



Rapport de stage

Institut de Recherche de l'École Navale (IRENav EA 3634), Lanvéoc

Fabien Le Chevanton, *Master 1 ILIADE 2018-2019*



Je tiens tout d'abord à remercier Eric Saux pour son encadrement sans faille, sa compétence et sympathie. Merci également à Wissame Laddada et Ludovic Moncla pour leur aide dans le bon déroulement de ce stage.



TABLE DES MATIERES

TABLE DES MATIERES	3
INTRODUCTION	5
2. CADRE DU STAGE : L'IRENAV	6
Description de la structure d'accueil	6
Des domaines scientifiques variés	6
Structuration de l'IRENav	6
L'équipe M2EN	7
L'équipe MoTIM	7
Encadrement et moyens techniques mis à disposition	7
3. Description globale de l'approche	8
4. Généralités sur les méthodes de Traitement Automatique du Langage	9
Préparation du Corpus	9
Mise en place de l'environnement informatique	11
Le traitement lexico-syntaxique	11
Traitement lexico-syntaxique sur un corpus de navigation maritime côtière	12
Le traitement sémantique en cascade par Unitex	15
Processus de traitement en cascade	15
Traitement en cascade dans le contexte de la navigation maritime côtière	17
Structuration des connaissances	17
Cascades pour l'extraction des motifs de navigation	19
5. Peuplement de l'ontologie de navigation	22
Qu'est-ce qu'une ontologie	22
Structure de l'ontologie de navigation maritime	24
Association des structures de l'ontologie et du fichier XML	25
Le concept "NavigationMark"	25
Le concept "Light"	27
Le concept "LeadingLine"	27
Instanciation de l'ontologie	28
Conclusion	29
BIBLIOGRAPHIE	30

Annexe A : Planification du stage	31
Mise en place (Deadline 17/05/19)	31
B. Traitement Automatique du Langage (Deadline 29/05/19)	31
C. Peuplement de l'ontologie (Deadline 31/08/19)	32
Annexe B : Le Traitement Automatique du Langage par PERDIDO	33
Annexe C : Correction des annotations et détermination des graphes	35
Annotation manuelle d'un texte	35



1. INTRODUCTION

Depuis quelques années, on note une émergence des systèmes dits “intelligents” fondés sur la sémantique pour des applications d’aide à la décision. Ces applications sont utilisées dans de nombreux domaines. Dans le cadre du stage, nous nous intéressons aux systèmes de navigation maritime côtière dont une des aides est de proposer aux navigateurs des routes maritimes sécurisées qu’ils peuvent emprunter. Parmi les différents modèles offrant de telles capacités, les modèles ontologiques sont prometteurs. En effet, en définissant des concepts et leurs relations à partir d’expressions logiques (formalisant le sens des concepts), l’utilisation de raisonneurs permet à partir de connaissances initiales stockées dans le modèle d’en déduire des connaissances nouvelles. Ainsi, il est tout d’abord nécessaire de recueillir ces connaissances initiales pour instancier le modèle (ou alimenter le système) soit à partir d’experts, soit à partir de documents textuels. Dans le cadre du stage, un modèle ontologique doit inférer des routes de navigation maritime à partir d’instances extraites des ouvrages nautiques que sont les pilotes côtiers. Ceci permettra à un navigateur de trouver plus rapidement une trajectoire et s’orienter dans l’environnement maritime côtier notamment lorsque ce dernier lui est non-familier.

D’un point de vue applicatif et plus précisément, le système doit permettre aux civils et militaires d’avoir accès, en saisissant deux points (de départ et d’arrivée), à un itinéraire maritime dont chacun des segments ou routes repose sur un relèvement ou un alignement d’amers. Un amer est un point de repère remarquable pour la navigation (phare, maison, clocher, rocher, etc.). D’un point de vue scientifique et technique, le stage vise tout d’abord à utiliser des outils de traitement automatique des langues (TAL) afin d’extraire les entités nommées dans les textes, leur positionnement géographique ainsi que les différentes relations spatiales (relations topologiques, de distance et d’orientation) les reliant. La seconde partie du stage consiste ensuite, à partir des connaissances extraites, à instancier le modèle ontologique afin que les raisonneurs puissent inférer les routes de navigation.

Dans une première partie, nous commencerons par une présentation du cadre dans lequel s’est déroulé le stage et des moyens techniques mis à disposition par l’IRENav. La deuxième partie sera consacrée au déroulement du processus d’analyse et de traitement des connaissances. La partie suivante traitera du peuplement de l’ontologie et de l’instanciation des concepts. Pour finir, nous concluerons ce rapport en rappelant les apports effectués et proposerons des extensions à notre travail.

2. CADRE DU STAGE : L'IRENAV

Description de la structure d'accueil



Figure 2.1 : Logo de l'IRENav

L'Institut de Recherche de l'École navale (IRENav) créé en 2000 est un centre de recherche qui touche au domaine maritime à travers

plusieurs disciplines sous la tutelle de l'École Navale et de l'École Nationale Supérieure d'Arts et Métiers. Il est localisé à l'École navale de Lanvéoc-Poulmic : dans un lieu militaire et accessible par bateau de Brest traversant la rade plusieurs fois par jour. Sa localisation en fait un lieu régi par un règlement strict.

Des domaines scientifiques variés

Les objectifs des activités de l'IRENav sont variés, mais toujours liés à l'environnement maritime :

- Proposer des enseignements scientifiques aux élèves ingénieurs de l'École navale (Hydrodynamique navale, signal et acoustique sous-marine, systèmes d'information géographique).
- Faire des activités de recherches afin d'avoir des retombées militaires ou civiles et imaginer les navires du futur.

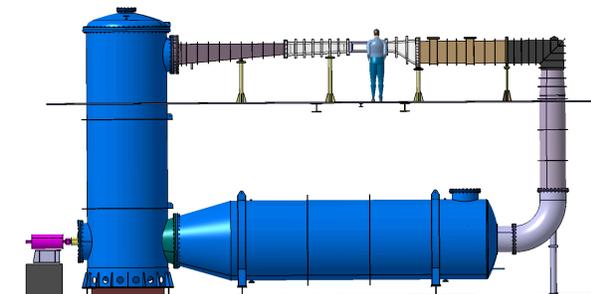
Les activités de recherche relèvent de la mécanique et de l'énergie sous l'angle de l'hydrodynamique et de l'électrotechnique, de l'observation du milieu maritime à l'aide de techniques d'acoustique sous-marine (ASM) et des systèmes d'information géographique (SIG).

Structuration de l'IRENav

L'institut est constitué d'une soixantaine de personnes dont une trentaine d'enseignant·e·s-chercheur·e·s, une quinzaine de doctorant·e·s ainsi que le personnel de soutien technique et administratif pour la recherche et la formation scientifique. La plupart de mes encadrants et collègues partagent leur temps entre la recherche et l'enseignement à l'école navale ou dans d'autres établissements. L'IRENav est ainsi composé de deux équipes :

L'équipe M2EN

Figure 2.2 : Le tunnel de cavitation



M2EN «Mécanique et Énergie en Environnement Naval», sous les aspects de l'Hydrodynamique (Interaction Fluide Structure et Ecoulements diphasiques) et de la conversion d'énergie sont destinés à la propulsion navale et aux énergies marines renouvelables. Pour cela, le laboratoire dispose notamment d'un tunnel de cavitation. Ce

matériel permet de mesurer l'hydrodynamisme de pièces (par exemple, pales d'hélice) et d'avoir des données expérimentales précises.

L'équipe MoTIM

MoTIM «Modélisation et Traitement de l'Information Maritime», sous les aspects de l'acoustique sous-marine et des sciences de l'information géographique dédiés à la définition d'outils et de systèmes favorisant l'aide à la décision et à leur mise en œuvre dans des systèmes opérationnels dédiés à la surveillance et/ou le contrôle. C'est dans cette équipe que j'effectue mon stage.

Encadrement et moyens techniques mis à disposition



Pour ce stage, je suis encadré par Eric SAUX (enseignant chercheur en informatique) et Wissame Laddada (docteur en informatique) avec Ludovic Moncla comme support, car l'approche déployée lors de mon stage s'inspire de ses travaux de thèse menés à l'Université de Pau en cotutelle avec l'Universidad de Zaragoza (Espagne) et soutenus en 2015. Au niveau technique, je dispose d'un bureau, utilise ma machine personnelle ainsi que le réseau WIFI de l'école navale.

Figure 2.3 : Le bureau mis à ma disposition à l'IRENav

3. Description globale de l'approche

Dans le cadre du stage, nous nous limiterons à l'exploitation des connaissances contenues dans les ouvrages des "Pilotes côtiers" en utilisant des méthodes de Traitement Automatique du Langage (TAL). Ces dernières nous permettent d'extraire les entités nommées spatiales (ENS) et les relations spatiales entre celles-ci. L'enjeu sera ensuite de les structurer dans un fichier XML, puis de réaliser un programme informatique permettant de peupler/instancier un modèle ontologique à partir des connaissances contenues dans le XML. Enfin, un raisonnement logique infèrera différentes trajectoires de navigation grâce aux concepts instanciés. Le processus ainsi décrit peut se résumer selon le schéma de la figure 3.1.

