1. Aims

In this document, we present aspects related to generativity and to polysemy from a long-term ongoing research whose aims are (1) to define the contents of an electronic dictionary that represents and organizes the syntax and the semantics of predicative forms so that they can directly be used for NLP applications and (2) to specify and to evaluate methods for creating such a dictionary from corpora of various forms, domain specific thesaurus and paper dictionaries. We are developing methods and tools in information retrieval and automatic construction of abstracts of texts which require the extraction of predicate-argument structures (Pugnaud et al. 94). A detailed description of syntactically and semantically organised classes of predicates is therefore particularly useful and crucial to us. This type of information is moreover useful in a number of other applications.

We basically take into account two major sources of knowledge to create lexical entries: dictionary definitions and corpora. We also consider thesaurus associated with specific domains which make more explicit, even in an incomplete way, certain forms of information. Besides creating a dictionary for NLP, our goal is also to show, within the scope of our study, how the different sources of information interfere, and how much of the work can be automated, at what cost, and what should be left to a human expert. The importance of these criteria clearly depends on the granularity of the lexicon one wants to attain.

In this document, we first introduce the way we have organized and reformulated for French the verb semantic class system defined by B. Levin (Levin 93) for English; then, we show that verb semantic classes form a very good framework for structuring the verb selectional restriction system and the Qualia structure system. Finally, we show that verb semantic classes is a powerful means for organizing and for restricting the different forms of type coercion. In this paper only verbs will be considered. A similar work is under way for predicative nouns. In (Grishaw 90) relations between the argument structure of verbs and the argument structure of their corre-

2. The verb semantic class system

This work is primarily based on B. Levin's work (Levin 93) for English, where she shows that the syntactic behavior of verbs is essentially predictable from some aspects of their semantics. By syntactic behavior, she means the way arguments are distributed in a syntactic form with respect to the predicate, how they can move (e.g. to define ergative or passive forms) and when they can be deleted. These movements and deletions are called alternations.

Although B. Levin's work corresponds to a certain level of granularity in the linguistic description which has a certain degree of stability, we do not think that it really introduces a new level with a theoretical status in syntax or in semantics. Rather, we consider this work as a very useful, practical and relatively comprehensive one, which can form a good perspective and a good practical basis for the extraction and for the organization of lexical data.

2.1 Alternations revisited: a more declarative verb class system

We have substantially reformulated B. Levin system in order to avoid a number of difficulties inherent to her approach, and to make the system more declarative and more usable in NLP systems. We have also added several types of semantic information at both the class and the verb level, namely thematic grids and fragments of the Lexical Conceptual Structures (LCS) (Jackendoff 90) described in (Daubèze, Saint-Dizier, Marrafà 95). Our research is applied so far to French, but the techniques can be used for other languages in a similar manner.

Instead of considering the alternation system of B. Levin,
we have reformulated this notion into a more declarative one: we consider possible syntactic contexts for verbs. Then, we can assign to each verb-sense the set of syntactic contexts it may have. By syntactic context, we mean a subcategorization frame where the category and some additional syntactic and semantic information are used to describe the nature and the form of the possible complements and the subject of a verb (see examples below).

Verb classes are then formed from verbs having similar sets of contexts. Compared to B. Levin system, our approach avoids having to define a basic form from which alternations are produced; moreover it avoids us to have to account for the changes in meaning provoked by alternations (e.g. by the deletion of an argument or by the adjunction of a preposition). It is also easier to specify very precisely the form and the contents of every element in a context. Contents should however remain quite general in order to avoid an explosion in the number of contexts. In particular, selectional restrictions and the specification of prepositions must remain general. More refined contexts can be specified at the verb level, whenever necessary, provided that they remain coherent with the contexts specified at the class level. The level of generality considered at the class level is defined a priori from data commonly agreed upon to be of general purpose (e.g. feature values such as human, animate and object). Similarly, very limited exceptions can be accepted in order to preserve the homogeneity and the 'completeness' of the class (i.e. verbs with very similar meanings should a priori belong to the same class). As shall be seen below, the context system that we have defined and implemented provides us with a very powerful tool for specifying the syntax and the semantics of verbs.

