Using Formal Concept Analysis to Acquire Knowledge about Verbs

Ingrid Falk[1], Claire Gardent[2] and Alejandra Lorenzo[1]

[1] INRIA/Nancy Universités, Nancy (France)
[2] CNRS/LORIA, Nancy (France)

Abstract. We use Formal Concept Analysis (FCA) to acquire information about verbs as required by Natural Language Processing (NLP) applications. In particular, we show that stable concepts permits creating verb classes with good generalisation power; and that association rules are useful for complementing incomplete verb information.

1 Introduction

Natural language processing (NLP) applications aim either to interpret (analysis) or to produce text data (generation). Because verbs are a central component of natural language sentences, detailed knowledge about their syntactic and semantic behaviour is an essential ingredient of many NLP applications. In particular, detailed subcategorisation information (that is, information about the number and the syntactic type of a verb's complements) has repeatedly been shown to be crucial in enhancing the linguistic coverage and the accuracy of NLP applications ([Briscoe and Carroll, 1993], [Carroll and Fang, 2004]).

To acquire and structure such knowledge, verb classifications have been proposed which group together verbs with similar syntactic and/or semantic behaviour. On the practical side, verb classes permit capturing generalisations about verb behaviour thus reducing both the effort needed to construct a verb lexicon and the likelihood that errors are introduced when adding new entries. On the theoretical side, [Levin, 1993] has shown that syntax reflects semantics and consequently, that verbs that belong to a syntactic class can be shown to often share a semantic component (meaning aspect).

For English, there exists several large scale resources providing verb classes (Framenet [Baker et al., 1998], Verbnet [Schuler, 2006] and to a lesser extend Wordnet [Fellbaum, 1998]) in a format that is amenable for use by natural language processing systems. For French however, existing verb classes are either too restricted in scope (Volem [Saint-Dizier, 1999]) or not sufficiently structured (the LADL tables [Gross, 1975], [Guillet and Leclère, 1992]) to be directly useful for NLP.

In this paper, we explore the use of Formal Concept Analysis (FCA) to acquire classes for French verbs from the available lexical resources. Additionally, we show that association rules can be put to work to extend and complement an existing subcategorisation lexicon. The paper is structured as follows. Section 2 shows how Dicovalence, a subcategorisation lexicon for French verbs, can be used to construct a lattice whose concepts are potential verb classes with objects being verbs and attributes being subcategorisation frames. Specifically, we show that the resulting set of stable concepts (i) achieves reasonably high coverage (77% of the verbs contained in the Dicovalence lexicon) and (ii) give rise to verb classes with good factorisation power in that most classes associate several frames with the verbs they contain. In Section 3, we extend the approach to construct verb classes that integrate both syntactic and semantic information. Finally, Section 4 shows how applying high confidence association rules derived from the Dicovalence formal context to a different lexicon, permits extending the coverage of Dicovalence.

2 Using formal concept analysis to acquire valency based verb classes

Formal concept analysis is one of many applicable classification and clustering techniques. We exploit it here to create concepts where the objects are verbs and the attributes, syntactic frames. Starting from a valency lexicon for French which associates each verb with a set of valency frames, we build a concept lattice and extract from it the most stable concepts. We show that these concepts form interesting verb classes in that (i) they permit grouping together verbs that share a common set of frames and (ii) they largely cover the verbs contained in Dicovalence.

We start by presenting the two lexicons used to build the lattice and evaluate the acquired verb classes namely, Dicovalence and VerbNet. We then describe the verb classification obtained using formal concept analysis and compare it to VerbNet.

2.1 Dicovalence, a valency lexicon for French verbs

The Dicovalence lexicon [van den Eynde and Mertens, 2003] lists the valency frames of 3 936 French verbs. A valency frame characterises the number and the type of the syntactic arguments expected by a verb. For instance, the valency frames for *maintenir* can be described as illustrated below. That is, each frame describes a set of syntactic arguments and each argument is characterised by a grammatical function[3] and a syntactic category (NP indicates a noun phrase, PP a prepositional phrase, CL a clitic ie a weak pronoun). The use of each frame is illustrated by an example.