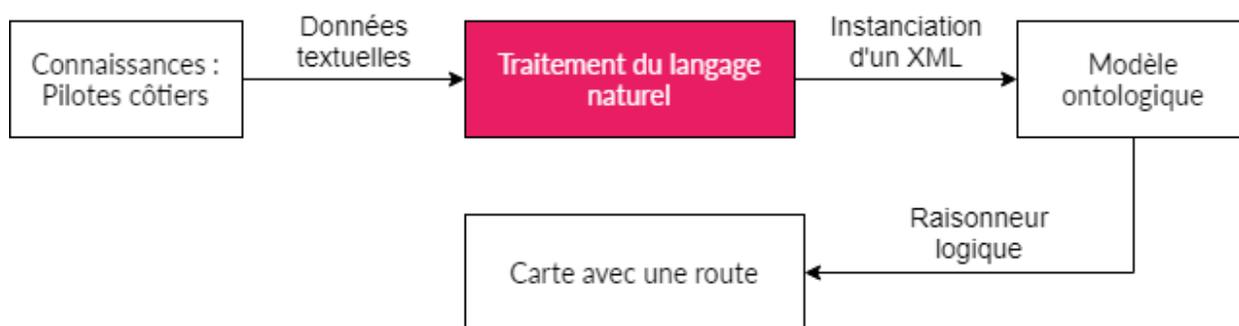


Figure 3.1 : Chaîne de traitement des connaissances

Ainsi, nous pouvons décomposer la chaîne de traitement en plusieurs étapes. Ces étapes sont à l'origine de la structure de ce rapport :

1. Former le corpus à partir des guides côtiers ;
2. Traiter automatiquement ces textes et annoter les relations ainsi que les entités spatiales nommées (ENS) associées à la navigation maritime côtière. Pour cela, nous utiliserons une chaîne de traitement, développée par Ludovic Moncla lors de sa thèse [1], permettant d'annoter les textes. Le but de cette chaîne de traitement linguistique est d'identifier des "motifs lexicaux" présents dans les textes puis de les structurer dans un fichier XML. Cependant les traitements proposés ne sont pas suffisants pour extraire les connaissances

nécessaires au modèle ontologique, ce qui nécessite de notre part des adaptations, c'est-à-dire la création de nouveaux "motifs lexicaux";

3. Peupler le modèle ontologique créé par Wissame Laddada dans le cadre de sa thèse [2] en créant un programme informatique instanciant les concepts de l'ontologie à partir de fichier XML précédemment généré.

4. Généralités sur les méthodes de Traitement Automatique du Langage

Le Traitement Automatique du Langage est un domaine scientifique qui implique la linguistique, l'informatique et l'intelligence artificielle. Son but est de créer des outils de traitement de la langue naturelle pour diverses applications.

Dans notre contexte, la chaîne de traitement linguistique est composée d'une étape de préparation du corpus initial, d'une étape d'analyse linguistique fondée sur différents outils de TAL, notamment Perdido et Unitez et d'une méthode de synthèse permettant de générer un fichier XML, là encore à partir de l'outil Unitez.

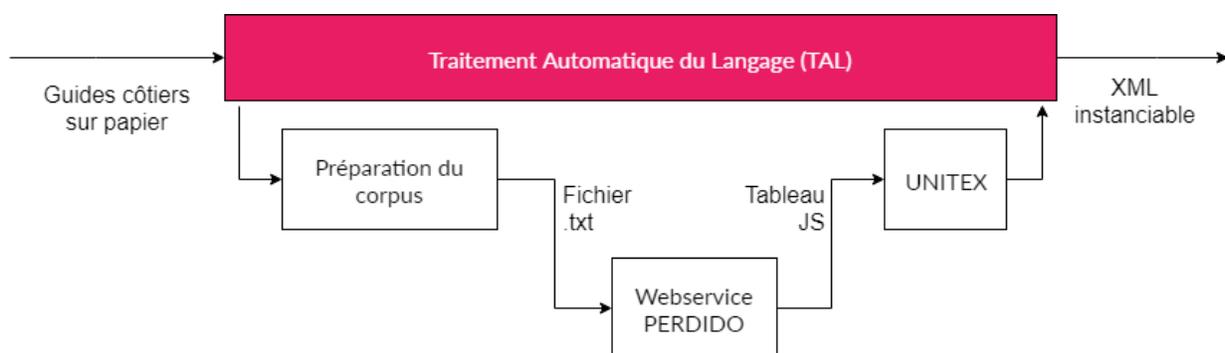
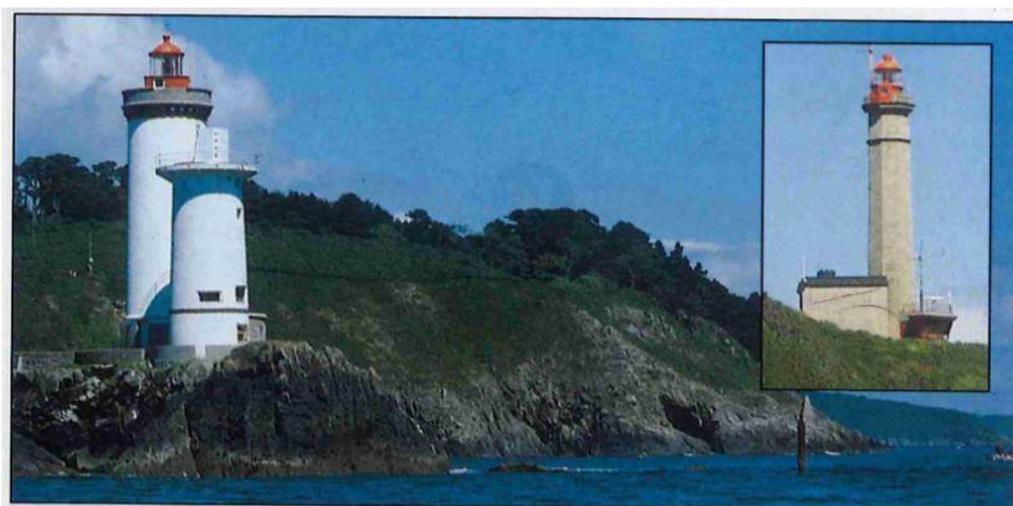


Figure 4.1 : Chaîne de traitement linguistique

Préparation du Corpus

Dans les “Pilotes côtiers”, les informations sur les trajectoires de la navigation maritime côtière sont répertoriées par des marins en intégrant les dimensions spatiales et temporelles. En effet, les dangers tels que les rochers à éviter y sont évoqués et positionnés géographiquement, et les itinéraires sont décrits pour une navigation de jour ou de nuit (avec les feux). De plus, certaines descriptions prennent en compte les conditions météorologiques qui réduisent la visibilité comme la brume ou encore les courants de marée qui influencent la navigation.



En venant du nord, du chenal du Four, on arrondira largement la pointe **St-Mathieu** qui porte la haute tour rouge et blanche du phare accolé à une abbaye en ruine, en venant virer par le sud la tourelle rouge des **Vieux Moines**. En se dirigeant vers l'Est sur la bouée bâbord **Coq Iroise**, on pare la roche isolée affleurante qui pointe à environ 0,2 mille au sud de l'éperon rocheux des **Respects** à 0,4 mille dans l'Est des Vieux Moines. De la bouée **le Coq**, une même route conduit sur la bouée **Charles Martel** en lisière du grand chenal. On notera que l'alignement à **275°** derrière soi des deux **Bozmen** pare ces dangers et conduit également sur le grand chenal. Les bouées **Charles Martel** et de la basse **Beuzec** peuvent

Figure 4.2 : Exemple de description textuelle issue de la numérisation des “Pilotes côtiers”



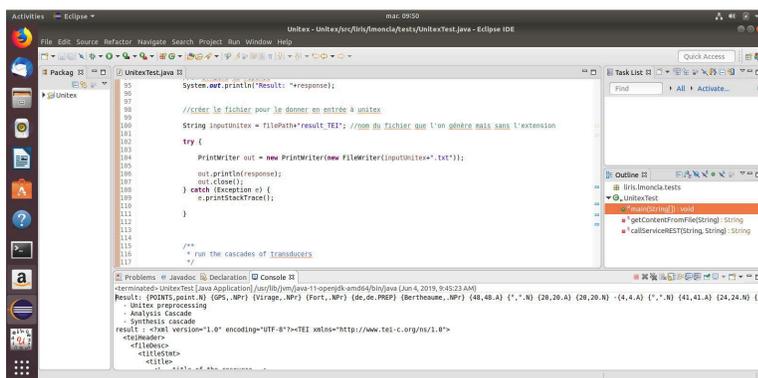
Figure 4.3 : Matière première pour peupler l'ontologie

Puisque ces ouvrages n'existent pas en version numérique, la première étape de création du corpus a été réalisée par Eric Saux. Il s'agissait de scanner les documents des guides des “Pilotes côtiers” et d'en extraire le texte avec des outils spécialisés de conversion disponibles en ligne. Ce processus nous a permis de composer notre corpus, avec quelques

erreurs liées à la reconnaissance textuelle à partir d'images. Ces erreurs, liées la plupart du temps à la non reconnaissance de certains caractères, ont été corrigées manuellement après relecture des textes originels.

Les données étant maintenant extraites en format texte (.txt) et accessibles par tout programme informatique, l'étape suivante est dédiée au traitement et à l'analyse.

