames and sets of semantic roles (semantic grids).

3 Related Work

As mentioned in the introduction, several large scale verb classifications exist for English including Verbnet, Framenet and the set of Propbank frames. However producing these classifications manually is costly and error prone. To overcome these shortcomings, there has been much work on automatically acquiring verb classes both for English [34] and to a lesser extent for German [4, 31, 32], Japanese [25], Italian [23] and French [10]. In what follows, we briefly review both the verb classifications that have been developed for French and the methods that have been proposed for the automatic acquisition of verb classifications. We also shortly survey other applications of FCA to natural language.

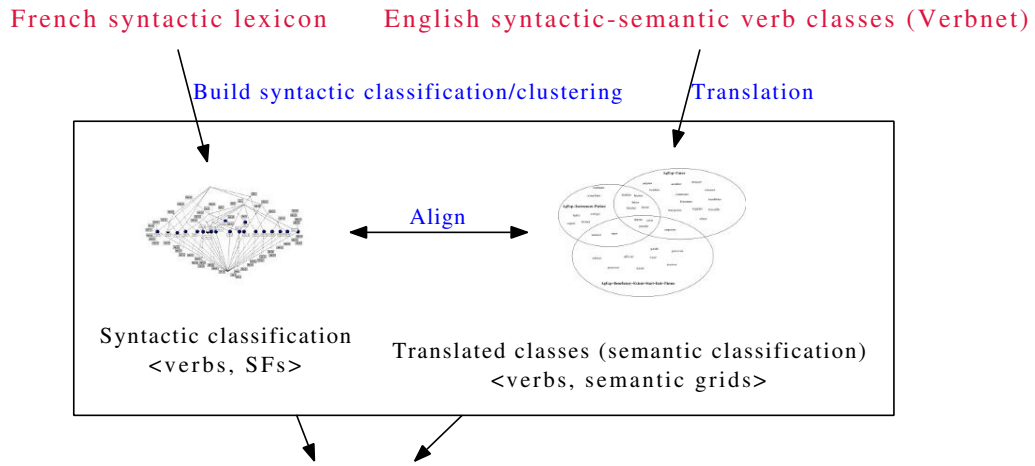

Fig. 2: Outline of the procedure for creating syntactic/semantic Verbnet-like classes for French verbs. Verbs are associated with syntactic frames (SFs) and semantic grids.

3.1 Verb Classifications for French

There are two main classifications which encode information about the syntactic frames and the semantic grids of French verbs: Volem [30] and the LADL tables [14]

Volem resembles the English Verbnet in that it defines for each verb described the set of syntactic frames and of semantic grids accepted by that verb. Volem has restricted coverage however and only describes around 1000 verbs against more than 6000 verbs in Verbnet. Because of this, Volem has limited scope for NLP applications and in particular, cannot be used to develop a training corpus for semantic role labelling since typically, in any given corpus chosen to build that training set, there will always be verbs which are not described by Volem.

The second verb classification available for French are the LADL tables which provide a systematic description of the syntactic and semantic properties of the syntactic functors of French including verbs, predicative nouns and adverbs. Each LADL table groups together all the verbs accepting a given syntactic frame. For instance, the first table groups together all verbs which take an infinitival complement but not a finite or a nominal one. Furthermore,

for each verb in a given table, a set of columns further specify its specific syntactic and/or semantic properties. The LADL tables contain around 5 000 verbs describing over a total of 15 000 verb usages. Although this resource is very rich and reasonably extensive, its format makes it difficult to use by NLP applications. As shown in [13], much of the information needed to derive the syntactic frames of a verb (i.e. the set of its syntactic arguments) is either implicit in the table headings or altogether absent. Similarly, information about semantic grids is either implicit or only derivable using knowledge about the conventions governing the construction of the tables. While there has been work on converting the syntactic information contained in the LADL tables into a format usable by NLP applications [8, 36, 7], semantic grid information remains unavailable thus hampering the exploitation of these tables in applications requiring knowledge about both the syntax and semantics of verbs.

3.2 Automatic Acquisition of Verb Classifications

For French, two main proposals for automatically acquiring verb classes have been put forward.

[34] applies a clustering approach developed for English to French. They exploit features extracted from a large scale syntactic lexicon (LexSchem [24]) acquired fully automatically from the Le Monde newspaper corpus and show that, as for English, syntactic frames and verb selectional preferences perform better than lexical cooccurence features. Their approach achieves a F-measure of 55.1 on 116 verbs occurring at least 150 times in Lexschem. The best performance is achieved when restricting the approach to verbs occurring at least 4000 times (43 verbs) with an F-measure of 65.4.

[10] present an approach to the automatic acquisition of a Verbnet like classification of French verbs which involves the use (i) of a neural clustering method which associates clusters with features, (ii) of several supervised and unsupervised evaluation metrics and (iii) of various existing syntactic and semantic lexical resources. Using a test set similar to that used by [34], the approach achieves an F-measure of 0.70.

More generally, existing methods for the automatic acquisition of verb classifications use hard clustering techniques thereby disallowing for verbs belonging to several classes. Moreover these approaches produce verb clusters but often fail to associate with these clusters the syntactic and semantic frames characterising the verbs belonging to those clusters[2]. By using FCA, the approach proposed here permits addressing both these shortcomings. It allows for polysemous verbs to appear in several clusters while the systematic association supported by FCA between objects and attributes provides a linguistic characterisation of each verb group which closely matches that provided by existing, manually written verb classifications such as Verbnet.

[2] [10] is an exception

3.3 FCA for Natural Language Processing

As shown in [27], several linguistic FCA applications have been proposed in the past. In particular, [33] describes an FCA based approach to build structured class hierarchies starting from unstructured lexicon entries while the features used for building classes in the approach presented in [6] are collected from a corpus. Our approach differs from these, in that the features we use are different. In addition we explore the use of concept selection indices to filter the concept lattices and finally relate the formal concepts we obtain to other classes obtained by a clustering approach based on different numeric features extracted from lexicons and English-French dictionaries.

Another related approach is presented in [37] where Formal Concept Analysis is used to model frame semantics, the theoretical foundations of Framenet [1]. Here the difference consists in the different target lexical representations (cf. Section 1).

4 Formal Concept Analysis

Formal Concept Analysis or FCA [2, 12] is a classification technique which permits creating, from a so-called formal context, a concept lattice where concepts associate sets of objects with sets of attributes.

Intuitively, a concept is a pair $\langle X, Y\rangle$ such that all the objects in X have exactly the attributes in Y and vice versa, all attributes in Y are true of exactly all the objects in X.

More formally, a formal context $\mathcal{K}$ is a triple $\langle \mathcal{O}, \mathcal{A}, R\rangle$ such that $\mathcal{O}$ is a set of objects, $\mathcal{A}$ a set of attributes and R a relation on $\mathcal{O} \times \mathcal{A}$. Given such a context, a concept is a pair $\langle X, Y\rangle$ such that

$$X = \{o \in \mathcal{O} \mid \forall a \in Y.\ (o, a) \in R\}$$
$$\text{and vice versa}$$
$$Y = \{a \in \mathcal{A} \mid \forall o \in X.\ (o, a) \in R\}.$$

Two operators, both denoted by $'$, connect the power sets of objects $2^{\mathcal{O}}$ and attributes $2^{\mathcal{A}}$ as follows:

$$' : 2^{\mathcal{O}} \rightarrow 2^{\mathcal{A}},$$
$$X' = \{a \in \mathcal{A} \mid \forall o \in X.\ (o, a) \in R\}$$

The operator $'$ is dually defined on attributes. For a formal concept $\langle X, Y\rangle$ we have $X' = Y$ and $Y' = X$. X is called the *extent* or *extension* and Y the *intent* or *intension* of the formal concept.