- SUJ:NP, (OBJ:NP)
  Manifester qu' il a les moyens de maintenir un cap.
- SUJ:NP, OBJ:NP, ATO:XP
  Le PDG d' Hachette s' est engagé à maintenir ouvert le petit robinet d' alimentation qui permettra à la Cinq de conserver une trésorerie minimale.

[3] SUJ refers to the subject grammatical function, OBJ to the object, P-OBJ, A-OBJ and DE-OBJ describes prepositional objects introduced by any preposition, *à* ou *de* respectively and ATO indicates an object attribute.

– SUJ:NP, A-OBJ:PP, refl:CL
La poursuite de la baisse de l' investissement productif se maintient à 2,5 % en rythme annuel depuis la mi-Novembre
– SUJ:NP, (OBJ:NP), P-OBJ:PP
L' écart entre taux des prêts et taux de refinancement leur permet de maintenir des concours suffisants aux entreprises demeurées solvables , puis d' accroître ce volume à mesure que les mauvais risques sont provisionnés.
– SUJ:NP, refl:CL
Le beau temps se maintient

2.2 VerbNet, a classification of English verbs

VerbNet ([Schuler, 2006]), is the largest electronic verb classification for English. It was created manually and classifies 3 626 verbs using 411 classes. Each VerbNet class includes among other things a set of verbs and a set of valency frames. For instance, the *Hit-18.1* class associates verbs and frames as follows[4]:

Verbs:	*batter, beat, bump, butt, drum, hammer, hit, jab, kick, knock, lash, pound, rap, slap, smack, smash, strike, tap*
Frames	SUJ:NP,P-OBJ:PP SUJ:NP,P-OBJ:PP,P-OBJ:PP SUJ:NP,OBJ:NP SUJ:NP,OBJ:NP,P-OBJ:PP SUJ:NP,DE-OBJ:Ssub

2.3 Verb classes as stable concepts

To construct verb classes that group together verbs sharing a set of frames, we first build a concept lattice[5]. The formal context K used to build this lattice is the triplet $\langle V, F, R\rangle$ such that V is the set of verbs contained in Dicovalence, F the set of valency frames used in Dicovalence and R the mapping defined by Dicovalence between verbs and frames: $(v, f) \in R$ iff Dicovalence associates the verb v with the frame f. The concept lattice of this context K contains 2115 concepts i.e., potential verb classes. Clearly however not all these concepts are interesting verb classes. Classes aim to factorise information and express generalisations about verbs. Hence, concepts with few (1 or 2) verbs can hardly be viewed as classes. Similarly, concepts with few frames are less interesting especially if many of the verb subclasses of the extension of these concepts have more frames than there are in their intension.

To select from the large set of concepts contained in the lattice those which are most likely to adequately characterise verb sets, we consider only concepts that are intensionally stable ([Kuznetsov, 2007]). The intensional stability of a concept (V, F) is defined as follows :

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

[4] The Verbnet format for valency frames does not mention grammatical functions. We have added them here to preserve notation consistency and facilitate reading.

[5] We used the Galicia Lattice Builder software (`http://www.iro.umontreal.ca/~galicia/`) to build the lattices

For instance, given the concepts C1 to C8 below, setting the stability threshold to above 0.5, will filter out all concepts except C1, C5, C6 and C7. If further we eliminate concepts whose extension is a singleton (classes with one verb only), then the only extracted verb class will be C1 $= \langle\{v_1, v_2\}, \{f_1, f_2, f_3\}\rangle$. That is, by retaining as verb classes only those concepts whose intensional stability is high, we produce classes which strike a good balance between the size of the frame set and that of the verb set.

