



**HAL**  
open science

## Modélisation et extraction de la sémantique des trajectoires à partir de données multi-capteurs

Frédéric Bisone, Laurent Etienne, Thomas Devogele

► **To cite this version:**

Frédéric Bisone, Laurent Etienne, Thomas Devogele. Modélisation et extraction de la sémantique des trajectoires à partir de données multi-capteurs. *Revue Internationale de Géomatique*, 2018, 28 (4), pp.461-483. 10.3166/rig.2018.00065 . hal-02110261

**HAL Id: hal-02110261**

**<https://hal.science/hal-02110261>**

Submitted on 25 Apr 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

---

# Modélisation et extraction de la sémantique des trajectoires à partir de données multi-capteurs

Frederick Bisone<sup>1,2</sup>, Laurent Etienne<sup>1</sup>, Thomas Devogele<sup>1</sup>

1. *Laboratoire d'Informatique Fondamentale et Appliquée de Tours (LIFAT)*  
64, avenue Jean Portalis, 37200 Tours France  
[prenom.nom@univ-tours.fr](mailto:prenom.nom@univ-tours.fr)

2. *Petit Picot by GRUAU*  
ZI n°2 – 18 rue de Prony, 37303 Joué-lès-Tours France  
[prenom.nom@gruau.com](mailto:prenom.nom@gruau.com)

---

*RÉSUMÉ. Les positions issues de capteurs GPS sont régulièrement employées pour analyser des déplacements d'objets mobiles. L'emploi exclusif des positions GPS s'avère insuffisant afin de détecter finement les comportements d'objets mobiles, pour des raisons d'imprécisions et de perte d'informations. Les objets mobiles sont maintenant équipés de plus en plus de capteurs complémentaires (génériques ou spécifiques à des types de véhicules). Cet article présente un modèle de trajectoires sémantique ainsi qu'une méthode pour extraire la sémantique liée à celles-ci à partir de données brutes issues de différents capteurs. Ces outils de fouille de trajectoires permettent d'identifier les activités des véhicules lors de déplacements ou d'arrêts. Ce processus a été appliqué aux déplacements d'ambulances connectées des sapeurs-pompiers d'Indre et Loire.*

*ABSTRACT. GPS positions are useful to analyze movements of mobile objects. Unfortunately, the outcome can be unsatisfactory due to imprecision and signal lost. Several sensors (generic or specific ones depending on the type of vehicle) are now included into mobile objects. This article describes a new model for semantic trajectories. This article also presents the movement pattern extraction process based on the raw data from GPS and other sensors, which help identified vehicle activities and movement patterns. This process was successfully applied to the Indre et Loire fire department and its connected ambulances.*

*MOTS-CLES : objet mobile, fouille de trajectoires, trajectoires sémantiques, capteurs GPS, analyse de mobilité, véhicule de secours*

*KEYWORDS: moving object, trajectory data mining, semantic trajectory, GPS sensor, moving object analysis, connected ambulance*

---

## 1. Introduction

Depuis plusieurs années, la miniaturisation des capteurs GPS ainsi que la baisse de leur coût de fabrication entraînent leur apparition dans un grand nombre d'objets mobiles (Smartphones, véhicules, balises autonomes...). Notamment, les capteurs GPS font leur apparition dans les véhicules de plus en plus connectés capables de transmettre en temps réel l'état d'un grand nombre de capteurs présents dans ceux-ci.

Par exemple, l'envoi de positions GPS autorise le suivi en temps réel de flottes de véhicules, pour définir des zones de couvertures optimales, annoncer des temps d'arrivée estimés (dépanneuse, livreur). Tandis que les informations issues d'autres capteurs se focalisent sur l'état du véhicule et ses capacités à fournir un service (niveau de carburant, taux de remplissage d'une benne pour un camion à ordures, etc.). De même, la combinaison des coordonnées GPS et des données issues des capteurs des véhicules permet la recherche de comportements de conduite : éco-conduite (Klotz et al., 2018), etc. Cependant, l'envoi par les véhicules de grandes quantités de données chaque seconde entraîne des problématiques classiques liées au big data. Le flux de données arrivant en continu produit une masse de données hétérogènes considérable à traiter.

La détection de comportements est spécifique à chaque type d'objet mobile : voiture individuelle, bus scolaire, ambulance, véhicule de police, taxis, ramassage d'ordures, etc. L'analyse de comportement se révèle pertinente dans le cas d'un ensemble homogène d'objets mobiles. Les véhicules d'urgences ont notamment des comportements liés à leurs caractéristiques. Ils répondent à des contraintes génériques : hauteur, largeur, capacité de franchissement de carrefour, etc. De même, leurs contraintes métiers : urgence vitale, patient traumatisé, etc. entraînent un aménagement spécifique du code de la route qui nécessite l'adaptation des itinéraires classiques. Les systèmes de navigation GPS classiques ne tiennent pas compte de ces spécificités. Par contre, des systèmes d'extraction de connaissances sur de gros volumes de données, portant sur les traces collectées d'un ensemble homogène d'objets mobiles, devraient autoriser l'identification de ces comportements propres (Andrienko et al. 2011 ; Ferrero et al. 2016).

Cet article s'intéresse aux problématiques suivantes : Est-il possible d'extraire automatiquement des comportements métier d'un type d'objets mobiles à partir d'un grand nombre de trajectoires sur une agglomération ? Ce processus est-il suffisamment générique pour être transposable à d'autres agglomérations ou à d'autres types de véhicules spécifiques ? Quelles sont les informations complémentaires aux données GPS nécessaires pour obtenir des résultats précis ?

La figure 1 représente le processus général de traitements des données brutes jusqu'à l'exploitation de ces extractions. A terme, les objectifs sont multiples. Premièrement, un système de navigation GPS spécifique à chaque type de véhicule professionnel doit être conçu. Ce dernier prendra en compte les données récoltées pour trouver un itinéraire adapté à chaque type de missions, ainsi qu'aux conditions de transport (horaires, qualité de la route). Deuxièmement, les gestionnaires de flottes doivent pouvoir bénéficier de tableaux de bord spécifiques, permettant l'exploitation

des données sémantiques à travers des requêtes simples. Par exemple, « Quelle est la durée moyenne entre la réception d'un appel au centre de secours et l'arrivée d'une ambulance sur les lieux ? », « Combien de temps passe un véhicule de secours à attendre aux urgences ? », « Quelles sont les zones où les ambulances ont du mal à circuler malgré l'utilisation de la sirène ? ». Troisièmement, le suivi des capteurs permettra également un suivi à distance de l'état des véhicules pour faciliter l'entretien et la prise en charge préventive de ceux-ci.

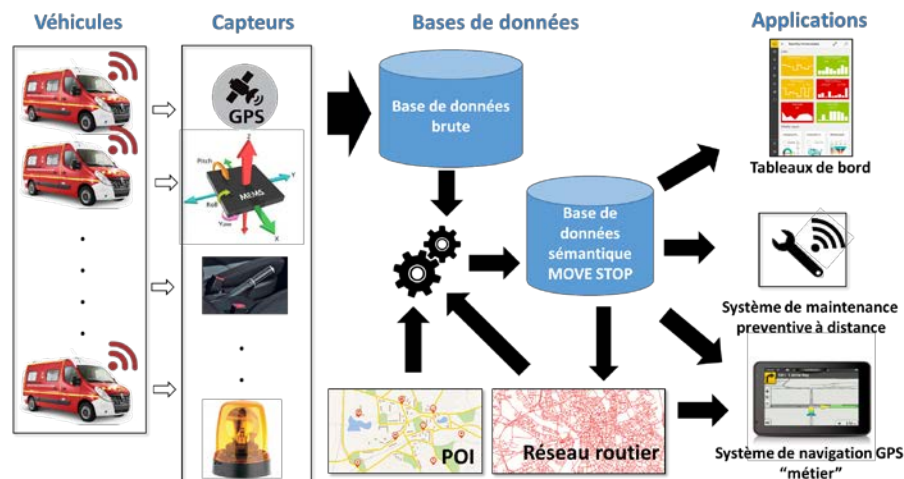


FIGURE 1. Aperçu du processus général du traitement des données

Les sections suivantes décrivent les processus génériques d'extraction des connaissances. Tout d'abord une extension du modèle de Parent et al. (2013) est détaillée. Une méthode pour extraire les arrêts (STOPS) et les déplacements (MOVEs) entre ces STOPS est alors présentée. Ensuite, la sémantique de ces objets mobiles est définie. Dans cet objectif, les STOPS et les MOVEs sont associés à des lieux. Une expérimentation sur des ambulances de secours d'une caserne de pompiers est proposée. Une discussion autour de l'apport de ce processus conclut cet article.

