

INAUT, a Controlled Language for the French Coast Pilot Books *Instructions nautiques*

Yannis Haralambous, Julie Sauvage-Vincent, and John Puentes

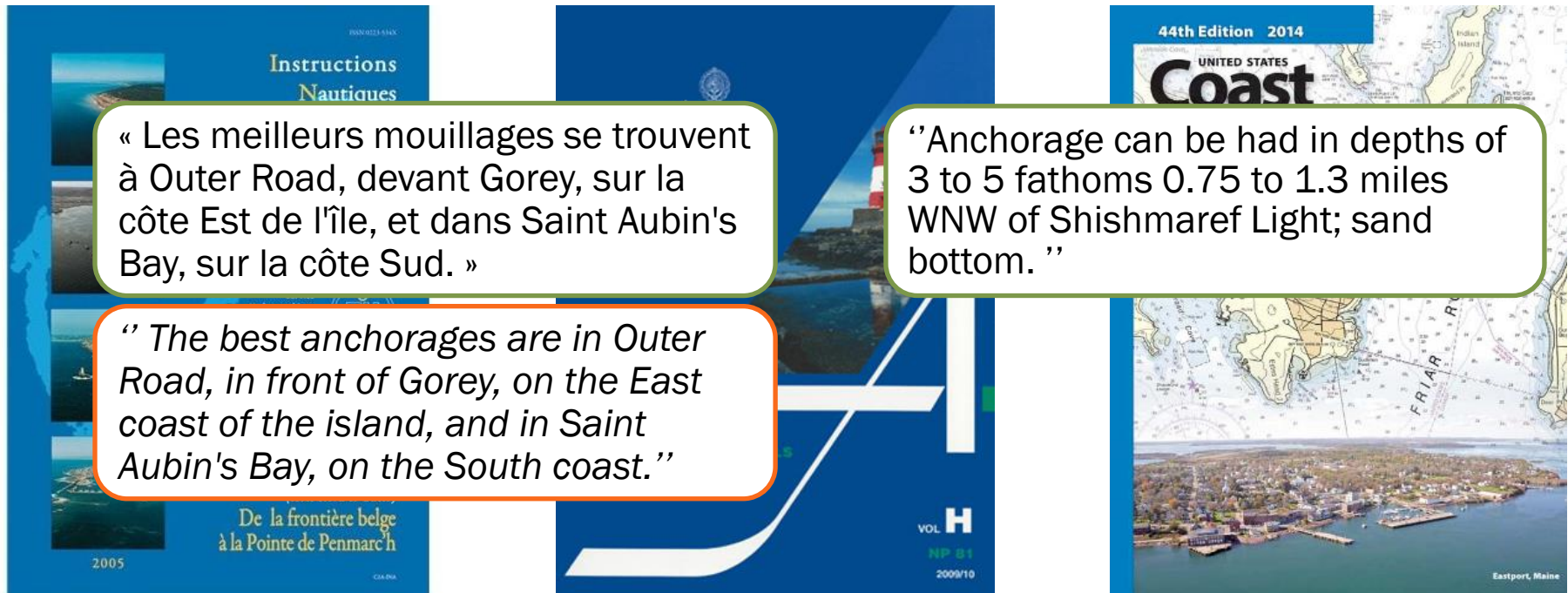
CNL 2014, Galway



Institut
Mines-Télécom

What are *Instructions nautiques* ?

- A series of nautical books
- Published by the French Hydrographic and Oceanographic Service
- A supplement to the nautical charts



The French Hydrographic and Oceanographic Service

SHOM

- Service Hydrographique et Océanographique de la Marine

Hydrography

- The branch of applied sciences which deals with the measurement and description of the physical features of oceans, seas, coastal areas, lakes and rivers, as well as with the prediction of their change over time

