

# Prepositions and the Names of Countries and Islands: A Local Grammar for the Automatic Analysis of Texts<sup>1</sup>

Mylene Garrigues

The expression of movement in French has given rise to a large amount of linguistic research centred on the attempt to attribute a particular meaning to a given syntactical form. Thus, for example, the use of alternatives such as the prepositions *à/sur* is explained in terms of a semantic opposition, with *à* marking "the conclusion or goal of a displacement", and *sur* "the direction of a rapid displacement" ("le terme d'un mouvement" and "la direction d'un mouvement rapide" respectively, G. Gougenheim, 1962). This view, illustrated by a number of examples, presupposes that it is enough to learn the "meaning" of a preposition in order to acquire the necessary competence for its production in any context.

What is presented here is an analogous study which, however, leads to a dismantling of this view. The systematic examination of circumscribed points, which is, after all, necessary to the formalisation of a linguistic phenomenon for automatic processing, shows a remarkable absence of any such correlation between meaning and form, whilst revealing a highly complex pattern of lexico-syntactic relationships. This reversal of the initial position leads us to envisage the learning of the usage of a preposition in terms of the memorisation not of a meaning, but of a multitude of grammars corresponding to the different contexts, or even micro-contexts, in which it occurs. These grammars, of which we shall provide graphical representations, reveal also, and in a tangible way, the extreme complexity of our cognitive linguistic system.

## 1. Definition of the Predicate

Along more rigorous lines, the object of our research was to develop a formal grammar of the use of prepositional complements with the names of countries, of continents and of islands, in the context of a semantic predi-

<sup>1</sup> This paper has been translated by Ivan Birks.

cate defined by the following scenario.

(1) A human being undertakes a displacement to an explicit destination ( $N_{dest} = N_{land}$ ,  $N_{cont}$  or  $N_{isl}$ ).

It was, in other words, a question of describing all the syntactic occurrences of this predicate by formulating a grammar of the constructions involving a verb of movement accompanied by these complements. We focused our attention on this particular type of complement because it enabled us:

- to reconsider the classic problem of the prepositions which accompany geographical nouns by listing them, and organising them much more rigorously than had been done previously.
- to describe two situations, of which one (countries and continents) is relatively regular, and the other (islands) is of exemplary complexity.

Over and above countries, continents and islands, we intend to extend our research into the domain of seas, regions, towns, etc. in order to cover, progressively, the geographical field in its entirety. We furthermore intend to work on the formulation of the complementary grammar, that of complements of origin (*venir d'Italie/to come from Italy*).

The elementary sentences (subject-verb-complement) which correspond to our scenario are, therefore, of the following type:

*Jo a regagné la France*

*Jo has arrived back in/regained France*

*Jo vole vers les Etats-Unis*

*Jo is flying to the US.*

*Jo se rend au Maroc*

*Jo is going to Morocco*

*Jo part en Afrique*

*Jo is leaving for Africa*

*La Sardaigne attire les touristes*

*Sardinia attracts tourists*

Since we were dealing with sentences which incorporate prepositions of place, it was necessary to describe all the other scenarios involving these prepositions in order to avoid confronting ourselves subsequently with a muddled situation, incompatible with the process of formalisation.

To illustrate this observation, we have provided three examples from a series of similar scenarios (M. Garrigues, 1993) which we were careful to distinguish from (1):

(2) A human being undertakes a displacement for which no limit is established:

*Jo parcourt la France*

*Jo is exploring France*

*Jo sillonne les Pays-Bas*

*Jo is travelling the length and breadth of the Netherlands*

*Jo survole les Etats-Unis*

*Jo is flying over the United States*

(3) A human being undertakes an action. The action may or may not involve displacement, but a displacement is implicit in relation to the activity in question. The prepositional complement is in this case called a complement of scenery (J. P. Boons, A. Guillet, C. Leclère, 1976):

*Jo travaille au Maroc*

*Jo is working in Morocco*

*Jo s'amuse en Angleterre*

*Jo is having fun in England*

*Jo a fait la guerre au Vietnam*

*Jo fought in Vietnam*

(4) A human being displaces an object or a human being from one place to another:

*Jo a emmené ses élèves en Angleterre*

*Jo has taken his pupils to England*

*Jo a envoyé un colis au Brésil*

*Jo has sent a parcel to Brazil*

It will be noted, furthermore, that the definition of the predicate requires that the field of application it operates on should be explicated (countries, continents or islands). We can observe that it is enough for us to change the field of application of our predicate, in order for a specific and unpredictable distribution of prepositions to emerge. If, for example, we apply the scenario of our predicate not to the geographical field of countries and islands, but to a highly limited spatial field, such as that of the rooms in a house, the result is the emergence of a particular grammar which governs its own prepositions and exceptions. We can compare the following examples:

<i>Jo va dans la chambre</i>	<i>* à la chambre</i>	<i>- Jo goes into the bedroom</i>
<i>Jo va dans la cuisine</i>	<i>à la cuisine</i>	<i>- Jo goes into the kitchen</i>
<i>Jo va dans le vestibule</i>	<i>* au vestibule</i>	<i>- Jo goes into the hall</i>
<i>Jo va dans le salon</i>	<i>au salon</i>	<i>- Jo goes into the living-room</i>

The same phenomenon can be observed in any other semantic field. If we take the case of a second field, just as restricted, that of "open spaces", the same arbitrariness reappears in the choice of the preposition, as the following examples indicate:

<i>Jo va sur la route</i>	<i>* dans la route</i>	<i>* à la route</i>
<i>Jo va * sur la rue</i>	<i>dans la rue</i>	<i>* à la rue</i>
<i>Jo va sur la place</i>	<i>* dans la place</i>	<i>* à la place</i>
<i>Jo va * sur le jardin public</i>	<i>dans le jardin public</i>	<i>au jardin public</i>
<i>Jo va sur le terrain vague</i>	<i>dans le terrain vague</i>	<i>* au terrain vague</i>

*(Jo goes onto/into/to/the road/street/square/park/waste ground)*

As we can see, no rule can be formulated which would enable us to predict globally these occurrences. There is, therefore, no alternative but to proceed by examining and describing each case one by one, within the context of the paradigm of each field taken into consideration. In view of this requirement, it is easy to appreciate that traditional grammars and manuals, whose purpose is not be used in the context of the automatic processing of a language, make an appreciable, but fundamentally inadequate contribution. As linguistic theory evolves, however, towards the formalisation of

the morpho-syntactic functions of a language, with a view to enabling the automatic processing of the latter, the development of such grammars will become necessary on a very large scale. The enormity of the task should soon become apparent.

Our method of formalisation consists in the representation of linguistic facts by means of graphs known as finite state automata (more specifically: DAGs or Directed Acyclic Graphs (M. Gross, 1989). We have, therefore, described and listed the verb-preposition-noun combinations resulting from scenario (1) using graphs. These graphs permit the regrouping of families of expressions according to their variations and common features. The graphs should be read from left (initial position) to right (final position). Each of the paths in the graph leading from the initial position to the final position corresponds to an acceptable expression. The degree of formalisation which results from this method of representation means that these grammars can be applied directly to the automatic analysis of texts. This is achieved by means of software applications which translate these graphs automatically into analysers capable of recognising in a text the expressions thus described.

## 2. The Syntactic Permutations

The syntactic permutations of our predicate can be represented by three possible structures, applicable to all fields.

(a) <i>Nhum V Prep Nloc</i>	<i>Jo va (dans le salon, en Italie)</i> <i>Jo goes into the living room/into Italy</i>
(b) <i>Nhum V Nloc</i>	<i>Jo regagne (sa chambre, l'Italie)</i> <i>Jo regains her bedroom/Italy</i>
(c) <i>Nloc V Nhum</i>	<i>(Ce restaurant, l'Italie) attire les touristes</i> <i>This restaurant/ Italy attracts tourists</i>

Our grammar will account for the realisation of syntactic structures corresponding to (a) and (b). The syntactic structures of type (c) are less specific, and can be processed and added at a later stage. Furthermore, for reasons we have already provided, we have momentarily limited the *Nloc* destination to *Nland*, *Ncont* and *Nisl*, taken from the strictly geographical

field which we propose to cover subsequently. Each field of application, even one as limited as that of the rooms of a house, possesses its own grammar, and, as a result, must be isolated and analysed in detail. We can conceive of no other satisfactory way of proceeding.