### 1. Modèle de trajectoire sémantique

Classiquement, les trajectoires sont modélisées à l'aide des concepts de STOPS et de MOVEs. Ces concepts sont issus de la Time Geography (Hägerstrand, 1970). Un STOP modélise un arrêt à un lieu durant une période. Il est souvent associé à un POI (Point Of Interest). Yan et al (2011) ont proposé un processus d'enrichissement sémantique de trajectoires brutes : SeMiTri., en utilisant des annotations de trajectoires. Parent et al. (2013) ont formalisé informatiquement ces concepts en décrivant les STOPS par une activité et les MOVEs par un mode de déplacement. Bogorny et al. (2014), ont étendu ce modèle pour y ajouter la sémantique des sous

trajectoires et des points, des places géographiques et des événements. Egalement, Beber et al. (2017) ont proposé une extension sur le concept des STOPS qui permet d'extraire les activités à partir de données GPS et réseaux sociaux.

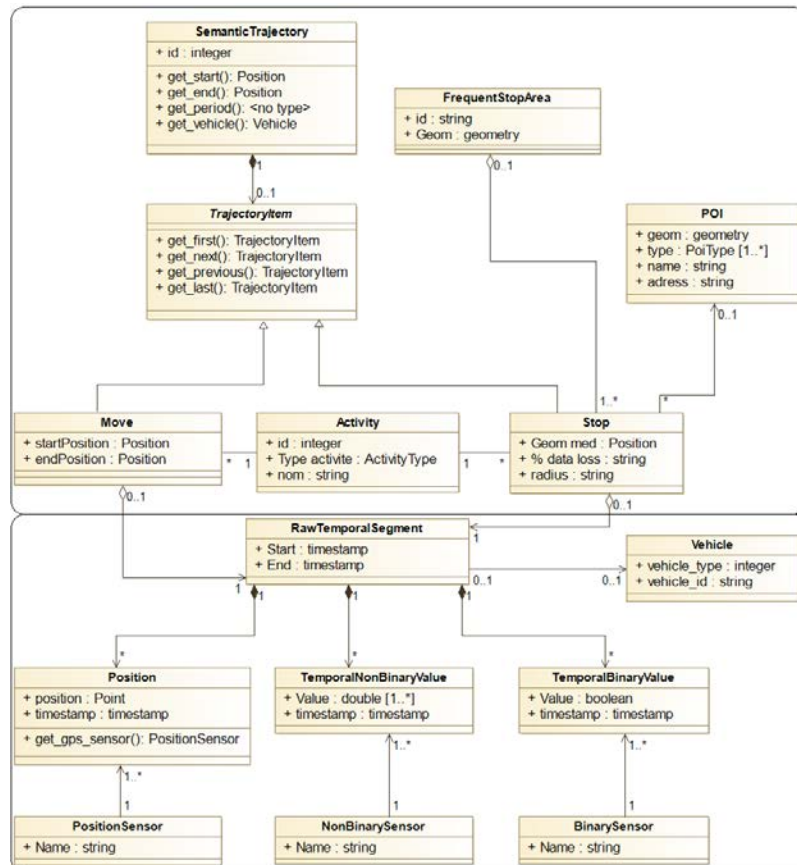


FIGURE 2. Modèle conceptuel d'une trajectoire sémantique

Dans le cadre de modélisation des déplacements de véhicules spécifiques multi-capteurs, une extension de Parent et al. (2013) est proposée dans cet article. La figure 2 décrit le diagramme de classe UML des trajectoires brutes et sémantiques. Il est décomposé en deux parties, la partie supérieure, qui contient les éléments sémantiques et la partie inférieure, les données brutes.

La trajectoire brute d'un véhicule est composée de trois types de listes de données ordonnées temporellement :

- 1) Liste de positions issues d'un système de positionnement par satellites (GPS, GALILEO...)

- 2) Liste de triplets (identifiant capteur, valeur, estampille temporelle) de capteurs binaires par exemple le capteur de frein à main dans un véhicule, le capteur de bouclage de ceinture.
- 3) Liste de triplets (identifiant capteur, valeur, estampille temporelle) de capteurs non binaires tels que le régime moteur d'un véhicule, le kilométrage, la valeur des gyroscopes.

Plusieurs listes du même type sont possibles. Qui plus est, il est également possible de bénéficier de plusieurs capteurs de type identique (Accéléromètre avant et arrière par exemple) dans un même véhicule.

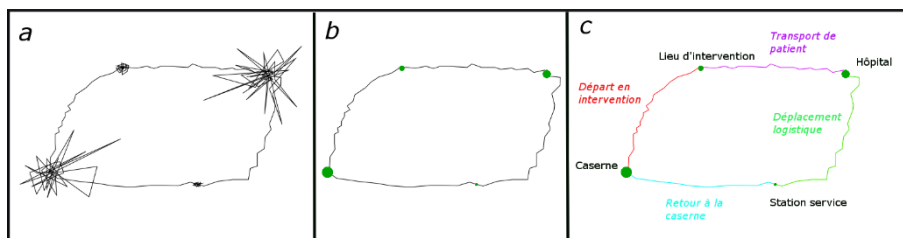
Les trajectoires sémantiques sont issues des trajectoires brutes. Chacune est composée d'une suite ordonnée d'items spatio-temporels. Un item spatio-temporel est soit un STOP soit un MOVE. Il est possible d'enchaîner plusieurs MOVEs ou plusieurs STOPs.

Le modèle présenté autorise une segmentation plus fine des trajectoires sémantiques. Deux MOVEs peuvent aussi se succéder si leurs propriétés changent. Par exemple, un véhicule de retour à la caserne peut être redirigé vers un lieu d'intervention. Ce déplacement doit être segmenté puis modélisé par deux MOVEs successifs.

Deuxièmement, dans les modèles classiques, les MOVEs n'avaient pour fonction que d'assurer un transit entre deux points. Dans le modèle proposé, la sémantique sur les déplacements est plus riche que sur les modèles existants. En plus du mode de déplacement, les MOVEs sont associés à une activité. Les véhicules spécifiques ont, en effet, des types de déplacement qui dépendent de l'application (Transfert rapide d'un patient, transfert délicat, déplacement logistique, dans le cadre de véhicules de secours). De même, un STOP peut être associé à une zone d'arrêts fréquents.

## 2. Processus de montée en sémantique

Dans cette section, le processus qui permet de définir les trajectoires sémantiques à partir des données brutes est présenté. Ce processus s'inscrit dans une démarche de fouille de trajectoires (Zheng, 2015). Le schéma de la figure 3 présente ces différentes étapes.



*FIGURE 3. Exemple de montée en sémantique : données brutes (a), extraction des lieux d'arrêts(b), labellisation des lieux d'arrêts et des trajectoires(c)*

La première étape est l'acquisition des données issues des différents capteurs. Les véhicules sont équipés de capteurs GPS classiques, de capteurs génériques liés au véhicule (frein à main, vitesse à l'ordinateur de bord, clignotants...) ainsi que de capteurs liés à l'activité du véhicule (compteur de taxi, sirène d'une ambulance, charge d'une benne à ordure...). Cette étape qui consiste à importer les données dans une base de données géographiques n'est pas détaillée dans cet article. Dans la deuxième étape, à partir de ces données, les STOPS et les MOVEs sont identifiés puis les lieux où les véhicules s'arrêtent fréquemment sont localisés à l'aide d'une méthode de clustering.

