

# Place names in Spanish Republican Life Stories: spatial patterns in locations and perceptions

Laurence Jolivet<sup>a, \*</sup>, Carmen Brando<sup>b</sup>, Catherine Domingues<sup>a</sup>

<sup>a</sup> LASTIG, Univ Gustave Eiffel, ENSG, IGN, F-94160 Saint-Mandé, France, [laurence.jolivet@ign.fr](mailto:laurence.jolivet@ign.fr); [catherine.domingues@ign.fr](mailto:catherine.domingues@ign.fr)

<sup>b</sup> École des hautes études en sciences sociales (EHESS), [carmen.brando@ehess.fr](mailto:carmen.brando@ehess.fr)

\* Corresponding author

**Abstract:** Places can be defined by different features including geographical coordinates, landscape elements and also subjective information like perception. The corpus collected by the *Réseau des acteurs de l'histoire et la mémoire de l'immigration* (RAHMI) contains oral interviews about Spanish Republicans journeys during Spanish Civil War and World War II. These life stories quote place names corresponding to proper name (Npr) or common name (Nc) and which can be qualified or associated to non-geographical nouns, adjectives and expressions. Methods from natural language processing have been used to automatically extract and characterize these pieces of informations into annotations. The objective of the presented work is to analyse the geographical aspects of places and associated features, and to propose cartographical displays illustrating life stories and supporting analysis. Distribution of quoted Npr place names was first studied. It enabled targeting relevant spatial scales and so mapping extents for the corpus. Second, features of Npr places about their type (administrative, topographic) and their spatial target based on associations with Nc have been synthetized. Then perception were allocated to Npr, falling into the three values of polarity positive, negative or neutral. Spatial patterns of each value have been displayed and interpreted. Summary of the annotations, analysis and created maps offer insights into the corpus driven by the mentioned places. Based on individual testimonies, it may help further interpretation and researches about this historical period.

**Keywords:** life story, place name, mapping, spatial analysis, local Moran's index

## 1. Introduction

Place names may be addressed as milestones in life stories. Knowing their locations enables highlighting the geography of individual journeys, i.e. how and potentially why the space was used and travelled. A single location is interesting to consider in the regional environment. A set of locations is interesting to address according to its spatial distribution (relative distances, global and local densities). A place is a location including coordinates and an extent, but not only (Purves et al., 2019). It is characterized by physical features (landscape, people) (Tuan, 1977), by its configuration and characteristics (building materials, colours), by a temporality (date when visited) (Huyghe, 2012), weather, people, sounds. A place is thus defined with a variety of features and can be considered as a multi-faceted object. It also corresponds to individual perceptions including opinions, emotions and sentiments (Bocagni et al., 2015; Zeile et al., 2015). Though in life stories, all places are not designated thanks to similar types of name and do not have same expressed and so available description.

Natural language processing (NLP) aims at defining and implementing tools so that to automatically annotate numeric written texts. An annotation process is generally modelled as an annotation schema that correspond to semantic or syntactic features (attributes or properties) added to a scope, i.e. start and end of the annotated expression, and rules to specify the features. In addition, manual annotations (features and scope) are made and compared with the automatic annotations in order to

assess the automatic process. Commonly the computed indicators for assessment are recall, precision and F-measure (or F-score) ratios. Precision (P) is equal to the number of true positives divided by the sum of true positives and false positives. It indicates the percentage of relevant annotations regarding the effective annotations. Recall (R) is equal to the number of true positives divided by the sum of true positives and false negatives. It indicates the percentage of relevant annotations regarding what should have been annotated. F-measure is a synthetic value combining P and R.

In linguistics, some researches are dedicated on place names as a category of proper names (Npr) (Leroy, 2004). Places can be designated by proper names only (Npr), e.g. *Catalogne* (Catalonia), by common names (Nc), e.g. *le front* (the front), possibly including Npr, e.g. *le front de Catalogne* (the front of Catalonia). Places can be distinguished by specific information which are added to the place names (Npr or Nc) and modify them. The additional information can be achieved by an adjective like in the modified Npr *Paris étendu* (extended Paris) or a prepositional phrase like in the modified Nc *le quartier de Barcelone de mon enfance* (the neighbourhood of Barcelona of my childhood). A challenge is to be able to determine the boundaries of place names including all the informative words, and only those.

