Classifying French Verbs Using French and English Lexical Resources

Ingrid Falk
Université de Lorraine
LORIA, UMR 7503
Vandoeuvre-lès-Nancy
F-54000, France
ingrid.falk@loria.fr

Claire Gardent
CNRS
LORIA, UMR 7503
Vandoeuvre-lès-Nancy
F-54000, France
claire.gardent@loria.fr

Jean-Charles Lamirel
Université de Strasbourg
LORIA, UMR 7503
Vandoeuvre-lès-Nancy
F-54000, France
jean-charles.lamirel@loria.fr

Abstract
We present a novel approach to the automatic acquisition of a Verbnet like classification of French verbs which involves the use of (i) various existing syntactic and semantic lexical resources; (ii) a neural clustering method which associates clusters with features; and (iii) several supervised and unsupervised evaluation metrics. We evaluate our approach on an established test set and show that it outperforms previous related work with an F-measure of 0.70.

1 Introduction
Verb classifications have been shown to be useful both from a theoretical and from a practical perspective. From the theoretical viewpoint, they permit capturing syntactic and/or semantic generalisations about verbs (Levin, 1993; Kipper Schuler, 2006). From a practical perspective, they support factorisation and have been shown to be effective in various NLP (Natural language Processing) tasks such as semantic role labelling (Swier and Stevenson, 2005) or word sense disambiguation (Dang, 2004).

While there has been much work on automatically acquiring verb classes for English (Sun et al., 2010) and to a lesser extent for German (Brew and Schulte im Walde, 2002; Schulte im Walde, 2003; Schulte im Walde, 2006), Japanese (Oishi and Matsumoto, 1997) and Italian (Merlo et al., 2002), few studies have been conducted on the automatic classification of French verbs. Recently however, two proposals have been put forward.

On the one hand, (Sun et al., 2010) applied a clustering approach developed for English to French. They exploit features extracted from a large scale subcategorisation lexicon (LexSchem (Messiant, 2008)) acquired fully automatically from Le Monde newspaper corpus and show that, as for English, syntactic frames and verb selectional preferences perform better than lexical cooccurrence 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.

On the other hand, Falk and Gardent (2011) present a classification approach for French verbs based on the use of Formal Concept Analysis (FCA). FCA (Barbut and Monjardet, 1970) is a symbolic classification technique which permits creating classes associating sets of objects (eg. French verbs) with sets of features (eg. syntactic frames). Falk and Gardent (2011) provide no evaluation for their results however, only a qualitative analysis.

In this paper, we describe a novel approach to the clustering of French verbs which (i) gives good results on the established benchmark used in (Sun et al., 2010) and (ii) associates verbs with a feature profile describing their syntactic and semantic properties. The approach exploits a clustering method called IGNUF (Incremental Growing Neural Gas with Feature Maximisation, (Lamirel et al., 2011b)) which uses the features characterising each cluster both to guide the clustering process and to label the output clusters. We apply this method to the data contained in various verb lexicons and we evalu-
ate the resulting classification on a slightly modified version of the gold standard provided by (Sun et al., 2010). We show that the approach yields promising results (F-measure of 70%) and that the clustering produced systematically associates verbs with syntactic frames and thematic grids thereby providing an interesting basis for the creation and evaluation of a Verbnet-like classification.

Section 2 describes the lexical resources used for feature extraction and Section 3 the experimental setup. Sections 4 and 5 present the data used for and the results obtained. Section 6 concludes.

2 Lexical Resources Used

Our aim is to acquire a classification which covers the core verbs of French, could be used to support semantic role labelling and is similar in spirit to the English Verbnet. In this first experiment, we therefore favoured extracting the features used for clustering, not from a large corpus parsed automatically, but from manually validated resources\(^1\). These lexical resources are (i) a syntactic lexicon produced by merging three existing lexicons for French and (ii) the English Verbnet.

Among the many syntactic lexicons available for French (Nicolas et al., 2008; Messiah, 2008; Kupš and Abeillé, 2008; van den Eynde and Mertens, 2003; Gross, 1975), we selected and merged three lexicons built or validated manually namely, Dicovalence, TreeLex and the LADL tables. The resulting lexicon contains 5918 verbs, 20433 lexical entries (i.e., verb/frame pairs) and 345 subcategorisation frames. It also contains more detailed syntactic and semantic features such as lexical preferences (e.g., locative argument, concrete object) or thematic role information (e.g., symmetric arguments, asset role) which we make use of for clustering.

