A Constraint-Based Model for Preposition Choice in Natural Language Generation

Véronique Moriceau and Patrick Saint-Dizier

Institut de Recherches en Informatique de Toulouse, IRIT,
118 route de Narbonne, 31062 Toulouse Cedex, France.
moriceau@irit.fr, stdizier@irit.fr

Abstract. In this paper, we show how a constraint-based approach influences the modelling of preposition lexicalization in natural language generation. We concentrate on the linguistic description, which is the most challenging. The CSP procedures themselves are then rather straightforward. Preposition choice depends on the verb and its requirements, on the one hand, and the characteristics of the NP the preposition heads, on the other hand. These dependencies may induce somewhat contradictory expectations. A Constraint-based approach introduces a declarative model of this complex relation, allowing to identify all the possible lexical choices which can be predicted from lexical descriptions.

1 Introduction

With the development of WEBCOOP (Benamara et al. 2003), a system that produces cooperative answers to queries, the production of accurate natural language (NL) responses becomes essential. This implies great care in the taking into account of both user parameters (e.g. the terms used in the query) and linguistic constraints to produce a response which is accurate in its contents and as fluid and as well-formed as possible from a language point of view.

1.1 WebCoop output forms

The outputs of the WEBCOOP reasoning component are logical formula that encode conceptual representations of a fine-grained level. These representations are true decompositional semantics representations; they are language independent, allowing thus a quite large freedom to the NL generator, so that it can take into account a variety of parameters, such as: user prefered terms, homogeneity of style, complexity of structures, etc. Representations are a conjunction of several types of predicates:

- those that type objects or variables (for example, *flight(X), city(paris)*),
- those that denote relations (such as prepositions: *from(loc,X,Y)*) and
- those that describe events (*visit(E,X,Y)*).

As an illustration, the output of WEBCOOP that says that *all hotels in Cannes have a swimmingpool* is:
$hotel(X) \wedge in(loc, X, cannes) \wedge equipment - of(X, Y) \wedge swimmingpool(Y)$
Within such a decompositional approach, preposition choice is, in general, really difficult to handle. The choice of the preposition *in* in the above example is not straightforward. Most prepositions are indeed heavily polysemic, making lexical choices highly contextual. Elaborating constraints to model lexical choice performances realized by humans requires a detailed analysis of the parameters at stake and how they interact. This is the main aim of this contribution.

1.2 Facets of lexicalization

Lexicalisation is the operation that associates a word or an expression to a concept. It is a major parameter in response production (Reiter and Dale, 1997), (a good synthesis can be found in (Cahill, 1999)). Lexicalisation is often decomposed into two different stages: **lexical choice**, which occurs during content determination, where the lexical term chosen, which may still be underspecified, is dependent on reasoning procedures, the knowledge base contents, and grammatical constraints; and **lexical variation** which is the choice of a particular word or form among possible synonyms or paraphrases. Lexical variation occurs in the surface realizer, and may have some pragmatic connotations, for example an implicit evaluation via the language level chosen (e.g. argotic entails low evaluation). In this paper, we are mainly concerned with lexical choice. Some simple forms of lexical variation occur when two or more prepositions have exactly the same behavior w.r.t. to the lexical description, which is rather unfrequent.

1.3 Scope of the investigation

In this paper, we propose a declarative model that handles the complex relation verb-preposition-NP and a correlate, the relation deverbal noun-preposition-NP, which has a quite similar behavior. We concentrate on the linguistic description, which is by far the most challenging. This modelling and the subsequent NL performances are obviously heavily dependent on the accuracy and the adequacy of the linguistic descriptions provided, in particular lexical, including a number of usage variations like metaphors, which abound in preposition uses. The power of our model mainly lies in the way linguistic descriptions are organized and processed, in a way that avoids making too early and definitive choices. Thus, in our approach, CSP brings a new way (1) to organize and formulate lexical structures so that they can be adequtely used, and (2) then to state co-occurence constraints between sets of potential candidates.

We investigate PP constructions which are compositional. We also consider verb constructions where the verb incorporates the preposition, as, e.g. *climb* which incorporates the preposition *up* by default. This verb requires an explicit preposition otherwise, as in *climb down*. We consider that in this type of construction the preposition is both part of the verb (as a particle) and of the head of the localization NP that follows. We exclude true verb-particle constructions,

however not common in Romance languages, but frequent in Germanic languages (e.g. move up, clear up), and fixed forms (e.g. boil down), which are best treated as single units.