The existence of a lexicon grammar of French verbs enabled us to select the simple verbs relevant to the task in hand. In building up an inventory of verbs of movement we relied on the following tables:

For the prepositional permutation (a):

- Table T2 (M. Gross, 1975)
- Table T35L (J. P. Boons, A. Guillet, C. Leclère, 1976)
- Table 38L0 (A. Guillet, C. Leclère, 1992)

This category also comprises sentences with a composed or periphrastic verb form, such as *faire un saut à* (*drop into*), *faire une excursion à* (*take a trip to*), *faire un tour à* (*have a wander in*). Our graph includes only a few of these forms. We intend to create a sub-graph which accounts for them at a later stage.

For the two non-prepositional permutations (b) and (c):

- Table 38L1 (A. Guillet, C. Leclère, 1992)

### 3. Semantic Compatibility in *VNloc*

Once we had made an inventory of these verbs we conducted a secondary selection based on factors of plausibility and semantic compatibility. The criterion of plausibility led us to discard, on the one hand, verbs such as *pagayer* (*paddle/row*) or *amerrir* (*to land on water*), since the plausibility of their occurrence in everyday language is slight (M. Garrigues, 1993). On the other hand we discarded verbs such as *se glisser* (*slide into*) and *se faufiler* (*to worm into*) — even though the plausibility of their occurring is considerable in everyday language, it is slight in the context of our predicate:

? <i>Jo pagaye vers l'Italie</i>	<i>Jo paddles towards Italy</i>
<i>Jo vogue vers l'Italie</i>	<i>Jo wanders towards Italy</i>
? <i>Jo s'est glissé en Italie</i>	<i>Jo slid into Italy</i>
<i>Jo s'est glissé dans la pièce</i>	<i>Jo slid into the room</i>

We then eliminated verbs with zero plausibility, due to their semantic incompatibility, or minimal compatibility with the following noun (G. Gross, 1992). *Marcher sur* (*to walk/march on*), meaning *aller vers* (*to go towards*) is logically incompatible with the name of an island, and is less plausible with the name of a country than with the name of a town. This meaning of the verb *marcher* is, in the event, generally linked to a military context:

<sup>?"</sup> <i>Les alliés marchent sur la Sardaigne</i>	<i>The allies march on Sardinia</i>
<sup>?</sup> <i>Les alliés marchent sur l'Italie</i>	<i>The allies march on Italy</i>
<i>Les alliés marchent sur Rome</i>	<i>The allies march on Rome</i>

Similarly, *monter* (*go up to*) and *descendre* (*go down to*) are generally used with the names of towns, rather than with the names of countries, although it does seem to us that the two verbs present certain differences:

<i>Monter à Paris</i>	( <i>to go up to Paris</i> )
<sup>?</sup> <i>Monter en Belgique</i>	( <i>to go up to Belgium</i> )
<i>Descendre à Toulouse</i>	( <i>to go down to Toulouse</i> )
<i>Descendre en Espagne</i>	( <i>to go down to Spain</i> )

#### 4. The Group *V Prep Nloc*

We have observed that in the case of the linguistic realisation of a phrase of type *V Prep Nloc*, the choice of the preposition is dependent on the prior selection of the verb and corresponding noun. For this reason we consider this group to constitute an indivisible phrasal unit. If, for example, we take the verb *aller* (*go*), and attempt to associate it with the three prepositions it is supposed to accept (*à, dans, en- to/in*), we realise that we must discover the noun, before producing the preposition retroactively. In other words the choice of the preposition is bound by a double constraint imposed simultaneously by the verb and the noun;

<i>Aller à la cuisine</i>	<i>dans la cuisine</i>	<sup>?</sup> <i>en cuisine</i>
<i>Aller à la chambre</i>	<i>dans la chambre</i>	<sup>?</sup> <i>en chambre</i>
<i>Aller à la salle de bains</i>	<i>dans la salle de bains</i>	<sup>?</sup> <i>en salle de bains</i>
<i>Aller à la salle d'opération</i>	<i>dans la salle d'opération</i>	<i>en salle d'opération</i>
<i>Aller à la banlieue</i>	<sup>?</sup> <i>dans la banlieue</i>	<i>en banlieue</i>

<i>Aller aux Etats-Unis</i>	• <i>dans les Etats-Unis</i>	• <i>en Etats-Unis</i>
<i>Aller aux Vosges</i>	• <i>dans les Vosges</i>	• <i>en Vosges</i>
<i>Aller à la Sardaigne</i>	• <i>dans la Sardaigne</i>	• <i>en Sardaigne</i>
<i>Aller à la Guadeloupe</i>	• <i>dans la Guadeloupe</i>	• <i>en Guadeloupe</i>
<i>Aller à la plage</i>	• <i>dans la plage</i>	• <i>en plage</i>
<i>Aller à la forêt</i>	<i>dans la forêt</i>	<i>en forêt</i>
<i>(To go [à/dans/en – to/in] the kitchen, the bedroom, the bathroom, the operating theatre, the suburbs, the US, the Vosge mountains, Sardinia)</i>		

Inversely, if we establish a noun, and try to use three prepositions (*à, vers, pour* – *to, towards, for*) apt to precede it, before we can decide on the acceptability of these prepositions we must know to which verb this noun is linked.

<i>Aller à l'école</i>	<i>vers l'école</i>	• <i>pour l'école</i>
<i>Se rendre à l'école</i>	• <i>vers l'école</i>	• <i>pour l'école</i>
<i>Partir à l'école</i>	<i>vers l'école</i>	<i>pour l'école</i>
<i>S'acheminer à l'école</i>	<i>vers l'école</i>	• <i>pour l'école</i>
<i>Rentrer à l'école</i>	• <i>vers l'école</i>	• <i>pour l'école</i>
<i>(to [go/leave/set out/return] [to/towards/for] school)</i>		

It would appear that we are in the presence of an operation in two stages. The first involves a latent programme of signification defined in terms of a semantic predicate. The second intervenes at the moment of the linguistic realisation of the predicate, and can be seen as an incredibly complex procedure which establishes a relationship between nominal and verbal groups. Out of the calibration of the noun and verb results the production of a preposition. Thus the latent programme of signification is entrenched in an indivisible lexico-syntactic unit *V Prep Nloc*, which constitutes the utterance. As we saw with *marcher sur* even the nature of the subject can interfere. We must, therefore, consider that the basic minimal unit determining the preposition is the simple sentence *NO V Prep Nloc*. The implications in terms of cognitive mechanisms which result from this hypothesis are currently unverifiable. We can say, however, that it is more in line with the requirements of a process of automatic formalisation than the results achieved by traditional grammars. It is also true that in the field of artificial intelligence (R.C. Schank 1982) the process leading to the generation of an

action has been described as a sequence of discriminatory operations proceeding from more abstract forms of representation to more concrete ones, resulting in the realisation of an action. It is true that in the field of infra-linguistics the ultimate goal is not, as with artificial intelligence, an action, but an utterance. This goal could, however, be achieved by a similar process, starting with a predicate 'schema' and, by means of a series of successive forks in a 'syntactic script', arriving at a final utterance.

In order to produce the 'syntactic script' of our predicate in the form of an automaton, we started with an inventory of selected verbs of movement, as we explained above. We then organised the different names of the countries, continents and islands into groups (cf. Fig. 1).