We use the French Verbnets as a resource for associating French verbs with thematic grids as follows. We translate the verbs in the English Verbnets to French using English-French dictionaries\(^2\). To deal with polysemy, we train a supervised classifier as follows. We first map French verbs with English Verbnets: A French verb is associated with an English Verbnet class if, according to our dictionaries, it is a translation of an English verb in this class. The task of the classifier is then to produce a probability estimate for the correctness of this association, given the training data. The training set is built by stating for 1740 (French verb, English Verbnet class) pairs whether the verb has the thematic grid given by the pair’s Verbnet class\(^3\). This set is used to train an SVM (support vector machine) classifier\(^4\). The features we use are similar to those used in (Mouton, 2010): they are numeric and are derived for example from the number of translations an English or French verb had, the size of the Verbnets, the number of classes a verb is a member of etc. The resulting classifier gives for each (French verb, English VN class) pair the estimated probability of the pair’s verb being a member of the pair’s class\(^5\). We select 6000 pairs with highest probability estimates and obtain the translated classes by assigning each verb in a selected pair to the pair’s class. This way French verbs are effectively associated with one or more English Verbnet thematic grids.

3 Clustering Methods, Evaluation Metrics and Experimental Setup

3.1 Clustering Methods

The IGNGF clustering method is an incremental neural “winner-take-most” clustering method belonging to the family of the free topology neural clustering methods. Like other neural free topology methods such as Neural Gas (NG) (Martinetz and Schulten, 1991), Growing Neural Gas (GNG) (Fritzke, 1995), or Incremental Growing Neural Gas (IGNG) (Prudent and Ennaji, 2005), the IGNGF method makes use of Hebbian learning product_info.php?products_id=666), Google dictionary (http://www.google.com/dictionary) and Dicovalence (van den Eynde and Mertens, 2003).

\(^1\)Of course, the same approach could be applied to corpus based data (as done e.g., in (Sun et al., 2010)) thus making the approach fully unsupervised and directly applicable to any language for which a parser is available.

\(^2\)For the translation we use the following resources: Sci-Fran-Euradic, a French-English bilingual dictionary, built and improved by linguists (http://catalog.elra.info/)

\(^3\)The training data consists of the verbs and Verbnets classes used in the gold standard presented in (Sun et al., 2010).

\(^4\)We used the Libsvm (Chang and Lin, 2011) implementation of the classifier for this step.

\(^5\)The accuracy of the classifier on the held out random test set of 100 pairs was of 90%.
(Hebb, 1949) for dynamically structuring the learning space. However, contrary to these methods, the use of a standard distance measure for determining a winner is replaced in IGNGF by feature maximisation. Feature maximisation is a cluster quality metric which associates each cluster with maximal features i.e., features whose Feature F-measure is maximal. Feature F-measure is the harmonic mean of Feature Recall and Feature Precision which in turn are defined as:

$$FR_c(f) = \frac{\sum_{v \in c} W_f^v}{\sum c' \sum_{v \in c'} W_{f'}^v}, \quad FP_c(f) = \frac{\sum_{v \in c} W_{f'}^v}{\sum c' \sum_{v \in c'} W_{f'}^v}$$

where $W_{f}^x$ represents the weight of the feature $f$ for element $x$ and $F_c$ designates the set of features associated with the verbs occurring in the cluster $c$. A feature is then said to be maximal for a given cluster iff its feature F-measure is higher for that cluster than for any other cluster.

The IGNGF method was shown to outperform other usual neural and non neural methods for clustering tasks on relatively clean data (Lamirel et al., 2011b). Since we use features extracted from manually validated sources, this clustering technique seems a good fit for our application. In addition, the feature maximisation and cluster labeling performed by the IGNGF method has proved promising both for visualising clustering results (Lamirel et al., 2008) and for validating or optimising a clustering method (Attik et al., 2006). We make use of these processes in all our experiments and systematically compute cluster labelling and feature maximisation on the output clusterings. As we shall see, this permits distinguishing between clusterings with similar F-measure but lower “linguistic plausibility” (cf. Section 5). This facilitates clustering interpretation in that cluster labeling clearly indicates the association between clusters (verbs) and their prevalent features. And this supports the creation of a Verbnet style classification in that cluster labeling directly provides classes grouping together verbs, thematic grids and subcategorisation frames.