Our study being carried out on French, most of our examples are in French with approximate English glosses.

2 The conflictual relation verb-preposition-NP

2.1 A motivational example

Let us first illustrate the problem by a simple example. Linguistic notions involved being quite complex, these will just be presented here, without any deep analysis and justifications. A more detailed analysis is given gradually in the next sections.

The formula:
$person(jean) \wedge sea(X) \wedge to(loc, jean, X)$,
which is a simplified form of the Lexical Conceptual Structure (Jackendoff 1990) used in WEBCOOP, conveys the idea of someone (jean) going to somewhere (the sea). If we want to construct a proposition out of this formula, then the expression R = $to(loc, jean, X)$ is selected as the only predicate that can be realized as a verb. This form is quite abstract, it characterizes a priori any verb from the class 'movement verb to a certain location', movement being induced from the conceptual domain 'loc'. Verbs of this class are, for example: *aller, marcher, se déplacer, monter, descendre, pousser*, etc. (go, walk, move, ascend, descend, push), most of which incorporate manners and/or means. The same expression, R also originates the production of the preposition required by the verb for its object, of type fixed localization (Cannesson et al. 2002). The two arguments required by the verb are a priori quite neutral, typed 'human' for *jean* and 'place' for *sea*.

The choice of the preposition depends on the verb, but also on more pragmatic factors. Let us examine the above example in more depth. Since there are no explicit manners or means in the formula, let us select a verb which is as neutral as possible. In the lexicon, a verb such as *aller* (go) basically requires a preposition of type 'fixed position' since *aller* already incorporates the notion of movement, but not all such prepositions are acceptable, e.g. *à* is acceptable but not *pour*. In contrast, a verb like *partir* accepts both. *Marcher* would rather select *sur, dans, à coté de* as fixed positions, characterizing the space in which movement occurs. This situation is not proper to movement verbs, in (Mari and Saint-Dizier 2003), we show that the same kind of problems are raised by prepositions denoting instrumentality.

2.2 Stability over deverbals

Note that, although there is a quite large stability, prepositions acceptable for the relation verb-preposition-NP are not systematically acceptable for the relation

deverbal noun-preposition-NP although the semantics of the phrase is the same. For example, the preposition *pour* is not acceptable in ** l'avion vole pour Paris* (the aircraft flies to Paris) whereas it is acceptable in *le vol pour Paris est complet* (the flight to Paris is full) - the semantics of both phrases conveyed by the preposition semantic representation being the same.

Conversely, the preposition *à* (to) is acceptable in *le vol va à Paris* (the flight goes to Paris) whereas it is not in ** le vol à Paris est complet* (the flight to Paris if full). In fact, the verb *aller* (go) incorporates a notion of destination which explains why the preposition *pour*, denoting a notion of trajectory towards a goal, is not acceptable because, among other considerations, of the redundancy it introduces. The preposition *à* is more neutral in this respect and can be combined with *aller*. In the case of the relation deverbal noun-preposition-NP, the preposition used must denote a destination this is why only *le vol pour Paris* is acceptable. Here *vol* simply denotes the event of a flight, it does weakly incorporate a notion of trajectory, but this notion needs some further lexicalization, as also shown in: *a climb up the hill*.

2.3 Derived uses

In addition, prepositions that indicate a direction, which can be interpreted (as a kind of metonymy) as denoting an area relatively well delimited in the far, are also acceptable: *aller vers la mer* (go towards the sea). We call this phenomenon semantic coercion. Finally, a number of movement verbs also accept a more complex object structure, of type trajectory, including source and vias, but this structure often needs to be fully instantiated.

Note that depending on the object type and on the type of movement induced by the verb, other fixed position prepositions are possible: *monter sur la chaise* (to go on the chair), *aller sous le pont* (to go under the bridge). The choice of the most prototypical preposition (which does not exclude others) crucially depends on geometrical parameters of the object, or on prototypical uses. Finally, pragmatic factors may interfere (pragmatic coercion). For example if Jean is far from the sea, a preposition denoting a destination, interpreted as defining an area, is possible: *marcher vers la mer* (walk towards the sea), whereas it is not so appropriate if Jean is close to the sea.