Concept	Extension	Intension	Stability	Decision
C1	v_1,v_2,v_3	f_1	$3/8 = 0.37$	$\times$
C2	v_1,v_2	f_1,f_2,f_3	$4/4 = 1$	$\surd$
C3	v_1,v_3	f_1	$2/4 = 0.5$	$\times$
C4	v_2,v_3	f_1	$2/4 = 0.5$	$\times$
C5	v_1	f_1 ,f_2,f_3	$2/2 = 1$	$\times$
C6	v_2	f_1 ,f_2,f_3	$2/2 = 1$	$\times$
C7	v_3	f_1	$2/2 = 1$	$\times$
C8	$\emptyset$	f_1,f_2,f_3		

As illustrated by this example, keeping only the more stable concepts potentially implies that some verbs may be excluded of the classification (here v_3). Figure 1 shows how the chosen stability threshold affects verb coverage that is, the proportion of Dicovalence verbs covered by the resulting classes. Varying the stability threshold (from 90 to 76) has little impact on coverage (from 3 025 verbs to 3043 verbs i.e., 18 verbs with the stability threshold decreasing from 90 to 76) but a strong impact on the number of classes (from 212 to 506). Overall keeping only stable concepts permits covering approximately 77% of the verbs in Dicovalence.

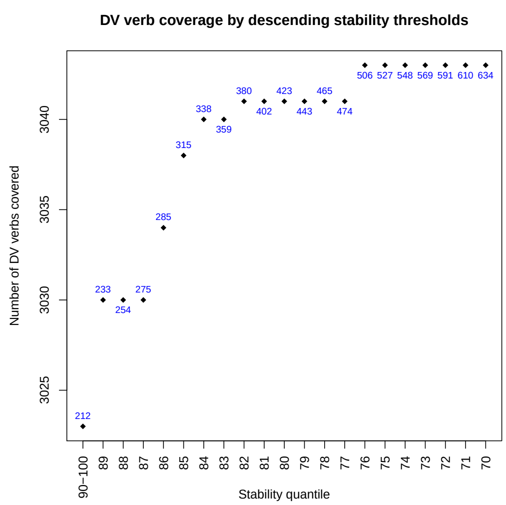


Fig. 1. Percentage of Dicovalence verbs contained in sets of concepts plotted against descending stability threshold. The numbers above the points are the number of concepts in a set.

To further assess the impact of the chosen stability threshold on the verb classes obtained, we compare these classification with respect to their number of singleton verb / frame classes, to the average number of frames / verb per class and to average harmonic mean of verb and frame size per class. Table 1 shows how these numbers vary with the chosen stability threshold and compare them with those for VerbNet. The graphs in Figure 2 compare the distribution of the verbs in classes wrt. the number of associated frames for these classifications and for VerbNet.

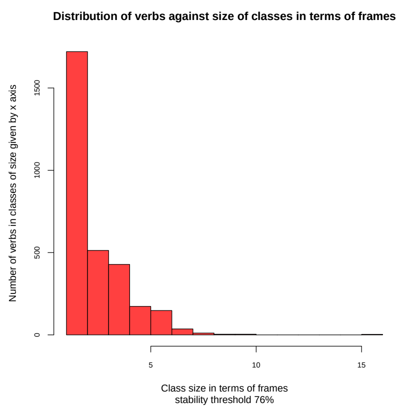
Focusing first on the graphs (Figure 2), we observe that the stability threshold has little impact on the number of verbs being in classes with 1 or 2 frames. With a stability threshold of 76%, approximately 56% of the verbs are in such classes against 57% with a threshold of 85 or 86% . Interestingly, these percentages are close to what is observed in Verbnet (54% of the verbs in classes with 1 or 2 frames). More generally, a stability threshold around 85% seems to offer a good compromise between the size of the frame sets (from 1 to 10 with 43% of the verbs having more than 2 frames), the overall verb coverage and the number of classes (315 for a threshold of 85% and 285 for a threshold of 86%).