A concept $C_1 = \langle X_1, Y_1\rangle$ is smaller than another concept $C_2 = \langle X_2, Y_2\rangle$ (written $C_1 \leq C_2$) iff $X_1 \subseteq X_2$ and $Y_1 \supseteq Y_2$.

Verb: *expédier*	
SF	**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,POBJ:PP	LA:38L

Table 1: Sample entries in syntactic lexicon for verb *expédier (send)*.

The set of all formal concepts of a context $\mathcal{K}$ together with the order relation $\leq$ form a complete lattice called $\mathbb{K}$, the concept lattice of $\mathcal{K}$. That is, for each subset of concepts there is always a unique greatest common sub-concept and a unique least common super-concept.

In the application we describe in this paper, the objects are the verbs and the attributes are syntactic frames and other syntactic and/or semantic features attributed to the verbs by lexicons. Thus, the resulting concept lattice represents a lexical resource showing exactly which groups of verbs may be used with a particular group of syntactic constructions and dually, which syntactic constructions may be used with which verbs. Therefore it perfectly fits the lexical representation model of Verbnet, which we are aiming at. However, as we will show in the following, we encountered some practical problems. In particular, possibly due to imperfect data, the resulting concept lattice is too large to be handled and reviewed manually, whence the need of methods to efficiently browse and traverse the concept lattice and/or to automatically select the most appropriate concepts.

5 Associating French Verbs with Syntactic Frames

To associate French verbs with syntactic frames, we use the FCA classification approach where the objects are verbs and the attributes are the syntactic frames associated with these verbs by the syntactic lexicon to be described below.

Syntactic lexicons. Syntactic information is retrieved from three existing lexicons for French: *Dicovalence* [38], *the LADL tables* [14], [15] and finally *TreeLex* [19]. Each of these was constructed manually or with an important manual validation by linguists. The combined lexicon covers 5918 verbs, 345 syntactic frames and has a total of 20443 ⟨verb, frame⟩ pairs. Table 1 shows sample entries in this lexicon for the verb *expédier (send)*. Using the Galicia Lattice Builder software[3], we first build a concept lattice based on the formal context $\langle V, F, R\rangle$ such that:

- V is the set of verbs in our syntactic lexicon. We ignore verbs with only one syntactic frame as they will result in classes associating verbs with a unique frame.

[3] `http://www.iro.umontreal.ca/~galicia/`

- F is the set of syntactic frames (SFs) present in the syntactic lexicon,
- R is the mapping such that $(v, f) \in R$ iff the syntactic lexicon associates the verb v with the SF f.

To build the formal context we only consider verbs having more than one syntactic frame and syntactic frames shared by at least two verbs. The resulting formal context is made of 2091 objects (verbs) and 238 attributes (frames), giving rise to a lattice of 12802 concepts. Clearly however not all these concepts are interesting verb classes. Classes aim to factorise information and express generalisations about verbs. Hence, concepts with few verbs can hardly be viewed as classes and similarly, concepts with only one frame are less interesting.

To select from this lattice those concepts which are most likely to provide the most relevant verb-frame associations, we explore the use of three indices for concept selection: *concept stability*, *separation* and *probability* which have been proposed and analysed in [17]. In the next sections we investigate which of these indices performs best in the context of our application : The indices are introduced in Section 5.1, Section 5.2 addresses computational issues and in Section 5.3 we evaluate these indices with respect to our data.

5.1 Filtering the Concept Lattice

As we saw in the previous section, the lattices we have to deal with are very large (about 13000 concepts) and many of the concepts do not represent interesting verb classes (for example many only contain one or two verbs). To select those concepts which are most relevant in the context of our application the concept lattice needs to be filtered. [20] and [17] propose three indices for selecting relevant concepts in concept lattices built from noisy data: *concept stability, separation* and *probability*. These indices are computed for each concept and are proposed as a measure to assess to what extent the concept may be affected by noise in the data. A concept less affected by noisy data is considered more relevant for our application.
Concept stability is a measure which helps discriminating potentially interesting patterns from irrelevant information in a concept lattice based on possibly noisy data. The stability of a concept $C = (V, F)$ is the proportion of subsets of the extent V which have the same attribute set F as V:

Definition 1 Concept stability [20]

$$\sigma((V, F)) := \frac{|\{A \subseteq V \mid A' = F\}|}{2^{|V|}} \quad (1)$$

Intuitively, a more stable concept is less dependant on any individual object in its extent and is therefore more resistant to outliers or other noisy data items. *Concept separation* indicates the significance of the difference between the

[4] Here and in the following $'$ represents the operator on the power sets of objects: $' : 2^{\mathcal{O}} \rightarrow 2^{\mathcal{A}}, X' = \{a \in \mathcal{A} \mid \forall o \in X.\ (o, a) \in R\}$ and dually on that of attributes.

objects covered by a given concept from other objects and, simultaneously, between its attributes and other attributes:

Definition 2 Concept separation [17]

$$\mathfrak{s}((V,F)) := \frac{|V|\,|F|}{\sum_{v\in V}|\{v\}'| + \sum_{f\in F}|\{f\}'| - |V|\,|F|} \tag{2}$$

Intuitively we expect a concept with high separation index to better sort out the verbs it covers from other verbs and simultaneously the frames it covers from other frames. Whereas concept stability is a measure concerned with either objects or attributes, separation gives information about objects and attributes at the same time.

Concept probability For an attribute $a \in A$, the attribute set, we denote by p_a the probability of an object to have the attribute a. In practice it is the proportion of objects having a: $p_a = \frac{|\{a\}'|}{|O|}$, where O denotes the set of objects. For $B \subseteq A$, we define p_B as the probability of an arbitrary object having all attributes from B:

$$p_B := \prod_{a\in B} p_a$$

This formulation assumes the mutual independence of attributes. Based on this, and denoting $n = |O|$ we obtain the following formula for the probability of B being closed:

Definition 3 Concept probability [17]

$$p(B = B'') = \sum_{k=0}^{n} p(|B'| = k, B = B'') \tag{3}$$

$$= \sum_{k=0}^{n} \left[\binom{n}{k} p_B^k (1-p_B)^{n-k} \prod_{a\notin B} (1-p_a^k) \right] \tag{4}$$

A small $p(B = B'')$ suggests a small probability of the attribute combination B to be a concept intent by chance only (and $p(B = B'') \approx 1$ that there is a high probability that the combination is a concept intent by chance). However, this reasoning is based on the independence of the attributes, which in our particular case can not be warranted.

5.2 Computing Stability, Separation and Probability Indices.

Stability. Calculating stability is known to be #P-complete [20], however [29] show that when the concept lattice is known it can be computed efficiently by a bottom-up traversal algorithm. This is the algorithm we used to compute concept stability.
Separation can be computed in $\mathcal{O}(|O|+|A|)$ time, where O and A are the object

and attribute sets respectively. Computing separation is the least prohibitive of the three indices.