### 3.2 Evaluation metrics

We use several evaluation metrics which bear on different properties of the clustering.

**Modified Purity and Accuracy.** Following (Sun et al., 2010), we use modified purity (mPUR); weighted class accuracy (ACC) and F-measure to evaluate the clusterings produced. These are computed as follows. Each induced cluster is assigned the gold class (its prevalent class, prev($C$)) to which most of its member verbs belong. A verb is then said to be correct if the gold associates it with the prevalent class of the cluster it is in. Given this, purity is the ratio between the number of correct gold verbs in the clustering and the total number of gold verbs in the clustering:

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

where Verbs$_{\text{Gold} \cap \text{Clustering}}$ is the total number of gold verbs in the clustering.

Accuracy represents the proportion of gold verbs in those clusters which are associated with a gold class, compared to all the gold verbs in the clustering. To compute accuracy we associate to each gold class $C_{\text{Gold}}$ a dominant cluster, i.e. the cluster dom($C_{\text{Gold}}$) which has most verbs in common with the gold class. Then accuracy is given by the following formula:

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

Finally, F-measure is the harmonic mean of mPUR and ACC.

**Coverage.** To assess the extent to which a clustering matches the gold classification, we additionally compute the coverage of each clustering that is, the proportion of gold classes that are prevalent classes in the clustering.

**Cumulative Micro Precision (CMP).** As pointed out in (Lamirel et al., 2008; Attik et al., 2006), unsupervised evaluation metrics based on cluster labelling and feature maximisation can prove very useful for identifying the best clustering strategy. Following (Lamirel et al., 2011a), we use CMP to identify the best clustering. Computed on the clustering results, this metrics evaluates the quality of a clustering with respect to the cluster features rather
than with respect to a gold standard. It was shown in (Ghribi et al., 2010) to be effective in detecting degenerated clustering results including a small number of large heterogeneous, “garbage” clusters and a big number of small size “chunk” clusters.

First, the local Recall ($R_f^c$) and the local Precision ($P_f^c$) of a feature $f$ in a cluster $c$ are defined as follows:

$$R_f^c = \frac{|v_f^c|}{|V_f^c|}, \quad P_f^c = \frac{|v_f^c|}{|V_c^f|}$$

where $v_f^c$ is the set of verbs having feature $f$ in $c$, $V_c$ the set of verbs in $c$ and $V_f^c$, the set of verbs with feature $f$.

Cumulative Micro-Precision (CMP) is then defined as follows:

$$CMP = \frac{\sum_{i=1}^{l} C_{inf}, C_{sup}}{\sum_{i=1}^{l} C_{inf}, C_{sup}} \frac{1}{C_{i+}} \sum_{c \in C_{i+}, f \in F_c} P_f^c$$

where $C_{i+}$ represents the subset of clusters of $C$ for which the number of associated verbs is greater than $i$, and: $C_{inf} = \arg\min_{c_i \in C} |c_i|, C_{sup} = \arg\max_{c_i \in C} |c_i|$

### 3.3 Cluster display, feature f-Measure and confidence score

To facilitate interpretation, clusters are displayed as illustrated in Table 1. The features used for clustering (cf. Section 4) are displayed in decreasing order of Feature F-measure and features whose Feature F-measure is under the average Feature F-measure of the overall clustering are clearly delineated from others. In addition, for each verb in a cluster, a confidence score is displayed which is the ratio between the sum of the F-measures of its cluster maximised features over the sum of the F-measures of the overall cluster maximised features. Verbs whose confidence score is 0 are considered as orphan data.

### 3.4 Experimental setup

We applied an IDF-Norm weighting scheme (Robertson and Jones, 1976) to decrease the influence of the most frequent features (IDF component) and to compensate for discrepancies in feature number (normalisation).

### 4 Features and Data

#### Features

In the simplest case, the features used for clustering are the subcategorisation frames (scf) associated with the verbs by our lexicon. We also experiment with different combinations of additional, syntactic (synt) and semantic features (sem) extracted from the lexicon and with the thematic grids (grid) extracted from the English Verbnet.

The thematic grid information is derived from the English Verbnet as explained in Section 2. The thematic features extracted from the lexicon are listed in Table 1(a). 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). As shown in Table 1(a), these
(a) Additional syntactic features.