Table 1) gives more details about the comparative properties of the various classifications. Two points give further support for a threshold around 85%. First, a lower threshold increases the number of classes while a stability threshold around 85% permits keeping this number down thereby improving the generalisation and factorisation power of the classification. Second, the harmonic mean of the verb set size and the frame set size increases with the stability threshold. In other words, the classes obtained with a higher threshold are overall better balanced and more populated.

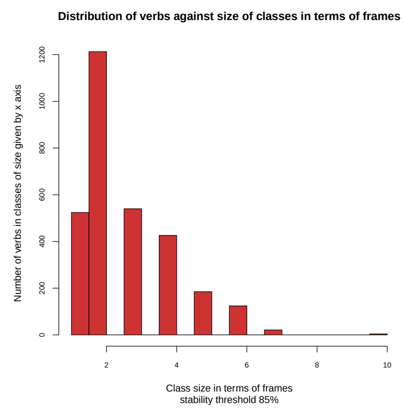
We observe an important difference in the average number of verbs per class. This difference is due to the different representation of the verb classes in VerbNet and the classification obtained by FCA: In VerbNet each verb belongs to exactly one (sub-)class and inherits the frames of the super-classes, whereas with FCA super- and sub-classes partly contain the same verbs. This is possibly also the reason for the big difference in the maximal class sizes in terms of verbs.

Stability threshold	75%	84%	85%	86%	VerbNet
Nb. of classes	506	338	315	285	411
Min. verbs	2	4	4	7	1
Max. verbs	1555	1555	1555	1555	383
Min. frames	1	1	1	1	1
Max. frames	16	10	10	7	25
Classes with 1 verb	0	0	0	0	53
Classes with 1 frame	20	17	17	16	76
Average class size (verbs)	53.06	70.78	75.03	78.48	12.14
Average class size (frames)	3.80	3.55	3.51	3.48	2.90
Average class size (harmonic mean)	5.90	5.98	5.98	6.01	3.38
Total number of verbs				3936	3626
Total number of frames				136	117

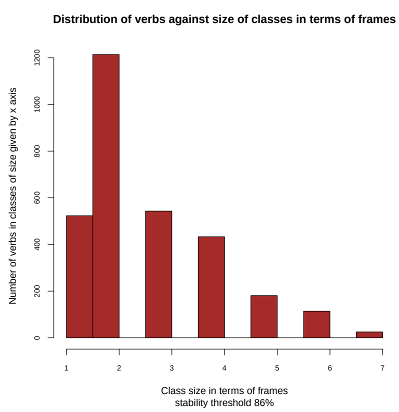
Table 1. Some features of the verb classification depending on the chosen stability threshold.



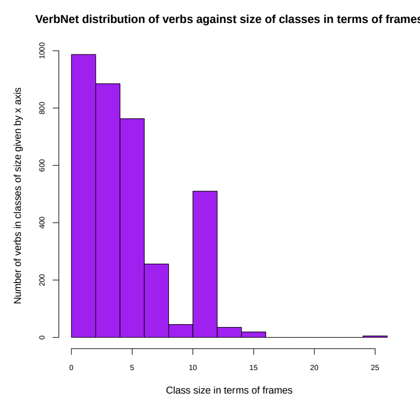
(a) *Class76*, stability threshold 76%



(b) *Class85*, stability threshold 85%



(c) *Class86* obtained with stability threshold 86%



(d) VerbNet

Fig. 2. Distribution of verbs against the size of their class in terms of frames for the classifications obtained with FCA with varying stability thresholds (fig. (a), (b), (c)) and for VerbNet.