Probability. [17] show that computing probability of only one concept involves $\mathcal{O}(|O|^2 \cdot |A|)$ multiplication operations which is computationally very costly. With the computational means at our disposal it was not possible for us to compute the concept probabilities. We therefore computed approximations derived as follows:

First, we consider $\prod_{a \in B}(1 - p_a^k) \approx 1$ for $k > 40$. In view of this, Equation (4) becomes:

$$p(B = B'') = \sum_{k=0}^{40} \left[\binom{n}{k} p_B^k (1 - p_B)^{n-k} \prod_{a \notin B} (1 - p_a^k) \right] \tag{5}$$

$$+ \sum_{k=41}^{n} \left[\binom{n}{k} p_B^k (1 - p_B)^{n-k} \right] \tag{6}$$

As $\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = 1$, Term (6) can be rewritten as:

$$1 - \sum_{k=0}^{40} \left[\binom{n}{k} p_B^k (1 - p_B)^{n-k} \right] = \tag{7}$$

$$1 - F(40; n, p_B) \tag{8}$$

$F(k; n, p) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-1}$ is the cumulative distribution function of the binomial distribution[5] and can be computed using various statistical software packages. Term (5) can also be computed more easily considering that $\binom{n}{k} p_B^k (1 - p_B)^{n-k}$ are binomial densities the computation of which is also provided by statistics software[6]. In our experiments we use these equations to compute approximate values for $p(B = B'')$ (for all concept intents B).

5.3 Evaluating the concept selection indices.

In the following we investigate the performance of the three concept selection indices with respect to our data. For this we proceed as follows. First we use each of these concept selection methods (or combinations thereof) to filter our concept lattice by selecting an arbitrary number of N concepts with highest index value. The selected concepts are then associated with semantic grids following our method which was sketched in the introduction and will be described in detail in Section 6. The resulting resource is compared to a reference

[5] Source Wikipedia: `http://en.wikipedia.org/wiki/Binomial_distribution`

[6] We used the R software environment for statistical computing (`http://www.r-project.org/`).

	cov.	**prec.%**	**rec.%**	F_2
stab only	39.88	18.96	32.55	26.27
sep only	34.25	28.37	21.52	23.41
prob only	35.53	26.60	20.73	22.38
w/o filtering	100	12.30	60.96	26.30

Table 2: F_2 scores and coverage for stability, separation and the 6th probability 10-quantile[7].

using measures originally introduced in the context of information retrieval, namely precision, recall and their F-measure. We consider those indices (or combination of indices) to work best in our application setting which achieve the highest F-measure.

The precision (P), recall (R) and F_β-measures were originally used to assess the performance of information retrieval systems by comparing the items retrieved by the system *Sys* to those given by a reference *Ref* and are defined as follows [39].

Definition 4 (Precision (P), Recall (R) and F_β measure for ⟨verb, semantic grid⟩ associations) Let *Ref* be the instances in a reference and *Sys* the instances produced by our system.

$$R := \frac{Ref \cap Sys}{|Ref|}$$
$$P := \frac{Ref \cap Sys}{|Sys|}$$
$$F_\beta := (1+\beta^2)\frac{P * R}{\beta^2 * P + R}$$

The F-measure was designed so that F_β measures the effectiveness of retrieval with respect to a user who attaches $\sqrt{\beta}$ times as much importance to recall as precision. As for our task recall is more important than precision, we use the F_2 measure, which gives more weight to recall, for comparison.

Since we only select roughly 10% of the total number of concepts we also have to make sure that the selected concepts cover at least a reasonable amount of verbs. Hence, we also measure in each configuration the coverage, i.e. the percentage of verbs covered by the selected concepts.

A further important element in this evaluation is the reference used for comparison. In our setting, this reference consists of the data used for translating the Verbnet classes. As we will see in Section 6, this translation is obtained through a supervised classifier. Our reference consists of the training data used to build this classifier. The comparison with this reference thus assesses which index selects the concepts most similar to the translated classes.

The number N of selected concepts was arbitrarily set to 1500. However, in order to assess the impact of this number we performed the same series of experiments by selecting the top 500 and 1000 concepts respectively.

We now present the result of these experiments.

Table 2 shows the F_2 scores and coverage when using only one index at a time. For stability and separation we applied the method above on the top ranking 1500 concepts. Regarding probability, at first sight, we should consider best the concepts with lowest probability – because the probability of their intents of being closed by chance only is accordingly low. However, looking at the data we found that these concepts have very few verbs and large intent (frame) sets - which rather suggest improbable or rare verb groups. On the other hand, the interpretation of concept probability suggests that a concept with a probability close to 1 could occur by chance only. For these reasons, to assess probability separately we settled on the values between the 5th and 6th 10-quantile[7]. In practice, this means we selected N concepts with medium probability value.

The results confirm the observations of [17]: stability alone gives F_2 scores close to an upper bound – the results obtained without filtering, ie. aligning the translated classes with all the concepts of the lattice. The results for separation and probability are several points lower.

The **cov** column gives the percentage of verbs in the lattice covered by the selected concepts. It shows that using only one index at a time the pre-selected concepts would contain only $35\% - 40\%$ of the verbs in the entire lattice, which is unsatisfactory.

In order to improve coverage, we adopt the reasoning of [17]. They introduced some noise in a context and investigated the performance of the stability, separation and probability indices on the resulting concepts at finding the concepts which would have been obtained based on the original context. The introduced noise was of two kinds: *Type I noise* is obtained by altering every cell in the context with some probability, *Type II noise* is obtained by adding a given number or proportion of random objects or attributes. According to this, our contexts are affected by Type I noise rather than Type II. [17] found that stability was most effective at sorting out Type II noise, but also proved helpful in the case of Type I noise, which is confirmed by our experiments shown previously. They suggested in contrast that separation and probability can not be used on their own but should rather serve as a normalising measure for stability and concluded that the most promising combination is of the form $\text{stability} + k_{sep} \cdot \text{separation} - k_{prob} \cdot \text{probability}$.

We hence start from the assumption that the most effective index for selecting relevant concepts is given by a linear combination of stability, separation and probability: $k_{stab} \cdot \text{stability} + k_{sep} \cdot \text{separation} - k_{prob} \cdot \text{probability}$. We then empirically determine the coefficients k_{stab}, k_{sep} and k_{prob} such that the selected concepts perform best with respect to our task.

We proceed as follows: We choose k_{stab}, k_{sep} and k_{prob}. We then compute the corresponding linear combination for the concepts and select the 1500 concepts ranking highest. As in the previous experiments, we measure the relevance of the selected concepts by aligning the concepts with the translated

[7] Quantiles are points taken at regular intervals from the cumulative distribution function (CDF) of a random variable. They are generalisations of, for example, the median, which is a 2-quantile, and the quartiles, the 4-quantiles.

(a) F_2 and coverage when $k_{stab}, k_{sep} \in \{0.5, 1\}, k_{prob} \in \{0.25, 0.5\}$.

k_{stab}	k_{sep}	k_{prob}	**cov.**	**prec.**%	**rec.**%	F_2
1	1	0.25	98.04	11.87	55.19	24.89
1	0.5	0.25	98.04	11.87	55.19	24.89
1	0.5	0.5	57.69	17.08	30.18	24.04
1	1	0.5	56.15	17.45	29.13	23.82
0.5	0.5	0.25	56.15	17.45	29.13	23.82
0.5	1	0.25	53.81	18.03	27.82	23.36
0.5	0.5	0.5	49.72	18.55	26.25	23.06
0.5	1	0.5	49.90	18.61	25.98	22.95

(b) F_2 and coverage when k_{stab} and k_{sep} are kept fixed and k_{prob} varies.

k_{stab}	k_{sep}	k_{prob}	**cov.**	**prec.**%	**rec.**%	F_2
1	1	0	98.04	12.05	55.12	25.16
1	1	0.05	98.04	12.05	55.12	25.16
1	1	0.005	98.04	12.05	55.12	25.16
1	1	0.0005	98.04	12.05	55.12	25.16
1	1	0.1	98.00	11.91	55.38	25.00
1	1	0.2	98.08	11.88	55.12	24.91
1	1	0.25	98.04	11.87	55.12	24.89
1	1	0.3	98.00	11.79	55.38	24.80
1	1	0.4	59.95	16.27	31.23	23.91
1	1	0.5	56.16	17.45	29.13	23.82
w/o filtering			100	12.30	60.96	26.30

Table 3: F_2 scores and coverage for various $k_{stab}, k_{sep}, k_{prob}$ combinations.

