

Extraction de connaissances pour la description de l'environnement maritime côtier à partir de textes d'aide à la navigation

Léa Lamotte (LaSTIG)

Nathalie Abadie (LaSTIG)

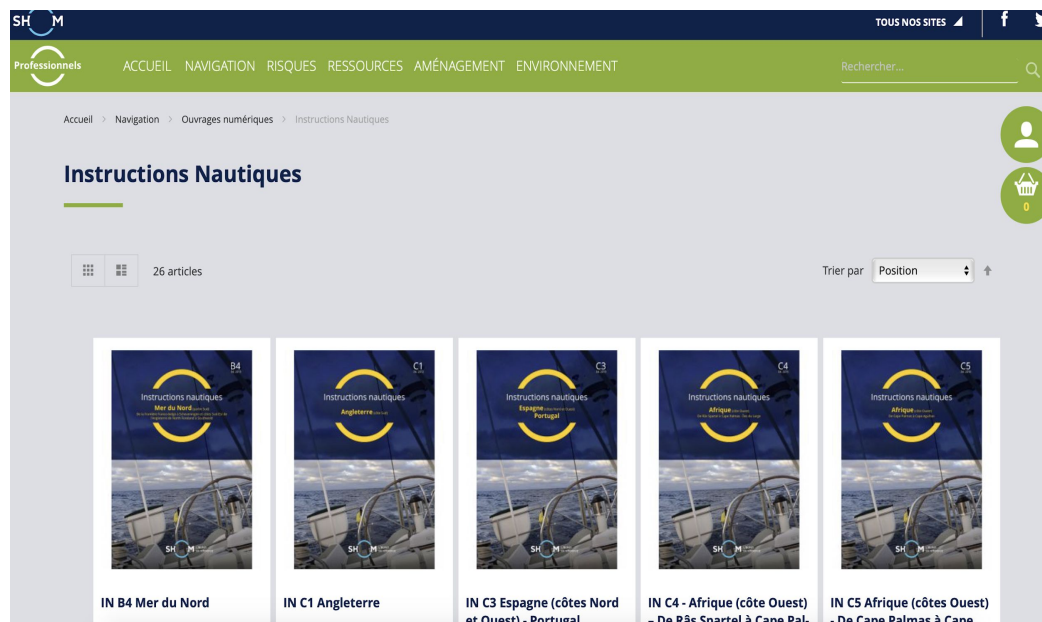
Eric Saux (IRENav)

Eric Kergosien (GERiiCO)

EGC 2020 - Bruxelles - Atelier GAST

Les Instructions Nautiques du Service Hydrographique et Océanographique de la Marine (SHOM)

- Ouvrage en 26 tomes : couverture de la zone côtière française, espagnole, africaine, etc.



SHOM TOUS NOS SITES

Professionnels ACCUEIL NAVIGATION RISQUES RESSOURCES AMÉNAGEMENT ENVIRONNEMENT Rechercher...

Accueil > Navigation > Ouvrages numériques > Instructions Nautiques

Instructions Nautiques

26 articles Trier par Position

IN B4 Mer du Nord

IN C1 Angleterre

IN C3 Espagne (côtes Nord et Ouest) - Portugal

IN C4 - Afrique (côte Ouest) - De Râs Spaertel à Cape Pal-

IN C5 Afrique (côtes Ouest) - De Cape Palmas à Cape

<https://diffusion.shom.fr/pro/navigation/ouvrages/instructions-nautiques.html>

- Notre cas d'étude : étude de la zone allant de la frontière belge à la pointe de Penmarc'h (instruction IN C22 France, côte Nord et Ouest)



IN C22 France (côte Nord et Ouest)

IN C22 France (côte Nord et Ouest) : Du cap de La Hague à la pointe de Penmarc'h Destinées à tous navigateurs, les instructions nautiques...

Les Instructions Nautiques du Service Hydrographique et Océanographique de la Marine (SHOM)

- Contiennent des connaissances sur l'environnement maritime et côtiers, complémentaires aux descriptions cartographiques
- Textes rédigés et mis à jour par différents auteurs et publiés par un établissement public → assure un français standard, une bonne diversité dans les formes présentes et une abstraction des habitudes langagières subjectives ou diachroniques.