The MATRICIEL project (PEPS 2016-2017, *Université Paris-Est*<sup>1</sup> and *Centre national de la recherche*

<sup>1</sup> became *Université Gustave Eiffel* in 2020: <https://www.univ-gustave-eiffel.fr/>

*scientifique* (CNRS)<sup>2</sup>) had the goal to study places and related perceptions during migratory journeys based on migrants' words. The studied corpus was about Spanish Republicans journeys from Spain to France during Spanish Civil War. The corpus was provided by the *Réseau des acteurs de l'histoire et la mémoire de l'immigration* (RAHMI)<sup>3</sup>. The project associated methods of NLP and geographic information science so that to report individual stories in the light of the collective story and memory. The presented work tackles the part and the perspectives of the project concerning the spatial analysis, and mapping of place names and associated features extracted from the life stories. Places are taken into account as objects described by a variety of features. A specific focus is about perception of places formulated by the interviewees. In this case, spatial analysis aims at highlighting geographical aspects of the features added to the texts. Spatial patterns and relations are sought. Mapping life stories deals with the display of geolocated elements from the stories. Maps can be illustrative or they can be a support for the spatial analysis (Caquard, 2017). Subjective information like perception is challenging to display since it does not have clear spatial boundaries, it differs between people and it may change over time (Griffin et al., 2012) (Olmedo, 2016).

The objectives are to analyse the spatial distribution of quoted places names, to characterize those places from associated information including perception and to determine potential spatial patterns in the perceptions of places. A transversal objective is to propose adapted displays for the located and thematic contents extracted from the corpus and for the spatial analysis results. Section 2 describes the corpus containing the life stories. Section 3 is about the results of spatial analysis and the proposed mapping. Section 4 is dedicated to envisaged improvements of the work, and to perspectives.

## 2. The corpus of life stories and method

### 2.1 The corpus of Spanish Republicans' life stories

The corpus was recorded in 2009 and is stored by the RAHMI. It contains 17 oral life stories told by the Republicans themselves or by their relatives. The covered period of time is from the childhood of the Republicans at the beginning of the 20<sup>th</sup> century to the contemporary time up to 2009. Focus is mainly at the end of the Spanish Civil War and during World War II. Interviews are lasting between 11 minutes for the shortest, up to 3 hours 25 minutes.

Transcription of the oral life stories of the corpus has been carried out manually (Dominguès et al., 2019). A version was provided by the RAHMI. This version was successively corrected and enriched by several transcribers and annotators. Besides formulated words by the interviewees, speaking turns, sounds, pauses, and translation of foreign words when possible (mainly Spanish and German language) were transcribed. The

number of tokens, i.e. words and punctuation marks, ranges from 2.300 to 52.000 per life story.

### 2.2 Annotation of place names and associated features

The transcribed corpus of life stories constitutes the material from which targeted pieces of information have been annotated. Annotations concerned specificities of formulations, place names and sentiment words. Details are as follows:

- Markers of oral speech: disfluencies (repetitions, revisions), accent, laugh, voice volume, etc.;
- Foreign words;
- Places designated by proper nouns (Npr);
- Places designated by common nouns (Nc);
- Sentiment words.

Life stories have been tackled as a set of mentioned places described by their names and by complementary features. Our focus is about Npr place names because they are the most easy to locate with geographical coordinates and so to map in order to study spatial distribution. Nc place names and sentiment words are considered in this article as complementary features to detail the spatial target of Npr place names and to determine perception of these places.

Inter-annotator agreements (IAA) between annotators have been computed in order to determine the relevance of automatic treatments for targeted pieces of information. A weak IAA would imply that human annotators cannot establish a consensual reference annotation. Defining automatic annotation processes would not be relevant as it would not be possible to assess them. The IAA has been calculated from two life stories and different annotators. Concerning the sentiment features that seemed the most difficult to automatically annotate, values of R, P and F-measure are respectively 0.72, 0.73 and 0.73 for one life story and 0.88, 0.89 and 0.89 for the second life story. Variations are mostly due to different interpretation, for instance the word *hiver* (winter) has either been annotated with negative polarity or not been annotated depending on the annotator. The results of IAA were good enough to proceed to automatic annotating methods.