Mise en place de l'environnement informatique



Les développements se faisant essentiellement en langage Java, j'avais initialement choisi de me munir d'un environnement Windows 10 avec l'IDE Eclipse. Cependant, suite à des problèmes de compatibilité avec UNITEX, j'ai rapidement mis en place une machine virtuelle avec un système

d'exploitation Ubuntu 18.04.

Figure 4.4 : Machine virtuelle sous Linux

Le traitement lexico-syntaxique

L'analyse lexicale consiste à segmenter un texte en entrée sous forme d'une chaîne de caractères en tokens (liste de symboles). Cette analyse est couplée à un traitement syntaxique qui met en évidence la structure du texte. Il en ressort un arbre syntaxique comme le montre la Figure 4.5.

{POINTS,point.N} {GPS,.Npr} {Virage,.Npr} {Fort,.Npr} {de,de.PREP} {Bertheaume,.Npr}

Figure 4.5 : Résultat d'un traitement lexico-syntaxique

La lemmatisation consiste à regrouper les mots d'une même famille dans un texte, afin de les réduire à leur forme canonique (le lemme), comme belle, beaux, belles. Certaines conjugaisons peuvent rendre cette tâche complexe pour le programme, comme retrouver la forme canonique «avoir» depuis «eussions eu». En revanche, « des avions » et « nous avions » n'ont pas le même lemme.

Traitement lexico-syntaxique avec PERDIDO

Nous avons utilisé le service web PERDIDO développé par Ludovic Moncla dans le cadre de ses travaux de thèse [1]. L'architecture logicielle est présentée en annexe B. Initialement conçu pour retracer un itinéraire de randonnée décrit sur un corpus textuel, nous l'avons choisi pour sa capacité à annoter les entités nommées ou spatiales, les relations spatiales ainsi que les expressions de déplacement ou de perception au niveau terrestre. Cet outil intègre également l'utilisation de l'analyseur syntaxique TALISMANE (Traitement Automatique des Langues par Inférence Statistique Moyennant l'Annotation de Nombreux Exemples)¹. La Figure 4.6 présente une démonstration de l'objectif du service web : le corpus résultant est annoté selon un code couleur (Figure 4.6 a) avec une modélisation graphique via Google Maps de l'itinéraire qu'il décrit (Figure 4.6 b).

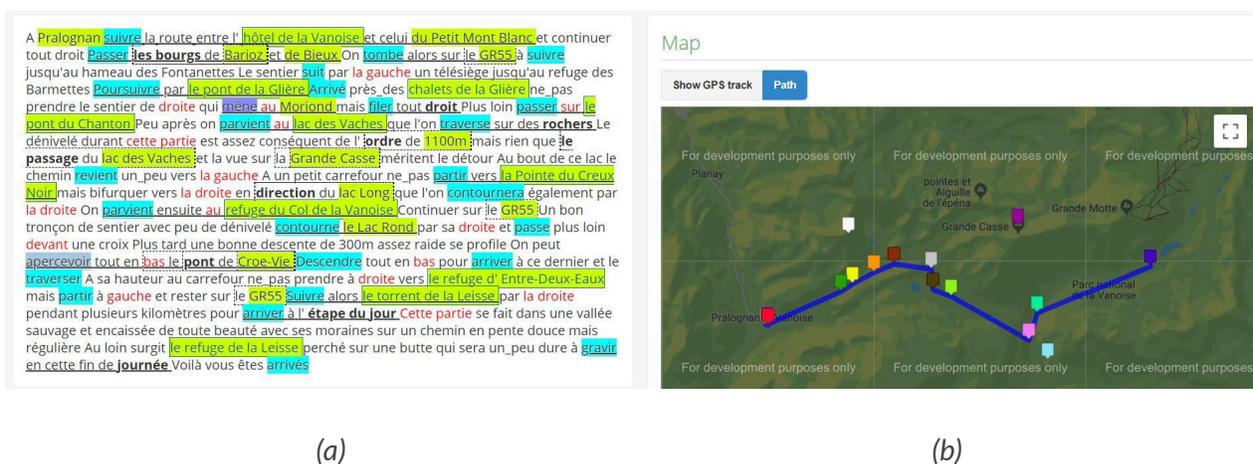


Figure 4.6 : Passage d'un document textuel à une carte présentant un itinéraire avec l'outil PERDIDO.

La principale difficulté est d'utiliser le service pour une description de l'environnement maritime alors qu'il est conçu pour le milieu terrestre.

Traitement lexico-syntaxique sur un corpus de navigation maritime côtière

Afin de vérifier que ce web service est performant avec des données maritimes, nous avons annoté manuellement un échantillon de corpus. (cf. Annexe C). Une fois cette tâche effectuée, nous avons

¹ <http://redac.univ-tlse2.fr/applications/talismane.html>

testé ces mêmes corpus avec le web service PERDIDO comme le montre la figure 4.7 et comparé les annotations.

Entre la sortie sud du chenal du Four et le Raz de Sein soit 18 milles la mer d'Iroise est largement ouverte vers l'ouest et ne présente aucun fond non balisé Elle se prolonge par le goulet de Brest au NE et la baie de Douarnenez au SE séparés par l'avancée de la pointe de Camaret que débordent quelques îlots et roches affleurantes jusqu'à la tourelle de la Parquette La Parquette une haute tourelle à bandes obliques noires et blanches est isolée à 3 milles à l'ouest de la balise du Petit Leac h plantée sur une roche découvrant de 90 m Le Corbeau m est la seule roche à craindre à 070 mille avant le Petit Léac h Elle qui signale au nord une roche de m La mer brise vite sur ces hauts fonds par vent frais d'ouest mais l'entrée du goulet de Brest n'en reste pas moins aisément praticable par un passage large de 3 milles Il suffit simplement de reconnaître les différentes bouées qui balisent la route des grands navires qui dans le goulet se divise en deux routes au nord et au sud du plateau des Fillettes

Figure 4.7 : Exemple visuel de prétraitement par PERDIDO

Après plusieurs essais, si PERDIDO s'est révélé assez fiable pour annoter les entités spatiales et nommées du corpus que nous avons préparé, il subsiste quelques erreurs au niveau des relations spatiales. Par exemple, dans la figure 4.8, le terme "haut" ne sert pas à donner une indication sur l'orientation de l'objet. En effet, un haut fond est un relief sous-marin du plateau continental où la profondeur de l'eau est faible par rapport à celles des points avoisinants. Il est donc nécessaire de vérifier la précision du traitement pour s'assurer que l'annotation est fidèle aux résultats attendus. Nous aurions pu faire une étude plus approfondie de fiabilité de PERDIDO (par exemple, entités détectées vs. entités attendues) mais cela n'a pas été un objectif du stage et aurait nécessité un corpus plus important.

```
<w type="CONJC" lemma="et">et</w>
<w type="ADV" lemma="ne">ne</w>
<w type="V" lemma="présenter">présente</w>
<w type="DET" lemma="aucun">aucun</w>
<w type="A" lemma="haut">haut</w>
<term type="offset" subtype="orientation">
  <w type="N" lemma="fond">fond</w>
</term>
```

Figure 4.8 : Erreurs d'annotation

En sortie, PERDIDO génère un fichier XML TEI (Text Encoding Initiative). La figure 4.9 illustre un exemple de résultat retourné (4.9 b) par le service web suite au traitement d'un texte brut (4.9 a) via une API dédiée à PERDIDO.



Ce fichier XML est composé de plusieurs types de balises :

- Les balises <w> contenant le token avec sa forme canonique ainsi que son type grammatical.

- Les balises <rs>, <name> et <term> qui vont introduire des relations entre les tokens (orientation, noms composés de plusieurs mots, inclusion...)

Une fois le prétraitement effectué par PERDIDO, la partie qui suit se concentre davantage sur la signification des termes afin d'annoter au mieux les textes pour peupler l'ontologie.

Le traitement sémantique en cascade par Unitex

Le processus de traitement sémantique est divisé en deux étapes :

1. La première étape, dite d'analyse, consiste à extraire les connaissances tout en désambiguïsant les unités lexicales, c'est-à-dire déterminer le sens d'un mot dans une phrase lorsque qu'il peut avoir plusieurs significations possibles et selon le contexte général. Par exemple, un avocat peut être un fruit ou une profession. Dans notre contexte, les connaissances à identifier auront un lien avec la navigation maritime côtière et seront associées aux concepts du modèle ontologique.
2. La seconde étape, dite de synthèse, structure les données extraites au format XML. Dans notre contexte, les balises XML correspondent aux informations modélisées dans l'ontologie.

Processus de traitement en cascade

Unitex est un logiciel permettant notamment d'exécuter un ensemble de cascades sur un texte. Celles-ci contiennent un ou plusieurs automates (appelés Graphs) qui s'exécutent successivement sur un corpus en entrée (figures 4.10 et 4.11). La matière en entrée des cascades Unitex sera pour nous le retour du service web PERDIDO. Un exemple de graphe initialement présent dans Unitex est présenté dans la figure 4.12 et permet de mieux en comprendre la structure. La symbologie utilisé lors de la création de ces graphes est détaillée dans le manuel utilisateur d'Unitex [6].