01 4.6.2.4. Saint-Cast

- 07 L'Anse de Saint-Cast, ouverte au NE entre la Pointe de Saint-Cast et la Pointe de la Garde, est en partie ensablée et n'offre qu'un mauvais mouillage. Elle est bien protégée des vents, sauf de ceux du Nord à l'Est.
- 13 Le **Port de Saint-Cast** (48° 38' N — 2° 15' W) est adossé à la Pointe de Saint-Cast, au Sud. C'est un port départemental, de pêche et de plaisance, géré par la chambre de commerce et d'industrie des Côtes d'Armor. Il comprend un bassin à flot et une zone d'échouage.
- 19 Le bassin à flot, accessible à toute heure, est protégé à l'Est et au SE par une digue courbe dont le musoir porte un feu, et au SW par une ligne de pontons lourds marquée à son extrémité par une balise lumineuse.

25



4.6.2.4. — Saint-Cast, au NW (2007, avant transformation).

- 31 CHENALAGE. — L'approche se fait en venant du NE, en passant soit dans le NW, soit dans le SE du **plateau des Bourdinots**, marqué au NE par une bouée cardinale Est mouillée à 1 M au NE du port. De nuit, on fait route en se tenant dans l'un des deux secteurs blancs (206° – 217° ou 235° – 245°) du feu implanté à l'enracinement de la digue.
- 37 Un court chenal, à peu près parallèle à la jetée, conduit à la passe d'accès au bassin. Il est balisé par deux paires de bouées latérales lumineuses.
- 43 Le bassin à flot se compose de trois zones :
– au NE, une zone réservée à la pêche ; elle est équipée d'un ponton pouvant accueillir une vingtaine de bateaux, par profondeur de 2 m ;
– au centre, une zone dite « de grande plaisance », par profondeur de 2 m, destinée aux bateaux de 8 à 14 m de long ;
– au SW, une zone « de petite plaisance », par profondeur de 1 m, destinée aux bateaux de 5 à 8 m de long.
- 49 Le port d'échouage s'étend au SW du bassin. 19 mouillages sont réservés aux bateaux de pêche, au Sud de l'extrémité de la jetée.

Les Instructions Nautiques du Service Hydrographique et Océanographique de la Marine (SHOM)

Les IN du SHOM décrivent :

- les amers et dangers pour la navigation côtière,
- les courants,
- les ports et mouillages,
- leurs chenaux d'accès,
- leurs équipements et les services proposés aux navigateurs,
- etc.

Étant destinés à être lus en complément des cartes marines lorsque les navigateurs préparent leur itinéraire, ces IN suivent avant tout une organisation géographique : décrivent la côte linéairement, d'un port à l'autre, du Nord vers le Sud et sauf mention contraire, vue depuis la mer.

L'extrait suivant présente une description du paysage typique de ce corpus :

"[...]La Pointe de Pléneuf est débordée vers le NW par l'îlot Le Verdelet, puis jusqu'à 1,5 M au large par le Plateau des Jaunes, découvrant ; celui-ci, ainsi que les hauts-fonds rocheux qui s'étendent dans l'Ouest du Verdelet, sont couverts par le secteur rouge (146°-196°) du feu de la tourelle La Petite Muette située à l'entrée de Dahouët.[...]"

