Aspectual Flexibility Increases with Agentivity and Concreteness
A Computational Classification Experiment on Polysemous Verbs

Ingrid Falk and Fabienne Martin
Universität Stuttgart
first.second@ling.uni-stuttgart.de

Abstract

We present an experimental study making use of a machine learning approach to identify the factors that affect the aspectual value that characterises polysemous verbs under each of their readings. The study is based on various morpho-syntactic and semantic features collected from a French lexical resource and on a gold standard aspectual classification of verb readings designed by an expert. Our results support the tested hypothesis, namely that agentivity and abstractness influence lexical aspect.

Keywords: lexical aspect; aspectual flexibility; French lexicon

1. Introduction

It is generally accepted that the event type of a sentence and in particular its aspectual value result from a complex interplay between lexical features of the predicate on one hand and its linguistic context on the other. In previous work by Siegel and McKeown (2000), Zarcone and Lenci (2008) and Friedrich and Palmer (2014), the event type of corpus utterances is predicted on the basis of syntactic and semantic features of the main predicate and the linguistic context thereof. In contrast, in this study, we investigate lexical (inner) aspect through a French lexical resource, "Les Verbes Français" (François et al., 2007), a valence lexicon of French verbs providing a detailed morpho-syntactic and semantic description for each reading (use) of a verb. Since the basic lexical units are the readings of the predicate rather than the (ambiguous) verbs, this resource allows one to study lexical aspect on an intermediate level between the coarse-grained predicate level and the very fine-grained corpus utterance level. Thanks to its detailed syntactic and semantic description, it also enables us to investigate the effect of particular complex semantic dimensions, as e.g. agentivity or abstractness, on the lexical aspect of verbs across their different readings.

2. Outline

In this work, we focus on the way the lexical aspect of a same verb varies across its different readings, using the lexical resource "Les Verbes Français" (François et al. (2007), henceforth LVF). We address this question through a supervised classification task. For this, we first extracted verbal readings from the LVF for a set of 167 verbs, chosen in such a way that each of the four Vendlerian aspectual classes are roughly equally represented. We manually annotated each of these readings on the basis of a fine-grained aspectual classification based on eight aspectual classes (see below for a rough description). This manual annotation provides the gold standard for our classification experiments. For each annotated reading, we then automatically collected morpho-syntactic and semantic features from the LVF. Based on these features, we trained classifiers to automatically predict the aspectual class of the (manually annotated) readings. We then evaluated the different classifiers used by comparing their accuracy in 10-fold cross validation on the basis of the gold annotation. We conducted experiments with three feature sets and their combinations. The first set of features ('syn/sem' features) provides information on the morphosyntactic make-up of the verb, its argument structure and potential adjuncts, and indicates which suffixes are used in nominalisations and adjectives derived from the verb, when those are available (note that derived nouns and adjectives are coupled to a particular reading only if they preserve the meaning of the verb on this particular reading). The second feature set indicates whether the reading is agentive or nonagentive, and the third one whether the reading is concrete or abstract.

The paper is organized as follows. Section 3 presents the resource we make use of. We first briefly review its design and theoretic principles and then describe the data used in our experiments. We detail our choices with respect to the analyzed instances (verbal readings) and their aspectual annotation. Finally, we list the features recorded in the lexical resource that we consider relevant for the aspectual classification. In Section 4, we present our experiments. We firstly describe the experimental approach and its technical implementation, and then discuss the results obtained.

3. The Data

3.1. The Resource – LVF

The LVF – "Les Verbes Français" (François et al., 2007) is a detailed and extensive lexical resource providing a systematic description of the morpho-syntactic and syntactico-semantic properties of French verbs. The basic lexical units are readings of the verbs, determined by their defining syntactic environment (argument structure, adjuncts) and a semi-formal predicate decomposition (provided through lexical templates called *opérateurs*, that use roughly the same inventory of labels and features as in the lexical templates found in Pinker (1989) or Jackendoff (1983)).

The LVF covers roughly 12 300 verbs with a total of 25 610 usages. Each reading is associated with an elaborate morpho-syntactic and semantic description.

In the following Table 1, we use sample entries for the verb *élargir* 'widen' to illustrate LVF's basic layout and give an idea of the underlying lexicographic and representation choices.

Syntactic description (Table 1b). Each reading of a verb is coupled with a schematic representation of its acceptable syntactic frames. In principle, a verbal reading can be coupled with a transitive use, a reflexively marked use and an intransitive use unmarked by the reflexive, or be coupled with only a subset of these syntactic frames. Note that the reflexive is not systematically disambiguated between its different readings (anticausative, passive, purely reflexive, etc). However, the reflexive use is systematically associated with a particular semantic class, that indirectly provides some information on the range of readings the reflexive can have.