Movement verbs, in particular, are subject to a large number of metaphorical uses on their object PP: *rentrer dans un projet (to enter into a project), passer sur un problème (to pass over a problem), monter dans les sondages (to rise in polls)*. In that case, the preposition choice heavily depends on the metaphorical interpretation of the object. For example, a project as an activity to be in, similarly to a place, or polls as a ladder to climb. We treat metaphors as a type coercion operation that changes the type of the object NP into a type appropriate for the preposition, that translates the meaning of the PP (Moriceau et al. 2003). In this paper, we consider the result of such type coercion operations, assuming they are specified in the system, in conjunction with the lexicon.

2.4 Cocomposition

A final point in the interaction verb-preposition-NP is the mutual compositional influence of the verb on the PP and vice-versa, called 'co-composition' in the generative lexicon (Pustejovsky 1995). This phenomenon remains compositional in its essence even if it requires non monotonic treatments of semantic representations. A case such as *aller contre un principe* (to go against a principle) preserves the idea of progression of the verb while *contre* overrides the fixed position naturally expected as the PP of the verb, to introduce a complex meaning which can be schematized as Opposition to a fixed, abstract position (here, a principle). Co-composition is still quite an open research field, essentially linguistic, which we will not discuss here.

3 Constraint domains

In this section, we present in more depth an analysis of the verb-preposition-NP complex relation, modelling constraints imposed by each protagonist in terms of domains of potential realizations. The basic idea is to produce sets of structures in which each element includes all the parameters necessary for expressing constraints and for making appropriate lexicalizations. In section 4, we show how constraints expressed as equations can be stated, so that a priori all potential acceptable preposition lexicalizations can be selected. Lexicalization decisions are thus delayed as much as possible avoiding useless backtracking and also allowing for the taking into account of long-distance dependencies.

Starting from semantic representations, let us first investigate the domains associated with each element in the formula, that corresponds to the following constructs to be determined and lexicalized: the verb (or the deverbal), the preposition and the NP. In section 4, we show how they interact.

3.1 Verbs

Our description of French verbs (Saint-Dizier 1998) heavily relies on the conceptual categories defined in WordNet. implemented a three level classification which seems to be sufficiently accurate to deal with preposition choice, as far as the verb is concerned. For example, for movement verbs, we have the following hierarchy:

1. 'local movement' (dance, vibrate)
2. 'movement': specified arrival, specified departure, trajectory (go)
3. 'medium involved': land, water, air/space, other types of elements (swim, skate)
4. 'movement realized by exerting a certain force': willingly or unwillingly (push)
5. 'movement with a certain direction': upwards, downwards, forward, backward, horizontal (rise, ascend, climb)
6. 'movement accompanied' (carry)

7. 'beginning/stopping/resuming/amplifying a movement' (stop, accelerate)
8. 'activity related to a movement' (drive).

We have a total of 292 verb classes organized around 17 main families borrowed from WordNet (e.g. verbs of body care, possession, communication, change, etc.). Each lower level class is associated with a by-default subcategorization frame (corresponding to the prototypical verb of the class) that describes the structure of the verb arguments, in particular the type of the preposition and of the argument normally expected. Each class also receives a by-default LCS representation, which can then be adapted to each verb, to capture some sense variations or specializations. For example, the movement verb with specified arrival subclass is represented as follows:

```
[ class: movement with specified arrival
prototypical verb: aller (go)
subcat: [X: NP(human), Y: PP(localization)]
lcs: to(loc, X,Y) ]
```

As defined in (Cannesson et al. 2001), `to` is a primitive that potentially covers a number of surface realizations.

Verbs often occur in different classes, even for a given sense: within a generative approach like ours, this characterizes the different facets the verb may take, as well as some of its frequent metaphoric uses. In fact, some very frequent metaphoric uses may get the status of a specific sense, with its own autonomy.

Verbs that incorporate a preposition by default (e.g. *climb, enter*) are marked as such in their subcategorization frame, informally:

```
enter: [NP, PP(prep: type: 'inside', incorporated]
```

resulting, in by-default situations, in the non-surface realization of the preposition. This phenomenon is implemented as a verb local constraint, and does not alter our model, since it is a lexical constraint.