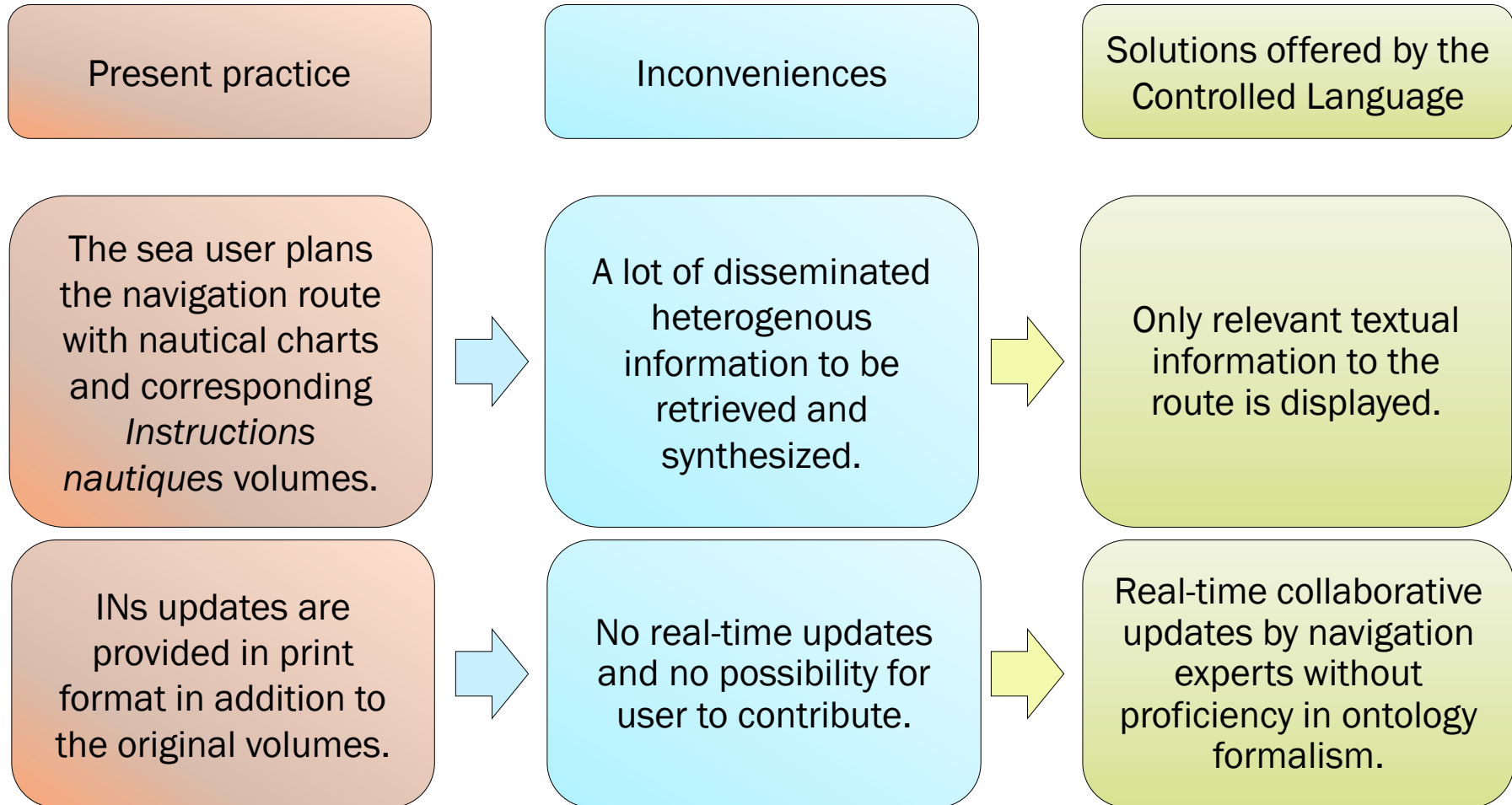
Oceanography

- The description of the marine environment

3 main aims :

- The provision of hydrography and maritime cartography
- Support to defence
- Support to government maritime and coastal policies

Context



A SHOM Knowledge Base (KB)

The SHOM aim

- Building a KB dealing with both nautical charts and nautical instructions.

Constraint 1

- The information contained in the KB has to be easily updatable by people not necessarily proficient in ontology formalism.

Solution proposal

- A CNL dedicated to the update and population of the SHOM KB.

Constraint 2

- The *Instructions nautiques* have to be generated out of the KB.

Solution proposal

- Generation of texts in INAUT out of the KB and of LitINAUT (=Literary INAUT) via aggregation and referring expressions generation operations.

Modellin

1/2

IN Volume D2.1 France
De la frontière espagnole a

Chap. 0
Introduction

Chap. 1
Rens. généraux

D

§2.1
Généralités

De la front
à Argel

§2.2.1
Cap Cerbère

§2.2.2
Port de Cerbère

§2.2.3
Du port de Cerbère au port
de Banyuls-sur-Mer

Général

01 3.5.5. PRINCIPAUX PORTS.

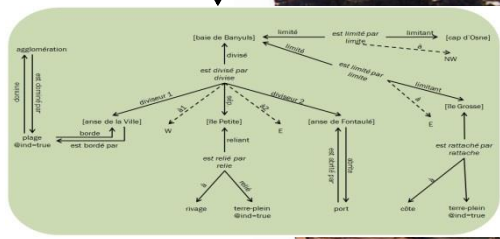
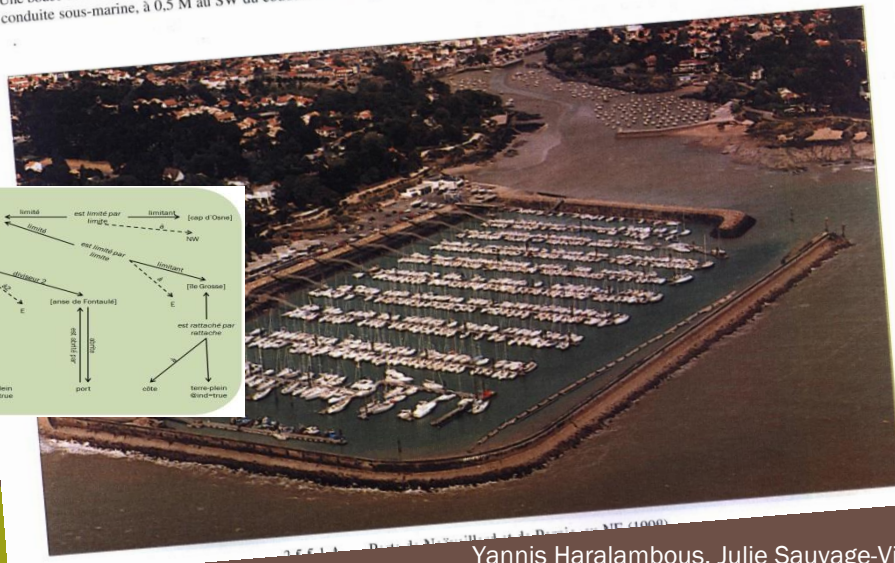
07 Carte 7394.

01 3.5.5.1. **Port de Pornic et port de plaisance de la Noëveillard** (47° 06,5' N — 2° 06,6' W) [vue 3.5.5.1.A et planche 3.5.5.1.B]. — Cet ensemble est établi sur la côte Nord de la Baie de Bourgneuf, à environ 6 M à l'Est de la Pointe de Saint-Gildas, entre la Pointe de Noëveillard à l'Ouest, et celle de Gourmalon à l'Est. Le port d'échancrure de la côte au fond de laquelle se trouve la ville ; le port de plaisance de la Noëveillard, en eau profonde, a été gagné sur la mer à l'Ouest de l'entrée à Pornic. Le port de Pornic est affecté principalement aux bateaux de pêche, de plaisance et de promenade en mer. Le port de plaisance de la Noëveillard peut recevoir 920 bateaux, dont 165 de passage, sur pontons, par 1,5 à 1,8 m d'eau. Du point de vue administratif, le port de Pornic englobe le port de plaisance dont la digue coude forme les limites Ouest et SW ; au Sud et au SE, le port est limité par la droite orientée environ à 069° depuis le musoir de la digue coude jusqu'à la **Pointe de Gourmalon**.

09 **COURANTS DE MARÉE.** — Devant l'entrée et au mouillage extérieur le courant de flot porte à l'Est et atteint 1,5 nœud en vive-eau, le courant de jusant porte à l'Ouest et atteint 2 nœuds en vive-eau.