Besides defining the argument structure (and potential adjuncts), the syntactic description also specifies some thematic features of the main arguments (e.g. whether the subject and direct object are animate and/or inanimate, whether the indirect object refers to a location, etc).

Semantic description (Table 1a). Each entry in the LVF is also characterized by a semi-formal semantic decomposition, called *opérateur*, which provides a rough approximation of the verbal meaning instantiated on each reading. Each entry is therefore paired with a finite set of primitive semantic features and labels which, in a subsequent step, enable the classification of verbal readings into 14 well-established semantic classes (*eg. communication verbs*, *psych-verbs*, etc.).

The semantic features and labels used in the semantic decomposition provide other cues about the type of verbs which is instantiated across each reading. For instance, for the reading 01 of *élargir* in Table 1a, 'r/d +qt [p]' roughly corresponds to BECOME(**more**(p)) ('r/d' stands for '(make) become'; '+qt' stands for an increase along a scale). From this, one can safely infer that under reading 01, *élargir* is a 'degree achievement' verb (a change of state verb with a multiple point scale in Beavers' (2008) terminology). The semantic decomposition may also contain abstract prepositions — see e.g. 'VRS' ('TOWARDS') in reading 04 of *élargir* — that stand for relations between arguments.

To summarize, each reading is associated with two semantic descriptions: the semantic decomposition (the *opérateur*) and the *semantic class*.

Additionally to the syntactic and semantic description, the LVF also indicates when a verb is formed through a derivational process, and in the positive case, provides information about the category of the verbal root, thus enabling one to identify deadjectival or denominal verbs. Finally, for each entry is specified which suffix is used for the available reading-preserving deverbal nominalisations and adjectives (*-ment, -age, -ion, -eur, -oir, -ure* or zero-derived nominalisations, and *-able, -ant, -é* or zero-derived adjectives).

3.2. The Annotation

We extracted 1199 entries (verbal readings) for the selected 167 verbs mentioned earlier. On average, each verb has roughly 15 readings, while 50% have more than 13. There are only 16 unambiguous verbs, while the largest number of readings is 37 and 19 (for 1 respectively 2 lemmas). Most lemmas (29) have 3 readings.

Interestingly, the average number of 15 readings per verb very closely matches the number of event categories per verb obtained in the experiment reported by Marvel and Koenig (2015), who propose a new method of automatically categorizing event descriptions.[1]

The extracted verbal readings were manually annotated according to a fine grained aspectual classification on a 'telicity scale' of eight values. At the bottom of the scale are readings that are unambiguously stative (i.e. for which any other aspectual value is excluded), rated with 1 (STA). At the top are found achievement readings for which any other aspectual value is excluded, rated with 8 (ACH). At the middle of the scale are found predicates of variable telicity that can *equally* take a telic or atelic value within the context of the evaluated reading, rated with 4 (ACT-ACC). For instance, *élargir* under reading 01 is rated with 4, because, among others, both *for-* and *in-* adverbials are *equally* compatible with *élargir* on this reading. These predicates of variable telicity, that show no preference for the telic or the atelic reading on the given reading, are distinguished from 'weak accomplishment' readings, rated with 5 (W-ACC). Weak accomplishment readings are *primarily* telic, but nevertheless share some properties with activity readings (see Piñón (2006)). For instance, *remplir* 'fill' under its reading 01 (*Pierre remplit le seau d'eau* 'Pierre is filling the bucket with water') is analysed as a weak accomplishment rather than a verb of variable telicity (and therefore rated with 5 rather than 4). This is justified by the fact that under reading 01, *remplir* is more natural with an *in*-adverbial than with a *for*-adverbial. However, it is nevertheless still acceptable with *for*-adverbial under its 'partitive' reinterpretation (described e.g. by Smollett (2005) or Champollion (2013)). This makes it a 'weak' accomplishment rather than a 'strong' one. 'Strong' accomplishments — like *tuer* in its reading 01 (*Le chasseur tue un lapin* 'The hunter is killing a rabbit') —-, which are incompatible with the partitive interpretation of *for*-adverbials, are rated with 6 (S-ACC). Those accomplishments that share some properties with achievements are rated with 7 (ACC-ACH).

The annotator evaluated each entry with a *definite* internal argument, in order to abstract away from the role of the determiner in the identification of the aspectual value of the VP (see e.g. Verkuyl (1993)). Additionally, sentences were systematically evaluated with a present tense, so as to avoid the possibility that lexical (inner) aspect is modified through grammatical (outer) aspect by coercion (see e.g. De Swart (1998)).[2]

In this first study, however, we used a coarser grained aspectual scale focusing on telicity. We therefore grouped the verb readings into the following classes: ATElic (rating

[1]Their finding that a verb is roughly associated to more than 15 event categories has been further confirmed by a subsequent rater study contrasting event categories and a native speaker judgment task validating the results of this rater study; see Marvel and Koenig (2016).