<table>
<thead>
<tr>
<th>Feature</th>
<th>related VN class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symmetric arguments</td>
<td>amalgamate-22.2, correspond-36.1</td>
</tr>
<tr>
<td>4 or more arguments</td>
<td>get-13.5.1, send-11.1</td>
</tr>
<tr>
<td>Predicate</td>
<td>character-29.2</td>
</tr>
<tr>
<td>Sentential argument</td>
<td>correspond-36.1, characterize-29.2</td>
</tr>
<tr>
<td>Optional object</td>
<td>implicit theme (Randall, 2010), p. 95</td>
</tr>
<tr>
<td>Passive built with se</td>
<td>theme role (Randall, 2010), p. 120</td>
</tr>
</tbody>
</table>

(b) Additional semantic features.

<table>
<thead>
<tr>
<th>Feature</th>
<th>related VN class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location role</td>
<td>put-9.1, remove-10.1, ...</td>
</tr>
<tr>
<td>Concrete object</td>
<td>hit-18.1 (eg. INSTRUMENT)</td>
</tr>
<tr>
<td>(non human role)</td>
<td>other_cos-45.4 ...</td>
</tr>
<tr>
<td>Asset role</td>
<td>get-13.5.1</td>
</tr>
<tr>
<td>Plural role</td>
<td>amalgamate-22.2, correspond-36.1</td>
</tr>
</tbody>
</table>

Table 2: Additional syntactic (a) and semantic (b) features extracted from the LADL and Dicovalence resources and the alternations/roles they are possibly related to.

features are meant to help identify specific Verbnet classes and thematic roles. Finally, we extract four semantic features from the lexicon. These indicate whether a verb takes a locative or an asset argument and whether it requires a concrete object (non human role) or a plural role. The potential correlation between these features and Verbnet classes is given in Table 1(b).

French Gold Standard To evaluate our approach, we use the gold standard proposed by Sun et al. (2010). This resource consists of 16 fine grained Levin classes with 12 verbs each whose predominant sense in English belong to that class. Since our goal is to build a Verbnet like classification for French, we mapped the 16 Levin classes of the Sun et al. (2010)’s Gold Standard to 11 Verbnet classes thereby associating each class with a thematic grid. In addition we group Verbnet semantic roles as shown in Table 4. Table 3 shows the reference we use for evaluation.

Verbs For our clustering experiments we use the 2183 French verbs occurring in the translations of the 11 classes in the gold standard (cf. Section 4). Since we ignore verbs with only one feature the number of verbs and (verb, feature) pairs considered may vary slightly across experiments.

5 Results
5.1 Quantitative Analysis
Table 5 includes the evaluation results for all feature sets.

In terms of F-measure, the results range from 0.61 to 0.70. This generally outperforms (Sun et al., 2010) whose best F-measures vary between 0.55 for verbs occurring at least 150 times in the training data and 0.65 for verbs occurring at least 4000 times in this training data. The results are not directly comparable however since (i) the gold data is slightly different due to the grouping of Verbnet classes through their thematic grids and (ii) (Sun et al., 2010) uses automatically acquired, rather than, manually specified resources.

In terms of features, the best results are obtained using the grid-scf-sem feature set with an F-measure of 0.70. Moreover, for this data set, the unsupervised evaluation metrics (cf. Section 3) highlight strong cluster cohesion with a number of clusters close to the number of gold classes (13 clusters for 11 gold classes); a low number of orphan verbs (i.e., verbs whose confidence score is zero); and a high Cumulated Micro Precision (CMP = 0.3) indicating homogeneous clusters in terms of maximising features. The coverage of 0.72 indicates that approximately 8 out of the 11 gold classes could be matched to a prevalent label. That is, 8 clusters were labelled with a prevalent label corresponding to 8 distinct gold classes.

In contrast, the classification obtained using the scf-synt-sem feature set has a higher CMP for the
Table 3: French gold classes and their member verbs presented in (Sun et al., 2010).

Table 3: French gold classes and their member verbs presented in (Sun et al., 2010).

clustering with optimal mPUR (0.57); but a lower F-measure (0.61), a larger number of classes (16) and a higher number of orphans (156). That is, this clustering has many clusters with strong feature cohesion but a class structure that markedly differs from the gold. Since there might be differences in structure between the English VerbNet and the thematic classification for French we are building, this is not necessarily incorrect however. Further investigation on a larger data set would be required to assess which clustering is in fact better given the data used and the classification searched for.