Annotations were done semi-automatically using NLP and manual treatments (Dominguès et al., 2019). Npr of places were annotated with gazetteers and sentiment words with a lexicon. Nc of places were annotated thanks to a machine learning approach using the Stanford-NER (Named Entity Recognition) tool (Finkel et al., 2005). The proposed method implements linear chain Conditional Random Field (CRF) sequence models (Brando et al., 2016). As with the evaluation of the IAA, gazetteer and lexicon-based annotations as well as results from the NER-trained model have been evaluated by comparison with a manual reference of one or two life stories. R, P and F-measure were calculated. Then corrections were applied manually on the whole corpus. Treatments and corrections were done in the GATE

<sup>2</sup> <http://www.cnrs.fr/>

<sup>3</sup> <https://www.rahmi.fr/>

platform (Cunningham et al., 2013) using the visual interface and by coding Java functions.

A gazetteer of Npr has been built. It lists Npr place names associated with point geographical coordinates. Foreign forms of Npr (e.g. *Francia* for France) were included, even if translation was carried out during transcription. It includes France and Spain place names, and the most populated cities in the world as well as important places (countries, famous mountains); they come from GeoNames<sup>4</sup> and French place names come from the detailed database BDNyme provided by *Institut national de l'information géographique et forestière* (IGN, the French mapping agency). Npr and modified Npr place names were annotated in the transcribed life stories. Disambiguation was done manually so that to select one location in the gazetteer in the case of several locations with the same entry (for example Paris, France and Paris, Texas). The type of place was also filled in Npr annotations. It discriminates between place names corresponding to administrative levels and to topographic features.

CRF-based annotation of Nc of place used the BIO labels (Ramshaw et al., 1995). In each quote, the head is marked as B-LOC and the remaining words as I-LOC. The O tag is used to indicate that the word does not belong to a named entity. For instance, the Nc quote *la ligne de démarcation* (the French demarcation line) is annotated as follow: *la* as O, *ligne* as B, *de* as I and *démarcation* as I. The focus was on Nc including Npr in close textual context, for instance the quote *un quartier de Barcelone* (a neighbourhood of Barcelona). Nc may be addresses with the names of streets, squares, etc. and sometimes the numbers of buildings, e.g. *vingt-trois rue Naujac à Bordeaux*. The Nc place names were geolocated with the locations if the included Npr. Modified Nc place names were also annotated. New annotations were created labelled with the BIO model. Perception of interviewees was addressed by formulated words expressing sentiments. Modified Npr, Nc which do not include Npr and modified Nc are not studied in this article.

Perception was set as a polarity positive, negative or neutral. A lexicon of sentiment words was created and each word was allocated with either positive or negative polarity. The transcribed life stories have been automatically annotated using this lexicon. Then Npr place names have been associated with a unique polarity value among the three defined values, based on annotated sentiment words and on manual interpretation.

### 2.3 Spatial analysis and cartographical display of place names

Textual and spatial analyses have been determined in order to characterize place names mentioned in the corpus, regarding their associated descriptive features. Analyses aimed at being combined with cartographical display so that to highlight spatial patterns. Analyses and mapping were run for the whole corpus and for each life

story. This part of the method relies on the previously annotated features.

First focus was about the geographical coordinates associated to place names. The spatial distribution of geocoded Npr place names was studied thanks to their map display. Displays were based directly on point locations of places or on heat maps of those locations. A heat map is a spatial density map based on distances between point geometries, i.e. here Npr place locations. Places are of different types and so various potential extents. For instance, the extent of a municipality is likely to differ greatly from the extent of a country and even an administrative or geographic region. To limit inconsistencies, heat maps have been computed considering the type of places. Geographical databases were used to describe spatial units and topographical features. The country borders and main administrative units come from the worldwide database Natural Earth<sup>5</sup>. The French administrative units and topographical features such as transport network, buildings, hydrography and vegetation cover were provided from IGN. They can be viewed in the French national geographical portal and downloaded from the dedicated services website<sup>6</sup>. Knowing the locations, occurrences were summarized. They concern the number of quotes of place names and the number of life stories in which place names are quoted.

Second, extracted places were characterized in order to complete the interpretation of their spatial distribution. Characterization of Npr place names focuses on their types and the spatial target in their quotes. Studying the types of place was about synthesizing places according to administrative levels or topographic shapes. The spatial target is related to annotations using the BIO labels of the Nc place names including Npr. The spatial target of a Nc place name including a Npr can match or be modified compared to the spatial target of this Npr considered independently. Modifications may entail decrease or increase of the spatial extent and a shift of location. Quotes in the corpus corresponding to Nc including Npr were studied in order to detail the spatial target expressed by the interviewees. Distinct cases of associations were extracted so that to be able to deal with them for cartographical display.