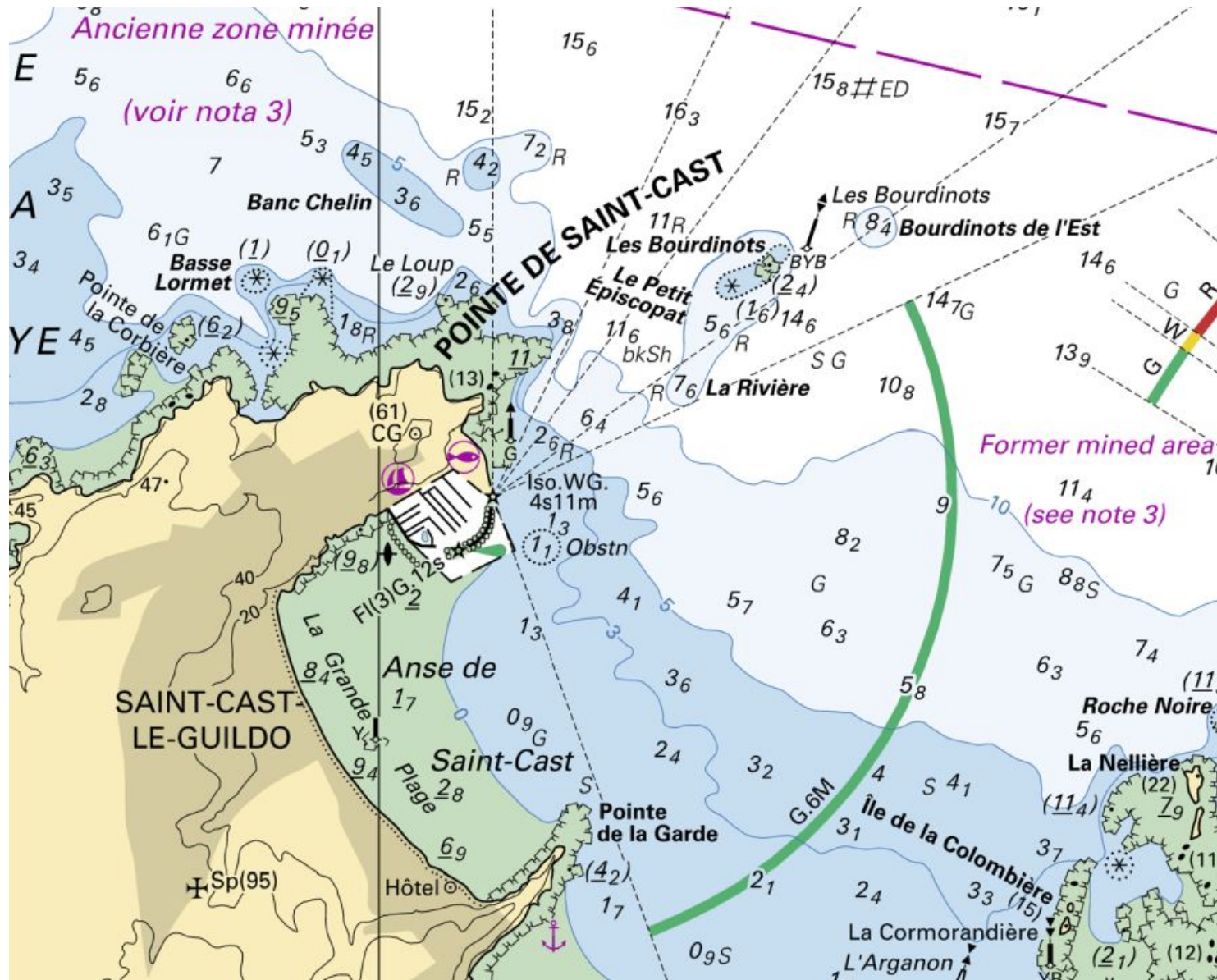
... un complément des “electronic navigational chart” (ENC)

Format standardisé (S57 - IHO) :

Object:	Buoy, cardinal
Acronym:	BOYCAR
Code:	14

- Bridge
- Building, religious
- Building, single
- Built-up area
- Buoy, cardinal
- Buoy, installation
- Buoy, isolated danger
- Buoy, lateral
- Buoy, safe water
- Buoy, special purpose/general
- Cable area
- Cable, overhead
- Cable, submarine
- Caim
- Canal
- Canal bank
- Cargo transhipment area
- Cartographic area
- Cartographic line
- Cartographic symbol
- Causeway
- Caution area
- Cemetery
- Chain/Wire
- Checkpoint
- Chimney
- Closing line
- Coastguard station
- Coastline
- Compass