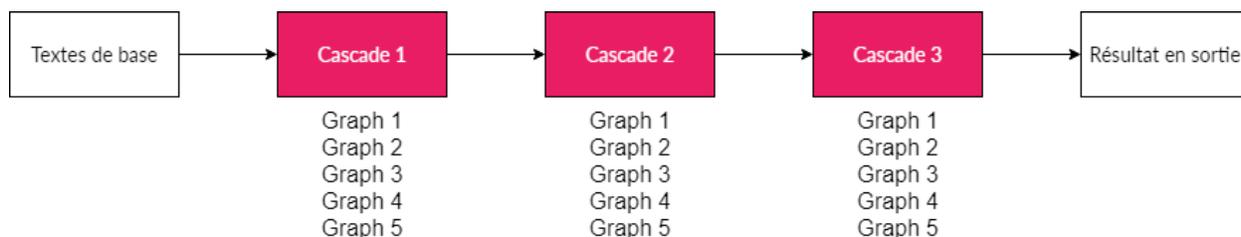


Figure 4.10 : Fonctionnement de Unitex en cascades et graphes

#	Disabled	Name	Merge
1	<input type="checkbox"/>	grfAddCodeW.fst2	<input type="checkbox"/>
2	<input type="checkbox"/>	grfAddCodeW2.fst2	<input type="checkbox"/>
	<input checked="" type="checkbox"/>	grfBalisageVerb.fst2	<input type="checkbox"/>
3	<input type="checkbox"/>	grfBalisage.fst2	<input type="checkbox"/>
	<input checked="" type="checkbox"/>	grfPc.fst2	<input type="checkbox"/>

Figure 4.11 : Exemple d'une liste de graphes dans une cascade

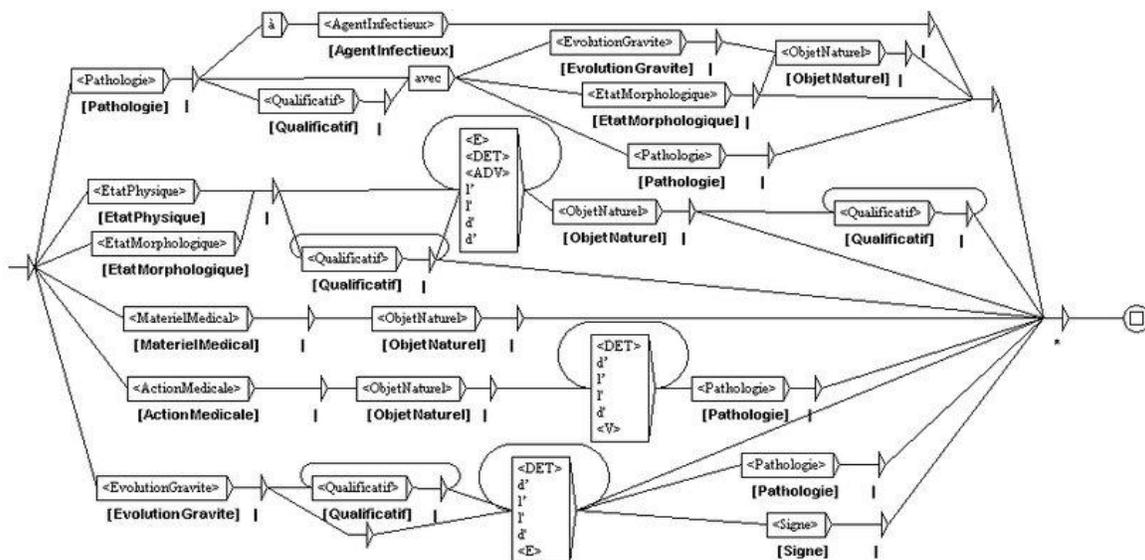


Figure 4.12 : Exemple d'un graphe Unitex à propos des pathologies

Ces graphes de traitement font également appel à des lexiques des termes qui nous intéressent et définissent des motifs à annoter dans le texte. Le système d'annotation se décompose de deux cascades :

- La cascade d'analyse "Perdido_analysis.csc" relative à l'étape 1 dont le but est d'identifier les informations dont nous avons besoin et d'annoter le texte.
- La cascade de synthèse "Perdido_synthesis.csc" relative à l'étape 2 qui organise les connaissances en un XML structuré.

Le premier objectif qui m'a été fixé est de modifier la cascade d'analyse "Perdido_analysis.csc" afin de définir des graphes et cascades permettant d'analyser les sorties de PERDIDO dans un contexte spatio-temporel maritime. Ceci sous-entend :

- i) l'identification/extraction de certains motifs (patrons) lexicaux relatifs à la navigation maritime côtière
- ii) la structuration de ces motifs selon une structure à base de balises XML.

La cascade de synthèse "Perdido_synthesis.csc", elle, ne sera pas modifiée et consistera à générer le fichier XML.

Traitement en cascade dans le contexte de la navigation maritime côtière

Dans cette partie, et en lien avec la chaîne de traitements proposé (c.-à-d. 1) extraction de motifs, 2) structuration de motifs), nous avons choisi de présenter tout d'abord la structuration du fichier XML attendu, et ensuite l'extraction de motifs lexicaux à partir des graphes Unitex.

Structuration des connaissances

Les connaissances à rechercher dans le corpus sont directement issues de l'ontologie de navigation que nous présenterons dans la partie 5. Le format pour importer les données annotées dans l'ontologie est le format de sortie d'Unitex, c'est-à-dire XML. Nous avons défini avec Wissame Laddada l'organisation des balises pour extraire au mieux les informations du corpus.

Le premier paramètre qui est décisif pour se repérer en mer est la visibilité. Nous avons donc introduit la dimension temporelle (jour/nuit) pour la navigation en créant une balise "Navigation" avec les paramètres suivants :

```
Navigation type="byDay | byNight"
```

Nous définissons également un alignement avec "LeadingLine". **De jour**, nous précisons qu'il est décrit par des "NavigationMark" (amers) et un angle de cap. Les "NavigationMark" sont constitués d'un nom, d'une position (latitude et longitude) et d'un type (bouée, phare, tourelle, etc.). **De nuit**, les alignements sont décrits par deux "Lights" (les feux provenant d'un phare ou d'une bouée). Les attributs d'un feu sont tels que montré dans la figure 4.13. Ces feux sont conceptualisés dans le modèle ontologique, et devront être extraits du corpus et annotés .

En venant du chenal du Four, on peut virer la pointe St-Mathieu au plus près en laissant sur bâbord le feu rouge (éclat 4 sec.) visible de 280 ° à 113° de la tourelle des Vieux Moines, puis le feu rouge (4 éclats 15 sec.) de la, bouée à sifflet Charles Martel que rase l'alignement lumineux à 68°.

Figure 4.13 : Exemple de description d'un feu rouge dans le corpus

Ainsi et selon les informations existantes dans le corpus et le modèle ontologique, nous avons choisi les paramètres qui suivent pour la balise "Light" :

- Name : Le nom du feu
- Location : Ses coordonnées GPS
- Color : La couleur de son feu
- LightCategory : La catégorie du feu : directionnelle, lumière pour orienter le trafic aérien, etc.
- LightCharacteristic : Les caractéristiques du feu : fixed, scintillante, flash...
- Sector : Les secteurs ou les angles d'où le feu est visible pendant la navigation

Les figures 4.14 et 4.15 présentent deux exemples d'instance pour les fichiers XML, de jour puis de nuit :

```
<Navigation type="byDay">
  <LeadingLine type="onNavigationMark">
    <Angle A="xxx°"/>
    <NavigationMark>
      <name value=""/>
      <location long="" lat=""/>
      <type value=""/>
    </NavigationMark>
    <NavigationMark>
      <name value=""/>
      <location long="" lat=""/>
      <type value=""/>
    </NavigationMark>
  </LeadingLine>
</Navigation>
```

Figure 4.14 : Structure XML décrivant un alignement de jour entre deux "NavigationMark".

```

<Navigation type="byNight">
  <LeadingLine type="onLight">
    <Angle A="xxx°"/>
    <Light>
      <name value=""/>
      <location long="" lat=""/>
      <color value="">
      <LightCategory value=""/>
      <LightCharacteristic value=""/>
      <sector number="1" value=""/>
      <sector number="2" value=""/>
    </Light>

    <Light>
      <name value=""/>
      <location long="" lat=""/>
      <LightCategory value=""/>
      <LightCharacteristic value=""/>
      <sector number="1" value=""/>
      <sector number="2" value=""/>
    </Light>
  </LeadingLine>
</Navigation>

```

Figure 4.15 : Structure XML décrivant un alignement de nuit entre deux "Light".

Les structures XML de base étant définies, l'étape suivante consiste à créer les motifs et les cascades afin de structurer la connaissance conformément au fichier modèle.

Cascades pour l'extraction des motifs de navigation

Dans le corpus des guides côtiers, les coordonnées GPS des amers mentionnés dans le texte sont annotés dans un encart "POINTS GPS" présent sur le côté des pages (cf. figure 4.3).

Les coordonnées GPS se présentent dans le texte comme suit :

```
Virage Fort de Bertheaume 48° 20 20 - 4° 41 24
```

En résultat, Perdido retourne ce qui suit :

```
{Virage, virage.N} {Fort, .NPr} {de, de.PREP} {Bertheaume, .NPr} {48, 48.N}
{20, 20.A} {20, 20.N} {-4, 4.N} {41, 41.N} {24, 24.A}
```

Dans chacune de ces accolades, nous avons le token, suivi de son lemme accolé par un point à sa catégorie grammaticale.

L'enjeu, pour des coordonnées GPS, est d'identifier les unités lexicales qui les composent, telles que le nom propre, les nombres représentant la latitude et ceux représentant la longitude. L'exemple qui suit illustre les mots clés à identifier.