Each group, or box, in our main graph bears a name indicating its morphosyntactic properties. The name in a shaded box (or node) refers to a sub-graph which contains a list of the relevant countries, and which would be consulted upon execution of the automaton. When the box is unshaded, as is the case with *Israël* (*Israel*), it means that the node is explicit, and does not refer to a sub-graph.

#### 4. 1. Names of Countries

— Graph *LandDetZ*. This graph contains the names of countries which are used without a determiner, i.e. with determiner 'zero' (cf. Fig. 2):

(*Aller à, partir pour*) (*Panama, Singapour, Bahrein*)  
(*to go to/leave for...*)

— Node *Israël*. Although *Israël* has determiner zero, it does not feature in the previous group of countries since it functions differently, as the following examples demonstrate:

<i>Partir</i>	<i>pour</i>	<i>Panama</i>	<i>Singapour</i>	<i>Israël</i>
<i>Aller</i>	<i>à</i>	<i>Panama</i>	<i>Singapour</i>	<i>* Israël</i>
<i>Aller</i>	<i>en</i>	<i>* Panama</i>	<i>* Singapour</i>	<i>Israël</i>

Furthermore, as we can see in our basic automaton *NVPrepN* (cf. Fig. 1),

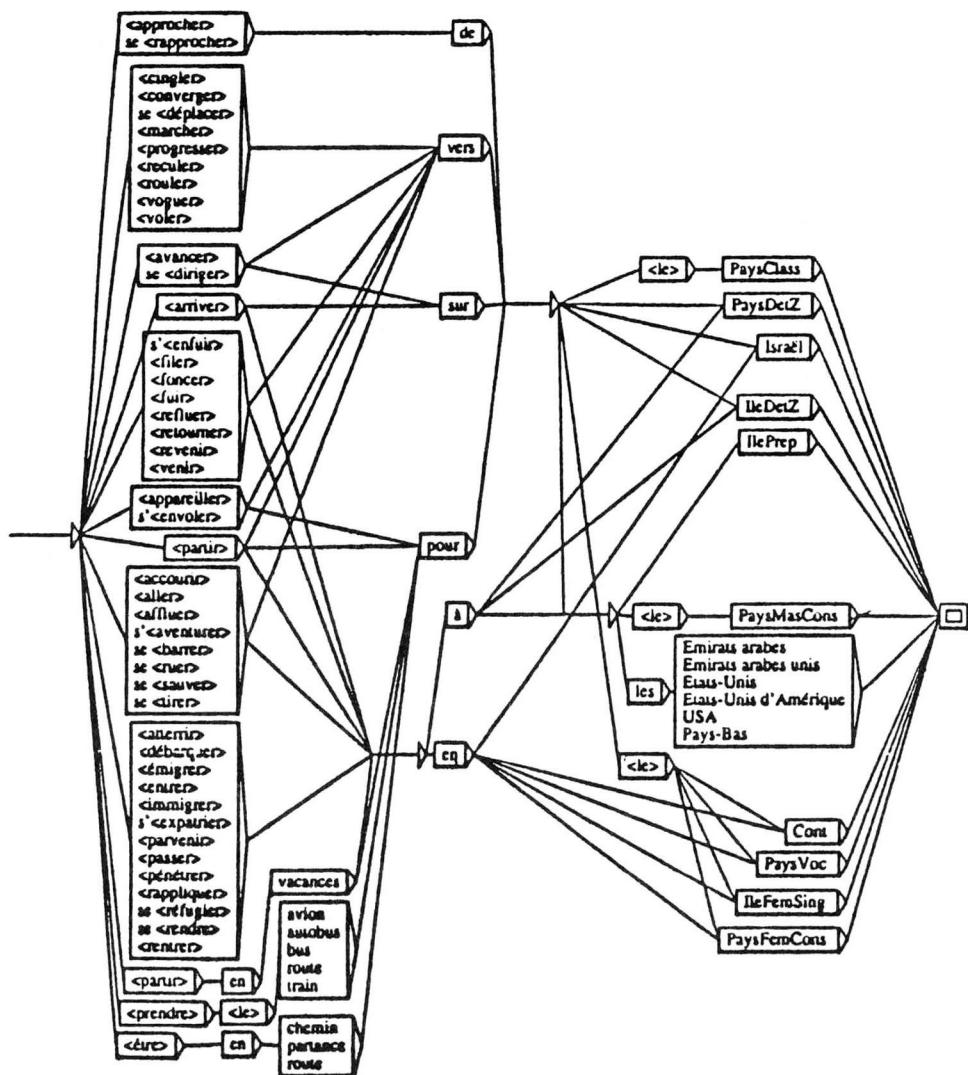


Fig. 1. NVPrepN



Fig. 2. PaysDetZ

*Israël* does not function like the names of countries whose article disappears when used with the preposition *en* (cf. graph *LandVowel* (Fig. 4) and graph *LandFemCons* (Fig. 5):

<i>Aller</i>	<i>en</i>	<i>Israël</i>	<i>Italie</i>	<i>Grande-Bretagne</i>
<i>Avancer</i>	<i>vers</i>	<i>Israël</i>	* <i>Italie</i>	* <i>Grande-Bretagne</i>
<i>Partir</i>	<i>pour</i>	<i>Israël</i>	* <i>Italie</i>	* <i>Grande-Bretagne</i>
<i>(to go to/advance towards/leave for...)</i>				

Given that the name of this country functions in a unique fashion, we placed it separately, in a specific node of the automaton.

Node *Etats-Unis* (*United States*). This node contains the country names which are plural, and used with the plural definite article:

*Aller aux* (*Etats-Unis, Pays Bas [Netherlands]*)

– Graph *LandMasCons*. This graph contains the country names which are masculine, begin with a consonant, and are used with the masculine singular definite article (cf. Fig. 3):

*Aller au* (*Bangladesh, Japon, Maroc [Morocco]*)

– Graph *LandVowel*. This graph contains the country names which begin with a vowel, and are used with the elided definite article, or without a determiner (cf. Fig. 4):

*Partir pour l'Angleterre/Aller en Angleterre*

*(To leave for/go to England)*

– Graph *LandFemCons*. This graph contains the country names which are feminine, begin with a consonant, and are used with the feminine definite article, or with no determiner (cf. Fig. 5):

*Partir pour la Pologne/Aller en Pologne.*

*(To leave for/go to Poland)*

– Graph *LandClass*. This graph contains the country names which are plural, with a classifier ('Etats'/States, 'Pays'/Countries), and which, unlike the node *Etats Unis*, do not take the preposition *à*, but do take the prepositions *de, vers, sur, pour* (cf. Fig. 6).