Third focus was about perception related to Npr of places. One Npr can be associated with as many polarity values as its number of occurrences. The spatial patterns of these polarity values were mapped and studied. Analyses of potential spatial similarities in some geographical areas regarding perception were carried out using the Moran's Local Index of Spatial Autocorrelation (LISA). The LISA values were computed to highlight spatial autocorrelation among Npr place names of the whole corpus regarding their polarities. For a Npr location, a positive computed LISA value means a similarity regarding polarity between this Npr and the

<sup>4</sup> <https://www.geonames.org/>

<sup>5</sup> <https://www.naturalearthdata.com/>

<sup>6</sup> <https://www.geoportail.gouv.fr/>, <https://geoservices.ign.fr/>

spatially close Npr locations. Negative LISA value corresponds to contrasted polarities in the spatial neighbour. The LISA can be put into perspective to the spatial distribution of the locations of Npr and in contrast with their associated polarities, i.e. for one Npr, the difference between the number of occurrences of the three defined values of polarity.

### 3. Spatial patterns of place names related to locations, associated features and perceptions

This section deals with spatial analysis and mapping of the corpus of life stories. Life stories were addressed from annotated Npr place names and associated features. The presented results focus first on the feature about geographical coordinates solely and the numbers of occurrences. Second, results about types and spatial target of Npr are gathered. Then annotations about perception are summarised and displayed by place names.

#### 3.1 Spatial patterns in locations and occurrences

The corpus contains 351 distinct Npr of places corresponding to 2787 occurrences. The automatic annotation of the Npr place names as been evaluated through two evaluated life stories. The R values are equal to 0.96 and 0.92. Precision and F-measure values were not calculated as disambiguation was done manually. Geographical distributions and text counts have been done for the entire corpus. Then each life story has been detailed.

##### 3.1.1 Spatial pattern of place name locations

When displaying Npr place names, three spatial scales emerged for the whole corpus: global, regional and local (Figure 1). They correspond to meaningful geographical extents for the corpus. They provide spatial patterns which can be located and characterized based on a historical highlight.

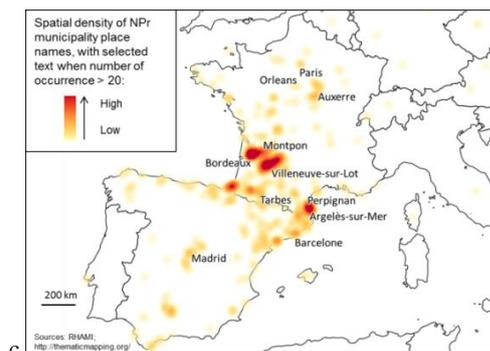
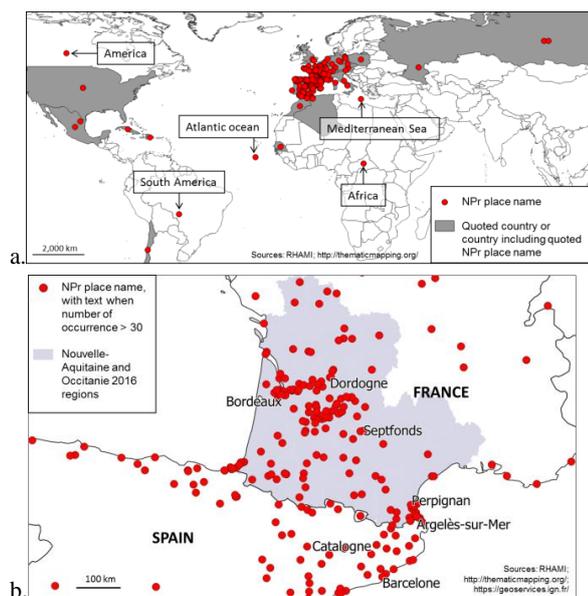


Figure 1. The spatial scales of the corpus based on place name locations: a. worldwide, b. regional in the south-west of France and north-east of Spain and c. local in several areas, mainly Bordeaux and surroundings, the Lot-et-Garonne department and the French-Spanish frontier.

The corpus has a global-scale extent since mentioned places are worldwide. Places are located in Europe, America and Africa, besides the Atlantic Ocean and the Mediterranean Sea. 22 countries are covered. However 95% of Npr place names are included in continental Europe and 85% in France ( $n=177$ ) and Spain (119).