13 **ACCÈS ET INSTRUCTIONS.** — Venant de Saint-Nazaire, on passe au Nord du Banc de Kerouras en longeant la côte Nord de la baie ; l'échancrure de Pornic se distingue d'assez loin sur cette côte sombre. En approchant, on aperçoit ensuite plus nettement le mur de béton qui forme la digue coude du port de la Noëveillard. Une bouée de marque spéciale, lumineuse, signale l'extrémité d'une conduite sous-marine, à 0,5 M au SW du coude de cette digue.

- 17 Du SW l'accès est compris entre l'extrémité Est du Banc de Kerouras et le **Plateau de Notre-Dame** signalé par une tourelle de danger isolé. Plus au SE une chaîne de hauts-fonds prolonge le plateau et porte une tourelle latérale bâbord marquant **Le Caillou**, roche découvrante. À 1 M environ dans le SW de cette tourelle, une bouée de danger isolé, à cloche, marque la **Pierre-du-Chenal**.
- 21 Dans l'accès du SW on est guidé par l'alignement à 055° du clocher de Pornic par le phare de la **Pointe de Noëveillard**, tour carrée verte et blanche et maison blanche (15 m). De nuit, on fait route sur le phare dans son secteur blanc 051°-079°.
- 25 La bouée de marque d'eaux saines, lumineuse, « 1 », est mouillée à environ 100 m dans le SSE du phare de la digue Est du port de plaisance. Le chenal de Pornic est ensuite balisé par des marques latérales.
- 29 Une balise de marque spéciale est implantée à l'aspiration d'eau de mer de la station de thalassothérapie de Pornic, à 116° et 1 395 m du phare de la Pointe de Noëveillard.
- 33 Le port de Pornic communique avec le canal, ou étier, de Haute-Perche par un vannage de décharge automatisé. L'évacuation des crues de cet étier crée un courant dont il faut tenir compte à l'entrée du port devant le Môle Leray et à l'amarrage à ce môle et au Quai de l'Écluse.
- 37 **MOUILLAGE.** — On mouille sur l'alignement d'accès par 3 à 5 m d'eau, fond de sable et vase.
- 41 **PORT DE PORNIC.** — On accède au port de Pornic entre la digue Est du port de plaisance à l'Ouest et la Pointe de Gourmalon à l'Est par un chenal asséchant de 0,3 à 1,6 m, marqué par des balises et des bouées, de marque latérale. Le port de pêche a une longueur de 240 m depuis le môle Leray, situé sur la rive Nord à 750 m dans le NE de la Pointe de Gourmalon, jusqu'au vannage de **Haute-Perche**. Sa largeur moyenne est de 70 m.



Modelling the *Instructions nautiques* 2/2

- 3 graphs
 - The hierarchical structure of the document
 - The geographic areas graph
 - The SHOM knowledge base

- 2 functions
 - A function which maps georeferenced nodes of the hierarchical structure of the document and of the KB to nodes of the geographic areas graph
 - A function which maps leaf nodes of the structure to subgraphs of the KB

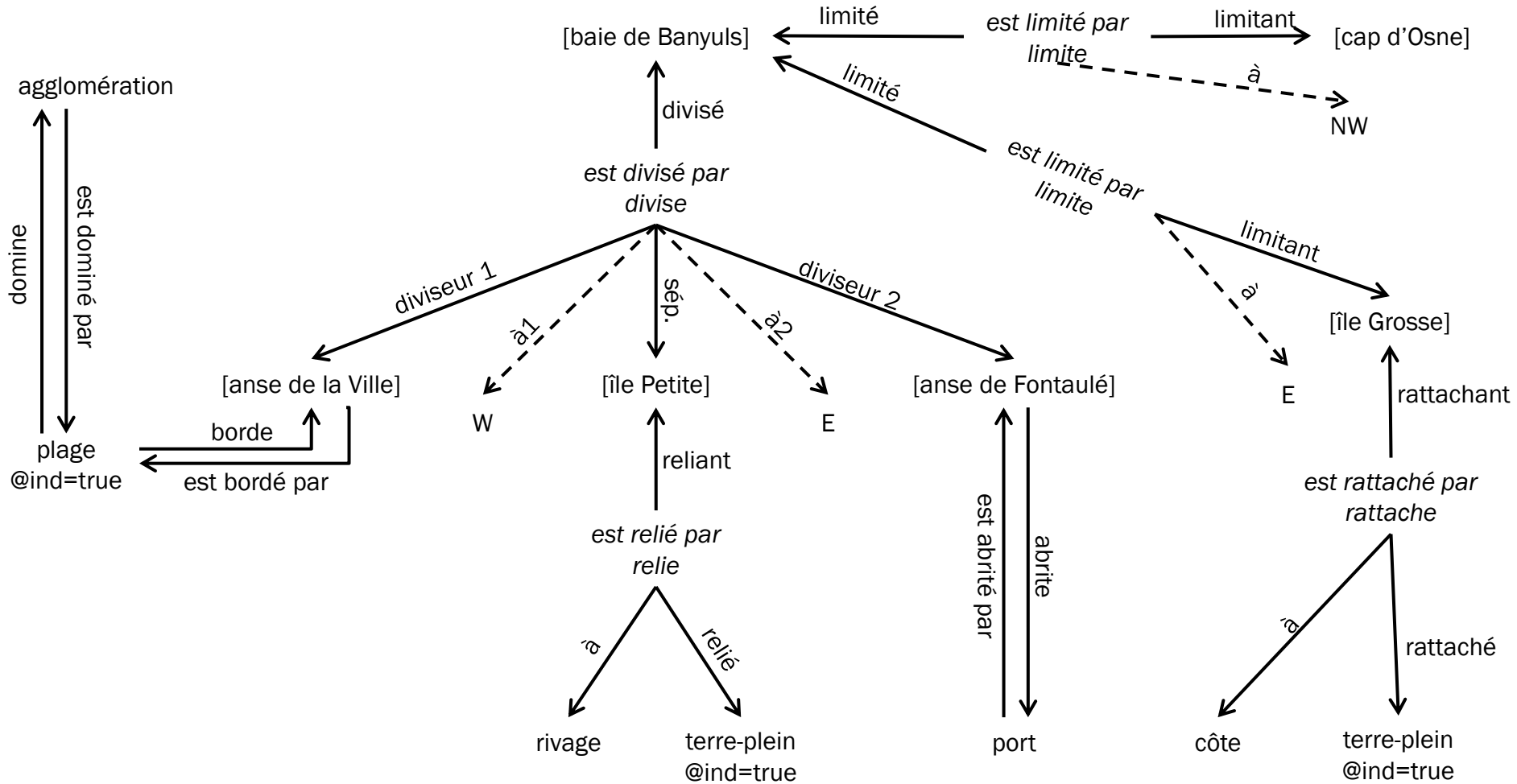
- 2 sets
 - A set of titles of hierarchical subdivisions
 - A set of geopolygons