Here is an example in Prolog of a verb class:

```
[202,914,51,33,1212,1112],
% lists of contexts (see below)
[abandonner,accorder,acheter,adjuger,allouer,
assigner,attribuer,ceder,choisir,conceder,
confier,decerner,distribuer,donner,octroyer,
offrir, preter, procurer, prodiguer, promettre,
racheter, remettre, transmettre],
[<np,th,tib],[ae,th,tiv]]).
% list of thematic grids
% exceptions: procurer +32, abandonner -914
```

**Thematic roles** are those defined in (Pugeault et al. 94), i.e.: ae (effective agent), th (theme), tib (beneficiary incremental theme), tiv (victim incremental theme).

**Contexts:** (for the sake of readability, only positive features are mentioned)

- **202:** `<np>, <Verb>, <np>, <pp(prep=a)>`
- **914:** `<np(+agent)>, <Verb(+refl)>, <np>`
- **51:** `<np>, <Verb>, <pp(prep=a)>, <np>`
- **33:** `<np>, <Verb>, <np> (pp not specified)`
- **1212:** `<np>, <Verb(+passive)>, <np>`
- **1112:** `<np(+theme)>, <Verb(+refl)>, <np>`

From the point of view of polysemy, we feel that a verb class permits us to define some of the syntactic and semantic boundaries of a verb-sense. A verb in a class corresponds to a unique sense, with a set of syntactic realizations. This approach provides verb-senses with a certain stability, even if the different contexts they may have could alter to some extent their meaning. These possible alterations define a family of 'variations' around the central sense of the verb. Different senses having a priori different syntactic realizations, they will be in different classes.

So far, we have defined semantic classes for about 2000 verbs for French (Daubee 94). These classes have been defined by a combination of a manual and an automatic analysis of texts, technical corpora and dictionary definitions. These classes have been defined on the basis of 48 different contexts.

### 3. Type selection, Qualia structures and verb semantic classes

The well-known, basic idea of the Generative Lexicon is to avoid long enumerations of possible arguments for predicates and to develop a general type shifting system, based on the notion of Qualia structure (Pustejovsky 93). These enumerations indeed hide the fundamental distinction between basic and derived meanings.

In the Generative Lexicon, the different aspects (or facets) of the meaning of a word (and of its associated type(s)) are summarized in a structure called the Qualia structure, composed of four main roles (which can possibly be further specialized): the formal, constitutive, telic and agentive roles. To each verb (and more generally, to each potential governor) are associated one or more subcategorization frames which indicate the type of the arguments the verb selects. In this document, we have directly adopted the Qualia system defined by James Pustejovsky. However, other, more or less derived, forms of Qualia structures could also be considered.

Type coercion can be viewed as a relation between a predicate and one of its arguments, where the argument does not fit exactly the subcategorization expectations of the verb (neither directly nor by type subsumption). In that case, the type of the argument must be coerced to another type, coherent with the subcategorization expectations of the verb, for the sentence to be well-
formed. This new type associated with the argument is derived from one of the predicates present in a Qualia role of the argument. Determining such a new type is called type pumping. This operation has received an operational semantics in logic programming in (Saint-Dizier 95).

3.1 Factoring out selectional restrictions

Verb semantic classes can be used to factor out the type selected by every verb in the class for each of their arguments. For example all the verbs of the aspectual class select for an object of type event, possibly further subdivided into process or transition, depending on the syntactic form of the verb (Raising or Control, see (Pustejovsky and Bouillon 94)). Similarly, all the verbs of the amuse class select for a subject of type event.

Verb semantic classes allow us to go beyond the aspectual types as exemplified above and any type of selectional restriction can be factored out at the class level. The selectional restrictions associated with each of the arguments of the transfer of possession verb class are for example: human for the external argument, object for the direct object and human for the third argument.