From an NL generation point of view, starting from a semantic representation such as $to(loc, jean, X)$ we get a set of verb lexicalizations (or equivalent deverbal nouns) or a more intentional specification such as a verb class. This is obtained via unification or subsumption of the representations in individual verb lexical entries or in the semantic representation specified at verb class level with the semantic representation R. The domain induced is set of tuples verb(A,B,C,D) that meet the subsumption constraint. Domain is the following:

$$Domain(verb(R)) = \{\cup\, verb(VClass, RR, scat(Type-Prep,\ Type-PP), Lexicalization) \wedge subsumes(RR, R)\}$$

where, in the verb lexical entry, RR is the verb semantic representation in the formula subsumed by R (subsumption in LCS is defined in (Saint-Dizier 2001)), which originates the construction of the domain. VClass is the verb class, Type-PP is the type normally expected for the object PP, and scat contains the subcategorization information. We have, in our example, the abstract form R = $to(loc, Y, X)$ and the set of verbs of the class 'movement, specified arrival', associated with their subcategorization frame, which is constructed via unification

(see above example). The same situation occurs with the corresponding deverbal nouns.

3.2 Nouns and NPs

The head noun of the NP must meet in some way with the verb subcategorization frame expectations. The domain of the head noun is its semantic type (and all its descendent, via subsumption in the domain ontology). Derived types may come from possible metaphorical and metonymic uses, implemented in our framework (Moriceau et al. 2003) via type coercion rules.

Type coercion Type coercion rules are modelled globally by means of constrained rewriting rules. They implement various forms of metaphors and metonymies. Let Σ be the set of such coercion rules σ, functions of the verb class and of the semantic types expected for the object PP:
$derived - type = \sigma(verb(VClass, Type - PP))$,
and let TC be the transitive closure of this operation on the verb at stake (i.e. the set of all elements which can be derived). TC is finite and is defined via recursive enumeration. In our approach, derived types are not so numerous: there is no recursive application of type coercion, resulting in a quite small and stereotyped set of derived types.

For example, a frequently encountered metaphorical construction such as *entrer dans un projet* (to enter in a project), where project is typed as 'epistemic artefact' requires a type coercion rule of the form:
σ_i: epistemic artefact = σ(verb(enter,localization)).
Coercion rule is here restricted to the subclass of 'enter' verbs, since the metaphor is not fully regular for all movement verbs with specified arrival. This means that an accurate description of metaphors involves a large number of rules if one wants to avoid overgeneration, a major problem in NL generation.

Then, if R1 is the semantic representation of the noun, the domain of the NP is therefore a set of triples of the form:
$Domain(NP(R1)) = \{\cup\ noun(R1, head - noun - type, Lexicalization),$
$noun(\_, derived - type, Lexicalization)\}$

3.3 Prepositions

Most prepositions are highly polysemic and are often involved in a large number of derived, unexpected or metaphorical uses. Analyzing and representing the semantics of prepositions and generating appropriate prepositions in natural language generation is a rather delicate task, but of much importance for any application that requires even a simple form of understanding. Spatial and temporal prepositions have received a relatively in-depth study for a number of languages (e.g. (Verkuyl et al. 1992)). The semantics of the other types of prepositions describing manner, instrument (Mari and Saint-Dizier 2003), amount or

accompaniement remain largely unexplored. In this section, for readability purposes, we introduce some background on preposition semantics, mainly on a general classification and semantic description we carried out a few years ago (Cannesson et al. 2002). Work is on French, but a number of elements are stable over a variety of languages, in particular Romance languages.

Identifying preposition senses Although prepositions have a number of idiosyncratic usages (probably much less in French than in English), most senses are relatively generic and can be characterized using relatively well-known and consensual abstract labels, as shown in (Cannesson et al. 2002) for French. To illustrate this point, let us consider the case of *par*. *Par* has the following 6 senses, which seem to be all approximately at the same level of abstraction:

- distribution: *il gagne 1500 Euros par mois* (he earns 1500 Euros per month),
- causality: as in passives but also e.g. in *par mauvais temps, je ne ne sors pas* (by bad weather I don't go out),
- origin: *je le sais par des amis* (I know it from friends),
- via: *je passe par ce chemin* (I go via this path),
- tool or means: *je voyage par le train* (I travel by train),
- 'approximate' value: *nous marchons par 3500m d'altitude* (we hike at an altitude of 3500m).