Dans la dernière étape, à l'aide de connaissances extérieures (objets géographiques d'OpenStreetMap (OSM), liste de points d'intérêt métier, etc.) et de règles métiers au cas prototypiques les lieux sont identifiés et les types de déplacements (objectifs du déplacement) sont déduits.

## **2.1. Détection des STOPS et des MOVEs**

### *2.1.1. Limitation des approches existantes*

Dans la littérature, plusieurs problèmes rendent la détection de STOPS (lieu et durée) complexe à l'aide de données GPS classiques (Cain et al., 2005 ; Duran et Earleywine, 2012 ; Knight et al, 2009 ; Ivanovic et al, 2016 ; Ordoñez et al. 2011).

Tout d'abord l'imprécision des capteurs GPS doit être prise en compte. Le signal est très perturbé aux abords de bâtiments ou dans des canyons urbains. Ces perturbations entraînent l'apparition de données aberrantes (rebonds, pelotes), voire d'une perte de signal totale si l'objet mobile se trouve à l'intérieur d'un bâtiment épais. L'analyse lorsque le véhicule stationne en garage fermé ou traverse des tunnels est donc problématique. La figure 4 montre une pelote de positions d'un véhicule stationné en un point fixe à l'intérieur d'un bâtiment (points en vert sur la figure).

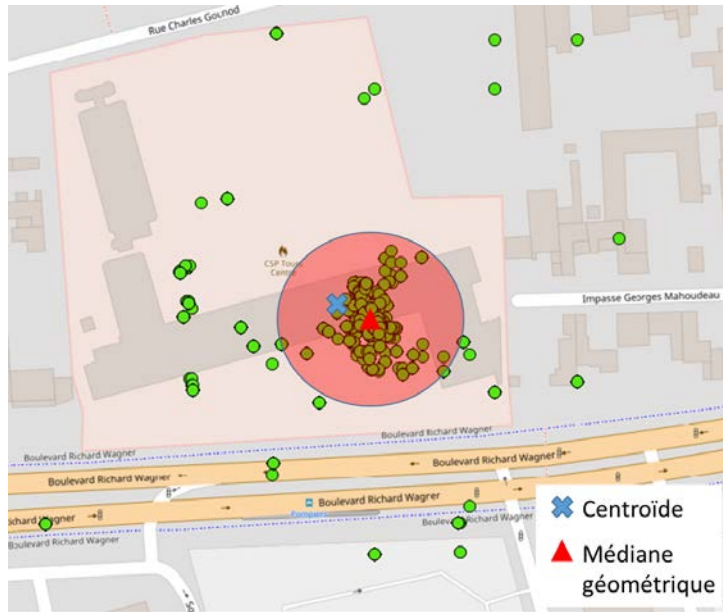


FIGURE 4. Positions d'un véhicule en stationnement à l'intérieur d'un bâtiment, mesure de centralité et de dispersion associées.

De plus, même lors de la sortie d'un bâtiment, la qualité des données est liée au temps dont le capteur GPS a besoin pour se synchroniser avec les satellites (TTF - Time To Fix). Qui plus est, afin d'extraire des informations propres aux véhicules et à leurs activités, il faut distinguer les micro-arrêts des arrêts. Pour cela, les contraintes usuelles de circulation (feux rouges, passages piétons) doivent être prises en compte. Pour répondre à ces problèmes, (Etienne et al., 2012) propose de définir des zones géométriques d'entrée et de sortie pour déterminer le passage d'un objet mobile et ainsi le début ou la fin d'un arrêt. Cependant, cette méthode implique de devoir connaître à l'avance les fronts géographiques d'entrée et de sortie pour chaque zone. Qui plus est, les effets de pelote et les pertes de signal entraînent des sorties virtuelles intempestives de l'objet mobile. Cette méthode efficace en zone ouverte est difficilement transposable en zone urbaine.

### 2.1.2. Définition d'une méthode de détection des STOPS à partir d'autres capteurs

Les valeurs issues des autres capteurs offrent de nouvelles opportunités pour définir les STOPS. En effet, pour déterminer le lieu d'arrêt d'un véhicule, le capteur de frein à main est pertinent. Le moment où le frein à main du véhicule est serré jusqu'à ce que celui-ci soit desserré permet de définir un arrêt de manière plus fiable que le capteur GPS.

Une fois un arrêt déterminé, il doit être modélisé spatialement par un point unique, représentatif de l'ensemble des coordonnées brutes acquises durant cette période.



(Small 1990) présente un état de l'art sur la représentation de médianes multidimensionnelles. La médiane géométrique proposée par (Weiszfeld et al., 2009) a été retenue. Cette dernière est moins sensible au bruit qu'un centroïde (croix bleue de la figure 4). Elle a donc été choisie pour représenter l'emplacement d'un STOP. En plus de la médiane géométrique, deux propriétés sont calculées pour représenter la qualité de la représentation. La première est le pourcentage de données nulles (%DP) qui constituent un stop. La deuxième est le rayon du cercle (Rmed) centré sur la médiane géométrique (triangle rouge de la figure 4). Elle décrit la dispersion des valeurs captées. Ce cercle contient au moins 95% des points. Ce seuil a été retenu par analogie à la précision des GPS (Chin, 1987 ; William J., 2013).

Pour filtrer les micros-arrêts, un seuil temporaire est aussi défini. Seuls les arrêts détectés à l'aide du frein à main et durant plus d'un certain temps sont considérés comme des STOPS. Ainsi, les micro-arrêts comme les arrêts à un carrefour, feux rouges, démarrage en côte ne sont pas retenus. Ce temps est un paramètre à définir suivant l'activité du type de véhicules étudiés. Pour des véhicules ayant des activités courtes, il faudrait aussi prendre en compte le contexte (présence de feux de circulation, etc.). Par exemple, un véhicule de distribution du courrier réalise de nombreux arrêts relativement courts, l'utilisation exclusive de ce seuil est donc insuffisante. Dans ce cas, des valeurs supplémentaires issues de capteurs devraient être aussi employées. Par exemple, le capteur d'ouverture du coffre où se trouve le courrier est pertinent pour détecter les STOPS des véhicules de distribution.

### 2.1.3. Détection des MOVEs

Un MOVE est le déplacement entre deux STOPS. Les positions d'un objet mobile entre deux STOPS sont donc regroupées et ordonnées temporellement pour former un MOVE. Dans un environnement complexe, il peut être nécessaire de filtrer les données aberrantes liées à une mauvaise réception temporaire de signal.

### 2.1.4. Segmentation des items spatio-temporels

Les items spatio-temporels (STOPS et MOVEs) sont associés à une seule activité. Si l'activité change, il faut segmenter cet item en items distincts. À l'aide de données supplémentaires fournies par divers capteurs, il est possible de segmenter les STOPS. Par exemple, à l'aide du capteur de rechargement électrique du véhicule, on peut différencier deux activités pour un STOP : en charge, puis en attente lorsque la batterie est chargée et la prise débranchée. Deux STOPS sont alors distingués, STOP en charge, puis STOP en attente.

De même, un véhicule peut changer d'activité pendant un MOVE. Dans le cas d'un véhicule d'urgence, celui-ci peut être appelé pour répondre à une intervention alors qu'il rentrait à la caserne. Ce changement est appelé re-routage. Le déplacement est alors segmenté en deux MOVEs : un retour caserne et un départ en intervention. Le point de re-routage est détecté par des changements importants et durables des valeurs des capteurs (augmentation de la vitesse et passage à une vitesse supérieure à la limite légale, allumage des gyrophares).

## 2.2. Association des STOPS à des lieux d'arrêts fréquents

Une fois les STOPS localisés, nous pouvons les associer à des lieux d'arrêt fréquents. Pour trouver les arrêts fréquents des véhicules, un clustering des STOPS est effectué en fonction de leur position géographique.