The meaningful regional scale extent for the corpus appears as the south-west of France and the north-east of Spain. 41% of the Npr of place names are located in the current regions Nouvelle-Aquitaine and Occitanie created in 2016 from the merge of former administrative units. The interviewees have moved to and then settled in this area, so it logically emerged.

Significant local-scale areas were identified thanks to the heat map of the places. They are of limited extent corresponding either to one specific place names or to a group of spatially close place names. When computing the heat map on municipality names only (256 distinct Npr), high spatial density areas have appeared. Main hotspots are located around the city of Bordeaux, within and in the surroundings of the Lot-et-Garonne department and the French Mediterranean and Atlantic coasts near the Spanish border. Explanations can be found: Bordeaux is an important economic centre and most of the interviews took place in this city and its surroundings; the Lot-et-Garonne department corresponds to places where interviewees and their relatives lived when they arrived in France as well as to settlement places of the Resistance, and the French-Spanish border is lined with crossing points and internment camps. Secondary hotspots are located in the areas of Barcelona, Madrid and Córdoba, in the French Pyrenees and in distant surroundings of Paris. The places in Spain correspond to where interviewees have grown and come from before going to France. They also included battlefronts during Spanish Civil War (*front de l'Èbre, bataille de l'Alto*). The French Pyrenees and surroundings of Paris contained other internment camps and working manufacturing sites.

In parallel to the display of all the places in the corpus, each life story was mapped. The minimum number of Npr occurrences in a single life story is equal to 15, and the maximum is equal to 73. No bias appeared, for instance regarding place names in a life story which

would have verified an outlier distribution. Each life story contains place names globally scarcely located, and locally more densely gathered into some clusters. These distributions contribute to the corpus geographical pattern with balanced weights.

### 3.1.2 Numbers of occurrences of place names

In the whole corpus, the most quoted Npr place names are *Espagne* and *France*, with a number of occurrences respectively equal to 378 and 224. They are the only Npr present in the 17 life stories. The following most quoted Npr are the cities of *Bordeaux* (174), *Barcelone* (77), *Argelès-sur-Mer* (65), *Madrid* (63), *Perpignan* (58) and *Villeneuve-sur-Lot* (58). The correlation between numbers of Npr quotes and number of life stories in which Npr are quoted is positive and significant with the Pearson's correlation value equal to 0.8 (p-value < 0.001). Some exceptions exist such as *Villeneuve-sur-Lot* corresponding to 58 quotes present only in 3 life stories, and *Allemagne* corresponding to 34 quotes present in 14 life stories. Figure 2 illustrates the number of life stories in which distinct Npr place names are quoted. 22 distinct Npr are included in at least 5 life stories, which is only 6%. 244 Npr, i.e. 70%, are present in one life story solely.

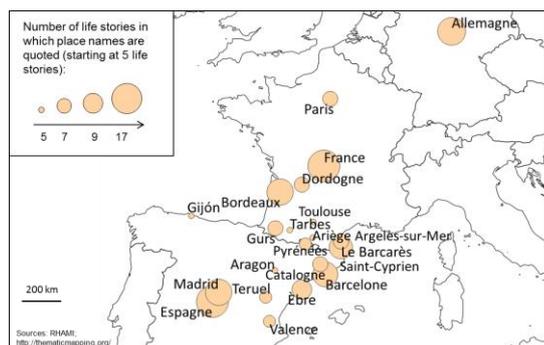


Figure 2. The number of life stories in which distinct Npr place names are quoted, starting from 5 life stories.

When considering each single life story, the minimum number of quotes of Npr place names is equal to 48 and the maximum is equal to 450. It is significantly correlated to the number of tokens (Pearson's value higher than 0.9, p-value < 0.001).

## 3.2 Associated features of place names

### 3.2.1 Types of place

Places are of two main types: administrative level and topographic feature. The administrative level contains continents (*Afrique*), parts of continent (*Amérique du Sud*), countries (*Italie*), municipalities (*Figueras*) and intermediate entities such as regions and departments (*Catalogne*, *Alpes-Maritimes*). Entities which are not strictly administrative were also included in the administrative level, such as historical or cultural areas (*Castilla*, *Médoc*, *Touraine*). Topographic features were divided into terrain elements (*Pointe de Grave*, *Sierra Morena*), hydrography (*Guadalquivir*, *Lac Saint-André*) and specific local sites (*fort de San Cristobal*, *chantiers*

*de l'Atlantique*). The count of the types and their details are given in Table 1.