Exemple : "Fort de Bertheaume" (nom propre) 48° 20 20 (3 nombres : latitude) - 4 ° 41 24 (3 nombres : longitude).

Il s'agit donc maintenant, comme expliqué précédemment, de faire deux cascades : la première va extraire un nom, des coordonnées GPS et un type et la deuxième qui formalise les informations selon le template XML.

De part l'identification de ce motif lexical, nous pouvons en déduire un premier graphe Unitex (figure 4.16) qui permet d'extraire les coordonnées GPS de notre corpus.

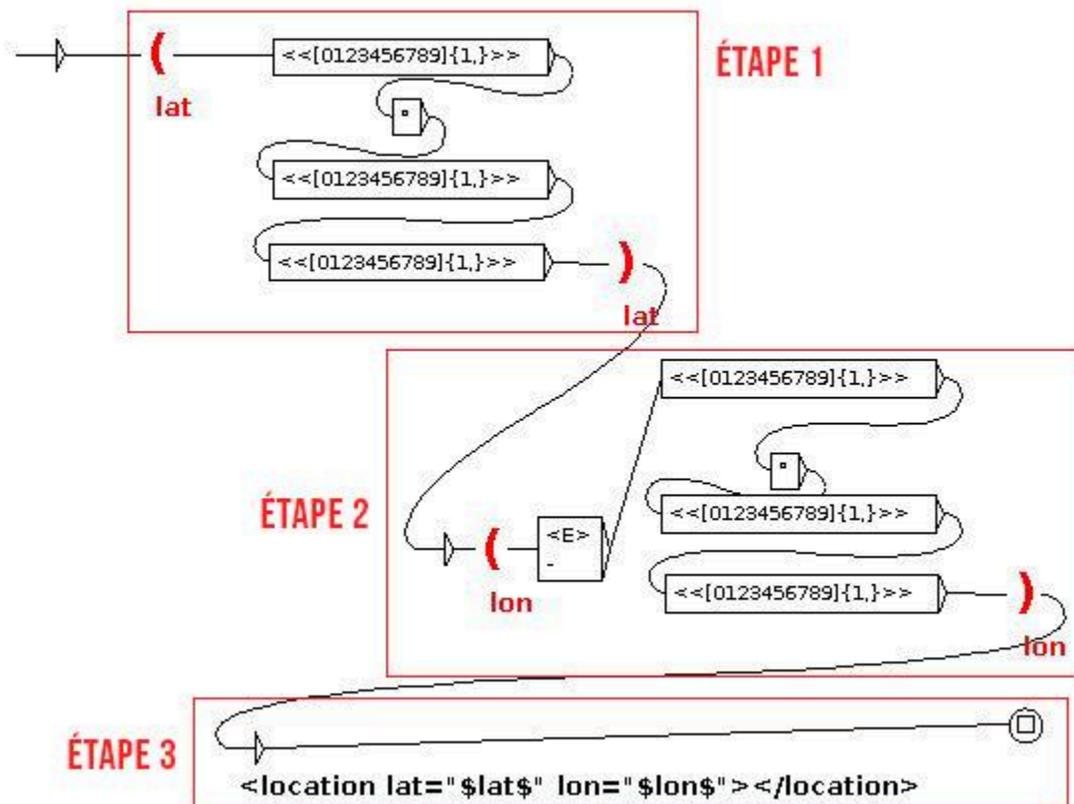


Figure 4.16 : Graphe reconnaissant les coordonnées GPS

Ce graphe se compose de trois étapes :

1. Extraction de la latitude avec la structure NB° NB NB et stockage de la valeur dans la variable lat ;
2. Extraction de la longitude avec la structure NB° NB NB et stockage de la valeur dans la variable lon ;
3. Mise en forme des données selon la structure définie par la cascade de synthèse.

Ainsi, plusieurs cascades doivent être définies pour extraire et structurer les données relatives à la navigation maritime côtière, en accord avec les connaissances du modèle ontologique, et avec l'objectif de le peupler.

5. Peuplement de l'ontologie de navigation

Qu'est-ce qu'une ontologie

Une ontologie constitue un modèle permettant de représenter des connaissances d'un domaine. On parlera alors d'ontologie de domaine, ou des connaissances de haut niveau plus générales comme par exemple les ontologies spatiale (GeoSPARQL) et temporelle (OWL-Time). Une des particularités est qu'un modèle ontologique permet d'inférer de nouvelles connaissances à partir de connaissances initiales par l'utilisation de raisonneurs (ou moteurs d'inférences).

Une ontologie de domaine, telle que celle de navigation maritime côtière définie par Wissame Laddada, est constituée d'une "terminological box" (Tbox) qui définit le vocabulaire du domaine, c'est-à-dire les concepts (ou classes) et leur hiérarchie et les relations (ou rôles) et leur hiérarchie, ainsi que d'une "assertional box" (Abox) qui définit les faits (ou instances). Les deux, Abox et Tbox, définissent alors ce que l'on appelle une base de connaissances. Une partie de l'ontologie de navigation maritime est présentée dans la figure 5.1. La fenêtre de couleur ocre (Class hierarchy) identifie les concepts et leur hiérarchie, la fenêtre de couleur mauve en bas à gauche (Direct instances) définit les instances des concepts et la fenêtre de couleur mauve à droite ("Property assertions") identifie ici la relation qui lie le *Cap de la Chèvre* au *Point345* (*Cap de la Chèvre*, hasGeometry, *Point345*). La hiérarchie de concepts de l'ontologie est plus précisément détaillée dans la figure 5.2.

Plusieurs étapes sont nécessaires afin de construire une ontologie. Premièrement, l'étape de conceptualisation vise à définir la Tbox, c'est-à-dire les concepts et leur hiérarchie et les rôles et leur hiérarchie. L'étape de formalisation vise ensuite à définir les concepts complexes à partir de concepts simples (ou atomiques), de relations et de constructeurs logiques issus des logiques de description ("Description Logic" ou DL). Un exemple de définition de concept complexe est présenté dans la figure 5.3. La troisième étape, appelée ingénierie ontologique, consiste à traduire cette formalisation dans un langage ontologique (Ontology Web Language (OWL)) avec une certaine syntaxe apte à être interprétée par une machine. Dans le cas de l'ontologie de navigation sur laquelle nous avons travaillé, toutes ces étapes étaient préalablement définies.



L'étape de peuplement de l'ontologie (Abox) constitue la seconde étape de mon stage et consiste à instancier les concepts de l'ontologie à partir du fichier XML créé à l'issue du traitement linguistique.

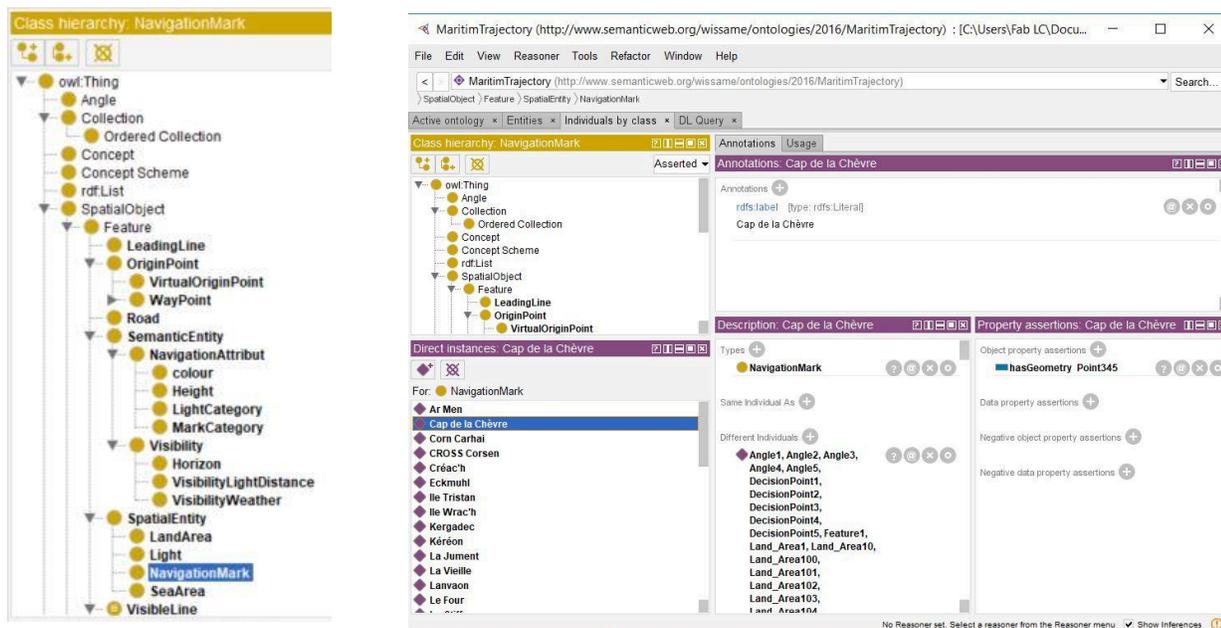


Figure 5.1 : Présentation de l'ontologie de navigation maritime et de l'éditeur d'ontologie Protégé

NavigationMark (is a) SpatialEntity (is a) Feature (is a) SpatialObject (is a) Thing

Figure 5.2 : Exemple de subsomption ou inclusion de concepts. Cas particulier de "NavigationMark"