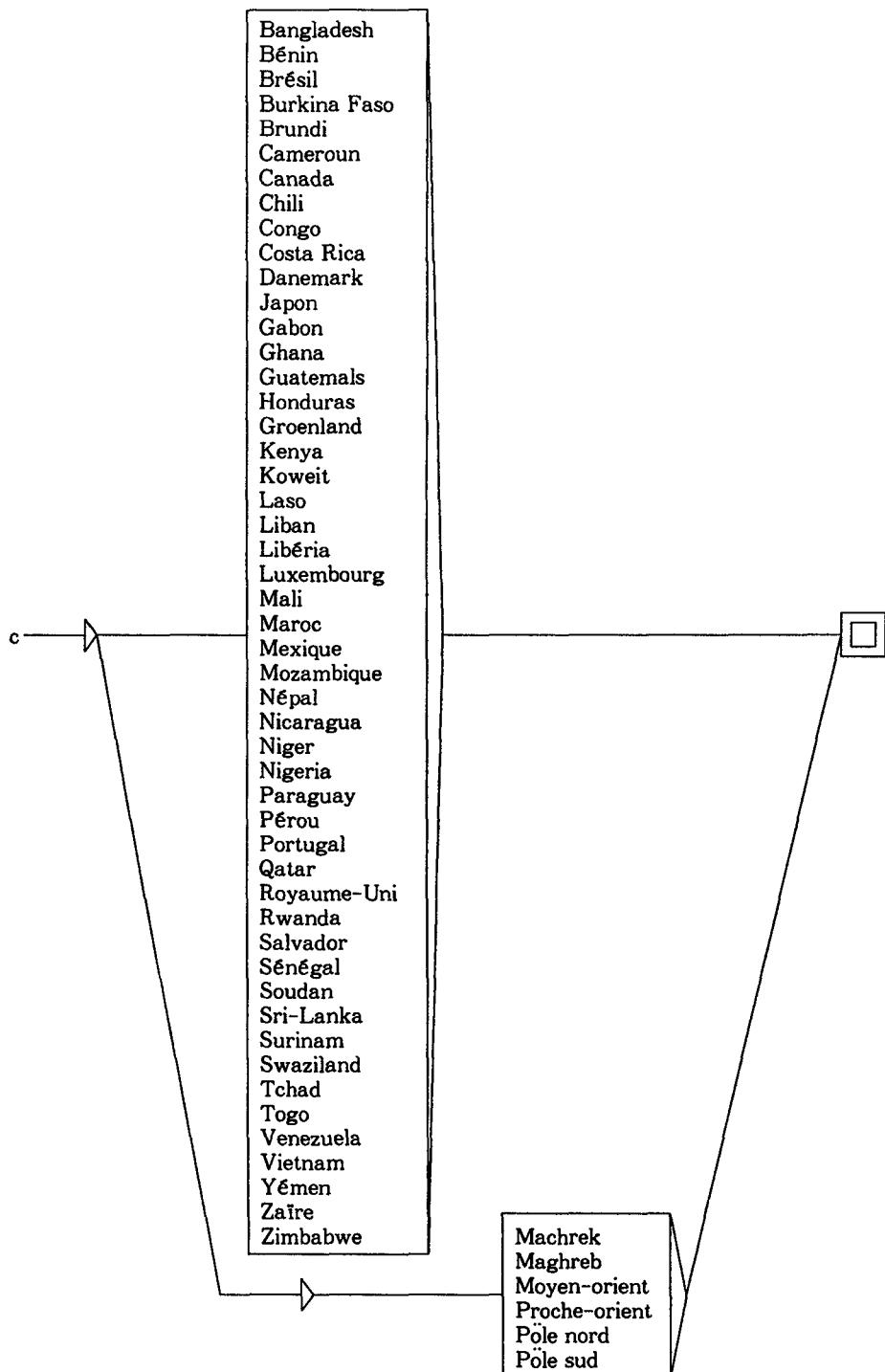


Fig. 3. PaysMasCons

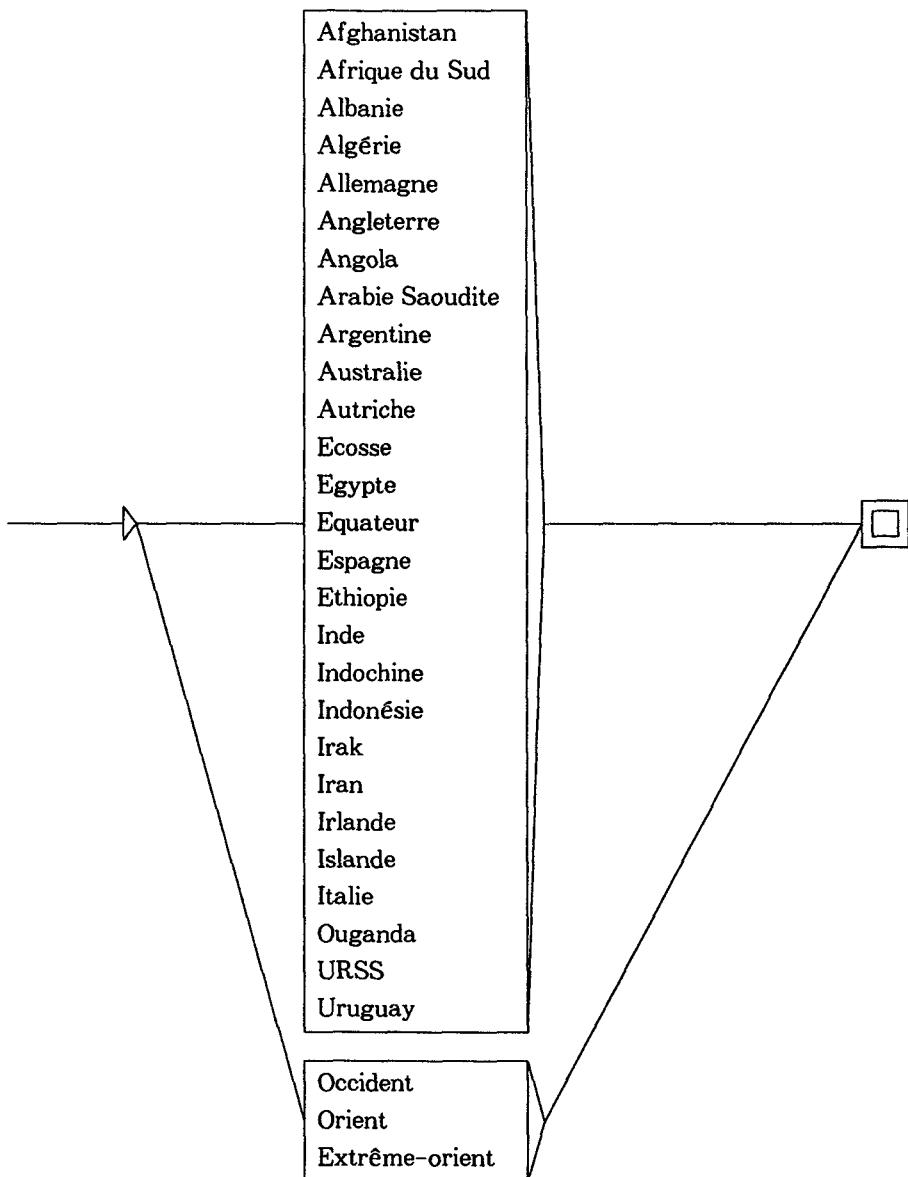


Fig. 4. PaysVoc

*Aller aux (Pays-Bas, \*Pays Baltes)*

*Aller dans les (\*Pays-Bas, Pays Baltes)*

This sub-set also corresponds to the only country names which function with the preposition *dans*, and, as such, has been re-used in the context of