[2]However, in order to check the compatibility of *for-* and *in-* adverbials, the annotator switched to perfective sentences, since those adverbials are not compatible with the present tense in their relevant interpretations.

id	example[a]	semantic decomposition	sem. primitive	sem. class
01	On *élargit* une route/ La route *s'élargit*.	`r/d+qt large`	become	Transformation
02	Cette veste *élargit* Paul aux épaules.	`d large`	become	Transformation
03	On *élargit* ses connaissances.	`r/d large abs`	become, figurative	Transformation
04	On *élargit* le débat à la polititque étrangère.	`f.ire abs VRS`	directed move, figurative	Enter/Leave

(a) The four readings illustrated by sample sentences and their semantic description

[a]Literal translations – 01: One widens a road. 02: This jacket widens Paul's shoulders. 03: One widens one's knowledge. 04: One extends the debate to foreign policy.

id	schema	encoded information
01	A30	intransitive with adjunct, inanimate subject
	T1308	transitive, human subject, inanimate direct object, instrumental adjunct
	P3008	reflexive, inanimate subject, instrumental adjunct
02	N1i	intransitive, animate subject, prep. phrase headed by *de* (*of*)
	A90	intransitive with adjunct, subject human or thing
	T3900	transitive, inanimate subject, object human or thing

(b) Syntactic descriptions

Table 1: LVF entries for *élargir*

1–3), with VARiable telicity for those with rating 4, and TELic (rating 5 or more).

Table 2 gives an overview of the distribution of the aspectual ratings. The first interesting finding is that verbs display a considerable aspectual variability across readings. The aspectual value of 2/3 of the 151 verbs with more than one reading (107) varies with the instantiated reading (on the 8 value scale). The greatest variation of 7 values on the telicity scale is attested for one verb. With respect to the coarser grained scale, roughly half of the verbs (69) has readings of only one of the three overarching aspectual classes.

3.3. The Features

For the design of the features for our classification we started from the features used in previous work (Zarcone and Lenci, 2008; Siegel and McKeown, 2000; Friedrich and Palmer, 2014) and looked for similar information in the LVF. Many of the features relevant for the aspectual value of utterances in context are obviously not encoded in the lexicon. This is in particular true of features related to outer/grammatical (i.e. non-lexical) aspect conveyed by tenses. This is also mostly true of the information conveyed by temporal adverbials, which also importantly contribute to the final aspectual value of an utterance in context. However, the LVF connects each verbal reading with specific morphological, syntactic and semantic features. Among such features, those that influence the lexical aspect of the verb in context are known to be pervasive: verbs encoding a result in their event structure are mostly accomplishment verbs; those encoding a manner but no result are mostly activity verbs (see e.g. Rappaport Hovav and Levin (1998) and subsequent work), verbs that subcategorise an indirect object are mostly bi-eventive, i.e. encode a result (see e.g. Pylkkänen (2008)) and are therefore accomplishments, intransitive verbs with no transitive use with a *-eur* nominalisation are most likely to be unergative verbs and therefore atelic, etc. Luckily, most of these features relevant for lexical aspect are coded in one way or another through the LVF. So for instance, if, in the semantic decomposition of the LVF, 'BECOME' ('r/d') is present, we know that a 'pure' activity reading is excluded; if a movement verb has a prepositional phrase that encodes a direction rather than a location, we know that it has at least an atelic reading, etc. Also, in some cases, the LVF's semantic decomposition even contains a temporal adjunct that determines the aspectual value of the verb under the given reading. For instance, 'tps long' or 'tps court' in the *opérateur* seems to systematically define an atelic reading.

Finally, some semantic classes give very clear hints to the lexical aspect of its members. For instance, readings instantiating the class of 'verbs of physical states and behaviors' are atelic, those instantiating 'enter/exit verbs' are telic, etc.

Overall we found that to a large extent, the information conveyed through the LVF features is comparable to Siegel and McKeown's (2000) and Zarcone and Lenci's (2008) linguistic indicators for lexical aspect. Most of the relevant linguistic information, in particular about the presence or absence of syntactic arguments and the nature of subject and/or objects could be derived from the descriptions.

Further semantic cues are conveyed by the semantic decomposition (the *opérateur*), in particular its main component (BECOME, DO, ITER, etc.) and features indicating the encoding of a scale or a manner component. Also, as we mentioned above, the semantic class to which the verb belongs on a particular reading sometimes suffices to aspectually disambiguate the verb in context.