- 2 functions
 - A function which maps nodes of the structure to the set of titles
 - A function which maps nodes of the geographical areas graph to the set of geopolygons

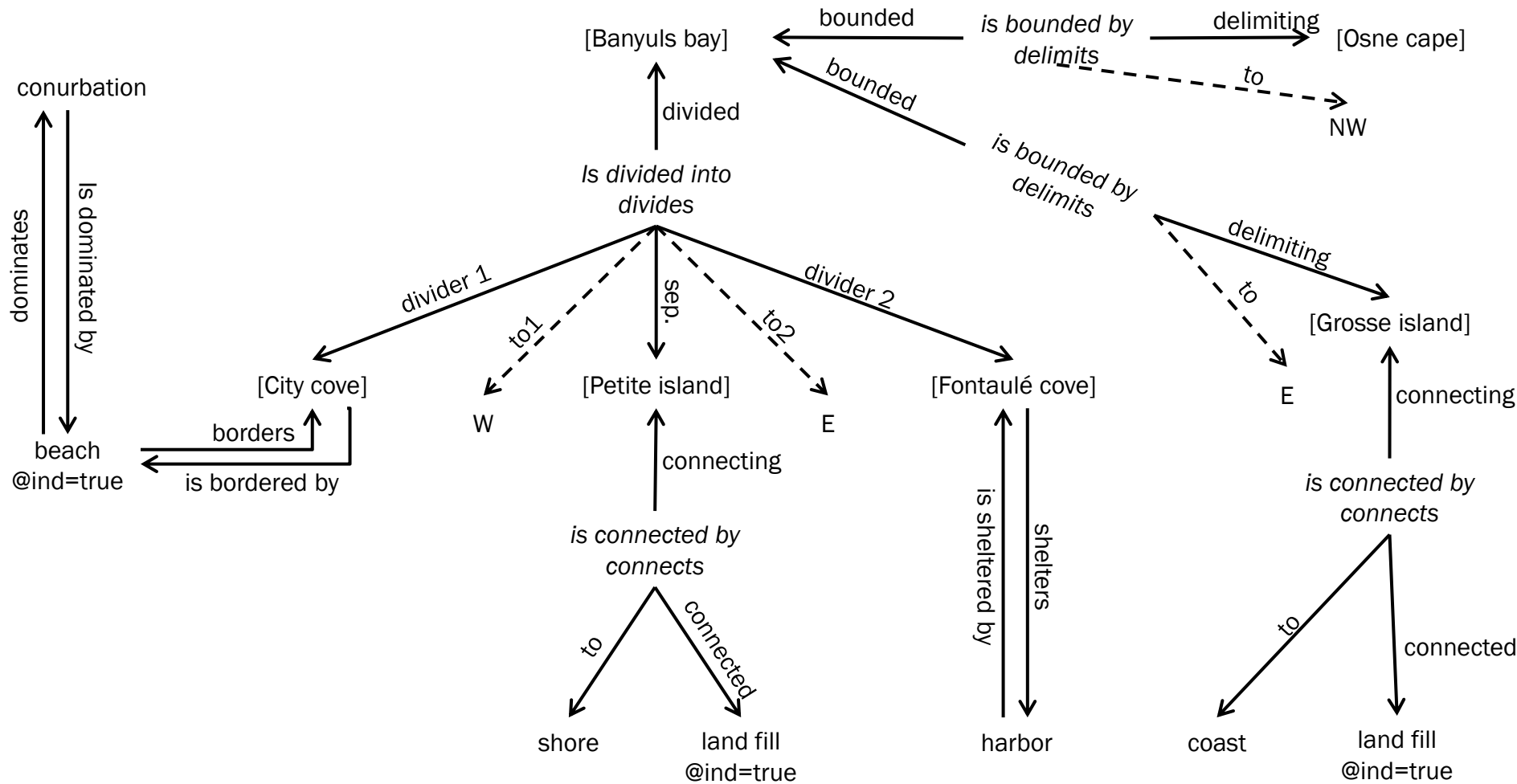
Knowledge Base 1/3

- Sets of:
 - Concepts
 - Attributes
 - Simple relations
 - Complex relations
 - Types
 - Instances
 - Values
- Sets of names:
 - Concepts
 - Attributes
 - Simple relations
 - Complex relations
 - Instances

Knowledge Base 2/3



Knowledge Base 3/3



The controlled languages INAUT and LitINAUT 1/3

A simple grammar:

- All nouns belong to the set of lexical references for instances of the KB and to the set of values of attributes of the KB
- All verbs belong to the set of names of complex and simple relations
- All adjectives belong to the set of values of attributes of the KB
- Modifiers, determinants and prepositions belong to a closed list

S → NP VP
NP → modif det NN | det NN | NN
NN → adj NN | NN adj noun
VP → verb NP | verb NP PP
PP → prep det NN | prep NN
...

The controlled languages INAUT and LitINAUT 2/3

Verbs:

- Always in 3rd person or infinitive
- Can be active or passive:
 - Voice can be changed
 - Permutation of the NPs in subject and object position

La [baie de Banyuls] est limitée par le [cap d'Osne] au NW.

The Banyuls bay is bounded by the Osne cape to the NW.

Le [cap d'Osne] limite la [baie de Banyuls] au NW.

The Osne cape delimits the Banyuls bay to the NW.

The controlled languages INAUT and LitINAUT 3/3

Articles

- Definite articles: all instances in the geographical areas graph if the name starts with the name of a concept to which belongs the instance

“baie de Banyuls” starts with “baie” (=bay)



“baie” is the name of a concept



Hence → “la [baie de Banyuls]”

- No article: all other instances in the geographical areas graph

[Notre-Dame de la Salette] est un amer remarquable à l’WSW du port.