L'algorithme de clustering DB SCAN (Ester et al., 1996) est retenu. Il utilise 2 paramètres en entrée : une distance  $\epsilon$  ainsi et un nombre minimum de points MinPts. Pour qu'un point soit considéré comme un centre de cluster (ou regroupement), appelé point centre, il faut que le point ait dans son voisinage de rayon  $\epsilon$  un nombre minimum de points MinPts (par exemple le point A de la figure 5).

Pour qu'un point soit inclus dans un cluster, il doit être un centre de cluster, soit être couvert par un centre de cluster de rayon  $\epsilon$  (points B de la figure 5). Les centres de clusters positionnés à une distance inférieure à  $\epsilon$  sont regroupés pour ne former qu'un seul cluster (les points rouges de la figure 5 sont tous des centres de cluster qui forment un ensemble plus grand). Les points ne remplissant aucune des deux conditions sont considérés comme des points isolés (Point M et N de la figure 5).

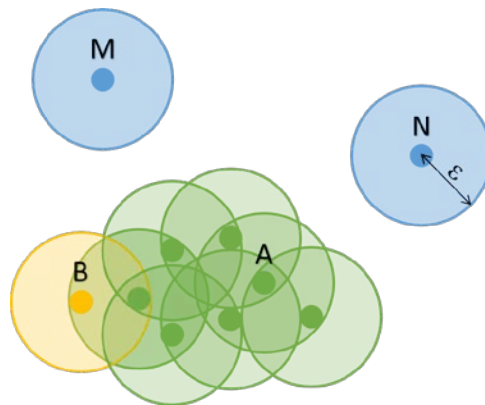


FIGURE 5. Représentation du fonctionnement typique de DBSCAN

La méthode DBSCAN a été choisie pour deux de ses propriétés. Elle ne nécessite pas de connaître le nombre attendu de clusters. Elle n'oblige pas à associer à un cluster les valeurs isolées. Cependant, elle nécessite de définir manuellement les valeurs  $\epsilon$  et MinPts.

## 2.3. Association des STOPS à des POIs

De même, les STOPS peuvent être associés à des centres d'intérêt (POI). Cette phase consiste à identifier le lieu d'intérêt à l'aide de connaissances externes. Des sources ouvertes telles qu'OpenStreetMap (OSM) ou professionnelles (BD TOPO @ IGN) sont employées à cet effet. Pour les sources ouvertes, il faut définir un ensemble de lieux potentiellement intéressant. Les objets géographiques d'OSM sont caractérisés par des couples clé et valeur. Le nom de la clé étant la catégorie de l'objet.

Dans notre cas, la clé amenity (<http://wiki.openstreetmap.org/wiki/Key:amenity>), qui regroupe différents types d'installation communautaire (Toilettes, téléphones, bâtiments...) est la plus utilisée. Les clés sont associées à des valeurs qui correspondent à des sous-catégories. Par exemple, la valeur taxi de la clé amenity identifie une station de taxis. Pour OSM, un ensemble d'objets géographiques pertinents peuvent être ainsi extraits et localisés à l'aide d'une liste de couples clé-valeur prédéfinis. Des POI externes sont ainsi extraits et localisés.

Un processus d'appariement est alors lancé afin d'associer les STOPS à des POI externes. (Beeri et al., 2004) décrit un grand nombre de processus d'appariement de points issus de deux jeux de données différents. Actuellement, un processus simple est employé. Il associe chaque STOP à un POI. Un seuil maximum de distance  $D_{max}$  est défini. Si le POI le plus proche est à une distance supérieure à ce seuil, alors le lieu n'est pas apparié à un POI. La figure 6 représente le résultat de l'appariement entre des STOPS (cercles bleus) et des POI externes (étoiles rouges). Les traits noirs définissent les appariements.

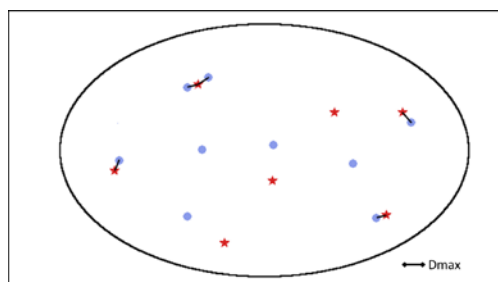


FIGURE 6. Appariement des STOPS avec des POIs

#### 2.4. Association des items spatio-temporels et des activités

L'étape suivante du processus consiste à associer une activité à un item spatio-temporel :

- Activité liée au déplacement pour un MOVE,
- Activité liée à l'arrêt pour un STOP.

Ce processus est un processus récursif, en effet, pour certaines activités associées à des items spatio-temporels il faut connaître l'activité de l'item précédent et/ou suivant. L'association commence par les STOPS, car la détection des activités des MOVES utilise toujours la sémantique des STOPS précédents et/ou suivant celui-ci. Si au moins un item spatio-temporel a été nouvellement associée, le processus reprend pour associer les items non définis. La figure 7 illustre ce processus.

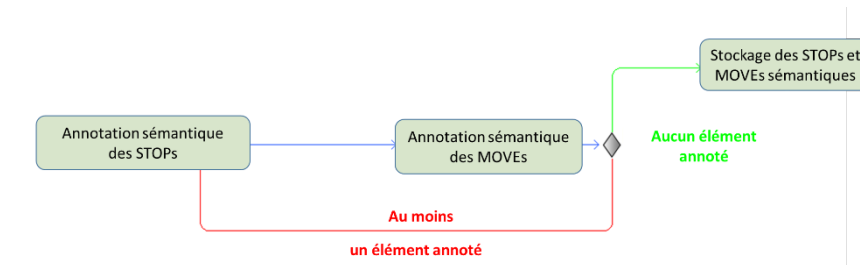


FIGURE 7. Cycle d'association des items spatio-temporels et des activités

#### 2.4.1. Association des STOPS

Pour associer une activité à un STOP, un ensemble de règles d'inférence métiers est défini comme le propose Browne (2009). À savoir, quand les conditions sont réunies, alors un cas précis est appliqué. Le modèle utilisé ne contient pas de règles contradictoires.

Un exemple de règle métier est le suivant : si un véhicule effectue un arrêt dans une station-service, et qu'entre le début de l'arrêt et sa fin, le capteur de la jauge de carburant montre une augmentation du volume de carburant, alors ce STOP est une activité logistique. De même, si un STOP est effectué dans un POI de type hôpital, si durant cet arrêt, un brancard est déchargé, et si les portes arrières sont ouvertes, alors ce STOP est une prise en charge du patient à l'hôpital.

#### 2.4.2. Association des MOVEs

Dans la plupart des représentations sémantiques de trajectoires, les activités sont associées uniquement aux STOPS. Pour les déplacements, la seule information sémantique est le mode de déplacement. Par exemple, la figure 8 décrit sémantiquement la trajectoire d'un touriste à l'aide du modèle de Parent et al. (2013). Pourtant, pour des véhicules spécifiques, l'objectif du déplacement est une information riche sémantiquement qu'il faut identifier.

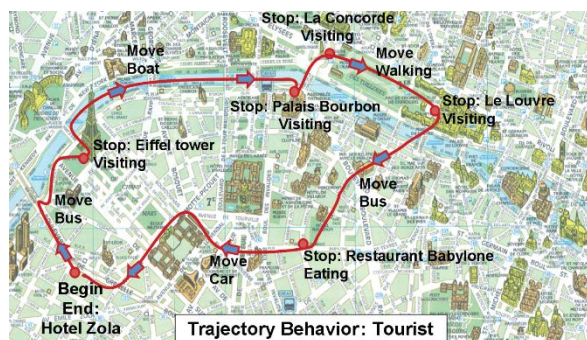


FIGURE 8. Trajectoire sémantique d'un touriste

Il est possible de monter en sémantique sur des trajectoires à l'aide de connaissances (ontologies, contexte géographique,) et d'un mécanisme de raisonnement. Par exemple, des règles, des ontologies et un contexte permettent la détection de comportements remarquables (Nogueira , 2017 ; Soltan et al., 2017 ; Klotz et al. 2017).