3 Acquiring syntactico-semantic verb classes

Beth Levin's hypothesis (cf. Section 1) states that syntax correlates with semantics. To create verb classes which capture both a shared syntactic behaviour (a shared set of valency frames) and a shared meaning component, we draw on another verb resource for French namely, the LADL tables ([Gross, 1975], [Guillet and Leclère, 1992]). These tables were specified manually over several years by a large team of expert linguists and contain syntactic and semantic information about French verbs. For instance, a table might state that the subject of all verbs in that table must be human; or that the object is a destination, etc. The classes created by the LADL tables however, are both too fine- and too coarse-grained to be useful for NLP. They are too coarse-grained in that at the table level, a single subcategorisation frame and a semantic description is associated with a large set of verbs – information about the syntactic sub-classes corresponding to different valency frame sets is not provided. They are too fine-grained in that within a table, detailed information is given about each individual verb but not about sub-groups of verbs.

To create verb classes that are characterised both by a set of valency frames and by semantic information, we apply the same method as described in Section 2 using as attributes both the valency frames contained in Dicovalence and the LADL tables identifiers. That is, the formal context used to build the lattice and extract stable concepts is the context $\langle V, F, R \rangle$ where V is the set of verbs contained in the intersection of Dicovalence and the LADL tables, F is the union of the set of valency frames used in Dicovalence with the set of LADL table identifiers and R the mapping such that $(v, f) \in R$ if either Dicovalence or the LADL tables associates the verb v with the frame/table f.

As before, we rank the concepts by stability. Additionally, we filter out concepts whose intension does not contain at least one table identifier and 2 valency frames. In this way, we ensure that each concept extracted from the FCA lattice assigns the verb group denoted by the concept extension both a semantic (LADL table description) and a syntactic characterisation (valency frames). We require that the concept intension contains at least 2 valency frames since each LADL table is associated with a defining valency frame.

Here is an example class extracted by this method. The class groups together verbs which indicate a change of state (mainly colour and age) and which can be used with and without object (*Jean rougit / Jean turned red ; Jean rougit le mur / Jean painted the wall red*) and with a sentential *de*-object (*Jean rougit de ce que Marie l'injure / Jean blushed that Marie insults him*).

Verbs:	blanchir (*to whiten*), bleuir (*to turn blue*), blêmir (*to turn pale*) pâlir (*to turn white*), rajeunir (*to become younger*), rosir (*to turn pink*), rougir (*to blush*), verdir (*to turn green*), vieillir (*to become old*)
LADL Table:	32RA (Make Adj$_v$), 8 (Verbs with sentential complement in *de*)
Frames	SUJ:NP SUJ:NP,OBJ:NP SUJ:NP,DE-OBJ:Ssub

Taking the top 500 concepts obeying the set constraints yields a set of classes such that each class is associated with one or more semantic label (i.e., LADL table) and between 2 and 6 valency frames. Furthermore, the resulting classes each contain between 9 and 237 verbs with an overall verb coverage of 62% . That is, the 500 classes cover 62% of the verbs present in the intersection of Dicovalence and the LADL tables. Overall thus, the classes obtained are interesting in that they are associated with an informative syntactico-semantic characterisation; they group together a satisfactory number of verbs; and they permit covering a majority of verbs covered by the verb resources used. Although coverage could be better, it is worth stressing that manual resources are always incomplete and imperfect. It is therefore likely that this incomplete coverage is due to missing and/or erroneous information either in the LADL lexicon (missing verbs in a table might prevent a syntactic class to be associated with that class thereby decreasing verb coverage) or in Dicovalence (missing frames might block a verb from being integrated in a class).