[Notre-Dame de la Salette] is a remarkable daymark on the West Southwest of the harbour.

- Indefinite articles: by default, all non-georeferenced instances. Object position only:

L’[anse de la Ville] est bordée par **une** plage. **La** plage est dominée par l’agglomération.

The [anse de la ville] is bordered by **a** beach. **The** beach is dominated by the conurbation.

Controlled language generation

Automatic production is done in two steps:

1. Produce INAUT text corresponding to a given leaf node of the hierarchical structure
2. Convert INAUT language into LitINAUT

According to Reiter & Dale [2000] the language generation is divided into seven **subtasks**:

1. Content determination
2. Document structuring
3. Lexicalization
4. Aggregation
5. Referring expression generation
6. Linguistic realization
7. Structure realization

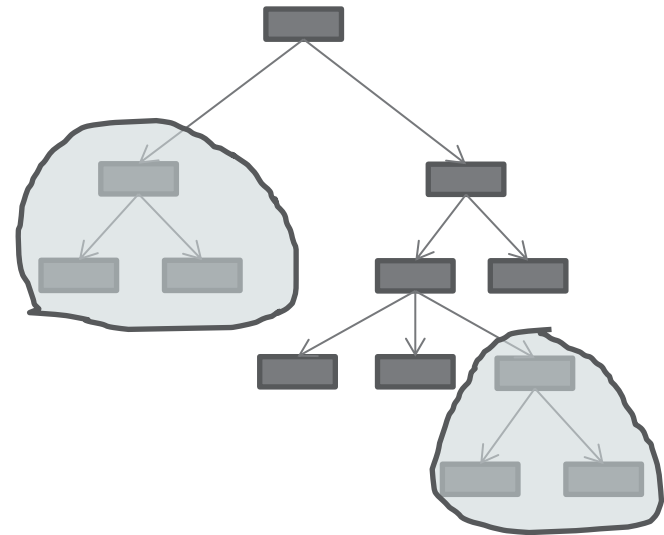
Controlled language generation: Content determination

Using a content determination algorithm:

- We find the subgraph in the KB the most relevant to the hierarchical structure of the document
- The connected components are tagged using a rule-based decision system

2 filterings:

- Geographical
- Thematic



Controlled language generation: Document structuring

To convert a subgraph into INAUT, we need to:

1. Sort the connected components of the subgraph:
 - i. The larger cumulated geographic areas will precede the smaller one
 - ii. Or, the path defined by the calculated barycenters of cumulated geographic areas should be roughly parallel to the guiding path of the volume
2. Find a starting node in each connected component and the order for the conversion of each component into INAUT.

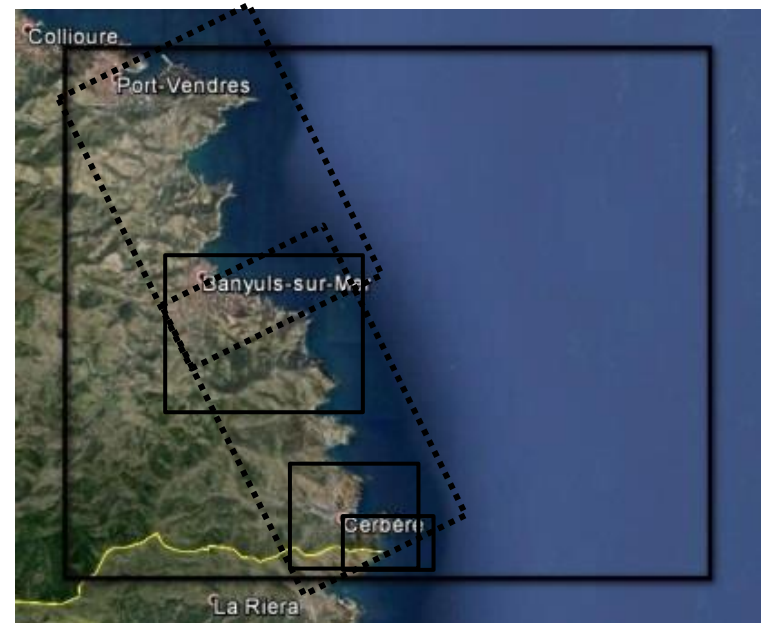
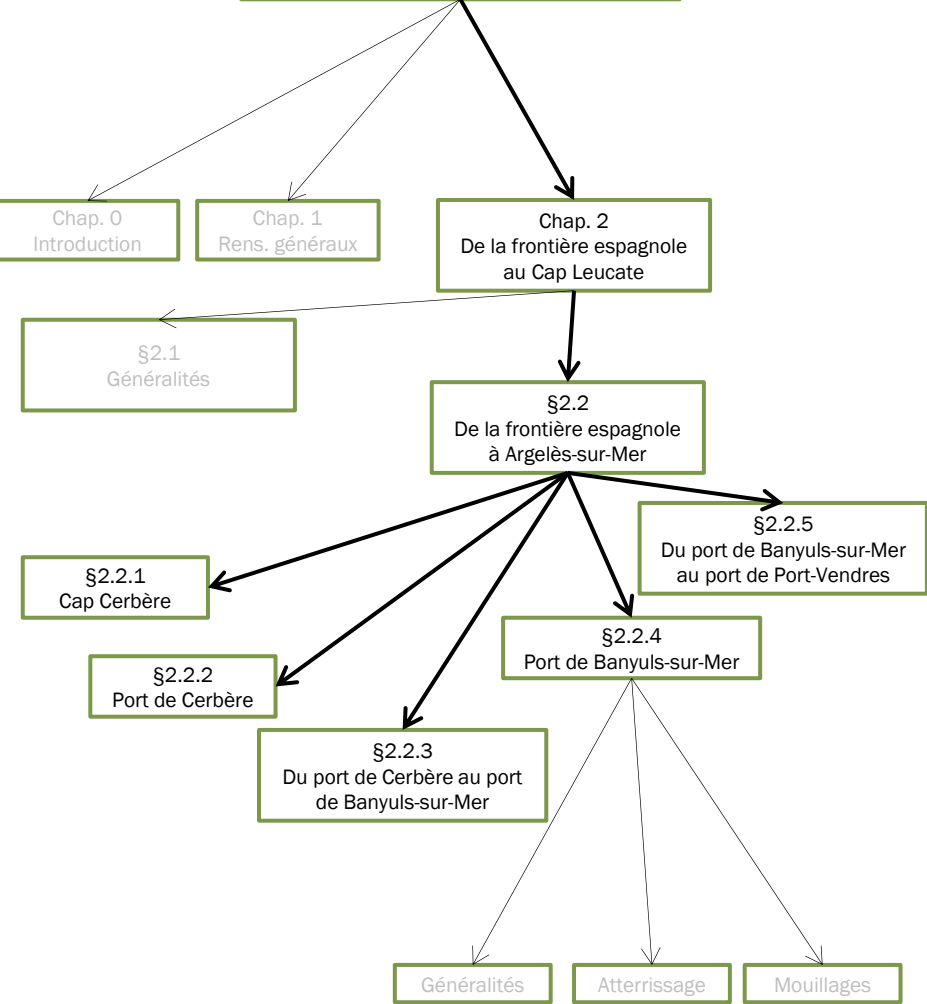
We need a weighting system:

- i. A logical method:
 - a. The closer the geographic area/name of a node to the one of the parent of the leaf node, the better
 - b. The position of the starting node in the geographical areas graph lattice
 - c. The interest for the navigator
- ii. A statistical method

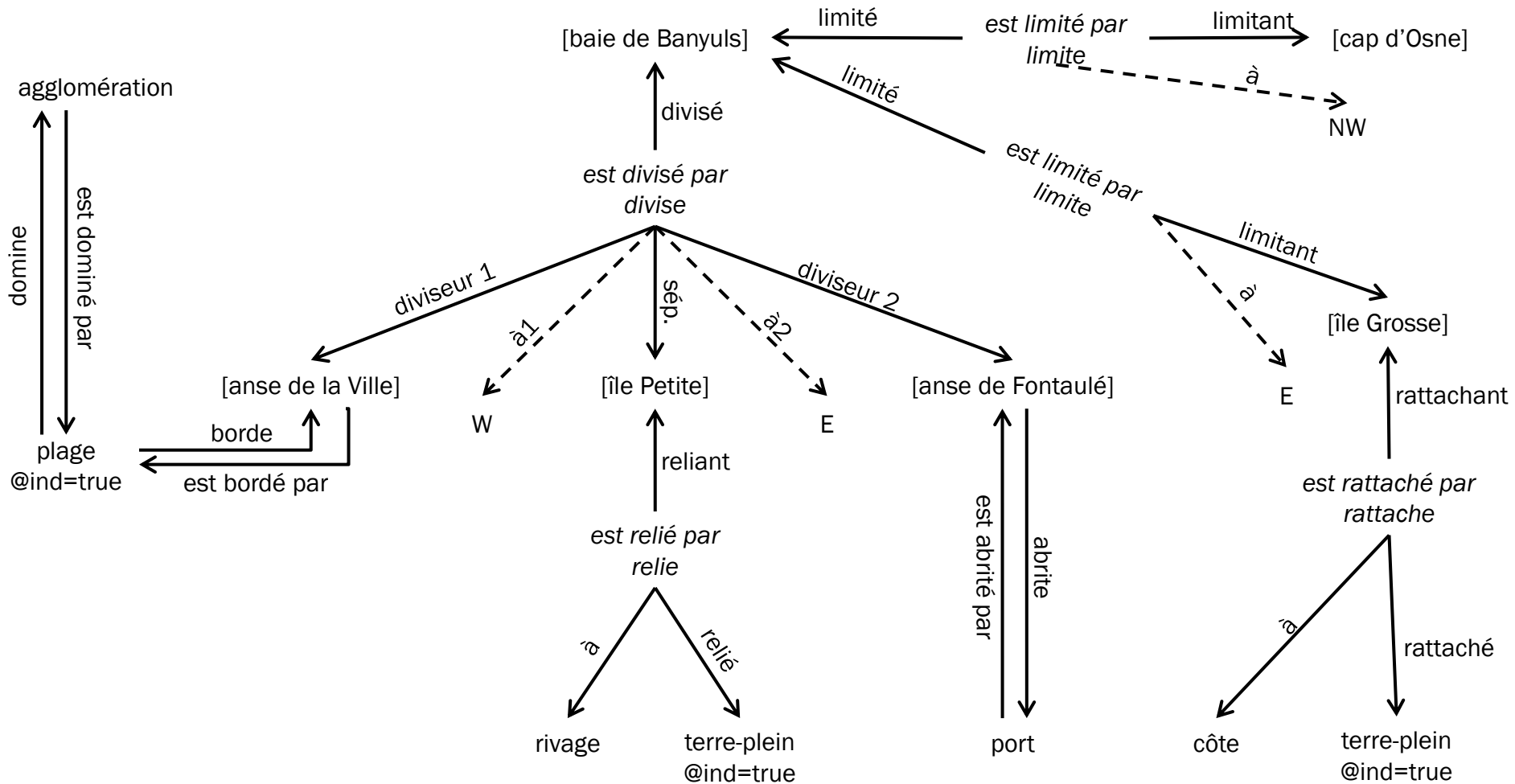




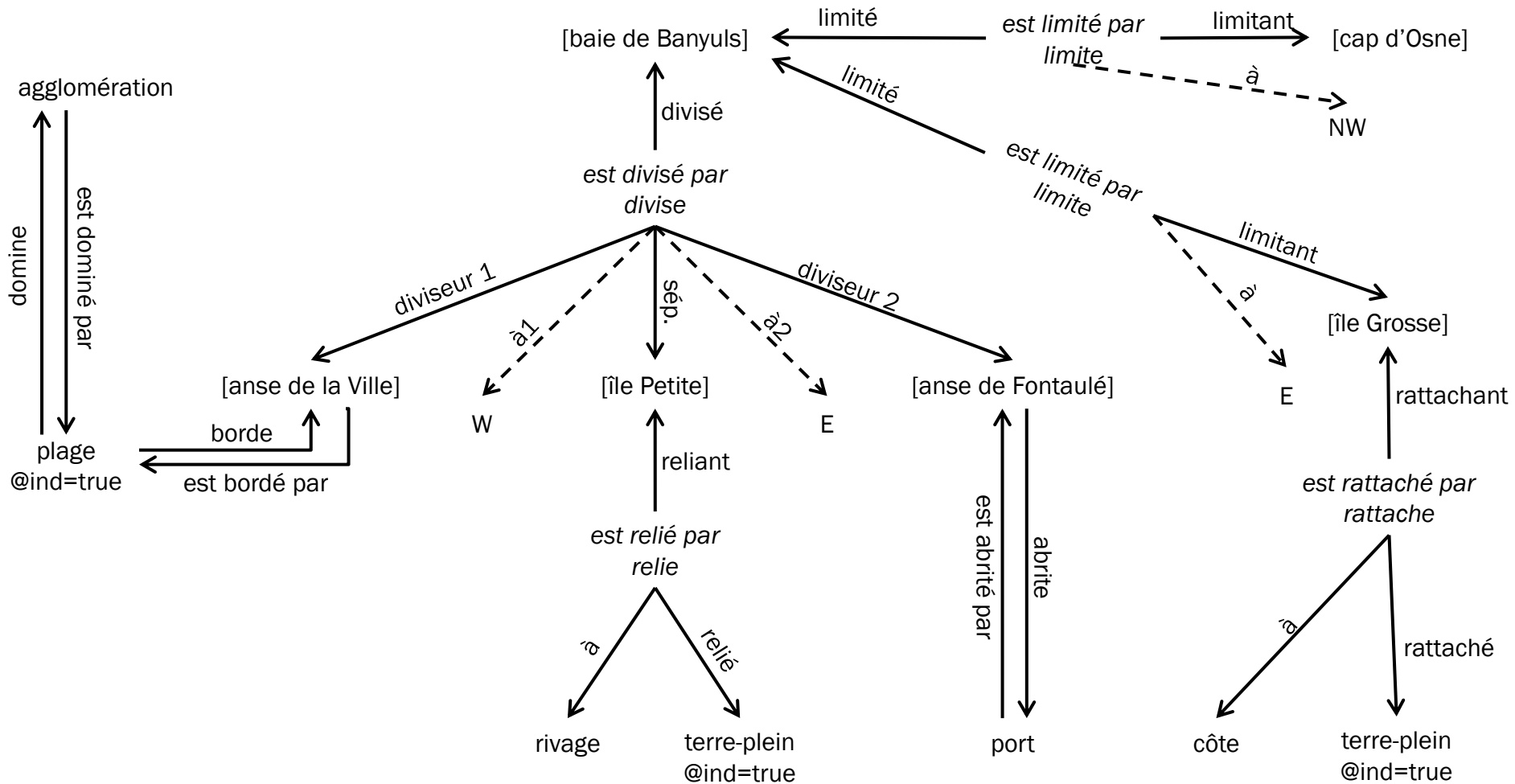
IN Volume D2.1 France (Côte Sud)
De la frontière espagnole au Cap de l'Aigle