The first theoretical hypothesis we aimed to test through the experiments reported here is that agentivity and abstractness influence lexical aspect. The second (related) one is that under their agentive and/or concrete (literal) readings, verbs are aspectually more flexible (i.e. show a higher range of aspectual values) than under their nonagentive and/or abstract (figurative) readings. Consequently, we were also interested in the interplay of lexical aspect with two other groups of semantic features, namely the agentivity of the external argument and the abstractness of the verb under its particular readings. The LVF does not allow

1	2	3	4	5	6	7	8
S-STA	ACT-STA	S-ACT	ACT-ACC	W-ACC	S-ACC	ACC-ACH	S-ACH
182	67	175	195	172	227	29	152

(a) 8-value scale

1-3	4	5-8
ATE	VAR	TEL
424	195	580

(b) 3-value scale

Table 2: Aspectual distribution of 1199 annotated verb readings

to directly determine the agentivity of an argument or the abstractness of a reading. However, it does provide some valuable cues that help to rate verbs along these two dimensions. Agentivity is for example suggested by the nature of the subject (animate or not). In addition, some of the semantic primitives used in the semantic decomposition as well as the semantic classes themselves suggest preferences for more or less agentivity.

Similarly, the nature of the internal argument also contributes to determine whether the verb is abstract or concrete for some verbs at least, that inanimate objects tend to be more abstract than animate ones (compare *kill Mary* and *kill time/ the bottle*). Also, the LVF systematically indicates whether the reading is literal and figurative, and often, figurative language positively correlates with abstractness. Finally, the semantic class is sometimes also indicative of more or less abstractness (for instance, when *tuer* 'kill' has a reading instantiating the class of psych-verbs, it is more likely to have abstract rather than a concrete sense, etc.).

To sum up, we extracted the following sets of features:

syn/sem "classic" features based on syntactic/semantic diagnostics. These include information about the syntactic frames, the argument structure and the potential adjuncts. These features also comprise the main components of the semantic decomposition and the general semantic class, as well as the information about the morphological make-up of the verb (including the information on the category of the verbal root and the suffixes used in derived nominalisations and adjectives), and finally some information about the degree of polysemy of the verb (through the relative number of entries for one lemma).

agent Features indicating (more or less) agentivity. Besides the animacy of the subject, we also use some of the main components in the semantic decomposition, as well as semantic classes positively or negatively relevant for agentivity. Additionally, we also consider the availability of *-age* and *-eur* nominalisations as indicative of agentivity.

abs Features relating to abstractness or concreteness. Here, we make use of the LVF's systematic classification in literal vs. figurative meanings, assuming that figurative readings are mostly abstract, while literal ones are mostly concrete. In addition, some components in the semantic decomposition indicate a preference for a concrete or abstract reading and some semantic classes (e.g. psych-verbs) indicate abstractness by themselves.

Table 3 presents some more details.

	#	sample feature
syn/sem	28	subcategorisation of an indirect object
agent	9	subject always inanimate
abs	10	literal or figurative sense

(a) Basic feature sets

	#features
agent-abs	19
syn/sem-agent	37
syn/sem-abs	38
syn/sem-agent-abs	47

(b) Combinations

Table 3: Feature sets.

4. Aspectual Classification

For the aspectual classification, we proceeded as follows. As described in Section 3, we extracted 1199 readings from the LVF. On one hand, these were manually assigned gold aspectual values on the telicity scale of eight values (as described in the previous section). On the other hand, we automatically collected features from the LVF for each of these readings. Based on these features and the gold aspectual values, we trained classifiers to predict the aspectual values of the readings, with regard to a coarser grained telicity scale of three values: ATElic, VARiable telicity and TELic. Finally, we evaluated the results by looking at the effects of the features on the overall classification accuracy and by analyzing their contribution to the identification of each of the classes.

Classifiers We assumed a supervised learning setting and applied the classifiers[3] shown in Table 4 with the implementation provided by Weka (Hall et al., 2009)[4] to predict the 3-way aspectual class of the 1199 verb readings (ATE, VAR or TEL). We measured the performance of the classifiers by assessing the accuracy in 10-fold cross-validation and compare it to the accuracy of a baseline classifier which always assigns the majority class (`rules.ZeroR`).

4.1. Results

Table 4 reports the classification results for the different feature sets (Table 3) and classifiers. The accuracy for the best performing feature set is marked in *italics* for each classifier (best accuracy in row) and in **bold face** for the best performing classifier for each feature set (best accuracy in column).

[3] We chose roughly one classifier from each class in the Weka framework.

[4] For libsvm (the SVM implementation) we used a linear kernel and normalization.