3.2 Specifying Qualia roles by means of verb semantic classes

Semantic verb classes can also be used to specify the contents of Qualia roles. Indeed, in most cases, a whole family of verbs, semantically related, can appear in a role instead of just one element of the class. For example, in the case of book, instead of just having the predicate write (écrire) in the agentive role of the QuaILs, or instead of having the whole list of possible verbs, it is more convenient, and more linguistically appropriate, to refer to verb classes, whenever possible. As a consequence, for the agentive Quaila role of book we have the class of Communication by message verbs (including verbs such as: cahier, écrire, rédiger, corriger, annoter, té légraphier, traver, transcrire, etc.). This can be noted in the Qualia structure as follows:

\[
\text{book}(X),
\]

\text{qualia:}

\[
\text{agentive: } \text{Pred}(Y,X) : \text{ 'communication by message'}
\]

where \text{Pred} is a variable typed 'communication by message'. All the predicates of this class bear the same type (e.g. event, or communication event); this makes type shifting simpler and more general since the type of the class is considered instead of the type of each verb in the class.

Similarly, the Qualia structure of the agentive Qualia role of the noun loan is of the form:

\[
\text{loan}(X)
\]

qualia:

\[
\text{agentive: } \text{Pred}(Y,X) : \text{ 'future having'},
\]

where future having verbs include verbs such as: accorder, garantir, assigner, adjuger, attribuer, ceder, confier, décerner, etc. (grant, guarantees, assign, cede, etc.). Therefore, in a construction such as:

\[
\text{accepts the loan,
}\]

it will possible, via type coercion to derive an appropriate type 'future having event' from the agentive role of loan and to paraphrase the above VP for example as follows:

\[
\text{accepts to grant/guarantee/assign... the loan.}
\]

4. Type coercion and verb semantic classes

4.1 Introduction

Depending on the semantic class the predicate belongs to, and for each argument, only some applications of type coercion are linguistically appropriate. This is what we call in (Saint-Dizier 95) the appropriateness function which restricts the scope of type shifting. For example, in a sentence like begin a novel only the telic and the agentive roles of novel are relevant. Begin is indeed an aspectual verb and it basically selects an object of type event. This consideration can be dealt with a certain degree of generality using the verb semantic classes. The example of begin can indeed be generalized to all verbs of the same class (such as cease, commence, end, finish, repeat, the begin verb class of B. Levin). Let us now characterize in a more general manner the relations between verb classes and type coercion.

We model the appropriateness function by means of qualia role selectional constraints. Let us integrate appropriateness in a way quite similar to appropriateness in type feature systems (TFS). Let \text{App} be a partial function mapping the set of shifting operations \(\Sigma^1(t_i)\) defined from type \(t_i\) of the argument A to the most general set of types \(\Gamma\) may be coerced to with respect to the semantic class \(s\) of the predicate P. Let \(Q\) be the set of all Qualia structures, then if \(S\) is the set of all semantic classes \(s\), then the appropriateness function is an application from \(T x Q x S\) into \(T\). Let us note the appropriateness function as follows:

\[
\Gamma^1(s,t_i) = \text{App}(\Sigma^1(t_i), s) \subseteq \Sigma^1(t_i)
\]

\text{App} selects aliases in \(\Sigma^1(t_i)\) for which coercion is appropriate.

The appropriateness function can be applied recursively when several type shifting operations are necessary. It is then recursively defined from the
appropriateness set of the preceding level and the
semantic type $s_l$ of the word being considered at this
level: $\Gamma^2(s_l,t_l) = \Gamma^2(s_1, \Gamma^1(s_1,t_l))$
and, more generally:
$\Gamma^n(s_l,t_l) = \Gamma^n(s_1, \Gamma^{n-1}(\ldots \Gamma^1(s_1,t_l)))$.

The construction of $\Gamma^0$ is monotone increasing. The set
of all possible appropriate types is noted $\Gamma(s_l,t_l)$. This set
defines the denotational semantics of type shifting in a
more restricted manner since we have:
$\Gamma(s_l,t_l) \subseteq \Sigma(t_l)$.