En effet, à partir de descriptions des lieux de départ et d'arrivée et des valeurs des capteurs, il est souvent possible de déduire l'objectif de chaque déplacement. Par exemple, si un MOVE part d'un POI de type A et arrive à un POI de type B et que durant ce déplacement les capteurs renvoient des valeurs significatives (Compteur en marche sur un taxi, gyrophare allumé sur une ambulance...) alors nous pouvons en déduire que l'objectif du déplacement est connu.

La difficulté de l'utilisation d'un moteur d'inférence classique est l'acquisition de ces règles. En effet, le cas des MOVEs est plus complexe que pour les STOPs. Il est difficile pour un expert de lister l'ensemble des règles pour définir l'activité d'un MOVE. A ce stade nous avons utilisé le raisonnement à base de cas (Bergmann, 2003 ; Kolodner, 2014). L'utilisation de ce mode de raisonnement permet de définir plus simplement des cas prototypiques. Un expert a plus de facilités à lister des cas prototypique qu'un ensemble de règles complexes. De plus, l'ajout de règles dans un système de raisonnement à base de cas ne nécessite aucune vérification. Au contraire, lors de l'ajout d'une règle dans la base de connaissances d'un moteur d'inférence classique, il est nécessaire de vérifier l'absence de conflits ou de redondances.

Un cas prototypique métier décrit les conditions qui caractérisent une activité. Il se rapproche d'un cas réel. Par exemple, un expert a décrit l'activité « départ vers intervention » comme « le départ d'un véhicule de secours de caserne avec les gyrophares allumés jusqu'à un lieu quelconque ».

Pour ajouter une activité à chaque MOVE, le cas réel est comparé aux cas prototypiques. Le cas prototypique le plus proche est retenu. Par exemple, le cas d'un véhicule de secours quitte sa caserne pour rejoindre un lieu quelconque en allumant les gyrophares qu'une minute après le départ ne correspond pas exactement à un cas prototypique. Cependant, le cas le plus proche infère que ce déplacement est un départ vers intervention.

Qui plus est, les valeurs d'un grand nombre de capteurs, couplés à un raisonnement à base de cas autorisent l'extraction d'activités plus précises. Dans le cas d'un transport de patient, si la vitesse est faible et les données des accéléromètres indiquent une conduite très souple alors l'activité plus précise de cette trajectoire est : Transport patient à risque où le confort de déplacement prime sur la vitesse vers un hôpital.

### **3. Expérimentation**

Ce processus d'extraction a été expérimenté à l'aide des données issues d'ambulances connectées de pompiers. Les véhicules envoient chaque seconde leur position GPS, les valeurs maximales issues d'un gyroscope 3 axes, des valeurs

capteurs spécifiques aux ambulances (Sirènes, gyrophares, taux de remplissage des bouteilles d'oxygène...) ainsi que 30 valeurs issues du bus CAN (Controller Area Network) du véhicule : Consommation du véhicule, frein à main, accélérateur, etc.

Notre étude préalable se base sur 30 jours de données issues des 3 premières ambulances connectées (Véhicule de Secours et d'Assistance aux Victimes - VSAV) de la caserne des Sapeurs-Pompiers de Tours-centre (37). Ce jeu de données regroupe un peu moins de 2 millions de lignes de données (50 attributs) par mois par véhicule, soit plus de 5 millions au total. La figure 9 représente l'étendue de ces données sur la ville de Tours.

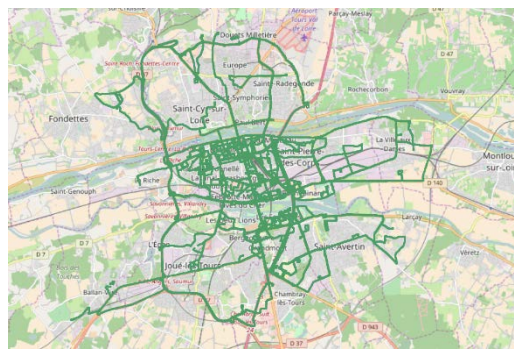


FIGURE 9. Représentation des données collectées

Le système est programmé pour qu'il envoie les données toutes les secondes avec une mise en mémoire tampon si le réseau n'est pas disponible. Le seul cas où le transfert est interrompu est lorsque le véhicule est arrêté (contact coupé) pendant plus de 30 minutes dans une caserne. Le système ré-émettra lors du redémarrage du véhicule.

Chaque ligne est composée des informations suivantes :

- |                           |                                   |
|---------------------------|-----------------------------------|
| Capteurs classiques       | – Voltage châssis                 |
| – Longitude               | – Vitesse de rotation du moteur   |
| – Latitude                | – Témoin problème moteur          |
| – Accéléromètre longitude | – Accélérateur                    |
| – Accéléromètre latitude  | – Odomètre                        |
| – Accéléromètre vertical  | – Contact moteur                  |
| Capteurs véhicule (CAN)   | – Consommation carburant          |
| – Vitesse                 | – Porte conducteur                |
| – Clignotant droit        | – Porte passager                  |
| – Clignotant gauche       | – Porte arrière                   |
| – Pédale de frein         | – Frein à main                    |
| – Veilleuse               | – Ceinture de sécurité conducteur |
| – Feux de croisement      | – Airbag                          |
| – Feux de route           | Capteurs spécifiques ambulance    |
| – Marche arrière          | – Feux de pénétration             |

- Gyrophares
- Sirène
- Chauffage cellule sanitaire
- Voltage cellule sanitaire

Le traitement de ces données a été réalisé en mode hors ligne. C'est-à-dire que les données ont été collectées directement sur les ambulances vers un fichier CSV, puis importées dans une base de données PostGIS indexée pour améliorer la vitesse de traitement des requêtes et des algorithmes.

### 3.1. Extraction des STOPS et des MOVES

La première étape du processus est l'extraction des STOPS et des MOVES. Pour le filtrage des micro-arrêts, la limite de temps a été définie à 60 secondes. En effet, l'activité la plus courte d'une ambulance selon les experts est une activité logistique (faire le plein du véhicule) durant au minimum une minute. Les véhicules de secours étant lourds, il est très fréquent que ceux-ci effectuent des arrêts dans la circulation (feux rouges, stops, ...) avec le frein à main, lorsque la pente est trop forte.

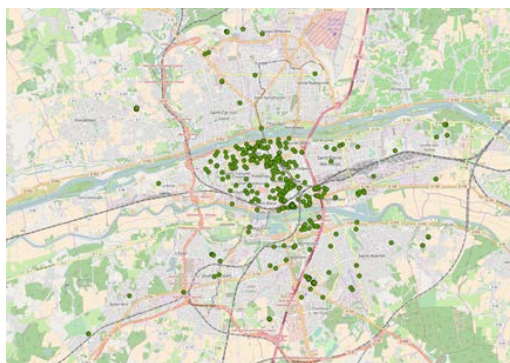


FIGURE 10. Représentation des arrêts sur une carte OSM

La détection a identifié 807 arrêts dont 140 micro-arrêts filtrés, soit 667 STOPS générés. La figure 10 représente la répartition des arrêts.

Sur les 667 STOPS détectés, 20 n'ont pu être localisés à cause d'une perte temporaire du signal GPS. Les 647 STOPS localisés peuvent se classer en deux catégories. La première catégorie, avec un rayon  $R_{med}$  compris entre 0 et 5 mètres contient 639 STOPS et traduit une bonne précision spatiale de ceux-ci. La deuxième catégorie avec un rayon  $R_{med}$  supérieur à 5 mètres qui contient 8 STOPS traduit une mauvaise réception du signal à l'endroit où le véhicule a effectué son STOP.

Les STOPS permettent ensuite d'extraire les MOVES entre ceux-ci. Ce processus a extrait 666 MOVES.