Type of Npr place name	Count
Continent	3
Part of Continent	2
Country	19
Intermediate entity	47
Municipality	256
Terrain	8
Hydrography	10
Local site	6

Table 1. Detailed types of Npr place names and counts, regardless their number of occurrences.

The main type present in the corpus is municipalities. Compared to larger entities, municipalities can be assumed to be landmarks which are more often mentioned and used to locate events in life stories. They are partitioning the geographical space at a scale compliant with space use and movement of the interviewees. They are structuring lived and perceived space. Their names are shared among interviewees, interviewers and then listeners. Quoted municipalities may have evolved since the period of the life stories. Though continuity in names offers hints to identify and locate targeted places, e.g. *Cosne* that is included in the municipality of *Cosne-Cours-sur-Loire* since 1973.

### 3.2.2 Spatial target of Nc including Npr

Between 50 and 400 Nc were annotated per life story. Evaluation of Nc annotations on one life story corresponds to R, P and F-measure values equal to 0.53, 0.85 and 0.66 for B-LOC (the head of the mark) and 0.28, 0.67 and 0.39 for I-LOC (the remaining words of the annotation). Nc including Npr correspond to 250 distinct annotations. The spatial target of the Npr place name solely could remain the same or be modified regarding the four main following situations: Nc matches with, enlarges, reduces or relocates the included Npr. In the corpus, these situations have been specified into five categories according to the quotes:

- The exact location (*ville de Bordeaux*);
- Inclusion into (*quartier entièrement rénové à Gijón, plages du Barcarès*);
- Outside nearby (*banlieue de Barcelone, front de l'Èbre*);
- A containing space (*région de Perpignan*);
- Unknown (*direction Bordeaux*).

Few Nc including Npr (around 10%) matches the exact location of the Npr. In this case, Nc (e.g. the city of) enhances the Npr by designating the type of the Npr place names. The most frequent case (around 50%) is Nc included into the Npr. Nc is used by the interviewees to indicate a part of a Npr place which might not be named by a dedicated Npr (e.g. the waterfront of a town) or which name might be ignored (e.g. an unknown name of a city district designated by the city name). Then the case in which Npr is located outside nearby Nc is also quite

frequent (around 30%). Such as the previous case about inclusion, it can concern places with no dedicated names because the boundaries are unclear (e.g. the territory covering several municipalities near a famous city) or with names unknown by the interviewees. The case about containing space may not be obvious to identify. Indeed, the gazetteer has entries about Npr designating areas, e.g. *province de Cordoue*. Though there are often administrative units and it might not be the place targeted by the interviewee (e.g. when using the word *region* which is either an administrative spatial unit or surroundings of a named place with no precisely defined boundaries). The additional fifth situation corresponds to the case when no conclusion could be given.

When summarizing by the distinct 351 Npr place names in the corpus, 96 distinct Npr place names (27%) has at least one quote corresponding to a Nc. Consistently to the numbers of occurrences, the most frequently quoted Npr place names are the ones with the most numerous Nc including them (*Bordeaux, Espagne, Barcelone, France*). Cartographical display in Figure 3 lists the content of quotes corresponding to Npr and Nc including Npr place names in the area of Bordeaux. The remaining 255 Npr place names (73%) have no identified quote corresponded to Nc.

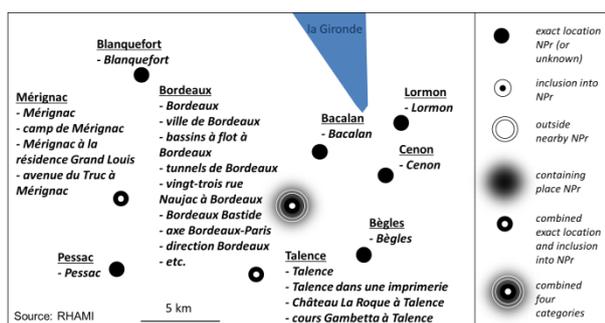


Figure 3. Cartographical display for spatial target of Npr place names: geolocated Npr and corresponded Nc in the area of Bordeaux with symbols for single or combined categories of spatial target (category “unknown” has been assimilated with exact location of Npr).

A cartographical semiology has been defined to display each category of quotes (Figure 3). Combinations of symbols were defined so that to cover cases where one Npr place names is associated to different categories. It aimed at conveying visually the assumed consequences of Nc regarding Npr place names geolocated based on the gazetteer.