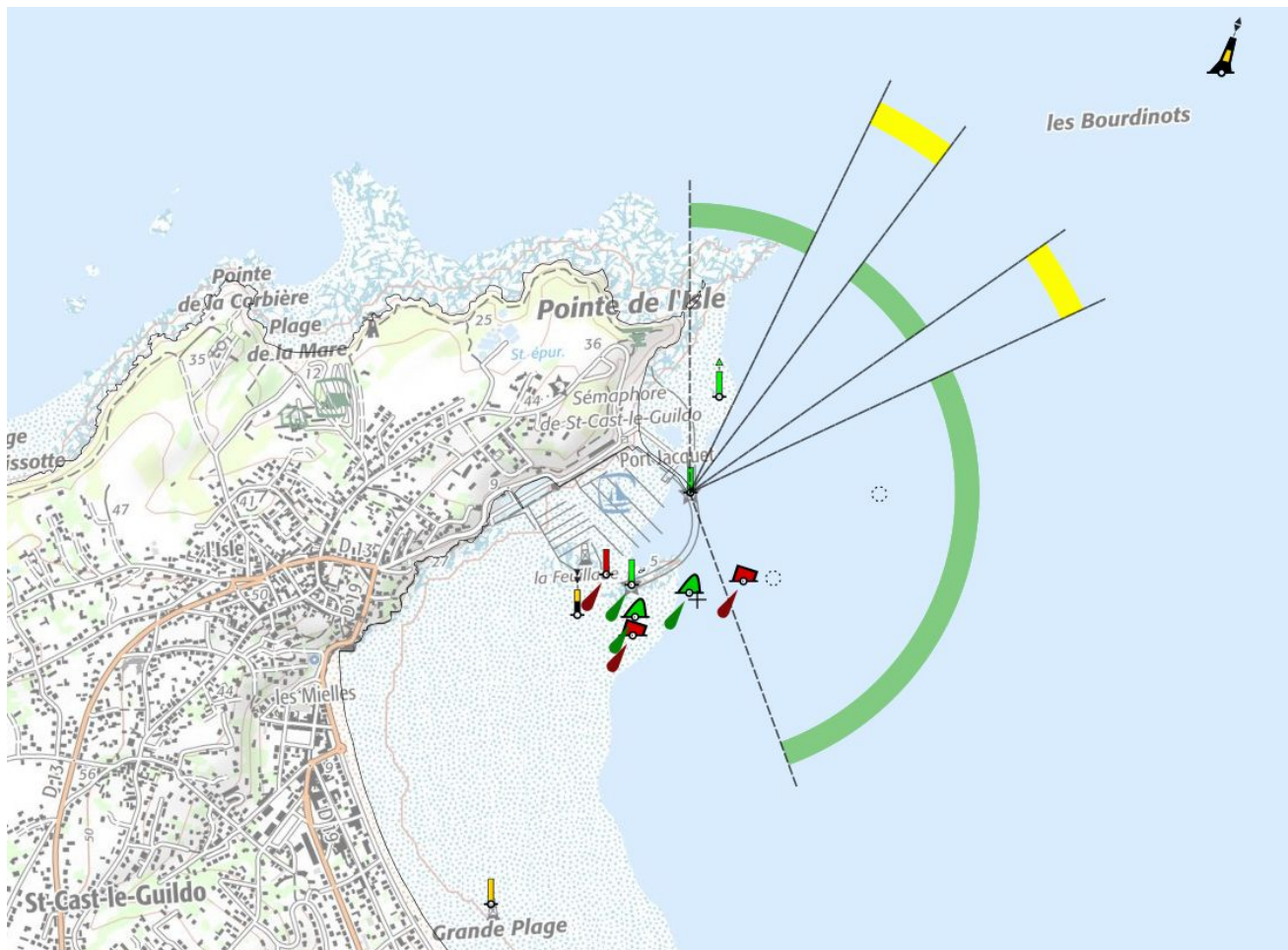
Source: <http://data.shom.fr>



Entre descriptions textuelles du terrain et bases de données géographiques vectorielles