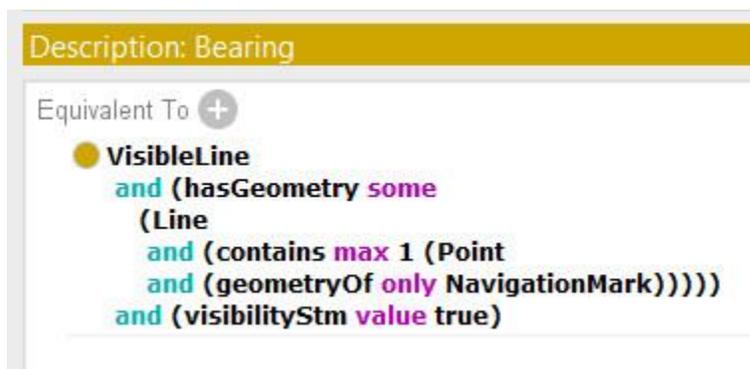


Figure 5.3 : Exemple de définition du concept complexe “Bearing” (Relèvement) en logique de description (DL)

Nous avons utilisé un éditeur d'ontologie nommé Protégé² (figure 5.1) développé par l'université de Stanford. Cet éditeur est communément partagé dans la communauté et permet de définir les concepts (atomiques et complexes), les relations et les instances. Cet éditeur intègre aussi des moteurs d'inférences tels que Fact, Pellet, etc. pour inférer de nouveaux faits et vérifier la cohérence des modélisations réalisées. Nous souhaitons à l'issue du stage utiliser un moteur d'inférence afin d'instancier de nouveaux concepts à partir des assertions nouvelles issues des textes des “Pilotes côtiers” et ainsi générer de nouvelles routes maritimes côtières. Cette dernière étape restera à confirmer.

Structure de l'ontologie de navigation maritime

La structure de l'ontologie, sur laquelle j'ai travaillé durant mon stage, a été réalisée par Wissame Laddada et optimisée pour raisonner avec des données maritimes issues des cartes électroniques de navigation (ou electronic navigational chart (ENC)). Ces cartes numériques respectent la norme S-57³ de l'organisation hydrographique internationale (OHI). Cette norme répertorie, entre autre, l'ensemble des catégories ou couches thématiques présentes dans les ENC. La figure 5.5 présente la couche thématique LIGHTS qui définit les objets maritimes ayant un feu. Cette catégorie est alors aussi définie comme un concept “Light” de l'ontologie de navigation (cf. figure 5.2).

² <https://protege.stanford.edu/products.php#desktop-protege>

³ <http://www.s-57.com/>

Object:	Light
Acronym:	LIGHTS
Code:	75

help

Geometric primitives: P

Set Attribute_A: (!?)CATLIT; (!?)COLOUR; DATEND; DATSTA; EXCLIT; (?)HEIGHT; (!?)LITCHR; LITVIS; MARSYS; MLTYLT; NOBJNM; OBJNAM; (!?)ORIENT; PEREND; PERSTA; (!?)SECTR1; (!?)SECTR2; (!?)SIGGRP; (!?)SIGPER; (?)SIGSEQ; STATUS; VALNMR; VERACC; (?)VERDAT;

Set Attribute_B: INFORM; NINFOM; NTXTDS; SGAMAX; SCAMIN; TXTDSC;

Set Attribute_C: REGDAT; RECIND; SORDAT; SORIND;

Definition:
A luminous or lighted aid to navigation. (adapted from IHO Dictionary, S-32, 5th Edition, 2766)

References
INT 1: IP 1-30.3, 40-65;
S-4: 470-473.5, 475-475.7, 476-478.5;

Remarks:
A light may be fixed on a buoy, beacon, tower etc. These are separate objects.

Distinction:
beacon, cardinal, beacon, isolated danger, beacon, lateral, beacon, safe water, beacon special purpose/general; buoy, cardinal, buoy, installation, buoy, isolated danger; buoy, lateral; buoy, safe water; buoy, special purpose/general; light vessel; light float;

Attribute type: E

Expected input:

ID	Meaning	INT 1	S-4
1	fixed	IP 10.1;	
2	flashing	IP 10.4;	
3	long-flashing	IP 10.5;	
4	quick-flashing	IP 10.6;	
5	very quick-flashing	IP 10.7;	
6	ultra quick-flashing	IP 10.8;	
7	isophased	IP 10.3;	
8	occulting	IP 10.2;	
9	interrupted quick-flashing	IP 10.6;	

Figure 5.4 : La couche thématique LIGHTS de la norme S-57

Il existe de nombreux concepts et relations dans l'ontologie de navigation. La première phase a été de comprendre la structure de cette ontologie afin de mettre en relation les deux objectifs du stage que sont le traitement linguistique et le peuplement de l'ontologie. Plus précisément, il nous faut comprendre la structure ontologique pour rechercher et extraire les connaissances contenues dans les ouvrages "Pilotes côtiers" à partir des éléments modélisés dans l'ontologie (concepts, attributs, relations) et entrant dans la composition du fichier XML à générer. Nous renvoyons le lecteur aux figures 4.14 et 4.15, afin de mettre en relation ces éléments ontologiques avec la structure du fichier XML. Nous ne présentons ici que les concepts principaux nécessaires à la réalisation du stage.

Association des structures de l'ontologie et du fichier XML

Le concept "NavigationMark"

Le concept atomique de "NavigationMark" rassemble les amers (phares, tourelles, pylônes, etc.) qui servent de points de repère (ou "landmarks") pour la navigation maritime. Ce concept est fondamental pour la définition des concepts que sont les alignements "LeadingLine" (un alignement est une ligne composée d'au moins deux amers) et les relèvements "Bearing" (un relèvement est une ligne composée d'un amer).

La figure 5.5 présente une instance "Pointe du Petit Minou" du concept "NavigationMark". Dans cette figure, la fenêtre "Annotations" présente l'ensemble des relations permettant de décrire une ressource. Ces relations font partie d'ontologies ou de vocabulaires existants tels que rdfs (rdfs:label, rdfs:comment) ou encore dcterms (dcterms:publisher, dcterms:provenance, etc.). Ici le "label" de l'instance "Pointe du Petit Minou" indique son nom commun. La valeur de la relation

“label” (c.-à-d. “Pointe du Petit Minou”) est de type “Literal”, qui est un terme du vocabulaire de rdfs (rdfs:Literal) . De la même façon, la fenêtre “Description” donne des informations sur l’instance. Ici, sont présentés les concepts auxquels l’instance appartient (“NavigationMark) ainsi que des formalisations logiques supplémentaires permettant de dire, dans le cas présenté, que l’instance “Pointe du Petit Minou” a une géométrie qui est exactement un “Point” (sous-entendu, et rien d’autre). Enfin, la fenêtre “Property assertion” décrit les relations que l’instance possède : “Pointe du Petit Minou” a pour géométrie le point “Point354”, qui est lui-même une instance du concept “Point” (cf. figure 5.6). En cliquant sur ce point “Point354”, nous pouvons avoir des informations complémentaires comme sa latitude et sa longitude.

En conclusion pour un amer, les balises <name value=""/>, <location long="" lat=""/>, <type value=""/> décrivant un “NavigationMark” dans le fichier XML (figure 4.14) sont en adéquation avec les descriptions faites dans l’ontologie et permettront ainsi son instanciation.



Figure 5.5 : Exemple d’une instance “Pointe du Petit Minou” du concept “NavigationMark”



Figure 5.6 : Présentation de l’instance “Point354” du concept “Point”

Le concept “Light”

Le concept “**NavigationMark**” peut être enrichi par des informations complémentaires comme les signaux lumineux (ou feux) présents la nuit ou par faible visibilité. Ce concept permet de différencier la navigation “par bonne visibilité” de la navigation “par faible visibilité” (nuit, brume, etc.). Ainsi, les attributs de la catégorie LIGHTS (cf. figure 5.4) utiles au projet se retrouvent modélisés au sein de l’ontologie notamment par les relations de la figure 5.7. On peut, par exemple, définir un signal lumineux par une couleur, un secteur angulaire de visibilité, un rythme d’occultation, etc.

Comme précédemment pour un “**NavigationMark**”, les balises `<name value=""/>`, `<location long="" lat=""/>`, `<color value="">`, `<LightCategory value=""/>`, `<LightCharacteristic value=""/>`, `<sector number="1" value=""/>`, `<sector number="2" value=""/>` décrivant un objet de type “Light” dans le fichier XML (figure 4.15) sont en adéquation avec les descriptions faites dans l’ontologie.

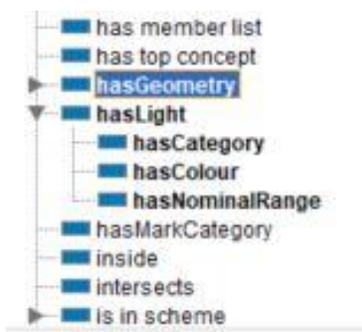


Figure 5.7 : Hiérarchie de relations dont le sujet de la relation est une instance du concept “Light”.

Le concept “LeadingLine”

L’objectif du projet étant d’élaborer un itinéraire en fonction d’alignements ou de relèvements maritimes, le concept “**LeadingLine**” (ou alignement) décrit donc simplement une ligne qui est elle-même constituée de deux amers ainsi que d’un angle. Dans la figure 5.8, l’instance “LeadingLine1” du concept “LeadingLine” a une géométrie “Line9” (instance du concept Line) et un angle “Angle4” (instance du concept Angle).