Type coercion is defined on this set.

4.2 Restrictions on type coercion application
and verb semantic classes

A first result of our investigations is that all the verbs of
a given class are subject to the same set of type
coercions. By 'type of coercion' we mean coercion on the
agentive, telic, formal or constitutive roles of the
argument. There are very few exceptions in French which
are expressions which are either totally or partly
incorrect. Let us consider the delay verb class, composed
of the verbs: ajouter, remettre, repousser and retarder. All
these verbs admit the following type coercions:
- coercion on the telic or agentive roles of the subject
argument:
  La neige ajoute le train --> la tombée de la neige
  ajoute le train (the (falling of the) snow delays the
  train).
Similarly, we can say:
La neige ajoute/remet/repousser/remet la réunion.

However, the use of remettre is slightly less usual than
the three other verbs in this context.
- coercion on the constitutive or formal roles of the
subject argument:
  La banque repousse/remet/ajouer/remet la réunion -->
  le directeur de la banque repousse/remet/ajouer/remet
  la réunion
  (the (director of the) bank postpones/delays... the
  meeting).
- coercion on the constitutive or formal roles of the
object:
  Jean repousse/remet/ajouer/remet la réunion -->
  Jean repousse/remet/ajouer/remet la date de la
  réunion (John postpones the (date of the) meeting).
- coercion on the telic or agentive roles of the object:
  Jean repousse/remet/ajouer/remet la décision -->
  Jean repousse/remet/ajouer/remet la prise de la
  décision. (John postpones the (taking of the) decision).

Let us now consider the array given in the annex. This
array thus defines in extension the appropriateness
function introduced above. On a sample of verb classes,
this array shows, for each verb class, the type of
coercions which are appropriate. For example, for the
first class (the grant verb class), the subject (S) and the
second object (O2) arguments can be subject to type
coevolution on either their constitutive or formal roles. The
telic and the agentive roles are relevant for the object.

In this array, we can first notice that almost any class
of verb accepts type coercion on the constitutive or
formal role of the subject argument and on one of the
object arguments (O1 or O2). This is not very surprising
since it is very frequent to use a part of an object or a
more general term to refer to that object. For example
an institution is used for its members, its staff or its
manager (typical cases are e.g. the following classes:
grant, read, talk, report, ask and debate). However, this
type of coercion is naturally not possible when the
semantics of the predicate requires the argument to
precisely indicate what is at stake. This is the case for
verbs of the following classes: forbid, save, differentiate
and refund (for the object(s) argument(s)).

Next, we can also notice that verbs having a certain
degree of aspectuality can have a coercion on the telic or
agentive roles of their object argument. Several classes of
verbs are concerned. This situation is due to the fact that
a number of verbs select an object NP which is either an
object or an event or state. This kind of ambiguity allows
a simple type coercion from the object type to the event
type. This is the case for example of the following verb
classes: report, ask, debate, forbid and organise. The same
situation occurs for subject arguments of a number of
other verbs such as: hasten, delay, benefit, save, change
and improve.

Finally, and more generally, we can notice that there
are basically two families of type coercion: coercions
involving agentive or telic roles and coercion involving
formal or constitutive roles of the verb's arguments.
These two families correspond to substantially different
forms of metonymies.

To conclude this section let us consider a few examples
that will illustrate the array given in the annex. Verbs of
the amuse class accept type coercion on the agentive or
the telic roles of the subject argument as in:
Le clown amuse les enfants

which should be read as:
Les actes du clown amusent les enfants

(The clown amuses the children --> The actions of the
clown amuse the children).

Put verbs accept type coercion on the constitutive or
formal role of their object argument (or container-
containee if there is such a role): for example:
mettre l'eau sur la table

--> mettre la carafe d'eau sur la table
(Put the water on the table -> put the jug of water on the table).

We have the same phenomenon for e.g. removing and push-pull verbs.