Verbnet classes and by comparing the alignments with the same reference as before. We consider the "best" $k_{stab}, k_{sep}, k_{prob}$ combination the one giving highest F_2 scores and good coverage.

Table 3a shows the results for a first series of experiments where k_{stab} and k_{sep} were assigned the values 0.5 and 1 and k_{prob} 0.25 and 0.5 (The lines are sorted by decreasing F_2 score). They suggest that the stability and separation coefficients had less impact on coverage and F_2 score than the probability coefficient. Interestingly the coverage is correlated with the F_2 score.

In the next series of experiments, shown in Table 3b, we kept the stability and separation coefficients fixed and varied only the probability coefficient. These results indicate that the probability coefficient is less helpful at selecting the most relevant concepts in our setting. This may be due first to the fact that our attributes are not independent (we assumed independence of attributes when setting up the formula for computing the probability index) and second to the fact that we had to approximate the probability index and this approximation may not be accurate enough. An interesting issue for future work would be to model the probability index in the context of correlated attributes and investigate if this way it could better contribute to the selection process.

In the next series of experiments we checked whether it suffices to select a smaller number of concepts (500). The results showed that with this smaller

number of concepts the selected concepts reached a slightly smaller F_2 score but a substantially lower coverage. Also, in this configuration the probability index did seem to be helpful. Preselecting 1000 concepts confirmed the previously observed tendencies: The F_2 score and coverage were only slightly lower than when preselecting 1500 concepts and again the probability index seemed to have only a small impact on the overall results.

From these experiments we conclude the following: First they suggest that the best linear combination is the sum of the stability and separation indices as the F_2 measure and the coverage for this combination are similar to those of an upper bound, ie. the alignment obtained without filtering. They show that selecting only approximatively 10% of the original lattice gives a verb/frame/semantic grid alignment which is close to the alignment obtained when using the entire lattice and that the pre-selected concepts also have a similar coverage.

Second, it does not seem evident that probability has a positive effect on the selected concepts. However, it does improve F_2-measure when the number of selected concepts is lower (500 or 1000 vs. 1500 in our experiments). Hence, for our application we concluded that it is a better strategy to select a larger number of concepts (1500) and not take probability into account. This is even more so as the probability index in our case should be taken with caution because first we had to use an approximation to compute it which may be too rough, and second the probability score is based on the independence of attributes which is not warranted in our case.

6 Associating French Verbs with Semantic Grids.

In Section 5 we described our method of associating French verbs with syntactic frames. As shown in Section 1, for a verb classification to be useful in a semantic role labeling task, the ⟨verb, syntactic frame⟩ associations need to be complemented with semantic information. In our setting the semantic information is provided by associating the concepts obtained as shown in Section 5 with semantic grids. To obtain this association we proceed as follows. In the first step (Section 6.1) the English Verbnet classes are translated to French. This effectively results in a "semantic" classification where groups of French verbs are associated with the semantic grids of English Verbnet classes. These "semantic" classes are aligned with the "syntactic" classes obtained by the previously described FCA approach, based on the verb members (Section 6.2). Finally the semantic grid of a "semantic" class is transferred to the aligned "syntactic" concepts.

In the following we describe this process more in detail.

6.1 Translating Verbnet classes

Verbnet, which is at the origin of this association process has been introduced in Section 2.2.

The Verbnet classes are translated to French using the following *English-French dictionaries*: Sci-Fran-Euradic, a French-English bilingual dictionary, built and improved by linguists, Google dictionary[8] and Dicovalence [38][9]. The merged dictionary contains 51242 French-English verb pairs.

As for this paper only the translated classes, but not the method to produce them is relevant[10] we only very briefly sketch the methodology. A more detailed description and evaluation is presented in [10] and [9].

The main problem encountered when translating the Verbnet classes is that such a translation is bound to be very noisy because verbs are polysemous and the dictionaries typically give translations for several readings of the verb: Thus the dictionary may give several translations v_{fr} which do not correspond to the meaning given by the ⟨english verb, Verbnet class⟩ pair or this meaning may even not be covered at all by the dictionary. To get more accurate translated Verbnet classes we use a machine learning method, namely Support Vector Machines (SVM)[11]. We follow a straightforward SVM application scenario: we build all the French verb, Verbnet class pairs $\langle v_{fr}, C_{VN}\rangle$ where v_{fr} is a translation of an English verb in C_{VN}. The classifier has to give a probability estimate about whether this association is correct or not.

For training the classifier we use ⟨verb, class⟩ pairs involving the 160 verbs appearing in the gold standard proposed by [35][12]. We built the pairs $\langle v_{fr}, C_{VN}\rangle$ where v_{fr} is a verb in the gold standard which is a translation of a verb in C_{VN}. We thus obtained a training set of 1740 $\langle v_{fr}, C_{VN}\rangle$ pairs. For each of these pairs we assessed whether or not there was a meaning of v_{fr} where the semantic roles involved in the event described by the verb were those given by C_{VN}. The features associated to the ⟨verb, class⟩ pairs are numeric and are extracted from the dictionaries and Verbnet[13].

The trained classifier is then used to produce probability estimates for all 28693 ⟨verb, class⟩ instances. We select the 6000 pairs with highest probability estimates[14] and finally obtain the translated classes by assigning each verb in a selected pair to the corresponding class.

To give an idea of the quality of the obtained classes: The accuracy of the classifier on the held out test set was 90%, compared to a maximum accuracy of 93.84% for five fold cross-validation on the development set. The frequency distribution of the translated classes obtained this way is much closer to the distribution of verbs in Verbnet classes than when using an approach based

[8] We obtained this data from `http://www.google.com/dictionary`, however this link is no longer available. We retrieved 13824 French-English verb pairs.

[9] The number of French-English verb pairs we obtained is 11351

[10] Of course better translated classes will result in a better performance of our method, but it is not straightforward to evaluate the quality of the translated classes.

[11] We used *libsvm*, the software package and methodology presented on `http://www.csie.ntu.edu.tw/~cjlin/libsvm/`, [5].

[12] In fact this is the only existing gold standard for French Verbnet style classes and we also use it for the overall evaluation of our system.

[13] They are described in detail in [9].

[14] In Verbnet there are 5726 verb, class pairs

only on translation frequencies, thus providing more accurate verb groups to guide the FCA concept $\leftrightarrow$ semantic roles associations.

6.2 Aligning Concepts and Translated Verbnet Classes

The concepts obtained with the FCA approach are aligned with the translated Verbnet classes based on the verb members. For each concept and each translated class we compute the F-measure of precision and recall defined as follows.