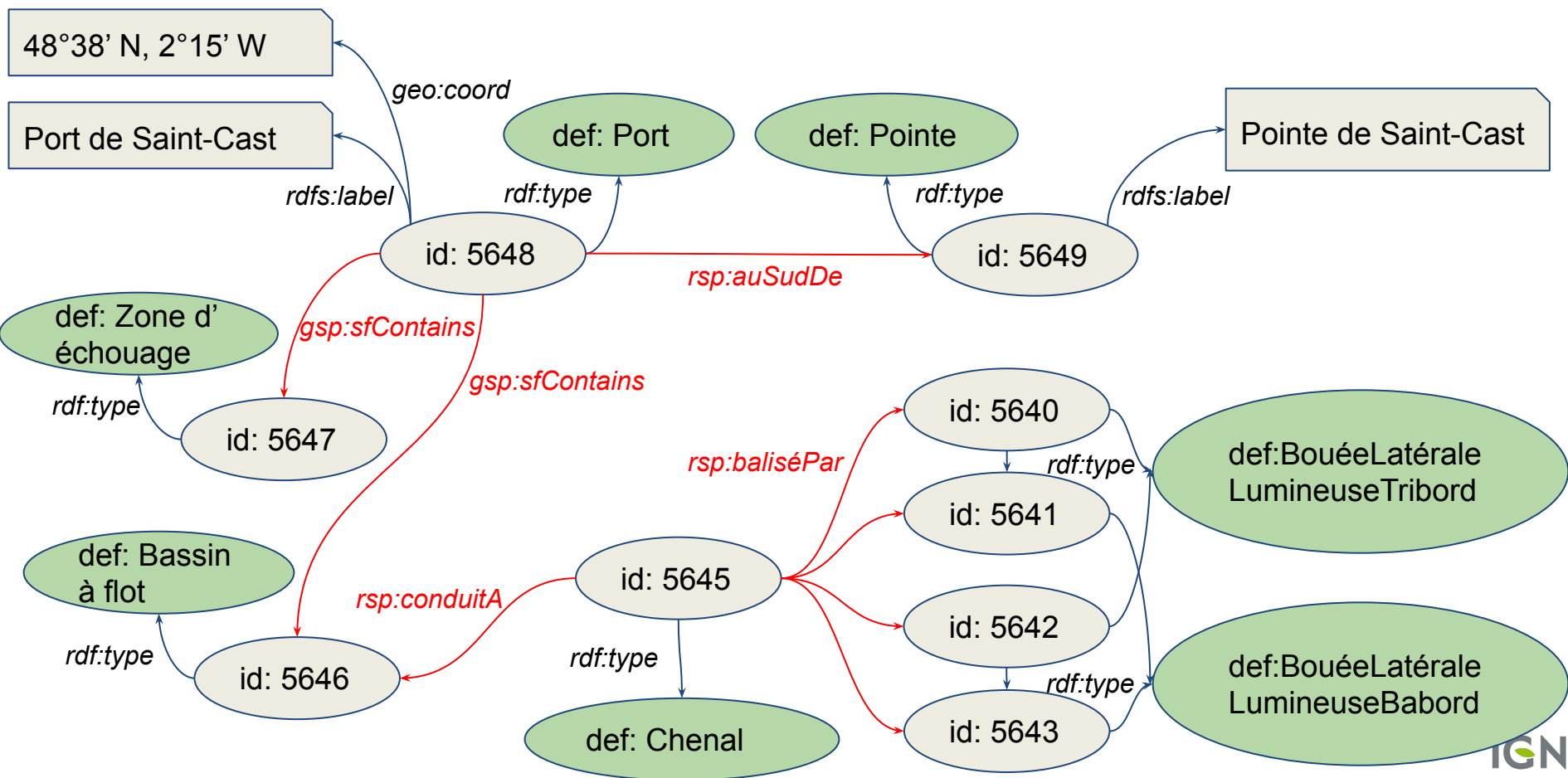
“Le Port de Saint-Cast (48°38' N, 2°15' W) est adossé à la Pointe de Saint-Cast, au Sud. [...] Il comprend un bassin à flot et une zone d'échouage. [...] Un court chenal, à peu près parallèle à la jetée, conduit à la passe d'accès au bassin. Il est balisé par deux paires de bouées latérales lumineuses.”

Source: <http://data.shom.fr>



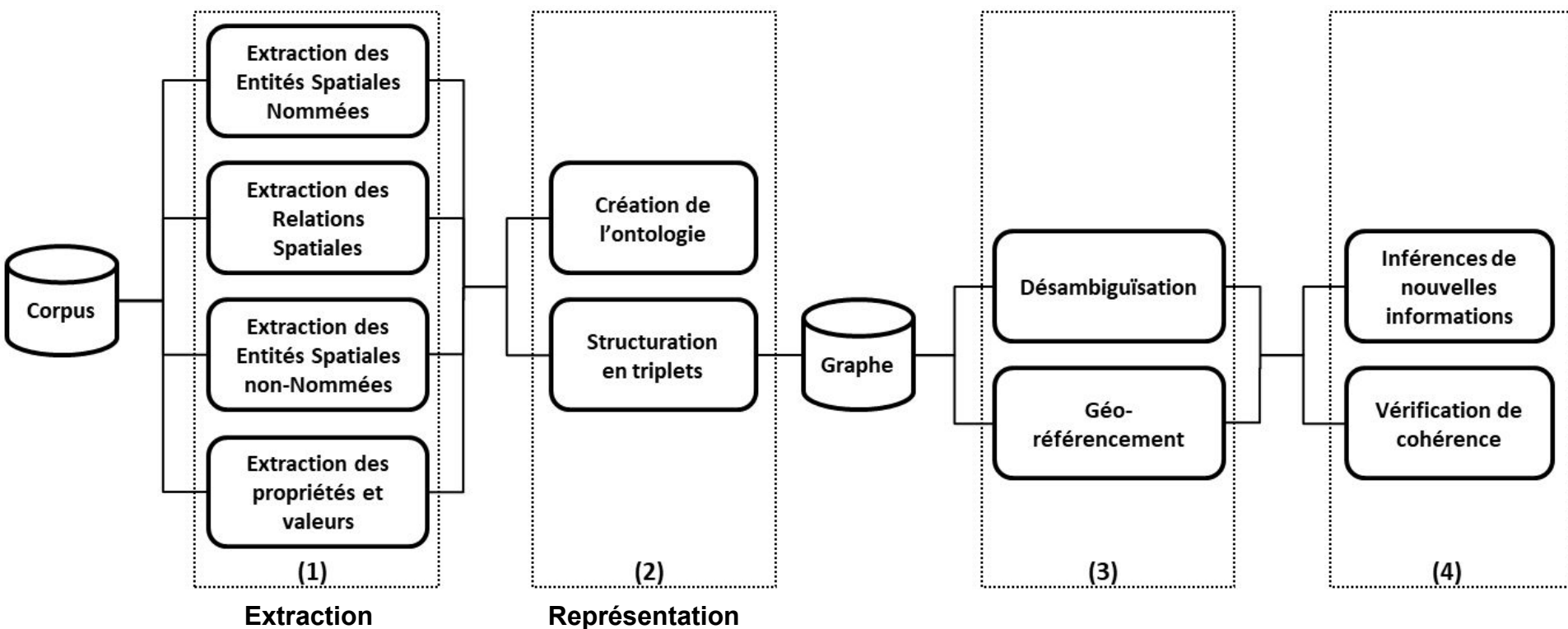
Une représentation du terrain dans un système de références spatiales indirect