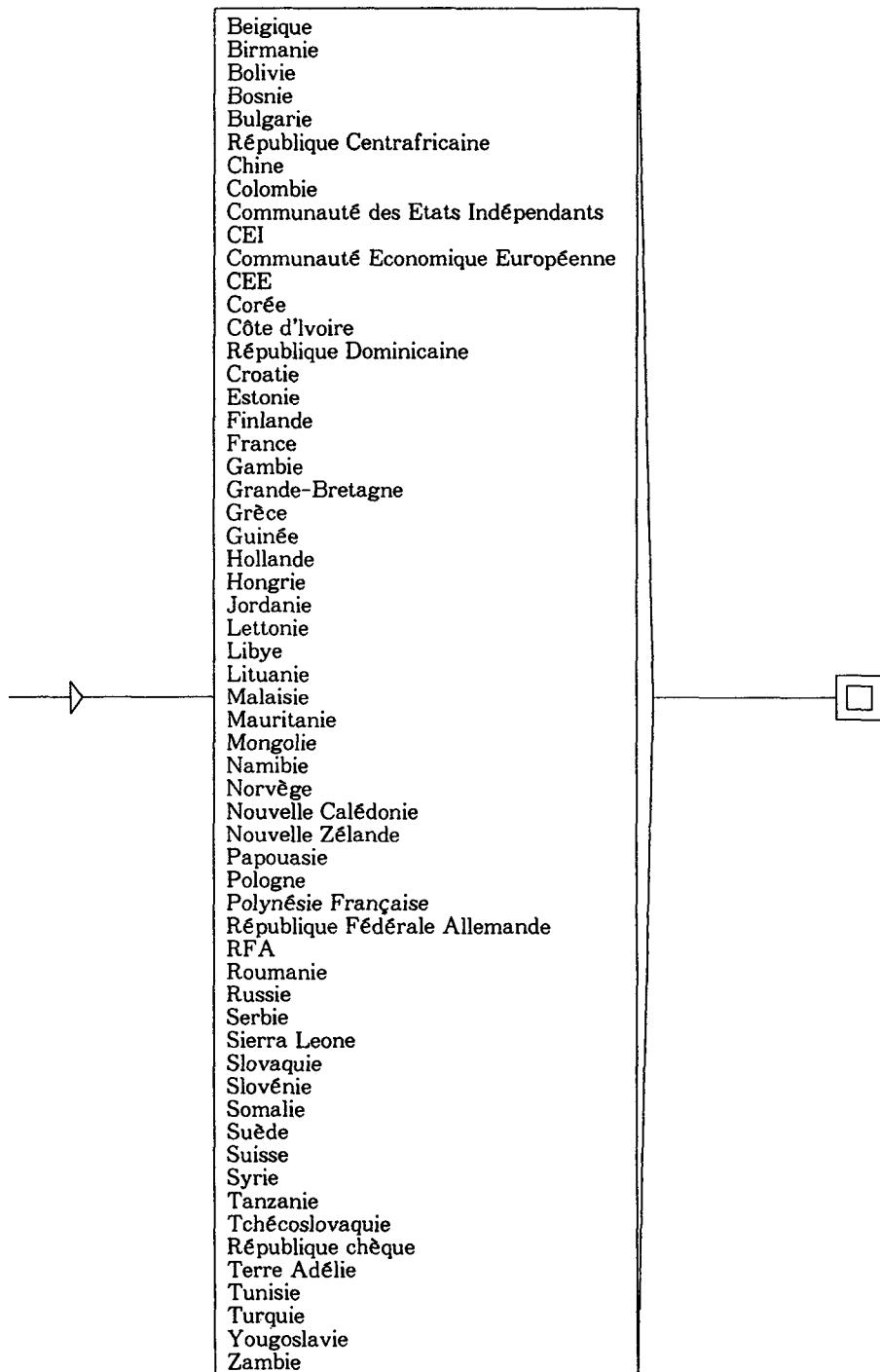


Fig. 5. PaysFermCons



Fig. 6. PaysClass

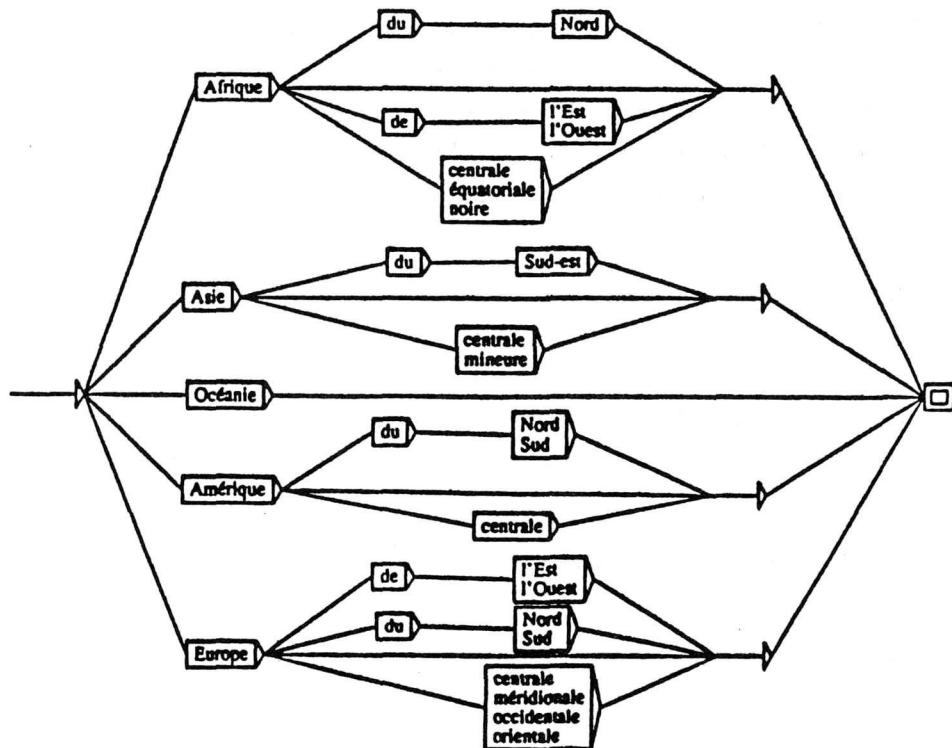


Fig. 7. Cont

the general graph *NVDansN* (cf. Fig. 11), devoted to this preposition.

– Graph *Cont*. This graph contains the names of the continents, and their variants (cf. Fig. 7).

The classification of country and continent names is facilitated by a cer-

tain number of identifiable, morphological criteria (masculine/feminine, vowel/consonant, etc.) which correspond to a relative regularity in their utilisation. The distributional organisation of the names of islands constitutes a problem of considerably greater complexity. It is this phenomenon that we intend to account for now.

#### 4. 2. Names of Islands

Independently of all syntactic considerations, the name of an island comprises a classifier of variable extension (*l'île de [the island of]/les îles de [the X isles/islands]*). The possible permutations are the following:

Singular constructions:

*Dét-'zero'-Nîle: Belle-Île est à une heure de bateau*

*Dét-Nîle: La Guadeloupe est très à la mode en ce moment*

*L'île-Nîle: L'île Maurice est idéale pour les vacances*

*L'île de-Nîle: L'île de Pâques est célèbre pour ses statues*

*L'île de la/le-Nîle: L'île du Diable est dangereuse*

Plural constructions:

*Dét-Nîle: Les Açores sont sous une dépression*

*Les îles-Nîle: Les îles Baléares sont plus fréquentable au printemps*

*Les îles de-Nîles: Les îles de Saint Pierre et Miquelon sont françaises*

*Les îles de Dét-Nîle: Les îles du Cap-Vert sont difficiles à situer*

*L'archipel des Nîles: L'archipel des Comores est au nord de Madagascar*

Classifiers can often be reduced, and sometimes even suppressed, although, as the following examples demonstrate, this procedure is unpredictable.

*(L'île de la, la) Guadeloupe est très courue.*

*(L'île de la, \*la) Tortue est peu fréquentée*

*(Les îles, les) Maldives sont devenues plus accessibles*

*(Les îles, \*les) Ioniennes sont connues*

If, on the other hand, we try to expand a classifier, the results are equally unpredictable.

*La (Corse, Crète) est très fréquentée*

*L'île de (\*Corse, Crète) est très fréquentée*

*Les (Comores, Philippines) sont intéressantes*

*Les îles (\*Comores, Philippines) sont intéressantes*

The system of classifiers for island names can, therefore, only be developed on the basis of a detailed, case by case, examination. This general description would, however, not be adequate for the purposes of creating the necessary graphs for our grammar. A further complication arises when, as is the case with our current research, these place names are combined with verbs of movement. The interference between our general description, and the preposition gives rise to new, and equally unpredictable, restrictions. Thus we find the following examples:

*Aller (à l'île de, \*à la) Crète*

*Aller (à l'île de, à) Chypre*

*(to go to (the island of) Crete/Cyprus)*

Unlike *Chypre*, *la Crète* is also to be found in the graph *IslFemSing* (cf. Fig. 10). However, as the following example shows, none of the other names contained in that graph could accompany *la Crète* in the graph *IslPrep* (cf. Fig. 8).