In a second stage, we have provided a conceptual representation of these senses, based on the Lexical Conceptual Structure (LCS, Jackendoff 90), which is based on a language of primitives, viewed as linguistic macros, which can be interpreted in various frameworks, such as the Euclidean geometry. The LCS also introduces a decompositional approach to meaning which allows for the development of an accurate and abstract theory of lexicalization, in particular for prepositions, which it represents particularly well.

A general typology for prepositions Here is a first classification proposal for French prepositions (see also (Cannesson et al. 2002)). We have identified three levels of decomposition which are quite regular and have an internal stability: family, facet and modality. Only the two first levels are introduced here. Labels for semantic classes are intuitive and quite often correspond to thematic role names (examples are direct translations from French), these are the labels specified in verb subcategorization frames:

- **Localization** with facets: **source, destination, via/passage, fixed position**. Destination may be decomposed into destination reached or not (possibly vague), but this is often contextual. Fixed position can either be vague (*he is about 50 years old*) or precise. From an ontological point of view, all of theses senses can, a priori, apply to spatial, temporal or abstract arguments.
- **Quantity** with facets: **numerical or referencial quantity, frequency and iterativity, proportion or ratio**. Quantity can be either precise (*temperature is 5 degrees above 0*) or vague. Frequency and iterativity: *he comes several times per week*.

- **Manner** with facets: **attitudes, means (instrument or abstract), imitation or analogy**. Imitation: *he walks like a robot; he behaves according to the law*.
- **Accompaniement** with facets: **adjunction, simultaneity of events, inclusion, exclusion**. Adjunction : *flat with terrace / steak with French fries/ tea with milk*, exclusion : *they all came except Paul*.
- **Choice and exchange** with facets: **exchange, choice or alternative, substitution**. Substitution : *sign for your child*, choice: *among all my friends, he is the funniest one*.
- **Causality** with facets **causes, goals and intentions**. Cause: *the rock fell under the action of frost*.
- **Opposition** with two ontological distinctions: physical opposition and psychological or epistemic opposition. Opposition: *to act contrary to one's interests*.
- **Ordering** with facets: **priority, subordination, hierarchy, ranking, degree of importance**. Ranking : *at school, she is ahead of me*.
- Minor groups: **About, in spite of, comparison**. The terms given here are abstract, and cover several prepositions. About: *a book about dinosaurs*.

Each of the subsenses described above is associated with a number of preposition senses, clearly distinct from other senses. Here is a brief description of the Ordering class:

Fig. 1 - prepositions of Ordering family

facet	prepositions
Priority	après (*after*), avant (*before*)
Subordination	sous (*under*), sur (*above*)
Hierarchy	devant (*in front of*), derrière (*behind*) avant (*before*), après (*after*)
Ranking	devant (*in front of*), derrière (*behind*)
Degree of importance	à côté de, auprès de (*close to, near*), par rapport à, pour, vis-à-vis de (*compared to*)

A general conceptual semantics for prepositions Each preposition facet or modality has a unique representation. For example, 2 major senses of the preposition *avec* (with) are:

- **accompaniement**, represented as, in the simplified LCS representation we developed:
  $with(loc,\ I,\ J)$,
  where 'loc' indicates a physical accompaniement (*I go to the movies with Maria*), while 'psy' instead of 'loc' would metaphorically indicate a psychological accompaniement (*Maria investigated the problem with Joana*).

– **instrument**, represented as:
$by-means-of(manner,\ I,\ J)$
(*they opened the door with a knife*). This is, in fact, a generic representation for most preposition senses introducing instruments, a more refined analysis can be found in (Mari et al. 2003).

In our framework, we defined 65 primitives encoding 170 preposition senses, which seem to cover most senses found in standard texts over a number of languages. Language being essentially generative and creative, this obviously does not exclude other senses, for which, most probably, a few additional primitives, or the composition of already defined ones, will be necessary.

Semantic and pragmatic coercion Semantic and pragmatic coercion on the type of preposition expected by the verb can be modelled as follows. Let Υ be the set of such coercion rules τ, functions of the verb class and of the semantic class of the preposition:
$derived-subclass = \tau(verb(VClass, Type-Prep))$,
For example, *parler sur le vague* (litt. to talk on vagueness) has a PP characterizing a topic of conversation. The default preposition in French, specified in the lexical entry of *parler* is *de* (about). The preposition *sur* introduces the idea of a general discussion, we have then the following pragmatic coercion rule:
$'sur' = \tau(verb(talk,'about'))$.
We limit here preposition classes to a precise preposition, to avoid potential overgeneration.