Definition 5 (Precision, recall and F-measure for verb classes) Let C_{VN} be a translated Verbnet class and C_{FCA} the extent (verb set) of an FCA concept. Precision (P), Recall (R) and their F-measure (F) are defined as follows :

$$R := \frac{|C_{VN} \cap C_{FCA}|}{|C_{VN}|}$$
$$P := \frac{|C_{VN} \cap C_{FCA}|}{|C_{FCA}|}$$
$$F := \frac{2RP}{R+P}$$

Each FCA concept is then aligned with the translated Verbnet class with best F-measure and is assigned the semantic grid of this class. Note that this way the same semantic grid can be assigned to several FCA concepts and also that some semantic grids may not be associated with any FCA concept.

7 Results and Discussion

Following the preliminary investigations in the previous sections we associated French verbs with syntactic frames and semantic grids according to the scheme listed below:

- We group the Verbnet semantic roles and assign to one class all the Verbnet verbs whose class have the same grid. We then translate the obtained classes using the methods described in Section 6.
- We use FCA to group French verbs and syntactic frames associated to these verbs by the lexicons described in Section 5. The concept lattices we create are based on the formal contexts consisting of French verbs as objects and syntactic frames as attributes.
- We then select the 1500 concepts where the sum of the stability and separation indices is highest because in Section 5.3 we found this combination of concept selection indices to work best for our application.
- For each translated Verbnet class we identify among the 1500 filtered FCA concepts the one(s) with best F-measure between precision and recall (as defined in Definition 5).

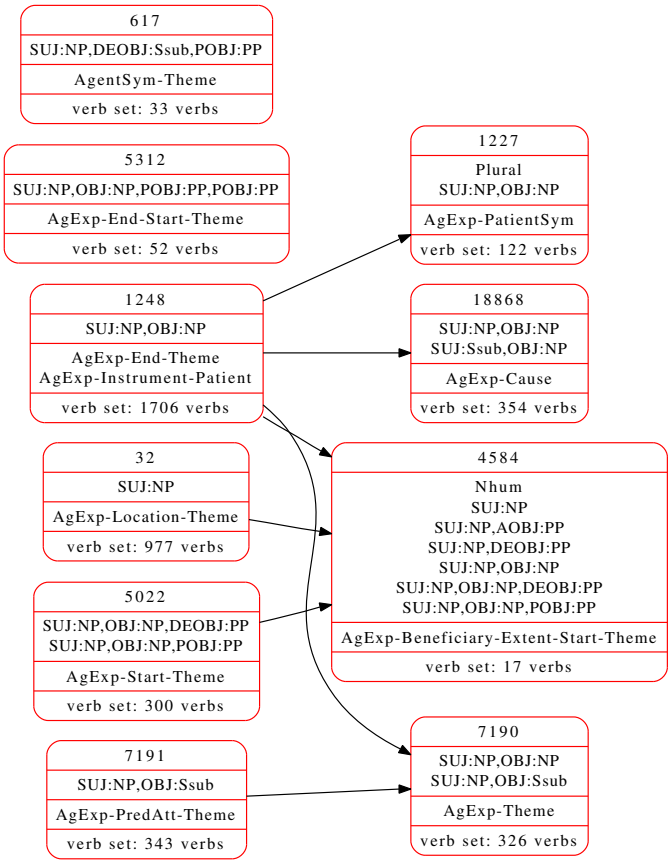

Fig. 3: French verb ↔ synt. frames ↔ semantic grid associations.

The translated Verbnet class is then associated with this FCA concept(s). Thus the verbs in the FCA concept are effectively associated with the semantic grid of the translated class and at the same time with the syntactic frames in the intent (attribute set) of the FCA concept. Figure 3 shows the associations between concepts, semantic grids and frames generated by our method for some Verbnet classes[15]. The figure shows the concepts associated to these semantic grids and for each of these concepts: their attribute set (syntactic frames), the associated semantic grid(s), the number of verbs in the concept and the hierarchical relations between the concepts as given by the concept lattice. For example one of the selected concepts is the concept with id 617. This concept has 33 verbs in its extension (*arguer/argue, bavarder/babble, convenir/agree, discuter/discuss, parler/talk, ...*) and the syntactic frame SUJ:NP,DEOBJ:Ssub,POBJ:PP (as in *J'ai parlé de ce qui s'est*

[15]These are the classes occuring in the gold standard proposed by [35], mentioned in Section 6.

passé avec le ministre./I talked about what happened with the secretary.) in its intension. The semantic grid associated with this concept is *AgentSym-Theme*, stating that the semantic roles involved in events described by these verbs are Agent and Theme and that these verbs enter a so-called "alternation", i.e. a systematic relation between a pair of syntactic frames and this semantic grid (as in *Le journaliste (Agent1) parle avec le ministre (Agent2)./The journalist (Agent1) talks with the secretary (Agent2).* $\leftrightarrow$ *Le journaliste et le ministre (AgentPlural) parlent ensemble./The journalist and the secretary (AgentPlural) talk together*). Many of the 33 verbs in the extent of this concept are indeed conversation verbs and were associated with the correct syntactic frame and semantic grid. We observe also that the concepts `1248` and `32` have very large extensions (1706 and 977 respectively) and are associated with very frequent syntactic constructions (the basic transitive frame SUJ:NP,OBJ:NP and the intransitive frame SUJ:NP). This indicates that these verb groups (and equally their association with semantic grids) are probably not correct. However, we may obtain more accurate information by descending in the lattice hierarchy: The daughter concepts `1227`, `18868` and `4584` have smaller extents and more specific intents (i.e. syntactic descriptions). Looking at these concepts we found that in many cases the generated ⟨verb, syntactic frame, semantic grid⟩ associations were indeed correct.

7.1 Quantitative evaluation.

We evaluate the obtained ⟨verb, semantic grid⟩ associations by a comparison with the gold standard proposed in [35], reproduced in Table 4. More specifically, we compare the ⟨verb, semantic grid⟩ associations engendered by the FCA classification with those extracted from this gold standard using precision, recall and F_1-score as defined in Definition 4[16].

According to this, precision is the proportion $(FCA \cap Gold)/FCA$ of ⟨verb, Verbnet class⟩ pairs found by our method that is correct. Recall is the proportion of ⟨gold verb, Verbnet class⟩ pairs that is found $(FCA \cap Gold)/Gold$. And F-measure is the harmonic mean of precision and recall.

As baseline we use clusterings obtained with K-means[17]. Verbs to be classified and features are the same as for the FCA classification. Similar to the FCA concepts, the obtained verb clusters are then associated with translated Verbnet classes and their semantic grid. The number of clusters is 67, because there are a total of 67 translated Verbnet classes. We performed 10 clusterings with K-means and produced the associations with the translated Verbnet classes and their semantic grids. The associations engendered for each clustering were then compared to the gold standard associations and we report the means of the resulting precision, recall and F-measure as baseline.

[16] Here we use the F_1-measure, where precision and recall are balanced.

[17] We used the *kmeans* function provided by the R software environment for statistical computing(`http://www.r-project.org/`).