Algorithm	*syn/sem*	*agent*	*abs*	*agent-abs*	*syn/sem-agent*	*syn/sem-abs*	*syn/sem-agent-abs*
`trees.j48`	62.72	56.96	**57.80**	58.72	61.80	63.05	***63.64***
`rules.jrip`	63.47	54.55	54.63	56.88	**63.80**	***63.97***	62.80
`lazy.kstar`	***64.64***	**57.72**	56.63	**59.05**	62.89	62.30	60.55
`functions.libsvm`	*63.30*	56.13	55.21	56.21	62.72	63.05	62.30
`bayes.naivebayes`	*60.80*	55.38	54.96	55.80	60.22	60.00	59.30
`baseline`	48.37	48.37	48.37	48.37	48.37	48.37	48.37

Table 4: Classification results for LVF entries, using different classification algorithms and features sets. *Italics* mark best performing feature set for each classifier, **bold face** best performing classifier for each feature set.

We see that all feature combinations and every classifier outperform the baseline where all readings are labeled with the majority class (the telic class). The best accuracy is achieved using the *syn/sem* features and `lazy.kstar` classifier. Using agentivity features and abstractness/concreteness features alone allows one to achieve similar results of 6-10 points above the baseline, suggesting that these features do indeed influence the aspectual value. However, the results improve considerably (about 8 points) when using the syn/sem diagnostic features. Again, when combining syn/sem features with agentivity and/or concreteness/abstractness features, we obtain results close to those using syn/sem features only.
With respect to the classification algorithms, the best results were achieved for the `lazy.kstar` classifier. On the other hand, `libsvm` and `naivebayes` classifier appear less appropriate for this task since, in contrast to `j48` and `jrip`, they did not outperform the other classifiers for any feature set.
From these first set of experiments, we draw the following conclusions. Firstly, we found that we could classify the LVF verb readings with respect to their aspectual value based on features extracted from the LVF with reasonable accuracy. Secondly, our initial hypothesis that lexical aspect varies along the dimensions of agentivity and abstractness/concreteness is confirmed. However, agentivity and/or abstractness/concreteness alone are not sufficient to determine the aspectual value reliably: unsurprisingly, the syn/sem diagnostic features are a decisive factor for more accurate predictions.

4.2. Features favouring the overall classification performance

In the next experiment, we used the Wrapper (Kohavi and John, 1997) method to select the feature subset producing the best classification accuracy. In our setting, this methodology consisted in using the prediction performance of the `naivebayes` learning algorithm to assess the relative usefulness of feature subsets obtained by forward selection[5]. In the following paragraph, we briefly describe the features selected this way for each of the three basic feature sets and the combination of all of these three sets.

syn/sem 7 selected features: temporal and manner complement or adjunct, evaluative meaning component, reflexive subcategorisation frame besides a transitive frame.

agent 4 selected features: (in)animacy of subject, components in the semantic decomposition suggesting more or less agentivity.

abs 5 selected features: *opérateur* suggesting abstractness and concreteness respectively, literal reading, (in)animacy of direct object.

syn/sem-agent-abs 10 selected features: temporal and manner complement or adjunct, iterative meaning component, subject always inanimate, components in the semantic decomposition indicating agentivity, features indicating abstractness.

Table 5 presents the accuracy in 10-fold cross validation based on the reduced feature sets. The results show that by using the reduced feature sets, the accuracy could be improved by 2-3 points. The best performing feature set and classification algorithm are the same as before, namely the syn/sem diagnostic features and the `lazy.kstar` algorithm (in this setting, the accuracy is of 67.56%). This leads us to the conclusion that the selected features are the most important for an accurate classification. As described above, among these selected features, many are related to agentivity and abstractness or concreteness. This, again, supports our initial hypothesis.
However, while these experiments confirm the influence of agentivity and abstractness/concreteness on aspectuality, they do not lead to any conclusion about the orientation of this influence. We return to this issue in the course of the following section, dedicated to the feature analysis by class.

4.3. Feature analysis by class

Presentation. In this section, we present our application of a feature filtering method proposed in (Lamirel, 2013), based on a feature maximization metric called FMC[6]. FMC allows us to analyze which features are the most helpful at the identification of specific classes. More concretely, it gives an indication of which features are mostly indicative of atelicity, telicity or variable telicity. Being "class oriented", this method also allows one to better tease out in which way agentivity and abstractness/concreteness influence lexical aspect.
Lamirel's (2013) method defines how characteristic each feature is for each class by introducing class based feature

[5]For sets consisting of less than 10 features, we performed an exhaustive search; for the larger feature sets, this turned out to be too time consuming.

[6]Feature Maximization Clustering or Classification (FMC) was initially developed to assess the quality of clusterings.