*Aller (en, \*à l'île de) (Corse, Sardaigne)*

*Aller (en, à l'île de) Crète*

*(to go to (the island of) Corsica/Sardinia/Crete)*

Furthermore, a change of preposition can result in a change of behaviour. The following example illustrates the consequences of the substitution of *à* (*to*) by *dans* (*in*):

*Aller à l'île de la Réunion, à la Réunion*

*Aller dans l'île de la Réunion, \*dans la Réunion*

These observations led us to adopt the following method in the construction of our graphs:

— Graph *IslPrep*. Graph of the island names which combine with prepositions other than *dans*. We placed the singular islands in the upper half of the automaton and the plural islands in the lower half (Fig. 8).

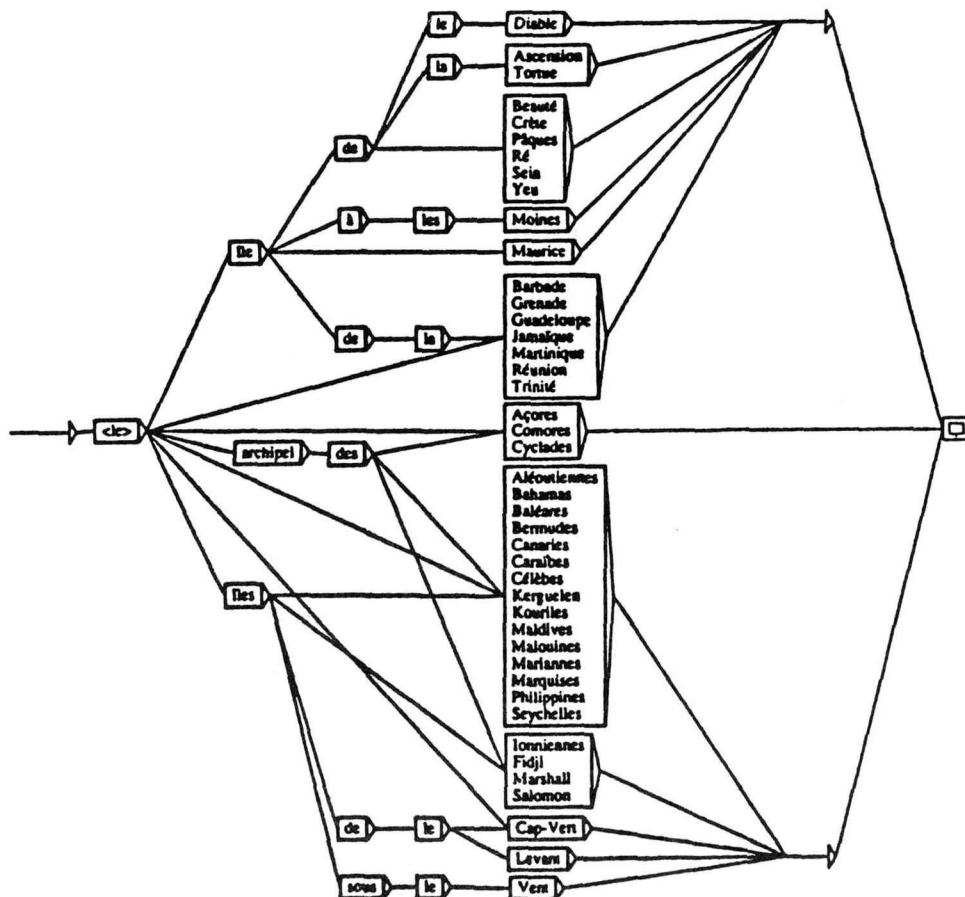


Fig. 8. HePrep

### *Débarquer à l'île de Ré*

### *Se diriger vers les îles Canaries*

- Graph *IslDetZ*. Graph of the island names which combine with prepositions other than *dans*, and which permit ‘zero’ determination (cf. Fig. 9).

### *Aller à (l'île de) Malte*

- Graph *IslFemSing*. Graph of the feminine island names which take the preposition *en* (*in/to*) (cf. Fig. 10).

Aller en Corse

Parallel to the graph  $NVP_{\text{Prep}}N$  (cf. Fig. 1), we devoted a general graph

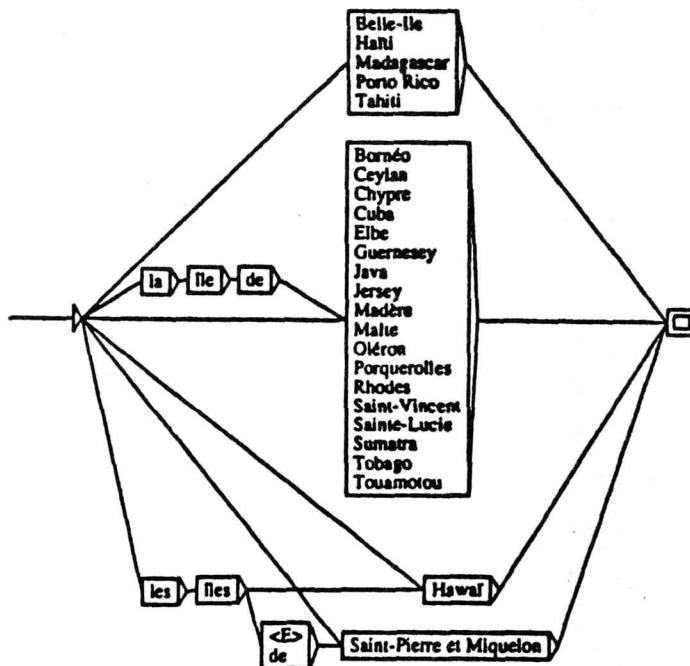


Fig. 9. IleDetZ

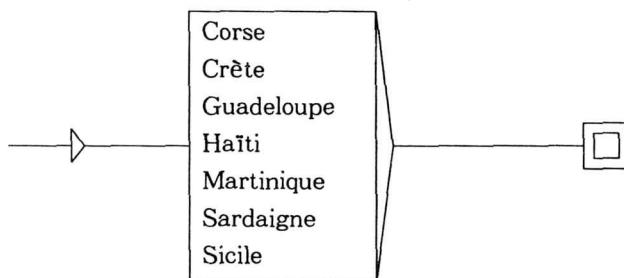


Fig. 10. IleFemSing

(*NVDansN*) to the use of the preposition *dans* (cf. Fig. 11).

It is worth noticing that, with the exception of the category of names in graph *LandClass*, country names will not take the preposition *dans*. On the other hand, this preposition is used extensively with island names. It is important to notice that the mere introduction of the preposition *dans* in the general description entailed a binary re-organisation of the context to the left (*V*), and to the right (*Nloc*), resulting in a specific description. For the verbs, we had to extract from the verbal paradigm of our basic automaton

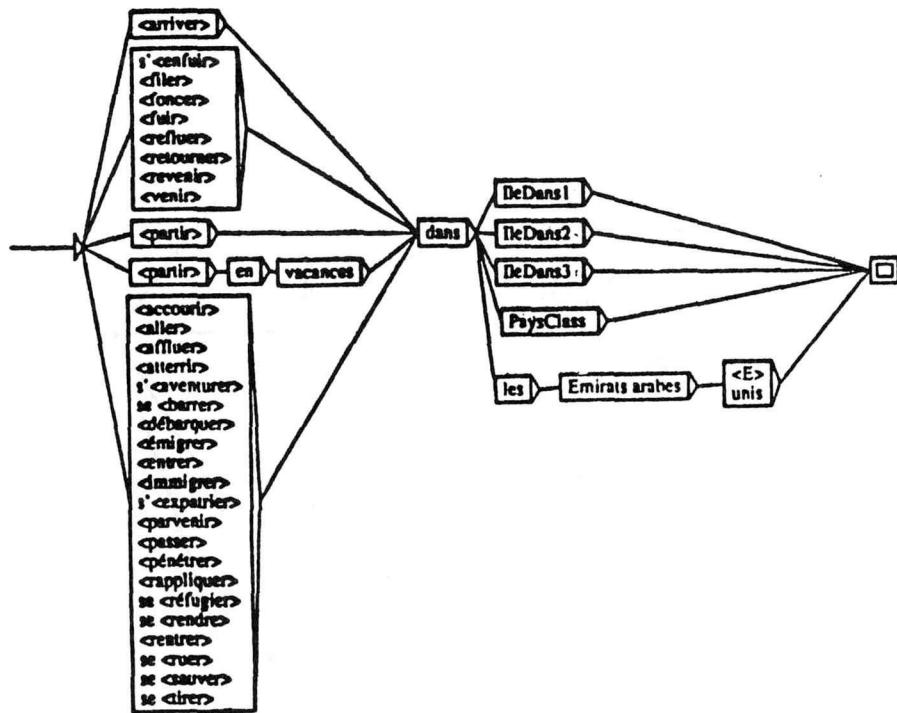


Fig. 11. NVDansN

*NVPrepN* (cf. Fig. 1) those verbs which function with the preposition *dans* accompanied by an island name