Variation of path depending on the weighting 1/3

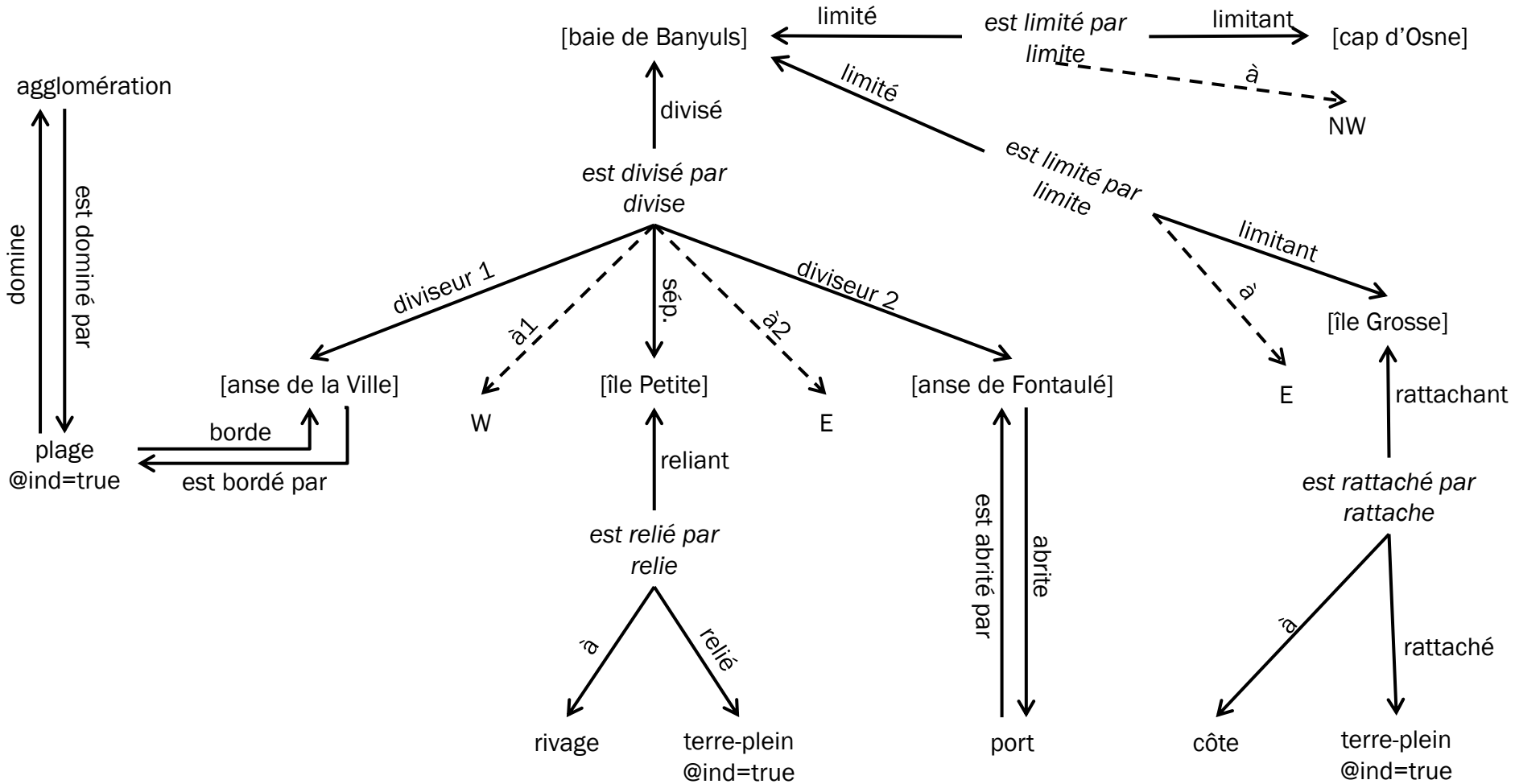


Variation of path depending on the weighting 2/3





Variation of path depending on the weighting 3/3



Controlled language generation:

Aggregation and referring expressions generation, LitINAUT language 1/2

LitNAUT

La [baie de Banyuls] est limitée au NW par le [cap d'Osne] et à l'Est par l'île Grosse] rattachée à la côte par un terre-plein.