From a more formal point of view and more generally, let TT be the transitive closure of this operation on the verb at stake (i.e. the set of all subclasses which can be derived). TT is finite and is defined via recursive enumeration from coercion rules and lexical descriptions. TT defines the semantic and pragmatic expansion of the verb-preposition compound.

Back to our example, the preposition is primarily induced by the semantic representation R2 = $to(loc, Y, X)$. More generally, the domain of potential lexicalizations is characterized by a set of triples as follows:

$$Domain(prep(R2)) = \{\cup\ (prep(Subclass, RestrNP, RR2, Lexicalization) \wedge subsumes(RR2, R2)),\ prep(derived-subclass, \_, \_, Lexicalization)\ \}$$

where, in the lexical entry 'prep', Subclass designates the modality or the facet level as exemplified above, RestrNP are the selectional restrictions imposed on the NP headed by the preposition, RR2 is the semantic representation which must be subsumed by R2, and Lexicalization is the surface realization(s) of RR2 (this implements lexical variation, when there is a strict lexical equivalence). The second set of triples prep is the transitive closure (produced by recursive enumeration) of all potential type coercions given the verb identified or its class.

4 Preposition Lexicalization as a CSP

In the previous section, we show how domains of lexical entries and associated lexicalizations can be constructed, via unification, subsumption and type, semantic and pragmatic coercion. The challenge is now to express and to manage the constraints of the triple verb-preposition-NP, so that all possible preposition lexicalizations can be made explicit, allowing for the generation of the VP.

4.1 Constraints between domains as equations

Let us first recall the 3 domains defined above:

$Domain(verb(R)) \ = \ \{\cup\ verb(VClass, RR, scat(Type - Prep,\ Type - PP),$
$\quad Lexicalization) \wedge subsumes(RR, R)\}$
$Domain(np(R1)) \ = \ \{\cup\ noun(R1, head - noun - type, Lexicalization),$
$\quad noun(\_, derived - type, Lexicalization)\}$
$Domain(prep(R2)) \ = \ \{\cup\ (prep(Subclass, RestrNP, RR2, Lexicalization)$
$\quad \wedge subsumes(RR2, R2)),\ prep(derived - subclass, \_, \_, Lexicalization)\ \}$

From these specifications, it is possible to state constraints under the form of equations, from which pairs:
(verb lexicalization, preposition lexicalization)
can be produced, assuming the noun has a fixed type, and a limited number of lexicalizations which cannot a priori be revised. We have the following equations:

1. Type-PP subsumes (a) head-noun-type or (b) an element s in the derived types via type coercion. In more formal terms:
   $subsumes(Type - PP, head - noun - type) \vee\ \exists \sigma \in \Sigma,$
   $\quad s = \sigma(Type - PP) \wedge subsumes(Type - PP, s)).$
2. RestrNP subsumes head-noun-type or s:
   $subsumes(Restr - NP, head - noun - type) \vee subsumes(Restr - NP, s).$
3. Type-Prep subsumes (a) Subclass in direct usages or (b) a derived Subclass via semantic or pragmatic coercion:
   $subsumes(Type - Prep, SubClass) \vee\ \exists \tau \in \Upsilon,$
   $\quad t = \tau(Type - Prep) \wedge subsumes(Type - Prep, t)).$

The satisfaction of these equations produces the lexicalization set composed of pairs (verb, preposition). Lexical choice can operate on these pairs to select one of these, possibly e.g. on the basis of user preferences. Preposition incorporation into verbs is also handled at this level.

As explained in the introduction, provided restrictions in various lexical entries are adequately described, and that type, semantic and pragmatic coercion rules are appropriate, our system does not overgenerate. Overgeneration is solely due to incomplete or inadequate linguistic descriptions. A priori, our system is sound and complete and produces all the admissible solutions, w.r.t. lexical specifications and coercion operations.