*(aller, partir, se rendre) dans les Baléares*

(*\*avancer, \*voler, \*appareiller*) dans les Baléares

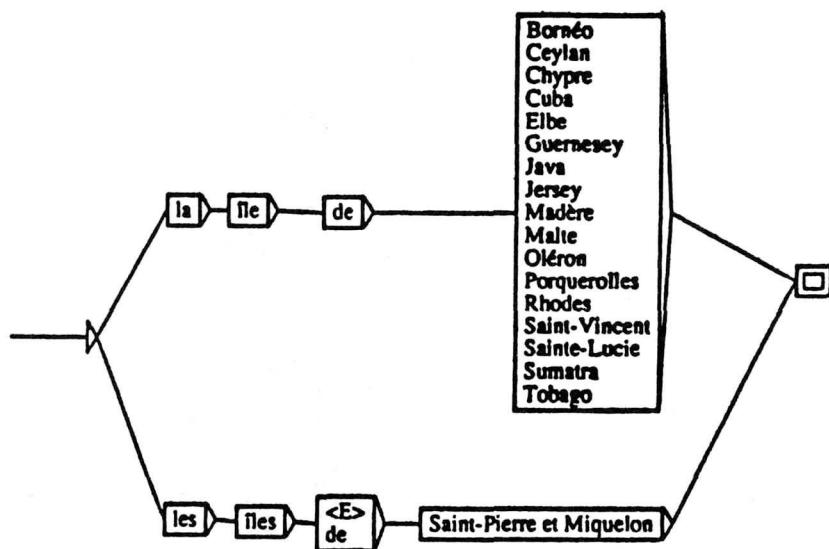
On the other hand, for the nouns, we had to re-organise the graphs of the island names in function of the modifications entailed by the use of the preposition *dans*. Thus, of those island names listed in the graph *IslFemSing* (cf. Fig. 10), some cannot be used with *dans*, whereas others can, but only with the extended classifier:

aller en (Corse, Crète, Sardaigne)

aller dans la (\*Corse, \*Crète, \*Sardaigne)

aller dans l'île de (\*Corse, Crète, \*Sardaigne)

Furthermore, among the island names listed in the graph *IslPrep* (Fig. 8), we notice that the group *Barbade*, *Martinique*, *Trinité*, etc. admits the possi-

Fig. 12. *ÎleDansI*

bility of two alternative classifiers, *Île de la/Det-Nisl*:

*aller à(l'île de la, la) (Barbade, Martinique, Trinité)*

*to go to [the island of (the)/φ(the)] (Barbados, Martinique, Trinidad)*

However, with the preposition *dans*, this alternation is no longer available:

*aller dans (l'île de la, \*la) (Barbade, Martinique, Trinité)*

On the other hand, for the group of plural island names (*Aléoutiennes, Bahamas, etc.*), from the same graph *IslPrep*, both alternatives for the classifier (*Les îles/Det-Nisl*) are acceptable, even with the preposition *dans*:

*Aller (à les îles, à les) (Bahamas, Baléares, Seychelles)*

*Aller (dans les îles, dans les) (Bahamas, Baléares, Seychelles)*

Our approach was, therefore, to take the graphs of the island names developed for use with prepositions other than *dans*, and to carry out in each those alterations made necessary by the new context on the left along with the preposition *dans*. Thus we arrived at the following graphs:

- Graph *NVDansN* derived from *NVPrepN* (cf. Fig. 11)
- Graph *IslDansI* derived from *IslDetZ* (cf. Fig. 12)

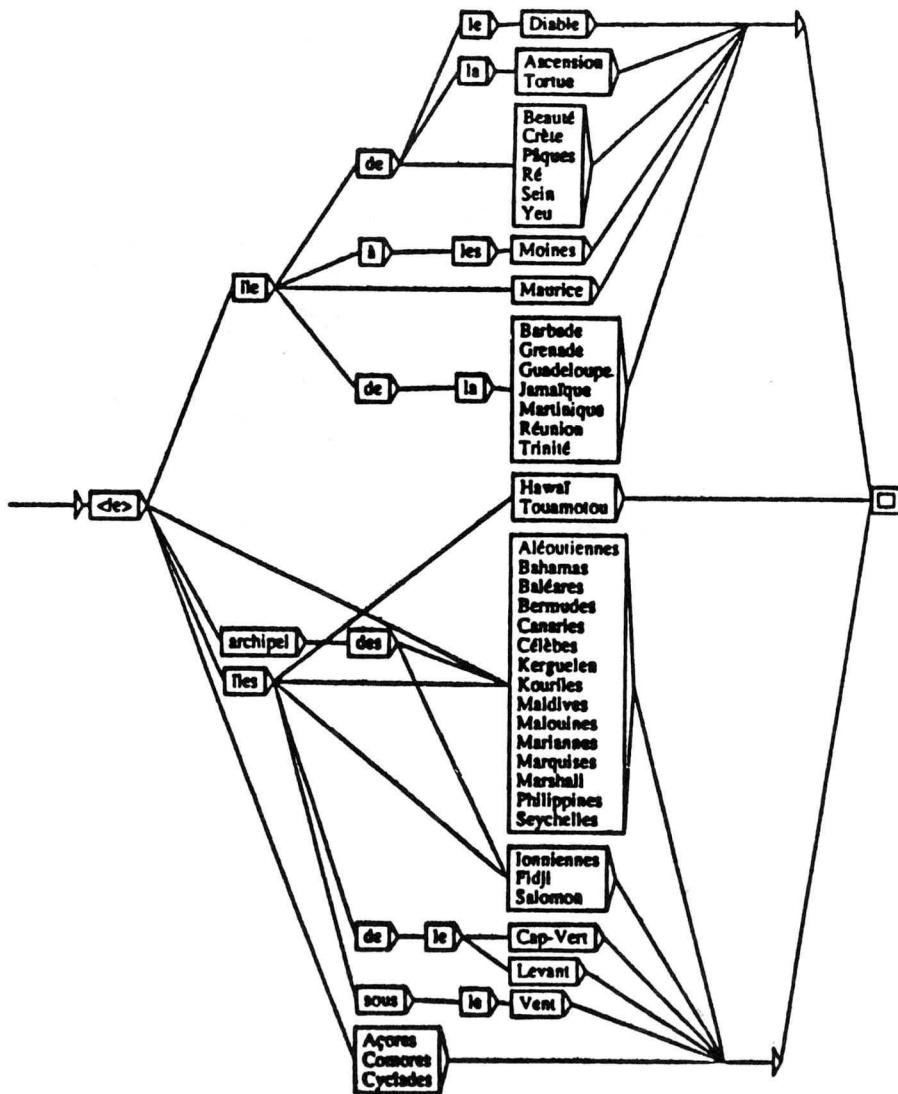


Fig. 13. *IleDans3*

- Graph *IslDans3* derived from *IslPrep* (cf. Fig. 13)

The graph *IslDans2* is not derived, and includes those island names which cannot be shortened (cf. Fig. 14).