### 3.1. Extraction des lieux d'arrêts fréquents

Afin d'identifier des lieux d'arrêts fréquents, le clustering de type DBSCAN a été utilisé. Dans notre étude, la valeur d' $\epsilon$  correspond à la valeur de l'imprécision du GPS

(5 mètres). Pour un mois de données, les experts métiers considèrent que 3 arrêts distincts au même endroit constituent un lieu d'arrêt fréquent. Aussi, le nombre MinPts a été fixé à 3.

Nous obtenons comme résultat l'association de 388 STOPS, répartis dans 13 zones différentes.

La première zone regroupe 199 STOPS. Elle est positionnée tout près de la deuxième qui en compte 14. Elles sont visualisées sur la figure 11(a). Le résultat montre qu'il est possible d'identifier deux zones d'arrêts fréquent (Ici deux lieux de stationnement habituel du véhicule) pour un même lieu d'intérêt (Cour avant et arrière de la caserne). Dans la figure 11 (b), la zone regroupant 13 STOPS est affichée, ces derniers sont associés à un POI station-service.

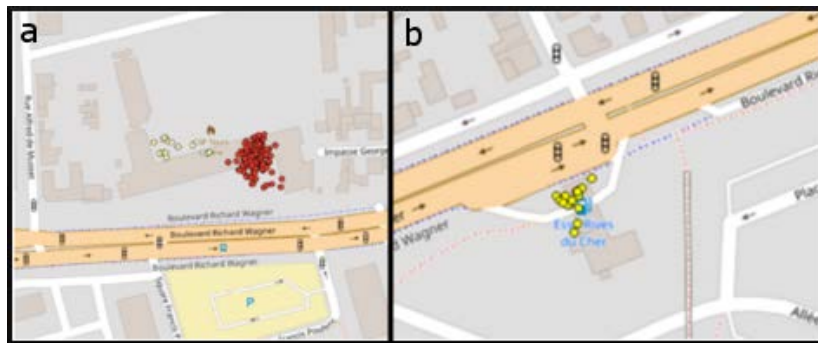


FIGURE 11. Zone n°1 regroupant les arrêts rouges(a), la zone n°2 en jaune clair(a) et la n°3 en jaune (b)

279 STOPS (plus de 40%) ne sont pas associés à un lieu d'arrêt fréquent par l'algorithme de clustering. Ce nombre important est logique. Une ambulance effectue généralement des interventions sur des lieux d'incident très divers. De plus, le transport du patient (et donc l'arrêt à l'hôpital) n'est obligatoire que si son état le nécessite. Pour des raisons d'anonymat, les données traitant des interventions (zone d'intervention fréquente ou répartition géographique) ne seront pas illustrées dans cet article.

### 3.2. Appariement des STOPS et des lieux d'intérêts

Afin d'associer les STOPS à des lieux d'intérêts, la méthode d'appariement présentée dans la section précédente est utilisée.

Préalablement, des POIs externes sont extraits d'OSM via un outil nommé Overpass turbo, disponible depuis [<https://overpass-turbo.eu/>] suivant un ensemble de couples clé valeur comme expliquée dans la partie 2.2.2. La figure 12 représente sur une carte les données extraites de l'outil, la recherche a porté sur la clé "amenity" avec pour valeur "fire\_station". Les points rouges sont les casernes de pompiers.



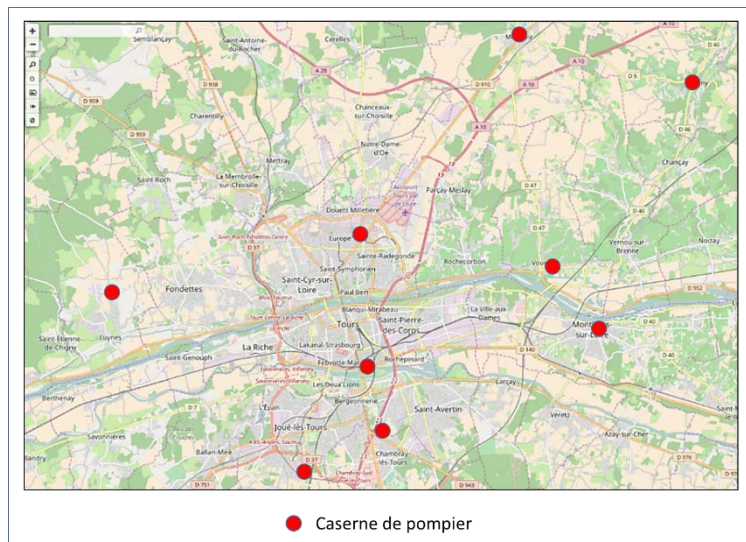


FIGURE 12. Représentation visuelle de l'extraction de données pour les casernes de pompiers issues de <https://overpass-turbo.eu/>

La figure 13 partie (a) représente un exemple de résultat de l'appariement de STOPS de la figure 11 (a). Ils sont appariés au POI le plus proche issu d'OSM (polygone jaune avec une bordure bleu) qui correspond à la Caserne des Sapeurs-Pompiers de Tours Centre. La partie (b) représente l'appariement de STOPS au POI le plus proche (polygone bleu) qui correspond à la station-service "Esso Rives du Cher". Le couple amenity="fuel" a permis l'extraction de ce POI.

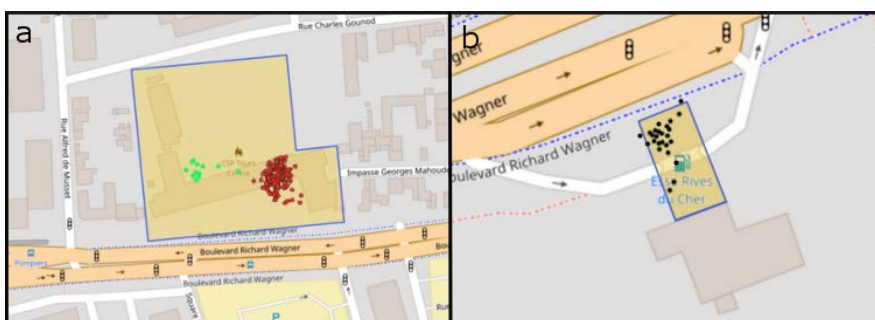


FIGURE 13. Représentation de l'appariement de STOPS et d'une caserne de pompiers (a) puis avec une station-service (b)

Néanmoins, il n'est pas possible d'associer tous les STOPS à des POIs. En effet, les lieux d'intervention se trouvent parfois dans les lieux quelconques, sans proximité avec un POI. Sur notre jeu de données, 337 STOPS ont été liés à des POIs.

### 3.3. Affectation des activités

#### 3.1.1. STOPS

Actuellement, seulement 4 activités sont identifiées pour les STOPS. Des règles métiers simples sont utilisées. Dans ces règles, des seuils sont définis pour prendre en considération l'erreur humaine.

Par exemple pour le gyrophare, pour supprimer les appuis involontaires sur un bouton, le seuil de 10% de marge d'erreur a été retenu. Les règles suivantes ont été obtenues :

- Si STOP avec les gyrophares allumés sur plus de 90% de la durée et POI différent d'un hôpital alors l'activité est : Prise en charge du patient sur lieu d'intervention
- Si STOP avec les gyrophares étaient allumés sur le MOVE précédent durant plus de 90% de la durée, et POI de type hôpital alors l'activité est : Prise en charge du patient à l'hôpital.
- Si STOP avec les gyrophares allumés moins de 10% du MOVE précédent et allumés moins de 10% durant la durée du STOP et POI de type logistique alors l'activité est : logistique.
- Si STOP associé à un POI de type caserne alors l'activité est : Attente caserne.

Les résultats suivant sont obtenus :

- 279 STOPS de type « Prise en charge du patient sur lieu d'intervention »
- 111 STOPS de type « Prise en charge du patient à l'hôpital »
- 13 STOPS de type « logistique ».
- 213 STOPS de type « Attente caserne ».
- 61 STOPS n'ont pu être liés à des activités.