The [Banyuls bay] is bounded on the NW by the [Osne Cape] and on the East by the [Grosse island] connected to the coast by a land fill.

INAUT

La [baie de Banyuls] est limitée au NW par le [cap d'Osne].
La [baie de Banyuls] est limitée à l'Est par l'île Grosse].
L'île Grosse] est rattachée à la côte par un terre-plein.

The [Banyuls bay] is bounded on the NW by the [Osne Cape].
The [Banyuls bay] is bounded on the East by the [Grosse island].
The [Grosse island] is connected to the coast by a land fill.

- When the object of a sentence is the subject of the following one → aggregation merges them into a single sentence:

La [baie de Banyuls] est limitée à l'Est par l'île Grosse].
L'île Grosse] est rattachée à la côte par un terre-plein.

The [Banyuls bay] is bounded on the East by the [Grosse island] connected to the coast by a land fill.

The [Banyuls bay] is bounded on the East by the [Grosse island].
The [Grosse island] is connected to the coast by a land fill.

La [baie de Banyuls] est limitée à l'Est par l'île Grosse],
rattachée à un terre-plein.

Controlled language generation:

Aggregation and referring expressions generation, LitINAUT language 2/2

- When consecutive sentences have the same subject and the same verb → we use conjunction:

La [baie de Banyuls] est limitée au NW par le [cap d'Osne].
La [baie de Banyuls] est limitée à l'Est par l'[île Grosse].

The [Banyuls bay] is bounded on the NW by the [Osne Cape].
The [Banyuls bay] is bounded on the East by the [Grosse island].

La [baie de Banyuls] est limitée au NW par le [cap d'Osne]
et à l'Est par l'[île Grosse]

The [Banyuls bay] is bounded on the NW by the [Osne Cape] and on the East by the [Grosse island].

- When we have subject identity but with different verbs, referring expressions are generated:

La [baie de Banyuls] est divisée en deux parties par...

The [Banyuls bay] is divided into two parts by...

It is divided into two parts by...

Elle est divisée en deux parties par...

- Text may be omitted from the realization because it is obtained from the context: when we have the name of a subdivision, we can omit the part implied by the subdivision title.

Interaction with Electronic Nautical Charts

Why the interaction with ENC's?

After specifying an area on an ENC, the system generates a LitINAUT text that has the advantage of being:

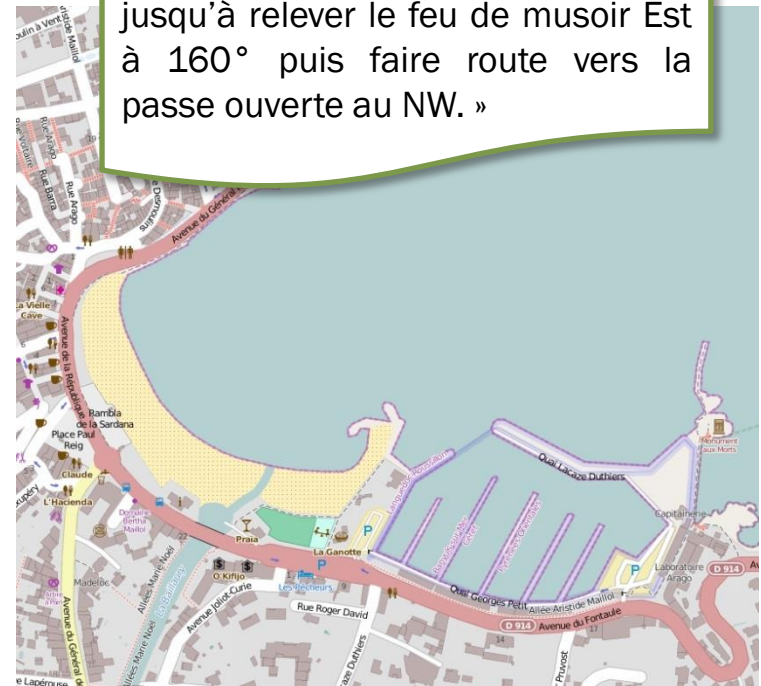
1. Limited to the zone of interest given by the user
2. Dependent on the local context
3. Up-to-date, since new information may be constantly provided by users

How?

1. The area given by the user is positioned in the set of geopolygons and in the geographical area graph.
2. We find the relevant nodes that induce a subgraph of the KB.
3. LitINAUT is generated as described above.

Take the route towards the white sector of the beacon of the Petite island (193° - 247°) until bearing the East breakwater beacon then take the route towards the open strait on the NW.

« Faire route dans le secteur blanc du feu de l'Île Petite (193° - 247°) jusqu'à relever le feu de musoir Est à 160° puis faire route vers la passe ouverte au NW. »



Collaborative updates of the knowledge base

The KB can be updated by trusted user.

- Updates are essential because of the risks involved.

The validation:

- The validation in the system operates on two levels:
 - The lexical and syntactic level: an interface analyzes segments written in INAUT and validates them
 - The semantic level:
 - Automatic validation for some segments
 - Human validation
- Final validation by human controller

Conclusion

Uses of INAUT (and its variant LitINAUT):

- Update of the SHOM maritime knowledge base
- Automatic generation of *Instructions nautiques* documents
- Interaction with ENCs

Perspectives:

- The extension of INAUT into a QA system
- The issue of dangerousity