4.2 A direct illustration

Let us illustrate the satisfaction of these constraints by a simple example:
$flight(5564) \wedge to(loc, 5564, Paris) \wedge city(Paris)$,
where *to(loc,5564,Paris)* is at the same time R (the semantic representation of the verb) and R2 (the semantic representation of the preposition); *city(Paris)* is R1 (the semantic representation of the noun of the PP).
The three domains are:

$$\begin{aligned} Domain(verb(R)) \;=\; \{\; & verb(movement-verb-to-destination, to(loc, X, Y), \\ & \quad scat(fixed-position,\ fixed-position), aller), \\ & verb(movement-verb-to-destination, to(loc, X, Y), \\ & \quad scat(fixed{-}position,\ fixed{-}position), monter), \\ & verb(movement-verb-to-destination, to(loc, X, Y), \\ & \quad scat(fixed{-}position,\ fixed{-}position), arriver), \\ & ... \\ & \wedge\ subsumes(to(loc, X, Y), to(loc, 5564, Paris))\} \end{aligned}$$

$$Domain(np(R1)) \;=\; \{\; noun(city(Paris), city, Paris)\}$$

$$\begin{aligned} Domain(prep(R2)) \;=\; \{\; & prep(fixed{-}position, fixed{-}position, to(loc, X, Y), \grave{a}), \\ & prep(destination, fixed-position, to(loc, X, Y), pour), \\ & prep(destination, fixed-position, to(loc, X, Y), en), \\ & \wedge\ subsumes(to(loc, X, Y), to(loc, 5564, Paris))\ \} \end{aligned}$$

Let us apply the constraints:

1. $subsumes(Type-PP, head-noun-type)$
   $\rightarrow\ subsumes(fixed-position, city)$
2. $subsumes(Restr-PP, head-noun-type)$
   $\rightarrow\ subsumes(fixed-position, city)$
3. $subsumes(Type-Prep, SubClass)$
   $\rightarrow\ subsumes(fixed-position, fixed-position)$

The solution domain is: {*(aller,à),(monter,à),(arriver,à), ...*} .

With respect to the lexical descriptions, these three solutions are acceptable, although *monter à* is les natural for a flight, than, e.g. for a car.

4.3 More complex situations

Our model deals with composition (or co-compositional) forms. It naturally discards constructions which do not meet the selection and possibly coercion constraints. For example, let $person(jean) \wedge to(loc, jean, Y) \wedge wall(Y)$ be the semantic representation of the metaphorical semi-fixed form: *Jean part dans le mur* (*Jean goes in the wall = john fails (in an enterprise), as a car going in a wall is going to be broken)*).
In this example, *partir* is a movement verb with destination specified, *dans* can

be acceptable provided the NP has an inside into which the subject can go, which is not the case for *mur* (wall).

The three domains are (without type coercion):

$$\begin{aligned}
Domain(verb(R)) \;=\; \{\; & verb(movement-verb-to-destination, to(loc, X, Y), \\
& \qquad scat(destination,\ fixed-position), partir), \\
& verb(movement-verb-to-destination, to(loc, X, Y), \\
& \qquad scat(destination,\ fixed-position), courir), \\
& \ldots \\
& \wedge\ subsumes(to(loc, X, Y), to(loc, jean, Y))\} \\
Domain(np(R1)) \;=\; \{\; & noun(wall(Y), compact-object, mur)\} \\
Domain(prep(R2)) \;=\; \{\; & prep(destination, fixed-position, to(loc, X, Y), \\
& \qquad vers), \\
& prep(destination, fixed-position, to(loc, X, Y), pour), \\
& prep(destination, fixed-position, to(loc, X, Y), a\ destination\ de), \\
& \ldots \\
& \wedge\ subsumes(from(loc, X, Y), to(loc, jean, Y))\ \}
\end{aligned}$$

Let us now apply the equations between domains:

1. $subsumes(Type-PP, head-noun-type)$
   $\rightarrow\ subsumes(fixed-position, object)$
2. $subsumes(Restr-NP, head-noun-type)$
   $\rightarrow\ subsumes(fixed-position, object)$
3. $subsumes(Type-Prep, SubClass)$
   $\rightarrow\ subsumes(destination, destination)$

The solution domain is:
{*(partir, vers), (partir, pour), (partir, à destination de), ...*} .
This solution domain does not contain the expected metaphoric solution *partir dans le mur* because of the type of wall, which is not an object with an inside. This is due to the fact that the preposition used *dans* (*in*) does not belong to the preposition class *destination*.

If we now add the domain engendered by the corresponding type coercion rule:
$'dans' = \tau(verb(partir, destination))$,
assuming *destination* is the type of preposition normally expected by the verb, the solution domain now includes the form (partir, dans), *mur* (wall) being correctly typed as a fixed position. In fact, this metaphor applies to most movement verbs with specified destination, each of them making more explicit a manner of going in a wall.