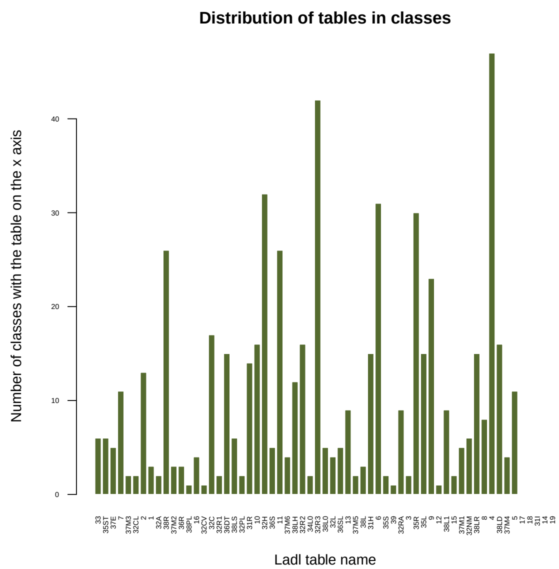


Fig. 3. Distribution of tables in classes. For 61% of the tables less than 5 classes are assigned that table. The last 5 tables are not assigned to any class.

Figure 3 shows for each LADL table the number of classes it includes. Interestingly, for most tables (61%), less than 5 classes are identified – this suggest a relatively strong association between the syntactic frames associated with these classes and the semantic component labelling the table. There are 5 tables which are assigned no class – these are all relatively small tables (around 20 verbs) for which no syntactic class could be found whose verbs were included in the set of verbs contained by the table.

4 Using association rules to extend the lexicon

Formal concept analysis provides another useful tool for developing verb resources namely, association rules. We first introduce them. We then show how association rules can be used to complement Dicovalence with frame information derived from another lexical resource.

4.1 Association rules, confidence and lift

Given a context $K = \langle V, F, R\rangle$ with attributes F, an association rule $A \rightarrow B$ with $A, B \in F$ relates *itemsets* of this context i.e., sets of attributes. Thus in our case, association rules describe dependencies between sets of frames.

Association rules can be evaluated using various metrics such as confidence and lift ([Szathmary, 2006]). The confidence of a rule $A \rightarrow B$ captures the probability of B given A. It is defined as the ratio between the number of objects having attributes A and B, and the number of objects having attributes A. Intuitively, it is the proportion of A that are also B. The *confidence* of an association rule $A \rightarrow B$ is defined as:

$$conf(A \rightarrow B) = \frac{P(A \cup B))}{P(A)} = \frac{sup(A \cup B)}{sup(A)}$$

where $sup(F_1)$, the *support* of F_1 for $F_1 \in F$ an itemset, is the number of objects including F_1.

The lift value of an association rule measures the strength of association between the antecedent and the consequent. It is defined as the ratio of the confidence of the rule and the relative support of the consequent.

$$lift(A \rightarrow B) = \frac{P(A \cup B))}{P(A) \times P(B)} = \frac{conf(A \rightarrow B)}{rsup(B)} = \frac{rsup(A \cup B)}{rsup(A) \times rsup(B)}$$

where the relative support $rsup(F_1)$ for $F_1 \in F$, is $sup(F_1)/ \mid V \mid$. The lift is a value between 0 and infinity. A lift value greater than 1 indicates that the antecedent and the consequent appear more often together than expected.

4.2 Using association rules to extend Dicovalence

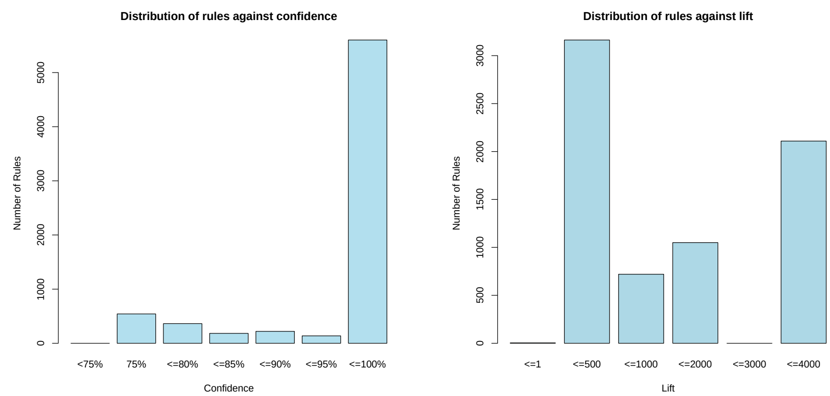
Dicovalence only covers the most frequent verbs of French. Using another verb lexicon (namely the LADL tables), we exploit association rules derived from the Dicovalence data to predict frames for verbs not in Dicovalence but that are partially described in the LADL tables. In this way, we complement Dicovalence with both the LADL table frame information (each table and thus each verb in that table is associated with a valency frame) and the information contained in the inferred frames.