- **Extraire** et **structurer** les informations à références spatiales (notamment **indirecte**) pour construire et peupler des référentiels de données géographiques :
 - pour assurer la médiation (complémentarité, redondance, incohérence) entre les données qualitatives et quantitatives



Vers une base de connaissances de l'environnement hydrographique côtier

FIG. 1 – De l'extraction d'information géographique à partir de textes à son exploitation au sein d'une base de connaissances.



Préparation du corpus

Version des IN au format XML : chapitres, sous-chapitres, paragraphes, etc.

Particularités :

- Chaque portion de côte est décrite par un même locuteur : pour se doter de corpus non biaisés par le style d'écriture d'une personne, un découpage non géographique du texte est donc préférable.
- Il faut conserver des portions de texte cohérentes pour permettre la résolution des coréférences.

Solution adoptée : Découpage au niveau des balises “paragraphes” et numérotation dans l'ordre. Répartition en 3 sous-corpus :

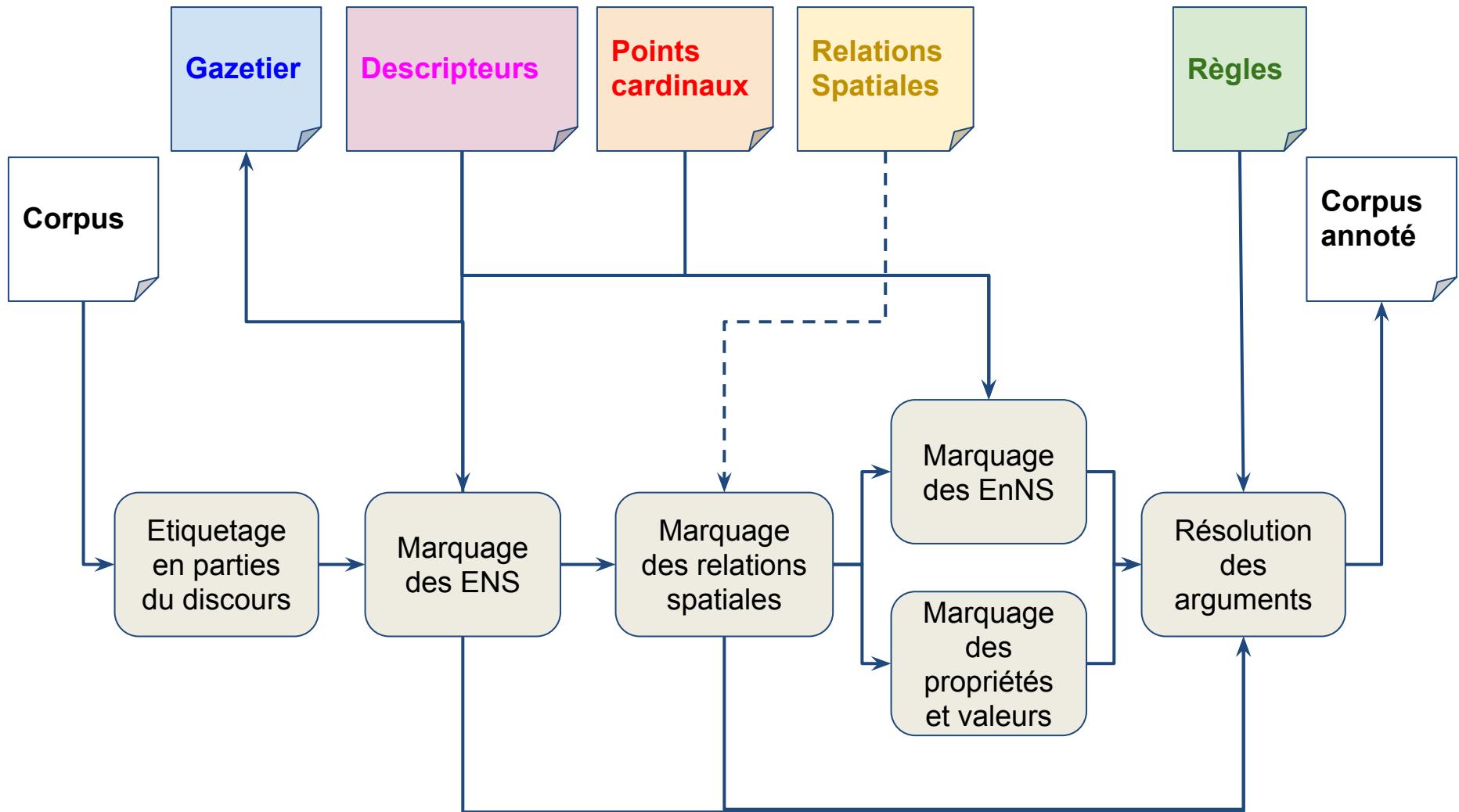
- si n multiple de 5 → corpus de test
- si $n + 1$ multiple de 5 → corpus de développement
- paragraphes restants → corpus d'entraînement