#### 3.1.2. MOVES

Des cas prototypiques métiers simples ont été définis afin d'identifier les activités liées aux MOVES. La méthode utilisée permet d'ajouter par la suite des cas prototypiques afin d'affiner la détection des activités. Le tableau 1 présente la liste des cas utilisés. Le signe \* signifie que toutes les valeurs sont possibles pour cet attribut.

Type	Préc. POI	Activité Préc.	Déplacement	Suiv. POI	Activité suivante
<b>Transport vers hôpital</b>	*	Arrêt sur intervention	Gyrophares 100%	Type Hôpital	Prise en charge
<b>Vers intervention</b>	Type caserne	Attente caserne	Gyrophares 100%	*	Arrêt sur intervention
<b>Trajet logistique</b>	*	*	0% gyrophares	Type logistique	Activité logistique
<b>Retour caserne</b>	*	*	0% gyrophares	Type caserne	Attente caserne
<b>Re-routage vers intervention</b>	∅	*	100% gyrophares	*	Arrêt sur intervention
<b>Re-routage annulation</b>	∅	*	0% gyrophares	Type caserne	Attente caserne

TABLEAU 1. Liste des cas prototypiques

La première ligne de ce tableau décrit le cas prototypique d'un transport vers l'hôpital. Pour ce cas, le POI précédent n'est pas pris en compte. L'activité précédente est « Arrêt sur intervention », l'utilisation des gyrophares tout au long du MOVE, le POI suivant est de type « Hôpital » et l'activité suivante est « prise en charge du patient ». Actuellement, une méthode de similarité simple est utilisée. Si au moins 3 éléments élément du cas réel (POI précédent, Activité précédente, ...) correspondent avec un cas prototypique, alors le cas réel est assigné à ce cas prototypique. Pour le capteur de gyrophare, la valeur la plus proche existante dans un cas prototypique est retenue pour décider de la correspondance (Si moins de 50% d'utilisation pendant le trajet, l'item est compté comme correspondant à 0%, sinon à 100%). Si plusieurs cas prototypiques correspondent, le cas avec le plus de correspondances est choisi.

Les cas prototypiques spécifiques au re-routage (ligne 5 et 6 du tableau) sont associés à des MOVEs précédés d'un autre MOVE (voir partie 2.1.4). Il n'y a donc pas de POI associé à l'item précédent. L'activité précédente n'est pas prise en considération.

La figure 14 présente les MOVEs associés à des « Retour caserne ». Les autres trajectoires ne sont pas représentées, car liées à des lieux d'intervention. On remarque que les chemins empruntés pour les retours des hôpitaux sont quasiment identiques. Quelques trajets inhabituels sont également identifiés. Ils sont représentés sur la figure par des lignes en tirets sur la carte (axe principal coupé, détour volontaire ...)

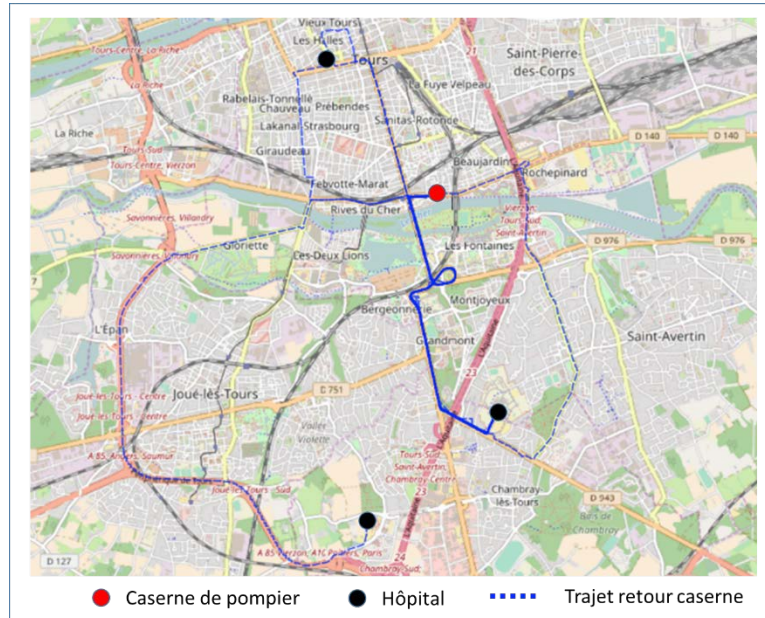


FIGURE 14. Représentation des MOVES liés à une activité « retour à la caserne »

Les résultats suivants sont obtenus :

- 78 Transports vers hôpital.
- 167 Départs vers intervention.
- 13 Trajets logistique
- 218 Retours caserne.
- 190 MOVES n'ont pas été liés à une activité, car ils n'étaient pas assez proches de cas prototypes. Ces MOVES sont pour la plupart des déplacements courts (changement de place dans le sas des urgences ou à la caserne, prise de renseignements pour trouver une adresse ...) et certains ne sont pas expliqués (MOVE partant et revenant à la caserne sans STOP).

La validation de ces premiers résultats est un processus complexe. En effet, il n'existe pas à l'heure actuelle de vérité terrain. Les pompiers disposent uniquement d'une liste des interventions. Ces données sont collectées à chaque appel déclenchant une intervention, puis enrichies sur la base des informations recueillies par l'opérateur du centre de secours (adresse géocodée, nature de l'intervention) et sur déclarations des pompiers (Estampilles temporelles). Qui plus est, des différences conceptuelles existent. Par exemple, la fin d'une mission, dans notre méthode, est le retour du véhicule à la caserne. Pour les pompiers, c'est lorsque le véhicule est à nouveau opérationnel (Nettoyage, réarmement). De même, la sémantique des interventions des

pompiers est plus concise. Quatre estampilles temporelles sont uniquement présentes dans cette liste (heure de départ caserne, heure d'arrivée sur le lieu d'intervention, heure de départ du lieu d'intervention et heure de retour caserne). Ce processus déclaratif entièrement manuel souffre donc d'inexactitudes. Néanmoins, une comparaison de notre extraction et de cette liste a été effectuée. Pour des raisons éthiques évidentes, seul un pompier est habilité à manipuler la liste des interventions. C'est donc un responsable du SDIS 37 qui a analysé les interventions détectées par notre processus pour deux des quatre véhicules de la caserne des pompiers de Tours centre.

Les résultats de ces analyses indiquent de bonnes détections pour une partie des interventions (plus de 40%). Une détection par le système est considérée comme correcte si les estampilles temporelles déclaratives et détectées (heure de départ caserne, heure d'arrivée sur le lieu d'intervention, heure de départ du lieu d'intervention et heure de retour caserne) se trouvent dans un intervalle de plus ou moins cinq minutes. Pour les 60% restants, une étude complémentaire doit être effectuée. Son objectif est double. D'une part, il faut mesurer la qualité des informations déclarées (inexactitude, incomplétude). D'autre part, une analyse des 17 % d'interventions extraites sans correspondance avec les interventions déclarées doit être réalisée afin de déterminer les fausses interventions extraites ou le cumul d'erreurs déclaratives. En effet, plusieurs erreurs présentes sur une même déclaration entraînent un non appariement avec une intervention extraite automatiquement. De même, la non détection d'un STOP à l'aide du frein à main entraîne une mauvaise extraction.

#### 4. Conclusion

Dans cet article une extension de modèle de trajectoires sémantiques classiques a été présentée. Elle est adaptée à des véhicules spécifiques disposant de nombreuses sources de données (capteurs, sources extérieures). Un processus d'extraction automatique a également été défini. Ce dernier s'appuie sur gros volumes de traces brutes GPS et de valeurs issues d'autres capteurs de véhicules spécifiques. Il détecte les arrêts (STOPS) et les déplacements (MOVEs) et les enrichit sémantiquement.