In general, data sets whose description includes semantic features (sem or grid) tend to produce better results than those that do not (sfc or synt). This is in line with results from (Sun et al., 2010) which shows that semantic features help verb classification. It differs from it however in that the semantic features used by Sun et al. (2010) are selectional preferences while ours are thematic grids and a restricted set of manually encoded selectional preferences.

Noticeably, the synt feature degrades performance throughout: grid,sfc,synt has lower F-measure than grid,sfc; sfc,synt,sem than sfc,sem; and sfc,synt than sfc. We have no clear explanation for this.

The best results are obtained with the IGNGF method on most of the data sets. Table 5a illustrates the differences between the results obtained with IGNGF and those obtained with K-means on the grid-scf-sem data set (best data set). Although K-means and IGNGF optimal model reach similar F-measure and display a similar number of clusters, the very low CMP (0.10) of the K-means model shows that, despite a good Gold class coverage (0.81), K-means tends to produce more heterogeneous clusters in terms of features.

Table 5a also shows the impact of IDF feature weighting and feature vector normalisation on clustering. The benefit of preprocessing the data appears clearly. When neither IDF weighting nor vector normalisation are used, F-measure decreases from 0.70 to 0.68 and cumulative micro-precision from 0.30 to 0.21. When either normalisation or IDF weighting is left out, the cumulative micro-precision drops.
(a) The impact of the feature set.

<table>
<thead>
<tr>
<th>Feat. set</th>
<th>Nbr. feat.</th>
<th>Nbr. verbs</th>
<th>mPUR</th>
<th>ACC</th>
<th>F (Gold)</th>
<th>Nbr. classes</th>
<th>Cov.</th>
<th>Nbr. orphans</th>
<th>CMP at opt (13cl.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>scf</td>
<td>220</td>
<td>2085</td>
<td>0.93</td>
<td>0.48</td>
<td>0.64</td>
<td>17</td>
<td>0.55</td>
<td>129</td>
<td>0.28 (0.27)</td>
</tr>
<tr>
<td>grid, scf</td>
<td>231</td>
<td>2085</td>
<td>0.94</td>
<td>0.54</td>
<td>0.68</td>
<td>14</td>
<td>0.64</td>
<td>183</td>
<td>0.12 (0.12)</td>
</tr>
<tr>
<td>grid, scf, sem</td>
<td>237</td>
<td>2183</td>
<td>0.86</td>
<td>0.59</td>
<td>0.70</td>
<td>13</td>
<td>0.72</td>
<td>67</td>
<td>0.30 (0.30)</td>
</tr>
<tr>
<td>grid, scf, synt</td>
<td>236</td>
<td>2150</td>
<td>0.87</td>
<td>0.50</td>
<td>0.63</td>
<td>14</td>
<td>0.72</td>
<td>66</td>
<td>0.13 (0.14)</td>
</tr>
<tr>
<td>grid, scf, synt, sem</td>
<td>242</td>
<td>2201</td>
<td>0.99</td>
<td>0.52</td>
<td>0.69</td>
<td>16</td>
<td>0.82</td>
<td>100</td>
<td>0.50 (0.22)</td>
</tr>
<tr>
<td>scf, sem</td>
<td>226</td>
<td>2183</td>
<td>0.83</td>
<td>0.55</td>
<td>0.66</td>
<td>23</td>
<td>0.64</td>
<td>146</td>
<td>0.40 (0.26)</td>
</tr>
<tr>
<td>scf, synt</td>
<td>225</td>
<td>2150</td>
<td>0.91</td>
<td>0.45</td>
<td>0.61</td>
<td>15</td>
<td>0.45</td>
<td>83</td>
<td>0.17 (0.22)</td>
</tr>
<tr>
<td>scf, synt, sem</td>
<td>231</td>
<td>2101</td>
<td>0.89</td>
<td>0.47</td>
<td>0.61</td>
<td>16</td>
<td>0.64</td>
<td>156</td>
<td>0.57 (0.11)</td>
</tr>
</tbody>
</table>

(b) Metrics for best performing clustering method (IGNGF) compared to K-means. Feature set is grid, scf, sem.