4.3 Type coercion and context variation

Another important result is that alternations do not a priori preserve the possibility of using type coercion on the agentive or telic roles of moved argument(s). Reformulated in our approach based on contexts instead of alternations, this means that type coercion does not apply uniformly to any context of a given verb. The results given in the annex are based on the form which is commonly admitted to be the ‘basic’ form, or the most usual form of the predicate.

Therefore, sentences which are well-formed because type coercion can be applied on their ‘basic’ form, may no longer be acceptable in other contexts. This is the case, for example, in:

Jean commence un livre. (John begins a book), standing for:

Jean commence l’écriture d’un livre, but
* Un livre commence is illformed, whereas:
L’écriture d’un livre commence is correct.

The above example shows that type coercion on either the telic or the agentive roles of the object argument of the verb may be no longer possible when that object is moved in subject position (See (Pustejovsky and Bouillon 94) for a study of the semantics of Begin). We cannot however draw any general conclusion from the above example. The relations between contexts and the possibility of applying type coercion seem to be very complex and related to ‘deep’ semantic issues of predicates and to the relation they have with their arguments. It is, for example, perfectly correct to say:

Le film commence,
whereas:
* Le livre commence.
This difference may be due to the fact that the projection of a film is a well-defined event, with a precise beginning, duration and end, whereas for a book we don’t know exactly what kind of event it is related to and what its exact duration is. According to (Pustejovsky and Bouillon 94) the rejection of a form like Le livre commence is due to the telicity of the event taken by the complement (book). This hypothesis sounds reasonable, but still remains to be characterized in more depth.

Type coercion on the formal or constitutive roles seems to be less sensitive to context variations. In particular, type coercion on subjects are not sensitive to context variations. Similarly, type coercion on agentive or telic roles of the subject does not seem to be sensitive to context variations, consider the above example:

Le clown amuse les enfants --> Les enfants sont amusés par le clown (passive alternation).

(Le clown amuses the children)
This latter sentence standing for:

Les enfants sont amusés par les actions du clown.
(Children are amused by the actions of the clown).

From that point of view, we may consider that this latter type of coercion is more superficial (in terms of distance between the original and shifted types) than the coercion based on telic and agentive roles. Besides alternations, it should be noticed that there are other syntactic effects that may also limit the application of type coercion.

5. Conclusion

In this document, we have proposed a more declarative reformulation of verb semantic classes as defined by B. Levin, based on the notion of syntactic context. We have applied these ideas to the construction of a lexical knowledge base of about 2000 verbs in French.

Next, we have shown that the notion of semantic class is a very strong criteria for organizing the semantics of verbs (and of predicates, more generally). Verb semantic classes can also be used to specify the contents of Qualia roles in the Generative Lexicon and to restrict the application of coercion to appropriate cases, linguistically motivated.

This paper contributes to the specification of what a word-sense is and how its boundaries can be practically defined. This specification allows for a certain flexibility in the definition of a ‘sense’; it also specifies its different syntactic uses.

Finally, this work is a contribution to the definition of Qualia structures and to the practical use of type shifting on a large scale. Defining precise methods for writing Qualia structures on a large scale and for controlling the power of type shifting remain indeed largely open problems. They deserve a lot of attention since they are probably two of the main corner stones of the Generative Lexicon principles. Their study is a necessary condition to the practical use of the Generative Lexicon in ‘real’ applications.

Acknowledgements

I am very grateful to James Pustejovsky, Federica Busa, and Palmira Marrafa for discussions on previous versions of this work, which helped improving the ideas and the results presented here.
Bibliography

Daubèze, S., Saint-Dizier, P., Marrafa, P., Constructing a knowledge base for representing the general semantics of verbs, under submission, 1995.


Levin, B., English Verb Classes and Alternations, the University of Chicago Press, 1993.