Lors de ce processus, les STOPS sont associés à des POIs, des lieux d'arrêts fréquents et des activités spécifiques à un type de véhicule. De même, les MOVEs sont associés à des activités spécifiques. Les résultats de la première expérimentation sont mitigés. Néanmoins, ils montrent l'intérêt d'extraire automatiquement les interventions pour éviter les erreurs déclaratives. De plus, l'utilisation du capteur GPS et du frein à main s'avère, dans certains cas, insuffisante pour extraire de manière précise les STOPS quand le frein à main n'a pas été serré. Il est donc nécessaire d'intégrer les valeurs d'un plus grand nombre de capteurs. Cette intégration devrait potentiellement permettre aussi de faire abstraction d'une durée minimale arbitraire d'un arrêt.

La solution proposée est suffisamment générique pour être employée pour d'autres types de véhicules (Taxi, Ramassage d'ordures...) ou d'autres zones d'activités. Cette

méthode démontre l'intérêt de disposer de multiples capteurs afin de compléter ou de corriger les informations issues d'un capteur GPS. Ce processus s'affranchit ainsi d'un grand nombre de problèmes liés aux données GPS classiques, notamment le bruit, l'imprécision ou les pertes de signal.

Qui plus est, cette extraction de connaissances permet la définition de nouvelles variables pour l'analyse de données. Par exemple, il est ainsi envisageable de déterminer automatiquement le temps moyen d'immobilisation d'un véhicule pour chaque hôpital. De même, elles pourront améliorer l'estimation du temps d'arrivée d'un véhicule de secours sur le lieu d'un accident. Plus globalement, elle facilite la visualisation et l'analyse des déplacements d'objets mobiles spécifiques à l'échelle d'une ville ou d'un département (point d'intérêts des véhicules, voies de circulation préférées suivant l'activité pratiquée par le véhicule, etc.) Cependant, le caractère sensible de ces données liées à la santé nous oblige à intégrer la dimension éthique lors de l'extraction et l'exploitation de ces connaissances.

À moyen terme, de nouvelles règles métiers seront ajoutées pour définir l'activité des STOPS. De même, de nouveaux cas prototypiques seront introduits afin de détecter des activités plus fines (Transport délicat ou rapide d'un patient à l'hôpital, ...). Ces nouveaux cas devront intégrer les activités très courtes des MOVEs comme les déplacements sur un même parking. L'algorithme de segmentation des MOVEs devra également être validé pour détecter les cas de re-routage. Enfin, ce processus sera testé avec un jeu de données issu d'autres types de véhicules par exemple des taxis.

À plus long terme, l'objectif est de développer des systèmes de guidage GPS adaptés à des métiers et des véhicules spécifiques. Ils proposeront des itinéraires optimisés en fonction des contraintes du véhicule et de l'objectif du déplacement.

### **Remerciements**

Ce travail a été financé et soutenu par Petit Picot by GRUAU, ainsi que l'ANRT (Association Nationale de la Recherche et de la Technologie). Nous remercions le SDIS 37 pour leur participation dans la validation des résultats et la définition des cas prototypiques.

### **Bibliographie et références**

- Andrienko, G., Andrienko, N., Bak, P., Keim, D., Kisilevich, S., & Wrobel, S. (2011). A conceptual framework and taxonomy of techniques for analyzing movement. *Journal of Visual Languages & Computing*, vol. 22, n° 3, p. 213-232.
- Beber, M. A., Ferrero, C. A., Fileto, R., & Bogorny, V. (2017) Individual and Group Activity Recognition in Moving Object Trajectories. *Journal of Information and Data Management*, vol. 8, n° 1, p. 50-66.

- Beeri, C., Kanza, Y., Safra, E., & Sagiv, Y. (2004) Object fusion in geographic information systems. In *Proceedings of the Thirtieth international conference on Very large databases*-vol. 30, p. 816-827.
- Bergmann, R., Althoff, K. D., Breen, S., Göker, M., Manago, M., Traphöner, R., & Wess, S. (2003). Case-based reasoning approaches. In *Developing Industrial Case-Based Reasoning Applications* p. 21-34. Springer, Berlin, Heidelberg.
- Bogorny, V., Renso, C., Aquino, A. R., Lucca Siqueira, F., & Alvares, L. O. (2014). Constant—a conceptual data model for semantic trajectories of moving objects. *Transactions in GIS*, vol. 18, n° 1, p. 66-88.
- Browne, P. *JBoss Drools business rules*. (2009). Packt Publishing Ltd.
- Cain III J.W., Krausman P.R., Jansen B. D., and Morgart J.R. (2005). Influence of topography and GPS fix interval on GPS collar performance. *Wildlife Society Bulletin* vol. 33, n° 3, p. 926-934
- Chin, G.Y. (1987). *Two-dimensional measures of accuracy in navigational systems*. Report DOT-TSCSPA-87-1. Cambridge, MA: U.S.Department of Transportation.
- Duran, A., Earleywine, M. (2012). GPS Data Filtration Method for Drive Cycle Analysis Application. In *SAE 2012 World Congress*, Detroit, Michigan
- Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Kdd*, vol. 96, n° 34, p. 226-231.
- Etienne, L., Devogele, T., & Bouju, A. (2012). Spatio-temporal trajectory analysis of mobile objects following the same itinerary. *Advances in Geo-Spatial Information Science*, vol. 10, p. 47-57.
- Ferrero, C. A., Alvares, L. O., & Bogorny, V. (2016). Multiple aspect trajectory data analysis: research challenges and opportunities. *GeoInfo*, p. 56-67.
- Hägerstrand, T. (1970). What about people in regional science?. In *Papers of the Regional Science Association*, vol. 24, n° 1, p. 6-21. Springer-Verlag.
- Ivanović, S. S., Raimond, A. M. O., Mustière, S., & Devogele, T. (2016) Detection of outliers in crowdsourced GPS trace, *12th international symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*
- Klotz, B., Troncy, R., Wilms, D., & Bonnet, C. (2018,). Generating Semantic Trajectories Using a Car Signal Ontology. *Companion of the The Web Conference 2018 on The Web Conference 2018*, p. 135-138.
- Knight et al. (2009). A comparison of outlier detection procedures and robust estimation methods in GPS positioning. *J. Geodesy*, vol. 62, n° 4, p. 699-709.
- Kolodner, J. (2014). *Case-based reasoning*. Morgan Kaufmann.
- Nogueira, T. P. (2017). *A Framework for Automatic Annotation of Semantic Trajectories* Thèse de doctorat, Université Grenoble Alpes.
- Ordoñez, C., Martínez, J., Rodríguez-Pérez, J., and Reyes, A. (2011). Detection of Outliers in GPS Measurements by Using Functional-Data Analysis. *Journal of Surveying Engineering*137, p. 150-155.

- Parent, C., Spaccapietra, S., Renso, C., Andrienko, G., Andrienko, N., Bogorny, V., Theodoridis, Y. (2013). Semantic trajectories modeling and analysis. *ACM Computing Surveys (CSUR)*, vol. 45, n° 4, p. 42.
- Small, C.A.: (1990) Survey of Multidimensional Medians. *International Statistical Review*, p. 263–277
- Soltan Mohammadi, M., Mougnot, I., Libourel, T., & Fagot, C. (2017). A Semantic Modeling of Moving Objects Data to Detect the Remarkable Behavior. In *AGILE 2017*.
- Weiszfeld, E., & Plastria, F. (2009). On the point for which the sum of the distances to n given points is minimum. *Annals of Operations Research*, vol. 167 n° 1, p. 7-41.
- William, J. (2007). Global positioning system (GPS) standard positioning service (SPS) performance analysis report. *Federal Aviation Administration, Washington*,
- Yan, Z., Chakraborty, D., Parent, C., Spaccapietra, S., & Aberer, K. (2011). SeMiTri: a framework for semantic annotation of heterogeneous trajectories. In *Proceedings of the 14th international conference on extending database technology*. p. 259-270. ACM.
- Zheng, Y. (2015). Trajectory data mining: an overview. *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 6, n° 3, p. 29.