Based on the context $\langle V, F, R\rangle$ introduced in section 2, we compute[6] the minimal non redundant association rules that is, the set of association rules $F_1 \rightarrow F_2$ such that F_2 is a closed itemset and F_1 is the minimal generator of F_2 . We then rank the rules according to both lift and confidence. Figure (4) shows the distribution of these rules. Most rules have a confidence between 98 and 100%. Moreover almost all rules have a lift above 1[7] indicating that the association

[6] We used the Coron system `http://coron.loria.fr/site/index.php` for computing the rules and the various metrics.

[7] Although the graph fails to show it, none of the rule with confidence between 0 and 50 has a confidence below 0.

between the frame sets related by the rules is higher than chance.



(a) *Rules distribution against confidence* (b) *Rules distribution against lift*

Fig. 4. Distribution of rules against two metrics: confidence (fig. (a)) and lift (fig. (b))

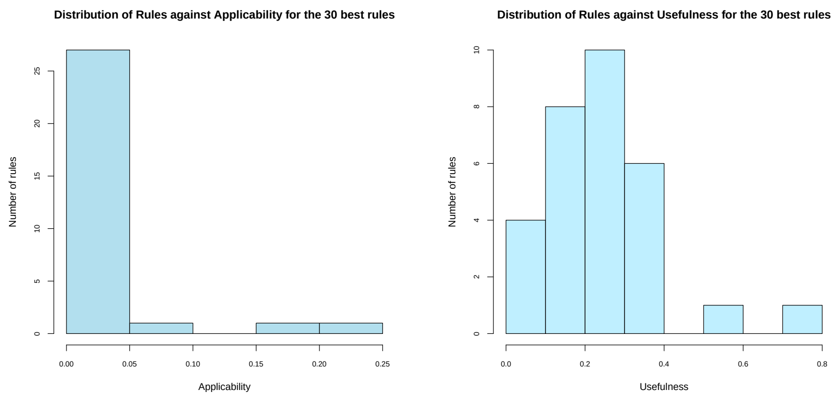
Next we apply these rules to the (verb, frame) pairs given by the LADL tables. For each rule, we then compute its applicability as follows. Let V_{ladl} be the number of verbs occurring in the LADL lexicon and V^r_{ladl} be the number of verbs in the LADL lexicon for which the rule r applies. Then the applicability of a rule r is the ratio between these two values.

$$applicability(r) = \frac{V^r_{ladl}}{V_{ladl}}$$

We also evaluate the usefulness of a rule i.e., its potential for discovering new frames. Let F^r_{ladl} be the number of frames present in the LADL for the verbs to which rule r applies and let $NewF^r_{ladl}$ be the number of frames inferred by the application of rule r and not present in the LADL lexicon, then the usefulness of a rule r is defined as the ratio between the number of discovered frames and the number of frames contained in the verb entries to which the rule applies:

$$usefulness(r) = \frac{NewF^r_{ladl}}{F^r_{ladl}}$$

Figure 5 plots both, the rule applicability and the rule usefulness against the number of rules for the best 30 rules according to the applicability criterion (i.e., picking the 30 rules with highest applicability). Although most rules apply to less than 5% of the LADL items, the usefulness score mostly ranges between 10 and 40% . Overall, applying these 30 best rules to the (verb,frame) pairs contained in the LADL tables, permits inferring 1435 (verb,frame) pairs. The confidence for these rules ranges from 0.762 to 1 with most rules having a confidence close to 1. Their lift ranges from 1.174 to 6.33, and their support from 2 to 586. That is, rules with high applicability are also reliable in that they display good confidence and lift score above 1. By comparison, when applying the 30 rules with best support, lift and confidence in that ranking order, we obtain an increase of 1157 verbs. In sum, to maximise both the number of frames inferred and their reliability, a good strategy is either to rank rules by support or applicability, and then take the n best rules wrt to the chosen ranking.