Algorithm	*syn/sem*	*agent*	*abs*	*agent-abs*	*syn/sem-agent*	*syn/sem-abs*	*syn/sem-agent-abs*
`trees.j48`	63.55	56.00	**58.13**	56.55	63.00	*63.72*	62.47
`rules.jrip`	59.80	55.30	55.38	56.21	*61.55*	60.55	61.13
`lazy.kstar`	***67.56***	55.88	55.38	57.38	**67.47**	**65.56**	65.39
`functions.libsvm`	59.30	55.80	55.47	55.96	61.13	59.88	*61.22*
`bayes.naivebayes`	64.55	**56.55**	56.71	**58.55**	65.80	64.97	***66.14***
`rules.zeror`	48.37	48.37	48.37	48.37	48.37	48.37	48.37

Table 5: Classification results for LVF entries, using different classification algorithms and reduced features sets. *Italics* mark best performing feature set for each classifier, **bold face** best performing classifier for each feature set.

precision and recall measures. The precision of a feature f measures how discriminating f is for a class c, when compared to other features. Feature recall shows how characteristic f is of class c. Feature precision and recall are combined into feature f-measure (the harmonic mean). Only features with an f-measure above a threshold are selected. This method is independent of the classification algorithm and allows to rank the features by f-measure (or recall or precision) with regard to each class.

In order to be able to take into account feature values for non-numeric features (so that we can analyze, for example, if a feature with a particular value is more characteristic of a class), we transform non-numeric features into binary features. The following example shows how this is done. For each initial feature F having *e.g.* two values V_1 and V_2 — for example the feature *subj-always-inanim*, which can be `TRUE` or `FALSE` —-, two new features are introduced, i.e. $F = V_1$ and $F = V_2$ (`subj-always-inanim=TRUE` and `subj-always-inanim=FALSE` in our example). For an instance, the new feature $F = V_1$ gets assigned the value TRUE (or 1) if for this instance F has the value V_1 and FALSE (or 0) otherwise. Respectively, $F = V_2$ is TRUE if for this instance F has the value V_2 and FALSE otherwise. The feature selection is then performed on these transformed features.

FMC works as follows. First, the *local Recall* (FR_c^f) and the *local Precision* (FP_c^f) of a feature f in a class c are defined as follows:

$$FR_c^f = \frac{|i_c^f|}{|I^f|}, \quad FP_c^f = \frac{|i_c^f|}{|I_c|}$$

where i_c^f is the set of instances (readings) having feature f in c, I_c the set of instances in c and I^f, the set of instances with feature f. The *local F-measure* FF_c^f is the harmonic mean of local recall and local precision. A feature f is said to maximize a class c iff its feature F-measure is higher for that class than for any other class and F_c designates the set of features maximizing c. Finally, the feature F-measure FF_c of a class c is defined as the average of the feature F-measure of the maximizing features for c:

$$FF_c = \frac{\sum_{f \in F_c} FF_c^f}{|F_c|}$$

The set S_c of features that are considered characteristic of a given class $c \in C$ results in:

$$S_c = \left\{ f \in F_c | FF_c^f > \overline{FF}(f) \text{ and } FF_c^f > \overline{FF} \right\}$$

where

$$\overline{FF}(f) = \sum_{c' \in C} \frac{FF_{c'}^f}{|C_f|}$$

and

$$\overline{FF} = \sum_{f \in F} \frac{\overline{FF}(f)}{|F|}$$

C_f represents the set of classes for which the feature f is present for some instances and F the set of features.

The two thresholds $\overline{FF}(f)$ and $\overline{FF}$ can be interpreted as follows. $\overline{FF}(f)$ is a threshold pertaining to a feature f and represents an average over the classes of the "usefulness" of f for each of the classes, where the "usefulness" is given by the feature F-measure FF_c^f. The second threshold $\overline{FF}$ is an average of the local thresholds over the whole set of features.

According to this, a feature f is considered characteristic of a class c if, on one hand, its feature F-measure is more "useful" for this class than the average over classes and if, on the other hand, its contribution is greater than the average contribution of all features.

Finally, the set of selected features is the union of the features selected for each class:

$$S = \bigcup_{c \in C} S_c$$

Results and discussion. Table 6 presents the accuracy of the classifications with the features selected this way. The best performing configuration in this setting are the syn/sem features used together with the agentivity features and the `rules.jrip` classifier, achieving an accuracy of 64.80% similar to the accuracy in our initial configuration (Table 4). We can therefore conclude that it is meaningful to investigate the respective relevance of the selected features for the aspectual classes through FMC. In the best configuration, 42 binary features were selected. Among the selected syn/sem features, the most noteworthy ones are the following.

1. Availability of a canonical (non reflexive) passive.

2. Several features related to the derivational family; in particular, features indicating that the verb reading is not coupled with derived adjectives/nouns formed with particular suffixes, as well as features indicating that the verb is itself formed by derivation.