Figure 5.8 : Description de l'instance "LeadingLine1" du concept "LeadingLine"

Dans la structure du fichier XML, on différencie une navigation "par bonne visibilité, où l'alignement est défini par un angle et un ensemble deux moins deux amers de type "NavigationMark" (figure 4.14), d'une navigation "par faible visibilité" où l'alignement est défini par un angle et un ensemble deux moins deux feux de type "Light" (figure 4.15).

Instanciation de l'ontologie

Pour réaliser cet objectif, il s'agira de développer un programme écrit en langage Java qui parsera les données contenues dans le fichier XML pour instancier les concepts de l'ontologie. Nous utiliserons des bibliothèques dédiées à la manipulation d'ontologie telles que Jena ainsi que des bibliothèques gérant la dimension spatio-temporelle.

Ce processus d'instanciation (figure 5.9) veillera également à adapter et effectuer des calculs sur les connaissances que Unitex ne peut pas faire comme, à titre d'exemple, transformer le format initial des coordonnées GPS (degrés, minutes et secondes) en degrés décimaux. Après avoir effectué tous les calculs et transformations de valeurs nécessaires, le programme instanciera un à un les concepts de l'ontologie afin d'appliquer en phase terminale un moteur d'inférences pour en déduire de nouvelles routes de navigation.

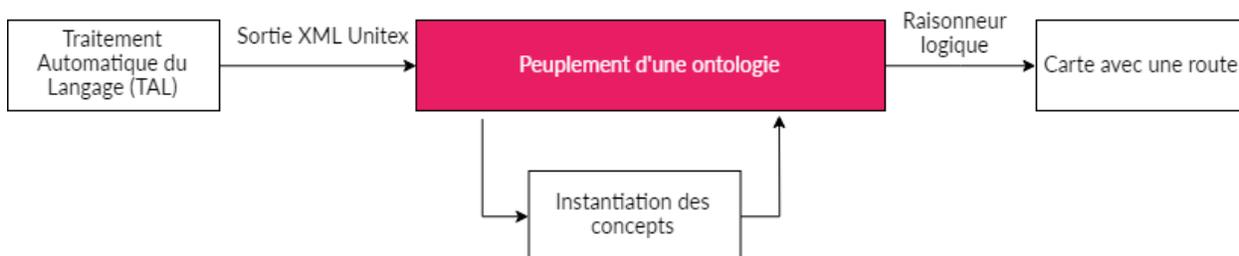


Figure 5.9 : Processus de peuplement de l'ontologie

Conclusion

De la numérisation des guides côtiers à l'instantiation de l'ontologie de navigation en passant par le traitement naturel du langage, nous avons travaillé sur une chaîne complète de traitements de connaissances. Ce système d'information permettra aussi d'inférer de nouvelles routes de navigation maritime côtière pour proposer aux navigateurs des routes maritimes sécurisées qu'ils peuvent emprunter.

Même si l'échéance du rapport et la soutenance interviennent relativement tôt dans le déroulement du stage à l'IRENav, les premières extractions de connaissances à partir du corpus ont pu être réalisées ainsi qu'une partie du traitement du langage naturel. Après avoir appréhendé le modèle ontologique et les outils de TAL, j'ai été amené à définir dans Unitex des graphes regroupés en cascade pour extraire les coordonnées GPS des entités nommées et les mettre en forme dans un fichier XML dont nous avons défini la structure. L'objectif suivant est d'extraire d'autres connaissances (alignements, amers, entités nommées, relations spatiales, etc.), par l'intermédiaire de cascades de graphes supplémentaires afin de les instancier dans l'ontologie. La liste des tâches restantes à accomplir dans le cadre de ce stage est proposée en annexe A.

Ce stage est enrichissant scientifiquement, car il me permet de découvrir le fonctionnement interne d'un laboratoire de recherche, ses acteurs, ses contraintes et il m'a aussi permis de contribuer au développement d'un projet de recherche très novateur. J'ai amélioré mes compétences en traitement du langage naturel et plus globalement en traitement et analyse de données. Sur le plan humain, j'ai été intégré dans une équipe de recherche MoTIM et en relation avec des enseignants chercheurs, des doctorants ou des stagiaires qui ont des projets et des compétences dans des domaines divers, mais complémentaires et toujours liés à l'environnement maritime.

L'application maritime de ce système "intelligent" fondé sur la logique et le raisonnement alimenté par des connaissances textuelles offre un potentiel prometteur pour créer un outil de navigation supplémentaire et fiable. La modélisation sémantique par des ontologies et l'intelligence artificielle démontrent clairement leur potentiel pour des applications d'aide à la décision actuelles mais aussi futures.



BIBLIOGRAPHIE

- [1] Ludovic MONCLA, Services Web pour l'annotation sémantique d'information spatiale. Conference: SAGEO Spatial Analysis and GEOMatics 2017, At Rouen, France, November 2017.
https://www.researchgate.net/publication/321018301_Services_Web_pour_l'annotation_semantique_d'information_spatiale_a_partir_de_corpus_textuels
- [2] Wissame LADDADA, Vers une émergence des systèmes d'information géographique maritime fondés sur la connaissance : application aux systèmes d'aide à la navigation. Thèse de Doctorat, Université de Bretagne Occidentale (UBO), France, Décembre 2018.
- [3] Mauro GAIO, Ludovic MONCLA. Extended Named Entity Recognition Using Finite-State Transducers: An Application To Place Names. The Ninth International Conference on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2017), Mars 2017, Nice, France. Ffhal-01492994
<https://hal.archives-ouvertes.fr/hal-01492994/document>
- [4] Ludovic MONCLA. Automatic reconstruction of itineraries from descriptive texts. Information Retrieval [cs.IR]. Université de Pau et des Pays de l'Adour; Universidad de Zaragoza, 2015. English. Fftel01249999
<https://tel.archives-ouvertes.fr/tel-01249999/document>
- [5] Nguyen V. T., Gaio M., Moncla L. (2013). Topographic subtyping of place named entities: a linguistic approach. In The 15th AGILE international conference on geographic information science, p. 1–5. Louvain, Springer
<https://hal.archives-ouvertes.fr/hal-00847505/document>
- [6] Sébastien Paumier (2016). MANUEL D'UTILISATION UNITEX 3.1. Université Paris-Est Marne-la-Vallée, France
<https://unitexgramlab.org/releases/latest-stable/man/Unitex-GramLab-3.1-usermanual-fr.pdf>
-
-

Annexe A : Planification du stage

A. Mise en place (Deadline 17/05/19)

1. Prise en main du sujet
 - Compréhension des enjeux et des rôles des technologies utilisées
 - Visio conférence avec L. Moncla pour éclaircir les points flous
 - Formalisation sur le rapport de stage -
 - Description de la structure sur le rapport de stage
2. Mise en place d'un environnement de travail
3. Exécution du programme fourni par L. Moncla -
 - Formalisation des difficultés rencontrées sur le rapport de stage -

B. Traitement Automatique du Langage (Deadline 16/07/19)

1. Annotation manuelle des textes -
2. Comparaison annotation manuelle vs. Perdido (rappel précision) - 0%
3. Analyse de l'ontologie de Wissame Laddada -
 - a. Formalisation sur le rapport de stage -
4. Graphes
 - a. Description des graphes de base -
 - b. Suppression des graphes inutiles -
 - c. Formalisation sur le rapport de stage -
5. Détermination des graphes qui en découlent - 50%
 - a. Proposition de graphes rectificatifs - 0%
 - b. Mail Eric instructions nautiques + template KML - Reçu
 - c. Proposition de modification des lexiques - 0%
 - d. Comparaison annotation manuelle vs. Perdido (rappel précision) - 0%
 - e. Adapter pour les instructions nautiques - 0% ⇒ Semaine du 27 Mai
 - f. Ordre Latitude + longitude - 0%
 - g. Sortir un kml pour visualisation dans Google Maps - 0%
6. Adaptation des graphes fournis et implémentation de graphes complémentaires dans les cascades de Ludovic Moncla - 0%

C. Peuplement de l'ontologie (Deadline 31/08/19)

1. Élaboration d'une cascade de synthèse compatible avec le logiciel - 0% ✗
 2. Formalisation de ce processus au fur et à mesure sur le rapport de stage - 0% ✗
-
-

Annexe B : Le Traitement Automatique du Langage par PERDIDO

Le Traitement Automatique du Langage par PERDIDO s'effectue selon le schéma suivant :