We intend to extend this set of graphs devoted to the use of the preposition *dans* subsequently to include graphs of the names of *départements* ( $\approx$  regions), since these are used regularly with this preposition.

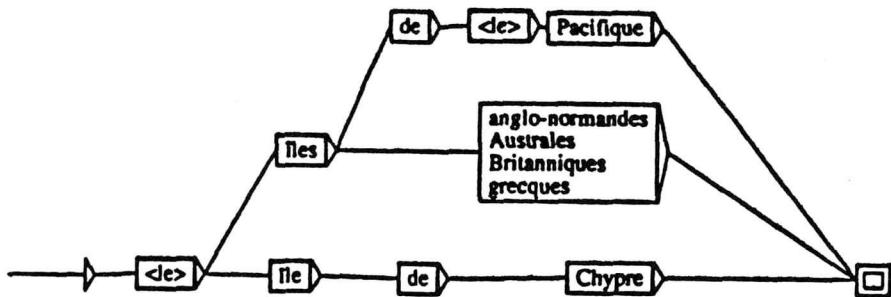


Fig. 14. IlleDans2

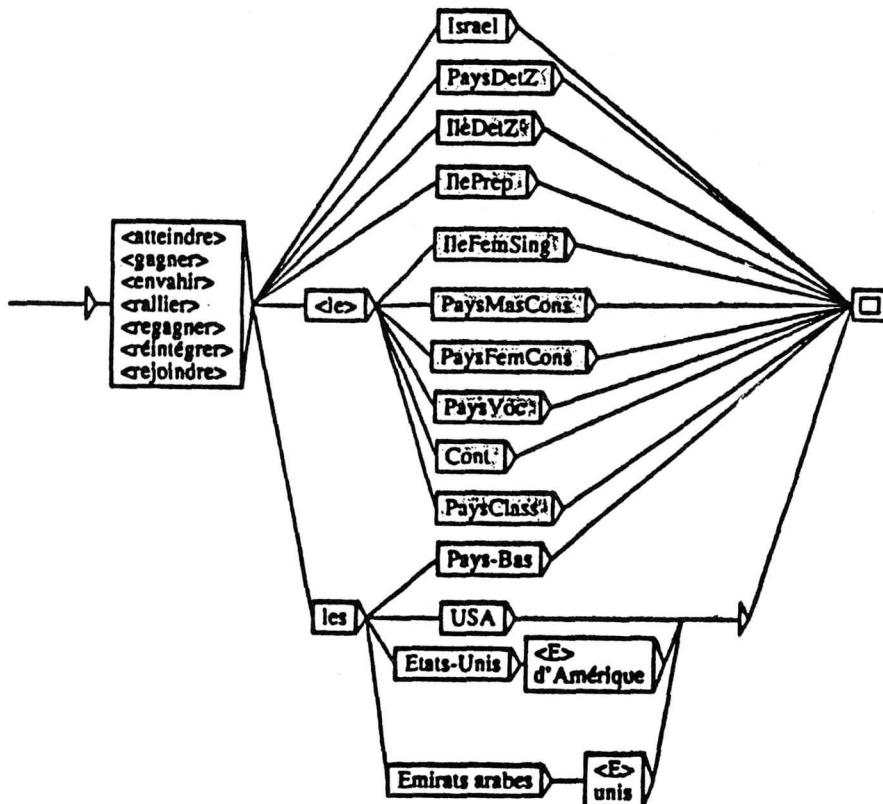


Fig. 15. NVN

The syntactic structure (b) of our predicate is represented in the graph NVN (cf. Fig. 15). This graph provides a formal representation of sequences such as:

*Atteindre (les Etats-Unis, la Barbade)*

*To reach (the United States, Barbados)*

Certain graphs (or modules) of our grammar could subsequently be re-injected into the grammars of certain neighbouring scenarios. It is therefore desirable, from a methodological point of view, to proceed in a highly structured, modular fashion. These selective, limited grammars will, by linking up with each other, progressively encompass all the scenarios which include a locative preposition. On a more general level, the accumulation of thousands of local grammars will ultimately encompass the morpho-syntactic system of our language in its entirety. This approach, as developed and used at the LADL, is made possible by the above use of finite automata for the representation of linguistic phenomena, and by the existence of applications software (INTEX system) which enables them to be put into practice.

## 5. Computerisation and Implementation

An automata editor (FSGRAPH), implemented under the INTEX system of textual engineering, enables the user to create graphs directly on the computer screen. This means of construction is very easy to use, and permits the immediate correction of the graph by means of the addition, or elimination of elements inside the boxes, or nodes, and also the addition or elimination of the paths linking them. The advantage of this tool is that it allows for the modular construction of the grammar, and for a direct representation of the linguistic classification of phenomena (cf. D. Maurel, 1989). It is also capable of the automatic conversion of a graph into a recognition automaton.

We have provided a list of examples of automatic recognition carried out on the basis of such grammars. The corpus consists of a day's AFP (Agence France Presse) dispatches (cf. Fig. 16). The system automatically lemmatises the inflected form of each verb, and furnishes these forms as a list of concordances.

It is also worth noticing that the *PrepNloc* elements of the automata we have provided can be found as the complements of various verbs (*téléphoner en Chine*), and as the complements of substantives derived from verbs (*l'arrivée au Japon, the arrival in Japan*) and of non-derivational forms (*l'ambassade de France en Chine, the French embassy in China*). This demonstrates the considerable generalising power of this form of representation.

Before	going to Asia	, where the first
after his	arrival in Ethiopia	.
against the	arrival in France	of Japanese indu
at the moment of her	arrival in Greece	, gave birth in a h
had	arrived in Israel	as a result of
managed to	reach Greece	and 35 people als
I intend to	return to Argentina	but Japan interest
He(...)	would return to Italy	to become manag
6 young Albanians	are taking refuge in Italy	.
was not able to	go to France	, said his spoke
are going to	go to Sweden	this summer
managed to	go to Syria	.
managed to	go to Yugoslavia	.
are invited to	come to Europe	”, declared M. A

Fig. 16. Concordances, dépêches de l'AFP Concordances, AFP dispatches

## References

Boons, Jean-Paul, Guillet, Alain, Leclère Christian (1976) *La structure des phrases simples en français I: Constructions intransitives*, 378, Geneva: Droz.

Garrigues, Mylène (1993) *Méthode de paramétrage des dictionnaires et grammaires électroniques*, Doctoral dissertation, Université Paris 7, LADL.

Gougenheim, Georges (1962) *Système grammatical de la langue française*, Paris: Editions d'Artrey.

Gross, Gaston (1992) *Classes d'objets et description des langues*, Report LLI, Université de Paris XIII, Villetaneuse.

Gross, Maurice (1975) *Méthodes en syntaxe*, 414, Paris: Hermann.

Gross, Maurice (1989) 'The use of finite automata in the lexical representation of natural language', *Electronic Dictionaries and Automata in Computational Linguistics*, 34–50, Berlin–New York: Springer.

Guillet, Alain, Leclère Christian (1992) *La structure des phrases simples en français. II. Constructions transitives locatives*, 445, Geneva: Droz.

Maurel, Denis (1988) 'Grammaire des dates, Etude préliminaire à leur

traitement automatique', *Linguisticae Investigationes*, 12.1, 101-128.

Schank, Roger C. (1990) *Tell Me a Story: A New Look at Real and Artificial Memory*, New York: Macmillan Publishing Company.

Silberztein, Max (1993) *Dictionnaires électroniques et analyse automatique de textes: Le système INTEX*, 233, Paris: Masson.

LADL Universite Paris 7  
2 place Jussieu, Tour Centrale,  
F 75251, Paris Cedex 05  
France  
E-mail : mgarrig@ladl.jussieu.fr