AgExp, PatientSym *amalgamate-22.2*: incorporer, associer, réunir, mélanger, mêler, unir, assembler, combiner, lier, fusionner
Cause, AgExp *amuse-31.1*: abattre, accabler, briser, déprimer, consterner, anéantir, épuiser, exténuer, écraser, ennuyer, éreinter, inonder
AgExp, PredAtt, Theme *characterize-29.2*: appréhender, concevoir, considérer, décrire, définir, dépeindre, désigner, envisager, identifier, montrer, percevoir, représenter, ressentir
AgentSym, Theme *correspond-36.1*: coopérer, participer, collaborer, concourir, contribuer, associer
AgExp, Beneficiary, Extent, Start, Theme *get-13.5.1*: acheter, prendre, saisir, réserver, conserver, garder, préserver, maintenir, retenir, louer, affréter
AgExp, Instrument, Patient *hit-18.1*: cogner, heurter, battre, frapper, fouetter, taper, rosser, brutaliser, éreinter, maltraiter, corriger *other_cos-45.4*: mélanger, fusionner, consolider, renforcer, fortifier, adoucir, polir, atténuer, tempérer, pétrir, façonner, former
AgExp, Location, Theme *light_emission-43.1*: briller, étinceler, flamboyer, luire, resplendir, pétiller, rutiler, rayonner, scintiller *modes_of_being_with_motion-47.3*: trembler, frémir, osciller, vaciller, vibrer, tressaillir, frissonner, palpiter, grésiller, trembloter, palpiter *run-51.3.2*: voyager, aller, errer, circuler, courir, bouger, naviguer, passer, promener, déplacer
AgExp, End, Theme *manner_speaking-37.3*: râler, gronder, crier, ronchonner, grogner, bougonner, maugréer, rouspéter, grommeler, larmoyer, gémir, geindre, hurler, gueuler, brailler, chuchoter *put-9.1*: accrocher, déposer, mettre, placer, répartir, réintégrer, empiler, emporter, enfermer, insérer, installer *say-37.7*: dire, révéler, déclarer, signaler, indiquer, montrer, annoncer, répondre, affirmer, certifier, répliquer
AgExp, Theme *peer-30.3*: regarder, écouter, examiner, considérer, voir, scruter, dévisager
AgExp, Start, Theme *remove-10.1*: ôter, enlever, retirer, supprimer, retrancher, débarasser, soustraire, décompter, éliminer
AgExp, End, Start, Theme *send-11.1*: envoyer, lancer, transmettre, adresser, porter, expédier, transporter, jeter, renvoyer, livrer

Table 4: French gold standard classes and their member verbs presented in [35]. Classes sharing the same set of semantic roles are grouped together and each group is displayed with its set of semantic roles.

We also report experiments using additional syntactic and semantic features extracted from our lexical resources, which are meant to help identify specific Verbnet classes and semantic roles. The syntactic features are listed in Table 5a. They indicate whether a verb accepts symmetric arguments (e.g., *John met Mary/John and Mary met*); has four or more arguments; combines with a predicative phrase (e.g., *John named Mary president*); takes a sentential complement or an optional object; or accepts the passive in *se* (similar to the English middle voice *Les habits se vendent bien / The clothes sell well*). The four semantic features extracted from the lexicon indicate whether a verb

Feature	related VN class
Symmetric arguments	*amalgamate-22.2*, *correspond-36.1*
4 or more arguments	*get-13.5.1*, *send-11.1*
Predicate	*characterize-29.2*
Sentential argument	*correspond-36.1*, *characterize-29.2*
Optional object	implicit theme [28], p. 95
Passive built with *se*	theme role [28], p. 120

(a) Additional syntactic features.

Feature	related VN class
Location role	*put-9.1*, *remove-10.1*, ...
Concrete object (non human role)	*hit-18.1* (eg. INSTRUMENT) *other_cos-45.4* ...
Asset role	*get-13.5.1*
Plural role	*amalgamate-22.2*, *correspond-36.1*

(b) Additional semantic features.

Table 5: Additional syntactic (a) and semantic (b) features extracted from the LADL and Dicovalence resources and the alternations/roles they are possibly related to.

Method	cov.	prec.%	rec.%	F
sem.	96.17	24.09	75.00	36.47
synt. & sem.	96.05	23.95	75.00	36.31
scf (frames only)	95.37	23.48	73.80	35.63
synt.	96.34	21.51	74.40	33.38
baseline (K-means)		24.29	25.99	25.11

Table 6: Verb coverage and precision, recall and F-measure for produced verb, semantic associations wrt. the gold standard. At the construction of the lattice we use *frames only* (scf), additional syntactic features *synt.*, additional semantic features *sem.* or both *synt. & sem.*. We pre-selected 1500 concepts with best sum of stability and separation indeces. Baseline precision, recall and F-measure result from associations based on K-means clusterings using the *frames only* feature set with 67 classes. We produced 10 K-means clusterings based on random initial sets and report the mean of precision, recall and F-measure for the resulting associations compared to the gold standard.

takes a locative or an asset argument and whether it requires a concrete object (non human role) or a plural role. Table 5b shows these features together with the Verbnet classes they may help to identify.

Table 6 shows the results for the various experiments, ordered by decreasing F-measure. A first observation is that verb coverage is sufficient in all cases, as it ranges from 95.37% to 96.17%. Second, results for all classifications obtained with FCA are above the baseline. According to Table 6 best results – an F-measure of 36.47 and 75.00 recall – were obtained when using the FCA lattice built with additional semantic features *sem.* In this setting the number of verbs in the gold standard and in our formal context is 111. Of these, 16 verbs are not in the translated Verbnet classes but still were assigned the correct Verbnet class in 11 cases thus illustrating the good generalisation power of our method. Noticeably, the *synt* features degraded performance. We have no clear explanation for this but a possible explanation may be that these features, being extracted from manually built lexicons, are not uniformly present for all

verbs. On the other hand, all but one of the syntactic features we use are also reflected by syntactic frames which possibly convey the same information more accurately.

We notice also, that the better performance of the FCA classifications compared to the baseline is a consequence of a much higher recall, whereas precision is lower for the FCA classifications. However, since the gold standard only gives a restricted number of ⟨verb, semantic grid⟩ associations and therefore possibly does not contain correct associations engendered by the FCA classification, recall is a more important indicator for the coherence of our classifications than precision and F-measure. The low precision is possibly due to the fact that the FCA classification is overlapping and thus associates verbs with several semantic grids, whereas the gold standard only accounts for one ⟨verb, semantic grid⟩ association per verb (in most cases). While the overlapping nature of the FCA classes arguably better reflects the polysemic nature of verbs, the evaluation with respect to this gold standard does not allow a verification of this supposition.

7.2 Qualitative discussion.

To evaluate the quality of the produced verb classes associated with syntactic frames and semantic grids we look at concepts which were associated to the Verbnet classes occuring in the gold standard and their hierarchic structure given by the concept lattice. We analyse the associations produced using the method which performed best compared to the reference, ie. the concept lattice is built based on syntactic frames and additional semantic features (cf. Table 6).

The Verbnet classes occuring in the gold standard are given by the following semantic grids: *AgExpBeneficiaryExtentStartTheme, AgExpCause, AgExpEndStartTheme, AgExpEndTheme, AgExpInstrumentPatient, AgExpLocationTheme, AgExpPatientSym, AgExpPredAttTheme, AgExpStartTheme, AgExpTheme, AgentSymTheme.*

Figure 3 shows that our method selected 10 concepts which were labeled with the 11 Verbnet classes (semantic grids) occuring in the gold. 9 FCA concepts were each mapped to exactly one Verbnet class and 1 was associated with two Verbnet classes.