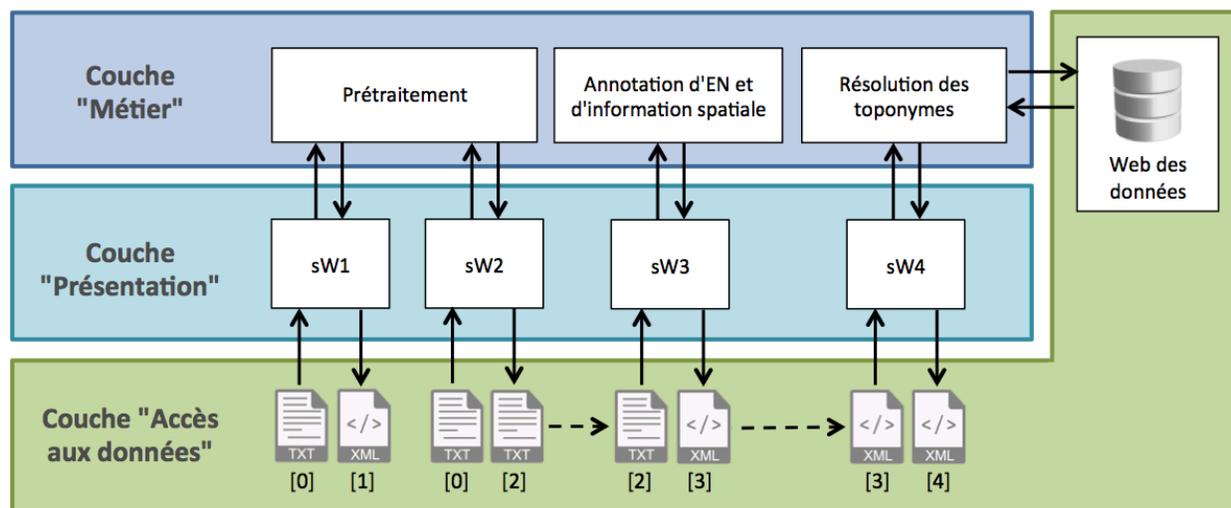


Figure B.1 : Architecture logicielle du Service Web pour l'annotation sémantique d'information spatiale (L. Moncla) [1]

Perdido permet d'interroger le module de prétraitement, propose un résultat au format texte adapté au module d'annotation automatique des entités nommées et des informations spatiales associées.

La cascade d'analyse exécute cinq transducteurs principaux et produit un résultat au format XML défini par le programme CasSys d'Unitex.

1. Le premier transducteur est appelé : il s'agit de *Talismane*⁷, un analyseur syntaxique fonctionnant en Java.

Il permet de passer d'un texte brut à un réseau de dépendances syntaxiques. Pour cela, Talisman utilise une analyse en cascade avec quatre étapes classiques : le découpage en phrases, la segmentation en mots, l'étiquetage (attribution d'une catégorie morphosyntaxique), et le parsing (repérage et étiquetage des dépendances syntaxiques entre les mots).

2. Le deuxième transducteur annote les relations spatiales exprimées dans le texte telles que les distances, les relations topologiques ou les directions.
 3. Le troisième transducteur annote les verbes, plus précisément les verbes de déplacement et de perception. Ce transducteur permet également d'associer une sémantique à l'annotation réalisée, en particulier pour les verbes de déplacement leur polarité.
 4. Le quatrième transducteur annote les entités nommées étendues composées de noms propres et de termes associés [3].
 5. Le cinquième transducteur annote les expressions de déplacement ou de perception, qui sont repérées à l'aide de la formalisation des structures VT [5] mettant en relation les différents éléments annotés par les transducteurs précédents.
-
-

Annexe C : Correction des annotations et détermination des graphes

Annotation manuelle d'un texte

Amer ⇒ PERDIDO le détecte très bien

Relation spatiale absolue ⇒ “sortie ... du ...”, “vers l'...”, “au ...”, “à l'... de”, “au ... une roche”, “au ... et au ...”, “au ... à”, “par le ...”, “dans ...”, “doubler par ... la”
 ⇒ [au / à l' / vers / sortie] ... [du / une roche / et / à]
 ⇒ [doubler par] ... [la / le]
 ⇒ “par vent frais d'ouest”

Relation spatiale relative ⇒ “sur tribord”

Verbe de déplacement - Perception - Topographique - Position

Entité spatiale

Distance ⇒ “soit ... milles”, “isolée à ... milles”, “à craindre à ... mille”, “dans le ...° à ... mille”, “à environ ... mille”
 ⇒ [soit / à / à environ] ... [milles / mille]

Hauteur ⇒ Dès qu'il y a des mètres : Profondeur (“découvrant de ... m”, “une roche de ... m”), hauteur de rocher ou longueur de navire (“de plus de ... m”)

Angle ⇒ “A ...° du”, “alignement à ...° du ...”, “se situe dans le ...° à ... mille”
 ⇒ [a / alignement à / se situe dans le] ... [du / à]

Danger potentiel au niveau du Traitement Automatique du Langage

Entre la sortie sud du chenal du Four et le Raz de Sein, soit 18 milles, la mer d'Iroise est largement ouverte vers l'ouest et ne présente aucun haut fond non balisé. Elle se prolonge par le goulet de Brest au N.E. et la baie de Douarnenez au S.E. séparés par l'avancée de la pointe de Camaret que débordent quelques îlots et roches affleurantes jusqu'à la tourelle de la Parquette.

La **Parquette**, une haute tourelle à bandes obliques noires et blanches, est isolée à **3 milles** à l'**ouest** de la **balise du Petit Léac'h** plantée sur une roche découvrant de **2.90 m**. Le **Corbeau** (**3.10 m**) est la seule roche à craindre à **0,7 mille** avant le **Petit Léac'h**. Elle se situe dans le **145°** à **0,7 mille** de la **bouée verte le Trépied** qui signale au **nord** une roche de **2.30 m**. La mer brise vite sur ces **hauts-fonds par vent frais d'ouest** mais l'entrée du **goulet de Brest** n'en reste pas moins aisément praticable par un passage large de **3 milles**. Il suffit simplement de **reconnaître** les différentes bouées qui balisent la route des grands navires qui dans le **goulet** se divise en deux routes au **nord** et au **sud** du **plateau des Fillettes**.

La ligne joignant la **pointe du Toulinguet** au **sud** à la **bouée Charles Martel**, en lisière de l'alignement à **68°** du **grand chenal** à **3,8 milles** avant le **phare du Petit Minou**, marque le début de la zone de pilotage où les navires de plus de **50 m** ont une totale priorité de manœuvre. Une veille doit en outre être assurée sur canal 16 en VHF.

À l'approche du **goulet**, le **phare de la pointe du Petit Minou** accolé à une **tour blanche** donne avec le phare de **Portzic**, une tour élancée sur la pointe à **gauche** de l'entrée dans la rade, l'alignement à **68°** du **grand chenal**. De jour, il n'est pas nécessaire de suivre exactement cet alignement car il ne manque pas d'eau à courir pour un voilier.

En venant du **nord**, du **chenal du Four**, on **arrondira** largement la **pointe St-Mathieu** qui porte la haute tour rouge et blanche du phare accolé à une **abbaye en ruine**, en venant virer par le **sud** la **tourelle rouge des Vieux Moines**. En se **dirigeant** vers l'**Est** sur la **bouée bâbord Coq Iroise**, on pare la roche isolée affleurante qui **pointe** à environ **0,2 mille** au **sud** de l'**éperon rocheux des Respects** à **0,4 mille** dans l'**Est** des **Vieux Moines**. De la bouée le **Coq**, une même route conduit sur la **bouée Charles Martel** en lisière du **grand chenal**. On notera que l'alignement à **275°** derrière soi des deux **Bozmen** pare ces dangers et conduit également sur le **grand chenal**. Les bouées **Charles Martel** et de la basse **Beuzec** peuvent être laissées à distance sur **tribord** en longeant le rivage rocheux et escarpé du **goulet**, les fonds étant accores, ce qui permet d'esquiver partiellement les courants forts et l'on trouve rapidement un mouillage dans l'anse de **Bertheaume**.

En venant du **sud**, si l'on **n'emprunte** pas le **chenal du Petit Léac'h** (voir p. 68) ou du **Toulinguet** par suite d'une trop forte houle ou d'une visibilité réduite, on viendra doubler par l'**ouest** la **tourelle de la Parquette** pour laisser ensuite sur **tribord** la **bouée verte du Trépied** et, à mi-parcours de l'entrée du **goulet**, on relèvera à petite distance sur **bâbord** la bouée de danger isolé de l'**épave du Swansea Vale**.

Points GPS

Pte St-Mathieu, phare 48° 19,78 - 4° 46,30

Vieux Moines, tourelle 48° 19,33 - 4° 46,61

Le Coq, bouée 48° 19,08 - 4° 43, 98

Pen Hir, bouée 48°20,01 - 4°39,56

ANNEXE D : Fonction des graphes dans les cascades

Cascade d'analyse

Graphes que j'ai ajouté

- grfCoord.fst2 : Sert à extraire les coordonnées GPS et les mettre sous la forme <location lat="" lon=""/>
- navigationMark : Extraction des NavigationMark et formulation selon la structure XML définie.
- grfNavigationMarkv2 : Extraction des NavigationMark et annotation à l'aide d'accolades et de variables.
- grfLeadingLine : Extraction des LeadingLine et annotation à l'aide d'accolades et de variables.
- grfEntitesSpatiale : Annotation d'une entité spatiale maritime.
- grfDistance : Annotation d'une distance.
-

Graphes concernant l'analyse des verbes

- grfVerbMotionInitial.fst2, grfVerbMotionMedian.fst2, grfVerbMotionFinal.fst2 : Détecte les verbes de mouvement.
- grfVerbLocation :
-
- grfVerbPerception :
- grfVerb : Annotation de l'ensemble verbes en faisant appel aux autres graphes précédents (Initial, Median, location...)

Autres graphes

- GPS_2.fst2 : Récupère et formalise la longitude directement en XML ⇒ Pas utile
- GPS : Formalisation des coordonnées GPS sous une forme différente encore ⇒ Pas utile