<table>
<thead>
<tr>
<th>Method</th>
<th>mPUR</th>
<th>ACC</th>
<th>F (Gold)</th>
<th>Nbr. classes</th>
<th>Cov.</th>
<th>Nbr. orphans</th>
<th>CMP at opt (13cl.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGNGF with IDF and norm.</td>
<td>0.86</td>
<td>0.59</td>
<td>0.70</td>
<td>13</td>
<td>0.72</td>
<td>67</td>
<td>0.30 (0.30)</td>
</tr>
<tr>
<td>K-means with IDF and norm.</td>
<td>0.88</td>
<td>0.57</td>
<td>0.70</td>
<td>13</td>
<td>0.81</td>
<td>67</td>
<td>0.10 (0.10)</td>
</tr>
<tr>
<td>IGNGF, no IDF</td>
<td>0.86</td>
<td>0.59</td>
<td>0.70</td>
<td>17</td>
<td>0.81</td>
<td>126</td>
<td>0.18 (0.14)</td>
</tr>
<tr>
<td>IGNGF, no norm.</td>
<td>0.78</td>
<td>0.62</td>
<td>0.70</td>
<td>18</td>
<td>0.72</td>
<td>180</td>
<td>0.15 (0.11)</td>
</tr>
<tr>
<td>IGNGF, no IDF, no norm.</td>
<td>0.87</td>
<td>0.55</td>
<td>0.68</td>
<td>14</td>
<td>0.81</td>
<td>103</td>
<td>0.21 (0.21)</td>
</tr>
</tbody>
</table>

Table 5: Results. Cumulative micro precision (CMP) is given for the clustering at the mPUR optimum and in parentheses for 13 classes clustering.

by up to 15 points (from 0.30 to 0.15 and 0.18) and the number of orphans increases from 67 up to 180. That is, clusters are less coherent in terms of features.

5.2 Qualitative Analysis

We carried out a manual analysis of the clusters examining both the semantic coherence of each cluster (do the verbs in that cluster share a semantic component?) and the association between the thematic grids, the verbs and the syntactic frames provided by clustering.

Semantic homogeneity: To assess semantic homogeneity, we examined each cluster seeking to identify one or more Verbnet labels characterising the verbs contained in that cluster. From the 13 clusters produced by clustering, 11 clusters could be labelled. Table 6 shows these eleven clusters, the associated labels (abbreviated Verbnet class names), some example verbs, a sample subcategorisation frame drawn from the cluster maximising features and an illustrating sentence. As can be seen, some clusters group together several subclasses and conversely, some Verbnet classes are spread over several clusters. This is not necessarily incorrect though. To start with, recall that we are aiming for a classification which groups together verbs with the same thematic grid. Given this, cluster C2 correctly groups together two Verbnet classes (other_c-45.4 and hit-18.1) which share the same thematic grid (cf. Table 3). In addition, the features associated with this cluster indicate that verbs in these two classes are transitive, select a concrete object, and can be pronominalised which again is correct for most verbs in that cluster. Similarly, cluster C11 groups together verbs from two Verbnet classes with identical theta grid (light_emission-43.1 and modes_of_being_with_motion-47.3) while its associated features correctly indicate that verbs from both classes accept both the intransitive form without object (la jeune fille rayonne / the young girl glows, un cheval galope / a horse gallops) and with a prepositional object (la jeune fille rayonne de bonheur / the young girl glows with happiness, un cheval galope vers l’infini / a horse gallops to infinity). The third cluster grouping together verbs from two Verbnet classes is C7 which contains mainly judgement verbs (to applaud, bless, compliment, punish) but also some verbs from the (very large) other_c-45.4 class. In this case, a prevalent shared feature is that both types of verbs accept a de-object that is, a prepositional object introduced by "de" (Jean applaudit Marie d’avoir dansé / Jean applaudit Marie for having danced; Jean dégage le sable de la route / Jean clears the sand of the road). The semantic features necessary to provide a finer grained analysis of their differences are lacking.

Interestingly, clustering also highlights classes
which are semantically homogeneous but syntactically distinct. While clusters C6 and C10 both contain mostly verbs from the amuse-31.1 class (amuser, agacer, énerver, déprimer), their features indicate that verbs in C10 accept the pronominal form (e.g., Jean s’amuse) while verbs in C6 do not (e.g., *Jean se déprime). In this case, clustering highlights a syntactic distinction which is present in French but not in English. In contrast, the dispersion of verbs from the other class over clusters C2 and C7 has no obvious explanation. One reason might be that this class is rather large (361 verbs) and thus might contain French verbs that do not necessarily share properties with the original Verbnet class.