3. Presence or absence of a direct or indirect object.

Algorithm	*syn/sem*	*agent*	*abs*	*agent-abs*	*syn/sem-agent*	*syn/sem-abs*	*syn/sem-agent-abs*
`trees.j48`	59.55	56.96	**57.72**	57.72	61.72	*62.97*	59.88
`rules.jrip`	**63.64**	55.21	55.13	56.55	***64.80***	**64.72**	**62.97**
`lazy.kstar`	60.55	**57.72**	56.63	**59.13**	*60.80*	58.47	57.80
`functions.libsvm`	55.00	56.21	55.13	56.46	58.04	58.88	*60.13*
`bayes.naivebayes`	51.80	55.38	54.96	*56.13*	52.63	52.21	54.13
`rules.zeror`	48.37	48.37	48.37	48.37	48.37	48.37	48.37

Table 6: Classification results based on feature selection with the FMC method. *Italics* mark the best performing feature set for each classifier, **bold face** the best performing classifier for each feature set

Among the features related to agentivity, the most important one is an 'at-least-one-true' agentivity feature, that was satisfied as soon as at least one feature of agentivity was satisfied. Another selected feature of agentivity is the `non-agentive` feature (satisfied whenever any of the agentivity features indicates non-agentivity).[7] Again, these results confirm the importance and effectiveness of the syn/sem features as well as the influence of the agentivity features. However, they again do not provide further insights about the direction of this influence.

In order to investigate the features most characteristic for each of the three coarse-grained aspectual classes (ATE, VAR, TEL), we looked at those features which were selected as relevant for one of those classes only (recall that the FMC selected features are based on the union of the features considered most characteristic for each class).

These features are presented in Table 7.

Line ATE in Table 7 shows that many of the features most helpful at predicting the atelic class are also those which suggest less agentivity. These features are the absence of any argument other than the subject (i.e. the presence of a unary predicate) and the absence of a passive form. This result is not surprising, given that non-passivable intransitive forms are mostly unergative or stative verbs. Among the best indicators of variable telicity (line VAR in Table 7) are the inanimacy of the subject, the availability of a deverbal adjectival derivation and the fact that the verb is itself formed by derivation. This last feature may simply be an experimental confirmation of the well-known correlation between verbs displaying variable telicity and deadjectival verbs (see *widen, deepen*, etc., cf. Kennedy and Levin (2008) and references therein). The TELic class (line TEL in Table 7) seems the hardest to predict, since only two features were found to be characteristic for this class only, namely the absence of an indicator for non-agentivity and the presence of an indirect object. We do not have an explanation for the first feature, but the second one is less surprising, given that verbs subcategorizing an indirect object are very often result (causative) verbs (see e.g. Pylkkänen (2008)) and that result verbs are very often accomplishment verbs (see e.g. Rappaport Hovav and Levin (1998)).

In Table 8, we look at those features which are found to be characteristic for two aspectual classes and therefore indicative for a greater degree of aspectual flexibility. Among the seven features characteristic for two classes, one is related to agentivity – `non-agentive=FALSE` (Table 8a) – and two – `conc=TRUE` and `dobj-inanim=TRUE` (Table 8b) reflect abstractness/concreteness. The agentivity feature helps to identify two opposite classes (TELic and ATElic). This confirms in some way our hypothesis that agentivity positively correlates with aspectual flexibility (or, in other words, that predicates are aspectually more flexible under their agentive than under their nonagentive reading). On the other hand, features indicative of concreteness are characteristic both of the VARiable and TELic classes. This, also, confirms our hypothesis that verbs are aspectually more flexible under their concrete readings than under their abstract ones. There were no features indicative of both the ATElic and VARiable class. These preliminary findings support the hypothesis that verbs show indeed more aspectual elasticity under their agentive and concrete readings than under their nonagentive and abstract readings. In future work, we plan to refine the features reflecting agentivity and abstractness/concreteness in order to better investigate finer grained aspectual variation across readings of the same verb.

5. Conclusion

In this work, we focused on the way the aspectual value of French verbs varies across their different readings, as listed and described in a syntactic-semantic lexical resource. We used a machine learning approach based on features extracted from the lexicon. We showed that the extracted information enabled a fairly accurate prediction of the aspectual class. In addition, our results based on feature selection suggest that agentivity and abstractness also contribute to the selection of the correct aspectual value. Our results can also be analyzed as confirming the additional hypothesis that aspectual flexibility is higher when verbs are used under their agentive and/or concrete (literal) readings than under their nonagentive and/or abstract (figurative) readings.

6. Bibliographical References

Beavers, J. (2008). Scalar complexity and the structure of events. In John Dölling et al., editors, *Event Structures in Linguistic Form and Interpretation*. De Gruyter, Berlin.