(a) *Rules applicability* (b) *Rules usefulness*

Fig. 5. Distribution of the 30 best rules against applicability (fig. (a)) and usefulness (fig. (b))

5 Conclusion

Much work on acquiring verb information for NLP has focused on identifying so called alternations i.e., pairs of valency frames that are often simultaneously true of a verb and classes that associate sets of verbs with syntactic and/or semantic information. The results presented in this paper suggest that FCA is an appropriate framework for modelling such knowledge acquisition process.

Concepts naturally model the association of verbs and syntactic and/or semantic information. Moreover, like fuzzy clustering, FCA permits "soft clustering" in that a data element may belong to several classes – a property of the produced classifications which is essential for our task since verbs (e.g., *to fly*) are highly ambiguous and may belong to several syntactic and/or semantic classes. Sections 2 and 3 show that stable concepts permit creating classes with good generalisation and factorisation power (e.g., a few hundred syntactic classes to cover roughly 3 500 verbs) and linguistically sound, empirical content (good average number of verbs and frames within the classes).

Association rules on the other hand, are a natural way to capture alternations while the various evaluation metrics proposed in the literature permit ranking them according to such criteria as reliability (confidence), strength of association (lift) and breadth of application (support). Section 4 illustrates this by showing how association rules can be used to extend an incomplete lexicon with additional valency information.

References

Baker et al., 1998. Baker, C. F., Fillmore, C. J., and Lowe, J. B. (1998). The berkeley FrameNet project. In *Proceedings of the 17th International Conference on Computational Linguistics*, volume 1, pages 86–90, Montreal, Quebec, Canada. Association for Computational Linguistics.

Briscoe and Carroll, 1993. Briscoe, T. and Carroll, J. (1993). Generalized probabilistic lr parsing of natural language (corpora) with unification-based grammars. *Comput. Linguist.*, 19(1):25–59.

Carroll and Fang, 2004. Carroll, J. and Fang, A. C. (2004). The automatic acquisition of verb subcategorisations and their impact on the performance of an hpsg parser. In *IJCNLP*, pages 646–654.

Fellbaum, 1998. Fellbaum, C., editor (1998). *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA.

Gross, 1975. Gross, M. (1975). *Méthodes en syntaxe*. Hermann, Paris.

Guillet and Leclère, 1992. Guillet, A. and Leclère, C. (1992). *La structure des phrases simples en franais. 2 : Constructions transitives locatives*. Droz, Geneva.

Kuznetsov, 2007. Kuznetsov, S. O. (2007). On stability of a formal concept. *Annals of Mathematics and Artificial Intelligence*, 49(1-4):101–115.

Levin, 1993. Levin, B. (1993). *English Verb Classes and Alternations: a preliminary investigation*. University of Chicago Press, Chicago and London.

Saint-Dizier, 1999. Saint-Dizier, P. (1999). Alternation and verb semantic classes for french: Analysis and class formation. In *Predicative forms in natural language and in lexical knowledge bases*. Kluwer Academic Publishers.

Schuler, 2006. Schuler, K. K. (2006). *VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon*. PhD thesis, University of Pennsylvania.

Szathmary, 2006. Szathmary, L. (2006). *Symbolic Data Mining Methods with the Coron Platform*. PhD Thesis in Computer Science, University Henri Poincaré – Nancy 1, France.

van den Eynde and Mertens, 2003. van den Eynde, K. and Mertens, P. (2003). La valence: l'approche pronominale et son application au lexique verbal. *Journal of French Language Studies*, 13:63–104.