Let us also note that while (partir, vers), (partir, pour) can be combined with a number of metaphors in abstract domains, (partir, à destination de) essentially expects an object NP of type physical location.

Finally, this approach allows us to treat the verb-preposition-NP constraints in a way independent from their syntactic realization. For example, the above example, with a left extraposition of the PP (due to a stressed focus) and a

nested proposition:
C'est pour Paris que part ce vol (this is for Paris that this flight leaves)
is treated in exactly the same way. This allows us to describe lexical choice of the pair verb-preposition at the highest level of abstraction, independently, a priori, of its syntactic realizations.

5 Conclusion

In this paper, we presented a model that describes in a declarative way the complex interactions between the verb, the preposition and the object NP. We focussed on the lexicalization of the preposition, which is a major difficulty in language generation.

Based on the notion of domain, we introduced equations that manage these interactions, and which, after resolution, propose a set of lexicalizations for the pair verb-preposition. We considered regular cases as well as derived ones, in particular a number of metaphors, modelled by means of type, semantic and pragmatic coercion rules.

We proposed in this paper a declarative model which has a priori completeness and soundness properties. These properties, as for e.g. grammars, entirely depends on the quality of the linguistic descriptions and restrictions, without which the system is just an empty shell.

In terms of implementation, this work is being integrated into a larger set of tools for language generation (including, among others: NP, PP, proposition and syntactic alternation generation). Such a set of tools, based on a decompositional representation of meaning, is of much interest and importance in language generation, where there are very few generic tools available. This work will be used and validated in the WEBCOOP project, where short statements with a large number of PPs need to be generated.

Implementations are under study. A simple solution is to manage set intersections, sets being viewed as finite domains. It is possible, furthermore, to define some of those sets (in particular those related to coercion) a priori, for each verb, by means, e.g. of partial evaluation techniques. We would like, on top of these set intersection constraints to be able to introduce linguistic preferences, be they general or related to a user. However, these types of heuristics can only operate after construction of the set of admissible lexicalizations. It would be of much interest to have them interact as early as possible in the resolution process in order to limit complexity.

At a linguistic level, this study involves only knowledge from lexical descriptions. It is abstract and not committed a priori to any grammatical form. This means that nothing prevents it from being integrated, with, obviously, some customization, into a linguistic theory such as HPSG, PP (via the projection principle), or LFG, and possibly TAGs, since these theories are centered around

lexical descriptions.

Acknowledgements. This project is partly supported by the CNRS TCAN program, we are grateful to its members for their advices. We also thank anonymous reviewers which helped improved this work.

References

1. F. Benamara and P. Saint-Dizier. WEBCOOP: a Cooperative Question-Answering System on the Web. EACL project notes, Budapest, Hungary, April 2003.
2. E.Cannesson, P. Saint-Dizier. Defining and Representing preposition Senses: a preliminary analysis. In ACL02-WSD, Philadelphia, July 2002.
3. L. Cahill. Lexicalisation in applied NLG systems, Research report, ITRI-99-04, 1999.
4. R. Jackendoff. Semantic Structures. MIT Press, 1990.
5. V. Moriceau and P. Saint-Dizier. A Conceptual Treatment of Metaphors for NLP. ICON, Mysore, India, December 2003.
6. J. Pustejovsky. The Generative Lexicon. MIT Press, 1995.
7. E. Reiter, R. Dale. *Building Applied Natural Language Generation Systems*, Journal of Natural Language Engineering, volume 3, number 1, pp:57-87, 1997.
8. P. Saint-Dizier. Alternations and Verb Semantic Classes for French. In Predicatives Forms for NL and LKB, P.Saint-Dizier (ed), Kluwer Academic, 1998.
9. P. Saint-Dizier and G. Vazquez. A Compositional Framework for Prepositions. IWCS4, Springer, lecture notes, Tilburg, 2001.
10. H. Verkuyl and J. Zwarts. Time and Space in Conceptual and Logical Semantics: the notion of Path, Linguitics 30: 483-511, 1992.
11. A. Mari and P. Saint-Dizier. A Conceptual Semantics for Prepositions Denoting Instrumentality. ACL-SIGSEM Workshop: The Linguistic Dimensions of Preposition and their Use in Computational Linguistics Formalisms and Applications, Toulouse, France, 2003.