**Syntax and Semantics.** We examined whether the prevalent syntactic features labelling each cluster were compatible with the verbs and with the semantic class(es) manually assigned to the clusters. Table 6 sketches the relation between cluster, syntactic frames and Verbnet like classes. It shows for instance that the prevalent frame of the C0 class (manner-speaking-37.3) correctly indicates that verbs in that cluster subcategorise for a sentential argument and an AOBJ (prepositional object in “a”) (e.g., Jean bafouille à Marie qu’il est amoureux / Jean stammers that he is in love); and that verbs in the C9 class (characterize-29.2) subcategorise for an object NP and an attribute (Jean nomme Marie présidente / Jean appoints Marie president). In general, we found that the prevalent frames associated with each cluster adequately characterise the syntax of that verb class.

### 6 Conclusion

We presented an approach to the automatic classification of French verbs which showed good results on an established testset and associates verb clusters with syntactic and semantic features.

Whether the features associated by the IGNGF clustering with the verb clusters appropriately characterise these clusters remains an open question. We carried out a first evaluation using these features to label the syntactic arguments of verbs in a corpus with thematic roles and found that precision is high but recall low mainly because of polysemy: the frames and grids made available by the classification for a given verb are correct for that verb but not for the verb sense occurring in the corpus. This suggests that overlapping clustering techniques need to be applied.

We are also investigating how the approach scales up to the full set of verbs present in the lexicon. Both Dicovalence and the LADL tables contain rich detailed information about the syntactic and semantic properties of French verbs. We intend to tap on that potential and explore how well the various semantic features that can be extracted from these resources support automatic verb classification for the full set of verbs present in our lexicon.

---

<table>
<thead>
<tr>
<th>Column 1</th>
<th>Column 2</th>
<th>Column 3</th>
</tr>
</thead>
</table>
| C0       | speaking: babiller, bafouiller, balbutier | SUINPOBJ:Suab,AOBJ:PP  
Jean bafouille à Marie qu’il l’aime / Jean stammers that he is in love |
| C1       | put: entasser, répandre, essaimer | SUINPOBJ:PP,DUMMY:REFL  
Loc, Plural  
Les déchets s’entassent dans la cour / Waste piles in the yard |
| C2       | hit: broyer, démolir, fouetter | SUINPOBJ:NP  
T-Num  
Ces pierres broient les graines / These stones grind the seeds, other_cos: agrandir, alléger, amincir |
| C4       | dedicate: s’engager à s’obliger à, SUINPOBJ:VPinf,DUMMY:REFL  
Cette promesse l’engage à nous suivre / This promise commits you to following us |
| C5       | conjecture: penser, attester, ag réter | SUINPOBJ:Sub  
Le médecin atteste que l’employé n’est pas en état de travailler / The physician certifies that the employee is not able to work |
| C6       | amuse: déprimer, déconterner, découvrir | SUINPOBJ:NP,Sub,OBJ:NP  
Jean enseigne l’anglais à Marie / Jean teaches Marie English |
| C7       | other_cos: dégager, vider, drainer, severer | SUINPOBJ:NP,DDEOBJ:PP  
vider le récipient de son contenu / empty the container of its contents |
| C9       | judgement | SUINPOBJ:NP,DDEOBJ:PP  
Jean blame Marie d’avoir cours / Jean blames Mary for running |
| C10      | amuse: agacer, amuser, enorgueiller | SUINPOBJ:NP,DEOBJ:XP,DP,OBJ:NP,DEOBJ:Ssub |
| C11      | Jean nomme Marie présidente / Jean appoints Marie president | Jean déprime Marie de ce que Jean parte / Marie depresses because of Jean’s leaving |
| C12      | Jean s’enorgueillit d’être roi / Jean is proud to be king | light: rayonner, clignoter, cliquer |
| C13      | Jean s’engueullit d’être roi / Jean is proud to be king | SUINPOBJ:NP  
DDEOBJ:PP  
glisser sur le trottoir verglacé / slip on the icy sidewalk |
| C14      | Jean s’enorgueillit / Jean is proud | SUINPOBJ:NP  
DDEOBJ:PP  
Les d échets s’entassent dans la cour / Waste piles in the yard |

Table 6: Relations between clusters, syntactic frames and Verbnet like classes.
References