### 3.3 Perceptions

#### 3.3.1 Spatial patterns in perceptions

Between 5 and 200 sentiment words have been annotated per single life story. Evaluation was carried out on one life story. R, P and F-measure values are respectively equal to 0.30, 0.80 and 0.4. 40% of Npr place names were related to one or several sentiment words. Among them, 30% were finally associated with positive polarity and 70% with negative one. The remaining 60% Npr which

have not been associated with sentiment words were set with a neutral polarity.

When polarities are summarised by Npr place name, 7% of Npr are associated with only positive polarity, 11% with only negative one, 40% with mixed polarities (positive, negative and neutral) and 42% with only neutral. The perception associated to a same Npr place name may vary during a life story due to a chronology of events, and between life stories due to different events, to the interviewee and to the context of collect. Spatial distribution of polarities was mapped focusing on municipalities in order to have consistency between Npr of places. Distributions of positive polarities and of negative polarities were compared using heat maps at a regional spatial scale (Figure 4) and maps of numbers of occurrences at a local scale (Figure 5).

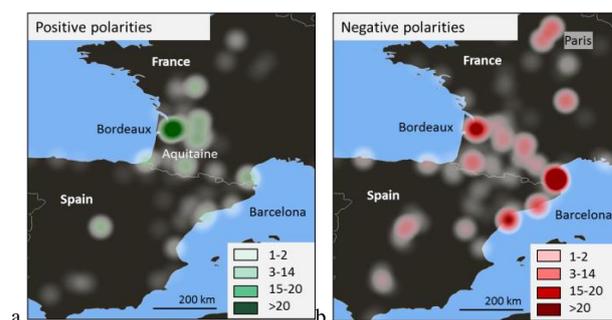


Figure 4. Spatial distribution of polarities associated to Npr place names: heat maps over Spain and France of a. positive and b. negative polarities. Numbers of occurrences were spatially merged according to map scale.

At regional scale, only the region of Bordeaux stands out with a high local density of places associated to both positive and negative polarity values. For positive values, the region of Barcelona and the south of Aquitaine also stand out even if with lower numbers of occurrences. For negative values, the French Mediterranean coast near the Spanish border and the region of Paris stand out highly.

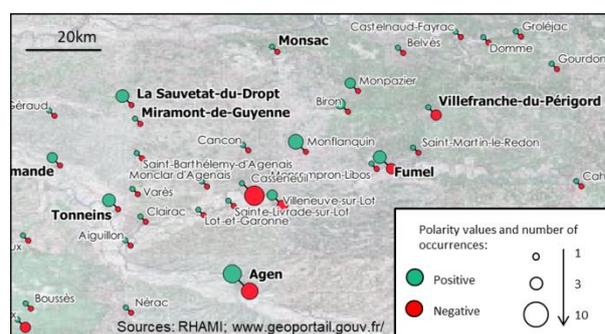


Figure 5. A count map of polarities associated to Npr place names zoomed in the south-west of France with numbers of occurrences of both polarities related to each place name.

At local scale, the balance between positive and negation polarities can be detailed and quantified. For instance in Figure 5, balanced polarities are observed associated to the Npr *Agen*, whereas unbalanced polarities with a majority of positive value are observed for the Npr *La Sauvetat-du-Dropt*.

### 3.3.2 Spatial similarities studied with local Moran's index

The LISA values were computed for all the Npr place names. Values are displayed in Figure 6. Regarding positive polarities (Figure 6.a), high spatial similarity is observed in the area of Bordeaux. Only medium values of dissimilarity were computed, except for France though not that relevant as the whole country is located by a central point. Regarding negative polarities (Figure 6.b), high spatial similarity was computed in the region of Barcelona. High dissimilarities are the most frequent for negative polarities. They are present about and around Toledo (due to mentions of living places and to battlefronts), the French Mediterranean coast near the Spanish border (e.g. *Saint-Cyprien*, *Collioure*), the Aquitaine (*Tarbes*, *Agen*), the area of Bordeaux (*Libourne*) and in the Loire region (*Nevers*). Taking into account both positive and negative polarities (Figure 6.c) indicates similar balance between both polarities in close Npr. High spatial similarities and dissimilarities are little less frequent than for negative polarities only. Same areas were outlined, except the Loire region and for some places in Aquitaine.