	Entraînement	Développement	Test
Paragraphes	108 (60%)	36 (20%)	36 (20%)
Types	9 893	4 547	4 516
Tokens	146 971 (66%)	34 963 (16%)	40 981 (18%)

Chaîne de traitement pour l'extraction

bases de données géographiques

vectérielles "BD NYME®" et "Toponymes"



Extraction des entités spatiales à base de patrons

- Patrons lexico-syntaxiques pour l'extraction des ENS :

(<descr><ADJ>*<DP>*)*(<maj><DP>*)*<topo>(<DP>*<maj>)*

```
<geogName>  
  <geogFeat>phare</geogFeat> de l'  
  <geogFeat>Île</geogFeat> de  
  <name>Sein</name>  
</geogName>
```

<descr><ADJ>*(<DP><descr><ADJ>*)*<DP>*<maj>(<DP>*<maj>)*

```
<geogName>  
  <geogFeat>roche</geogFeat> détachée de la  
  <geogFeat>Pointe</geogFeat> des  
  <name>Grottes</name>  
</geogName>
```

- Patron lexico-syntaxique pour l'extraction des EnNS :

<DET>*<NOM>*<Car>*<DP>*<ADJ>*<descr><Mod>*(<DET><WORD>)*

```
<ESnN> angle Sud-Est de la baie </ESnN>
```

Extraction des entités spatiales : résultats

- Implémentation sous UniTEX

Corpus de développement Corpus de test	Précision	Rappel	F-Mesure
ENS : CasEN	0.16	0.19	0.17
ENS : Perdido	0.73	0.16	0.26
ENS : Notre approche	0.84	0.61	0.7
ENS : Notre approche	0.79	0.72	0.75
EnNS : Notre approche	0.32	0.57	0.41

- CasEN détecte comme ENS :
 - les points cardinaux
 - les articles définis au pluriel en début de phrase
- Perdido et CasEN : lexiques généralistes, peu adaptés au domaine maritime côtier.