Verbs in the gold standard class *AgExp-PatientSym* performed best (F-measure 57.14 and recall 80). This class is characterised by the "Symmetric patient alternation":

The merger associated company A with company B.	AGENT V PATIENT1 PATIENT2
The merger associated company B with company A.	AGENT V PATIENT1 PATIENT2
The merger associated the two companies.	AGENT V PATIENT[plural]

The syntactic frame associated with this concept is the basic transitive frame. This concept also has the *Plural* feature set in its attribute set (see Figure 3) showing the benefit of the additional semantic features. Thus, one

major component of the "symmetric patient alternation" is indeed reflected by the feature set of this concept: The basic transitive syntactic frame together with the *Plural* feature represent the construction illustrated in the third row above. However, the syntactic constructions used in the first two rows above were not associated to this concept.

Several verbs which in the reference are not associated to this class were labeled with it by our system: *accrocher, battre, dire, empiler, identifier, mettre, répartir*. Arguably, of these some do have an *AgExp-PatientSym* reading:

Example	Source
Il accroche son manteau au crochet	DV
Les branches battent contre les vitres	DV

They also accept the basic transitive construction, but arguably not with the same reading:

Example
Il accroche son manteau au crochet.
?Il accroche manteau et crochet.
Il accroche son manteau.
Il bat les branches contre les vitres.
?Il bat branches et vitres.
Il bat le chien.

The classes performing worst were *AgExp-Beneficiary-Extent-Start-Theme* and *AgentSym-Theme* classes. None of the verbs associated to these classes by the reference were associated to them by our method. In both cases the associated concepts contain relatively few verbs (17 and 33 respectively). The corresponding translated classes have 224 and 229 members respectively and are the smallest classes of those occuring in the gold standard. There may be two reasons for the fact their being associated to such small concepts: First, in French verbs with this semantic grid might not share enough syntactic frames to build larger groups or second, our lexicons might not contain the relevant syntactic constructions.

Looking at the class *AgExp-Beneficiary-Extent-Start-Theme*, this is the Verbnet class *get-13.5.1*. As the name of the class shows, the verbs in this class may have a large number of roles. Several of the reference verbs labeled with this class clearly have the roles given by the semantic grid, and thus may be labeled correctly. For most of these verbs however, not all of the semantic roles are realised in one syntactic frame. Most (all) syntactic frames in our lexicons have less then 5 arguments, so this class may be hard to characterise adequately by groups of syntactic frames only.

Considering now the *AgentSym-Theme* class, member verbs of this class are characterised by the "Symmetric subject alternation"[18]:

He collaborates with his friend about how to solve the problem.	Actor1, Actor2, Theme
His friend collaborates with him about how to solve the problem.	Actor1, Actor2, Theme
They collaborate about how to solve the problem.	Actor[Plural], Theme

[18] Examples are taken from Verbnet.

It contains the translations of several Verbnet classes, eg. *correspond-36.1*, which is in the gold standard but also *cooperate-73*, *talk-37.5* and *battle-36.4* which are each characterised by different syntactic constructions. It is associated with the syntactic frame SUJ:NP,DEOBJ:Ssub,POBJ:PP and to 33 verbs, most of which are communication verbs which do undergo the "symmetric subject alternation":

Il parle avec son copain de ce qui se passe au monde.	Actor1, Actor2, Theme
Son copain parle avec lui de ce qui se passe au monde.	Actor1, Actor2, Theme
Ils parlent de ce qui se passe au monde.	Actor[Plural], Theme

Thus, for many of these verbs, the associated frame and semantic grid appear to be correct. The verbs in the reference assigned to this class are translations of the English verbs in the *correspond-36.1* class, which do not accept the syntactic frame of the concept selected by our method. It appears that our method was able to correctly label a concept corresponding to some of these classes (as for example that of communication verbs as *talk-37.5*) but, as it selects only one concept per class, failed to identify a concept relating to the *correspond-36.1* class in the gold standard. A possible solution could be to further explore the concept lattice, ie. to look at super- or sub-concepts for concepts also mapping well to this translated class.

As the results in Table 6 show and as noted in Section 7.1, in general precision is low. That is, there are many verb/Verbnet class pairs produced which are not in the gold. A more detailed investigation of the data shows that this stems from several reasons.

First, the gold standard sometimes fails to include a verb in a class. For instance, the verb *accrocher* (to link) is associated in the gold standard only with the *put-9.1* Verbnet class with the semantic grid *AgExp-End-Theme*. However *accrocher* also has an "emotion verb" reading (to attract, cf. the LADL table 4, [22]). Since the gold standard classes only include verbs whose predominant sense belong to that class, the *amuse-31.1* class with the semantic grid *AgExp-Cause* however does not include this verb.

Second, as can be seen in Figure 3, there is one case where the classification maps the same concept to two distinct Verbnet classes. Thus, concept 1248 is mapped to the *AgExp-End-Theme* and *AgExp-Instrument-Patient* class. Although there are verbs which belong to several Verbnet classes, in this case, the double mapping is not warranted because no verb is shared by both classes. Therefore the double class association supported by this mapping of a concept to two Verbnet classes is systematically erroneous. This may stem from the fact that the difference in the syntactic realisation of the *End* and *Instrument* roles and the *Theme* and *Patient* roles may not be reflected by our representation of syntactic frames. These observations also suggest that we might obtain more accurate mappings by choosing the grids and groupings more carefully.

The third, numerically most important, source of errors is the inclusion of the same verbs in several FCA concepts either through inheritance or simply, by inclusion in two distinct concepts.

Of the 111 verbs in the gold standard and in the selected concepts, 24 were in one FCA concept only and the resulting association was correct in 18 cases.

7 verbs were in two FCA concepts and 4 were associated with the gold Verbnet class. The remaining verbs were in 3 or more concepts and introduce most of the errors.

For instance, zooming in on the verbs contained in concept 7191, we observe that this concept contains 29 gold standard verbs, of which 4 are correctly mapped to the *AgExp-Theme* (Verbnet *say-37-7*) class. However 6 verbs in this concept should be mapped to the Verbnet class associated with the super concept of 7191 namely, the *AgExp-PredAtt-Theme* (Verbnet *characterize-29.2*) class.

This suggests that our method did group together related FCA concepts but failed to appropriately discriminate these concepts wrt. to the 2 Verbnet classes *AgExp-PredAtt-Theme* (*characterize-29.2*) and *AgExp-Theme* (*say-37.7*): Indeed, the *characterize-29.2* and *say-37.7* Verbnet classes are related in that they both have semantic roles often realised as clauses which is reflected by the frames associated with the concepts 7191 and 7190. But the frames in the attribute sets are not specific enough to capture the characteristic difference in the classes. This observation highlights the necessity to also evaluate our verb classes with respect to the associated syntactic frames, but, since this association is not provided by the gold standard at hand, it can not be performed based on this data.

8 Conclusion

We introduced a new approach to verb clustering which involves the combined use of the English Verbnet, a bilingual English-French lexicon and a merged syntactic lexicon for French. Using these resources, we built two classifications, one derived from the English Verbnet by translation and the other, from the syntactic lexicons via the construction of a formal concept lattice. We then use the translated Verbnet to associate FCA concepts with Verbnet classes and thereby associate verbs with both syntactic frames and a semantic role set. We explored the performance of the concept selection indices introduced by [20, 17] which are *stability*, *separation* and *probability* at selecting most relevant concepts with respect to our data and found that the sum of stability and separation gave best results in the setting of our application. These results were similar to those obtained without filtering, showing that this combination of the indices did indeed allow to select the most relevant concepts with respect to our data. Finally we showed the French verb, syntactic constructions and semantic grids associations we obtained and performed a preliminary evaluation with respect to a gold standard. Thus Formal Concept Analysis in combination with the concept selection indices, translation and set mapping methods proved an adequate method in this knowledge acquisition process.