Champollion, L. (2013). The scope and processing of *for*-adverbials: A reply to Deo and Piñango. In Todd Snider, editor, *Proceedings of Semantics and Linguistic Theory (SALT) 23*, pages 432–452. Linguistic Society of America.

De Swart, H. (1998). Aspect Shift and Coercion. *Natural Language and Linguistic Theory*, 16:347–385.

[7] However, for this feature, there is no preference with regard to the value (both the features `non-agentive=TRUE` and `non-agentive=FALSE` were selected).

François, J., Le Pesant, D., and Leeman, D. (2007). Présentation de la classification des Verbes Français de Jean Dubois et Françoise Dubois-Charlier. *Langue française*, 153(1):3–19, March.

Friedrich, A. and Palmer, A. (2014). Automatic prediction of aspectual class of verbs in context. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2:517–523.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18, November.

Jackendoff, R. (1983). *Semantics and Cognition*. MIT Press, Cambridge, Mass.

Kennedy, C. and Levin, B. (2008). Measure of change: The adjectival core of verbs of variable telicity. In Louise McNally et al., editors, *Adjectives and Adverbs: Syntax, Semantics, and Discourse*, pages 156–182. Oxford University Press, Oxford.

Kohavi, R. and John, G. H. (1997). Wrappers for feature subset selection. 97(1):273–324, 7.

Lamirel, J.-C. (2013). Enhancing classification accuracy with the help of feature maximization metric. In *2013 IEEE 25th International Conference on Tools with Artificial Intelligence, Herndon, VA, USA, November 4-6, 2013*, pages 569–574.

Marvel, A. and Koenig, J.-P. (2015). Event categorization beyond verb senses. In *Proceedings of the 11th Workshop on multiword expressions, NAACL 2015*, pages 569–574.

Marvel, A. and Koenig, J.-P. (2016). Microsenses: exploring multi-word event categorization. Paper presented to the workshop *Events in Language and Cognition*, held in conjunction with the 29th Annual CUNY Conference on Human Sentence Processing, University of Florida, March 2016.

Piñón, C. (2006). Weak and strong accomplishments. In Katalin Kiss, editor, *Event structure and the left periphery. Studies on Hungarian*, pages 91–106. Springer, Dordrecht.

Pinker, S. (1989). *Learnability and Cognition: The Acquisition of Argument Structure*. MIT Press, Cambridge, Mass.

Pylkkänen, L. (2008). *Introducing Arguments*. MIT Press, Cambridge, MA.

Rappaport Hovav, M. and Levin, B. (1998). Building verb meanings. In Miriam Butt et al., editors, *The Projection of Arguments: Lexical and Compositional Factors*, pages 97–134. CSLI Publications, Chicago.

Siegel, E. V. and McKeown, K. R. (2000). Learning Methods to Combine Linguistic Indicators: Improving Aspectual Classification and Revealing Linguistic Insights. *Computational Linguistics*, 26(4):595–628, December.

Smollett, R. (2005). Quantized direct object don't delimit after all. In Henk Verkuyl, et al., editors, *Perspectives on Aspect*, pages 41–59. Kluwer Academic Publishers, Amsterdam.

Verkuyl, H. (1993). *A Theory of Aspectuality : the Interaction between Temporal and Atemporal Structure*, volume 64 of *Cambridge studies in linguistics*. Cambridge University Press, Cambridge; New York.

Zarcone, A. and Lenci, A. (2008). Computational Models for Event Type Classification in Context. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).

class	features selected
ATE	dobj=FALSE, passive=FALSE, subj-only=TRUE, literal=FALSE
TEL	indobj=TRUE, non-agentive=FALSE
VAR	der-ment=TRUE, is-derivation=TRUE, subj-inanim=TRUE, has-adj-deriv=TRUE, is-derivation=TRUE

Table 7: Features characteristic for each of the three coarse-grained aspectual classes: ATElic, TELic and VARiable.

Feature set	Selected features	Feature set	Selected features
syn/sem		*syn/sem*	has-adj-deriv=TRUE, reflexive=TRUE
agent		*agent*	
abs		*abs*	conc=TRUE, dobj-inanim=TRUE
agent-abs		*agent-abs*	conc=TRUE, dobj-inanim=TRUE
syn/sem-agent	has-derivation=FALSE, non-agentive=FALSE	*syn/sem-agent*	reflexive=TRUE
syn/sem-abs	has-derivation=FALSE	*syn/sem-abs*	reflexive=TRUE
syn/sem-agent-abs	conc=FALSE, has-derivation=FALSE	*syn/sem-agent-abs*	reflexive=TRUE

(a) Features indicative of both the TELic and ATElic classes

(b) Features indicative of both the VARiable and TELic classes

Table 8: Features indicative of two aspectual classes