Sample verbs of the class considered

<table>
<thead>
<tr>
<th>Telic role</th>
<th>Agentive role</th>
<th>Constitutive role</th>
<th>Formal role</th>
</tr>
</thead>
<tbody>
<tr>
<td>(grant) accorder, acheter, assigner, allouer, attribuer</td>
<td>O1</td>
<td>O1</td>
<td>S, O2</td>
</tr>
<tr>
<td>(provide) fournir, livrer, procurer, servir</td>
<td>O1</td>
<td>O1</td>
<td>S, O2</td>
</tr>
<tr>
<td>(receive) accepter, acquérir, recevoir, attendre, obtenir</td>
<td>O1</td>
<td>O1</td>
<td>S, O2</td>
</tr>
<tr>
<td>(withdraw) enlever, extraire, ôter, placer, prélever</td>
<td>S, O1, O2</td>
<td>S, O1, O2</td>
<td>S, O1</td>
</tr>
<tr>
<td>(read) lire, déclamer, dicter, narrer, prêcher</td>
<td>S, O1</td>
<td>S, O1</td>
<td>S, O1</td>
</tr>
<tr>
<td>(talk) bavarder, discouvrir, discourir, papoter, argumenter</td>
<td>O1</td>
<td>O1</td>
<td>S, O1</td>
</tr>
<tr>
<td>(report) annoncer, communiquer, rapporter, relater</td>
<td>O1</td>
<td>O1</td>
<td>S, O1</td>
</tr>
<tr>
<td>(ask) demander, quêmander, implorer, solliciter</td>
<td>O1</td>
<td>O1</td>
<td>S, O1</td>
</tr>
<tr>
<td>(debate) débattre, discuter, marchander, négocier</td>
<td>O1</td>
<td>O1</td>
<td>S, O1</td>
</tr>
<tr>
<td>(forbid) défendre, interdire</td>
<td>O1</td>
<td>O1</td>
<td>S, O1</td>
</tr>
<tr>
<td>(organize) organiser, renouveler, structurer</td>
<td>O1</td>
<td>O1</td>
<td>S, O1</td>
</tr>
<tr>
<td>(amuse) amuser, charmer, absorber, captiver</td>
<td>S</td>
<td>S, O1</td>
<td>S, O1</td>
</tr>
<tr>
<td>(please) plaire, déplaire, aimer, apprécier</td>
<td>S</td>
<td>S, O1</td>
<td>S, O1</td>
</tr>
<tr>
<td>(admire) admirer, adorer, aimer, envier, haïr, vénérer</td>
<td>O1</td>
<td>O1</td>
<td>S, O1</td>
</tr>
<tr>
<td>(hasten) accélérer, activer, hâter, presser</td>
<td>S, O1</td>
<td>S, O1</td>
<td>S, O1</td>
</tr>
<tr>
<td>(delay) ajourner, remettre, repousser, retarder</td>
<td>S, O1</td>
<td>S, O1</td>
<td>S, O1</td>
</tr>
<tr>
<td>(benefit) bénéficier, profiter</td>
<td>S, O1</td>
<td>S, O1</td>
<td>S, O1</td>
</tr>
<tr>
<td>(contribute) compter, contribuer, influer, dépendre</td>
<td>S</td>
<td>S</td>
<td>S, O1</td>
</tr>
<tr>
<td>(save) épargner, renouveler, économiser, amortir</td>
<td>S</td>
<td>S</td>
<td>S, O1</td>
</tr>
<tr>
<td>(change) changer, transformer, métamorphoser</td>
<td>S</td>
<td>S</td>
<td>S, O1</td>
</tr>
<tr>
<td>(improve) améliorer, amender, améliorer, redresser</td>
<td>S</td>
<td>S</td>
<td>S, O1</td>
</tr>
<tr>
<td>(differentiate) différencier, discriminer, distinguer</td>
<td>S</td>
<td>S</td>
<td>S, O1</td>
</tr>
<tr>
<td>(refund) dédommager, indemniser, rembourser, payer</td>
<td>S</td>
<td>S</td>
<td>S, O1</td>
</tr>
</tbody>
</table>

Annex : Sample of verb classes and appropriateness of type coercion w.r.t. Qualia roles of the arguments

1st column: between brackets is an English translation of the most representative verb of the class
2-5th columns: S = subject, O1 = object1, O2 = object2 of 'basic' forms.