Several open issues remain for future work. First, it would be interesting to apply the FCA approach described in this paper to other resources available for French. In particular, [3] has recently made available a large scale, very detailed lexicon of French verbs containing 12 310 verbs which it would be interesting

to use as source data for the approach we propose. Second, a more detailed comparison of the FCA approach with numerical clustering approaches would permit a better assessment of the pros and cons of the two methods and in particular, of how well each of these methods can account for polysemy. Finally, it would be interesting to explore ways of exploiting the structure of the lattices produced for instance, by using association rules to mine for dependencies between concepts and more specifically, to identify so-called "alternations" i.e., systematic relations between syntactic frames and semantic grids.

References

1. Collin F. Baker, Charles J. Fillmore, and John B. Lowe. The Berkeley FrameNet Project. In *Proceedings of the 17th International Conference on Computational Linguistics*, volume 1, pages 86–90, Montreal, Quebec, Canada, 1998. Association for Computational Linguistics.
2. Marc Barbut and Bernard Monjardet. *Ordre et Classification*. Hachette Université, 1970.
3. Paul Bédaride. Raffinement du lexique des verbes français (Resource Refining : « Les Verbes Français ») [in French]. In *Proceedings of the Joint Conference JEP-TALN-RECITAL 2012, volume 2: TALN*, pages 155–168, Grenoble, France, June 2012. ATALA/AFCP.
4. C. Brew and S. Schulte im Walde. Spectral Clustering for German Verbs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 117–124, Philadelphia, PA, 2002.
5. Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology*, 2, 2011. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.
6. P. Cimiano, S. Staab, and J. Tane. Automatic Acquisition of Taxonomies from Text: FCA meets NLP. In *Proceedings of the PKDD/ECML'03 International Workshop on Adaptive Text Extraction and Mining (ATEM)*, pages 10–17, 2003.
7. M. Constant and E. Tolone. A generic tool to generate a lexicon for nlp from lexicon-grammar tables. *arXiv preprint arXiv:1005.5596*, 2010.
8. I. Falk, G. Francopoulo, C. Gardent, et al. Evaluer synlex. *TALN 2007*, 2007.
9. Ingrid Falk. *Making Use of Existing Lexical Resources to Build a Verbnet like Classification of French Verbs*. PhD thesis, Université de Lorraine, June 2012.
10. 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 Association for Computational Linguistics (ACL'12)*, Jeju, Republic of Corea, July 2012.

11. Charles J. Fillmore. The case for case. In Emmon Bach and Robert T. Harms, editors, *Universals in Linguistic Theory*, pages 1–88. Holt, Rinehart and Winston, New York, 1968.
12. Bernhard Ganter and Rudolph Wille. *Formal Concept Analysis: Mathematical Foundations*. Springer, Berlin-Heidelberg, 1999.
13. C. Gardent, B. Guillaume, G. Perrier, I. Falk, et al. Extracting subcategorisation information from maurice gross' grammar lexicon. *ARCHIVES OF CONTROL SCIENCE*, 15(3):309, 2005.
14. Maurice Gross. *Méthodes en syntaxe*. Hermann, Paris, 1975.
15. Alain Guillet and Christian Leclère. *La Structure des phrases simples en français. 2 : Constructions transitives locatives*. Droz, Geneva, 1992.
16. Karin Kipper-Schuler. *VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon*. PhD thesis, University of Pennsylvania, 2006.
17. 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.
18. Anna Korhonen and Ted Briscoe. Extended lexical-semantic classification of English verbs. In *Proceedings of the HLT-NAACL Workshop on Computational Lexical Semantics*, CLS '04, pages 38–45, Stroudsburg, PA, USA, 2004. Association for Computational Linguistics.
19. Anna Kupść and Anne Abeillé. Growing TreeLex. In Alexander Gelbkuh, editor, *Computational Linguistics and Intelligent Text Processing*, volume 4919 of *Lecture Notes in Computer Science*, pages 28–39. Springer Berlin / Heidelberg, 2008.
20. Sergei O. Kuznetsov. On Stability of a Formal Concept. *Annals of Mathematics and Artificial Intelligence*, 49(1-4):101–115, 2007.
21. Beth Levin. *English Verb Classes and Alternations: a Preliminary Investigation*. University of Chicago Press, Chicago and London, 1993.
22. Yvette Mathieu. A Computational Semantic Lexicon of French Verbs of Emotion. In James G. Shanahan, Yan Qu, and Janyce Wiebe, editors, *Computing Attitude and Affect in Text: Theory and Applications*, volume 20 of *The Information Retrieval Series*, chapter 10, pages 109–124. Springer-Verlag, Berlin/Heidelberg, 2006.
23. P. Merlo, S. Stevenson, V. Tsang, and G. Allaria. A multilingual paradigm for automatic verb classification. In *ACL*, pages 207–214, 2002.
24. C. Messiant. A subcategorization acquisition system for French verbs. In *Proceedings of the ACL-08: HLT Student Research Workshop*, pages 55–60, Columbus, Ohio, June 2008. Association for Computational Linguistics.
25. A. Oishi and Y. Matsumoto. Detecting the organization of semantic subclasses of Japanese verbs. *International Journal of Corpus Linguistics*, 2(1):65–89, october 1997.
26. M. Palmer, P. Kingsbury, and D. Gildea. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31(1):71–106, 2005.

27. Uta Priss. Linguistic Applications of Formal Concept Analysis. In Bernhard Ganter, Gerd Stumme, and Rudolf Wille, editors, *Formal Concept Analysis*, volume 3626 of *Lecture Notes in Computer Science*, pages 149–160. Springer Berlin / Heidelberg, 2005.
28. Janet H. Randall. *Linking*. Studies in Natural Language and Linguistic Theory. Springer, Dordrecht, 2010.
29. Camille Roth, Sergei A. Obiedkov, and Derrick G. Kourie. Towards Concise Representation for Taxonomies of Epistemic Communities. In *Concept Lattices and their Applications (CLA'06)*, pages 240–255, 2006.
30. P. Saint-Dizier. 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, 1999.
31. S. Schulte im Walde. *Experiments on the Automatic Induction of German Semantic Verb Classes*. PhD thesis, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, 2003. Published as AIMS Report 9(2).
32. S. Schulte im Walde. Experiments on the automatic induction of german semantic verb classes. *Computational Linguistics*, 32(2):159–194, 2006.
33. C. Sporleder. A Galois Lattice based Approach to Lexical Inheritance Hierarchy Learning. In *15th European Conference on Artificial Intelligence (ECAI'02): Workshop on Machine Learning and Natural Language Processing for Ontology Engineering, Lyon, France*, 2002.
34. L. Sun, A. Korhonen, T. Poibeau, and C. 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.
35. 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.
36. E. Tolone. Les tables du lexique-grammaire au format tal. *Actes de MajecSTIC 2009*, 2009.
37. Francisco J. Valverde-Albacete. Extracting Frame-Semantics Knowledge using Lattice Theory. *Journal of Logic and Computation*, 18(3):361–384, June 2008.
38. Karel van den Eynde and Piet Mertens. La Valence : l'approche pronominale et son application au lexique verbal. *Journal of French Language Studies*, 13:63–104, 2003.
39. Cornelis Joost van Rijsbergen. *Information Retrieval*. Butterworth-Heinemann, 1979.