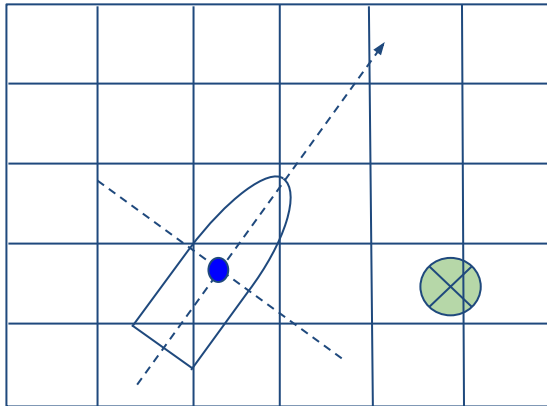
Extraction des relations spatiales

- Un corpus (IN C22 France, côte Nord et Ouest)
 - très riche en relations spatiales : 1618 relations spatiales sur le corpus de test
 - avec une grande variété de formulations!
Ex. pour les relations de distance qualitative :
au large de, au voisinage de, autour de, aux abords de, aux environs du, dominant, déborder, en bordure de, en face de, enraciné au, environnant, est bordé de, le long de, le plus au large, nettement détaché de, plus au large, proche de, près de, border, émerger de, se détacher nettement de, voisins l'un de l'autre, à l'écart de, à l'entrée de, à l'extérieur de, à la limite de, à la lisière de, à proximité de, à toucher, etc.
- Mise en place de patrons lexico-syntaxiques pour rechercher des expressions exprimant des relations de :
 - distances qualitative et quantitative
 - direction
 - adjacence
 - traversée
 - inclusion (“contient” ou “à l’intérieur de”)
 - “entre”

Référencement indirect d'entités spatiales

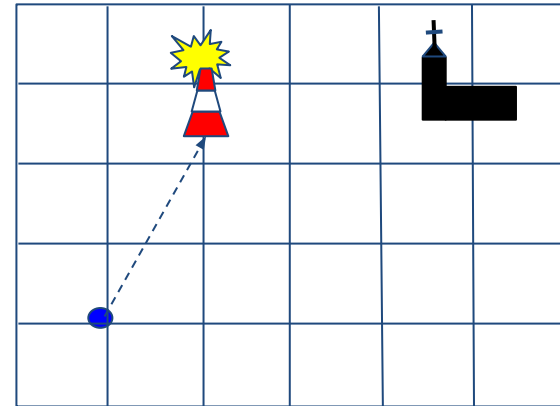
- Différents référentiels spatiaux existants (relatif, absolu, intrinsèque)

En référentiel intrinsèque



“Laisser la bouée à tribord”

En référentiel relatif



“Suivre la direction du clocher à droite du phare”

- Définition d'un 5-tuple $\langle \text{Locatum}, \text{Relatum}, \text{Relation}, \text{Origine}, \text{Orientation} \rangle$ afin de caractériser une relation spatiale entre deux objets

	locatum	relatum	relation	origine	orientation
relatif	✓	✓	✓	✓	✓
absolu	✓	✓	✓		
intrinsèque	✓	✓	✓		✓

Extraction des relations spatiales : résultats

	Précision	Rappel	F-mesure
Distance	0.78	0.76	0.77
Direction	0.82	0.93	0.87
Adjacence	0.67	0.39	0.49
Traverse	1.0	0.6	0.75
Contient	0.72	0.71	0.72
A l'intérieur de	0.56	0.42	0.48
Entre	0.57	0.76	0.65
Relations : marquage	0.78	0.74	0.76
Relations : classification	0.74	0.71	0.72

Résolution des arguments

But : identifier, pour chaque relation, son relatum et son locatum.

Approche : représentation simplifiée du texte : ponctuations fortes, conjonctions de coordination, entités et relations spatiales, dans l'ordre.

Exemple :

“Au Nord du clocher de Ploubazlanec se trouve un château d'eau remarquable”

[“Au Nord de”, “clocher de Ploubazlanec”, “château d'eau remarquable”]

locatum = "château d'eau remarquable",

relation = "Au Nord de ",

relatum = "clocher de Ploubazlanec"

Résolution des arguments : résultats

	Précision	Rappel	F-mesure
Triplet (marquage manuel)	0.69	0.63	0.66
Triplet (marquage automatique)	0.25	0.16	0.2
Locatum	0.35	0.31	0.33
Relatum	0.62	0.44	0.48

Conclusion

- Traitement des problématiques de l'extraction :
 - d'entités spatiales nommées (ESN)
 - d'entités spatiales non-nommées (ESnN)
 - de relations spatiales (RS) et des éléments nécessaires à leur interprétation hors-contexte
- Particularité du corpus maritime côtier moins favorables aux approches Perdido et CasEN
- Résultats encourageants fondés sur
 - des patrons prenant en compte la casse
 - le contexte lexical
 - des lexiques ("BD NYME®" de l'IGN et "Toponymes" du SHOM)

Perspectives :

- Désambiguïsation lexicale
- Détection et correction d'éventuelles incohérences spatiales dans les informations extraites, inférence de relations entre ES
- Exploitation des relations spatiales pour le géoréférencement

Merci!