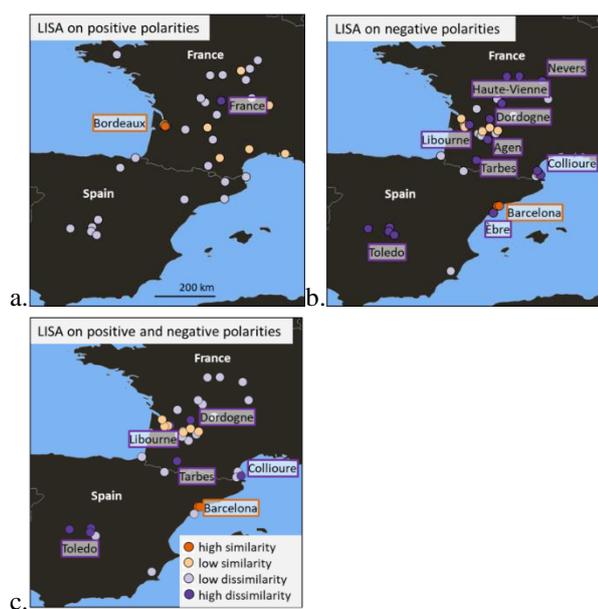


Figure 6. Similarity between Npr place names and spatially close Npr regarding polarity: maps of Npr locations with significant Moran's LISA, computed for a. positive polarities, b. negative polarities, c. both.

Spatial analysis based on LISA highlights some Npr which correspond to high numbers of occurrences (e.g. *Saint-Cyprien*). In these cases, local spatial similarities or dissimilarities are associated to numerous formulated sentiment words and so various perceptions related to place names. Contrasted perceptions may exist among life stories, along a life story and in spatially close places.

## 4. Discussion

This section deals with improvements and perspectives about the input, i.e. annotations of places names and

associated features, and the implementation of spatial analysis and mapping.

### 4.1 Perspectives for improving annotations of Npr and Nc place names and sentiment words

Automatic approaches for annotating target pieces of information in the transcribed life stories may be improved. Concerning Npr place names, the corpus was quite limited and so manual correction remained feasible. The main improvement would be a disambiguation algorithm for selecting one entry in the gazetteer among several entries which share the same name and differ in their locations. Many contributions have been proposed (for example (Hoffart et al., 2011) or (Nguyen et al., 2016)) which must, in all cases, be adapted to the corpus and its context.

The CRF-based annotation for Nc place names gave relatively good results. Remaining errors may be due to the limited training corpus, to scarcity of some Nc forms in the training corpus and to ambiguous words which can designate a geographical place or not. Improvements must focus on existing confusions between B-LOC and I-LOC tags. For instance in *usine d'armements*, *usine* was set as B-LOC (true positive) and *d'armements* was not annotated whereas it is I-LOC (false negative). Improvements may also deal with false negatives and false positives. Selected examples are respectively: *arènes de Madrid* (Madrid arenas) which was not detected and *équipe d'Agonac* (Agonac's team) which was annotated as Nc. Besides Npr and Nc place names, spatial relations would be additional interesting pieces of information to extract, e.g. *près de* (close to), *au nord de* (north of) (Clementini et al., 1993; Egenhofer et al., 1991). When a Nc form contained Npr, the Npr location was automatically propagated to the Nc. When Nc form did not contain Npr, no location was allocated to Nc. In this last case, challenge remains to define a propagation algorithm.

Concerning sentiment words and allocation of polarities to Npr, improvements would apply for the lexicon and for propagation. The reference lexicon of sentiment words may be enriched with specific vocabulary, in particular about World War II and the Spanish Civil War. Propagation of perception polarities from sentiment words to Npr is done manual for the moment. Defining an automatic process for propagation would benefit from taking into account negations (possible reversal of polarity) and elements of context (e.g factual elements such as *on avait faim* (we were hungry)).

### 4.2 Ad hoc life story mapping

#### 4.2.1 Creating local scale maps

The three identified spatial scales of the corpus give interesting overviews of the distribution of place names. Zooming out enables to contextualize some local distributions, and zooming in enables to detail distribution where the density of place names is high. Ad hoc maps can then be created on specific geographical areas and adapted extents and levels of detail can be

defined. Regarding the local spatial scale, a zoom in has been experimented on the city of Bordeaux. Historical city maps have been produced based on Npr place names located using the gazetteer as well as on Nc of precise addresses (e.g. *huit rue Constantin à Bordeaux*) and of specific buildings and monuments (e.g. *bassins à flot à Bordeaux*) located using the IGN geocoding service. Background display to the map was set alternatively as the 1950's IGN map and as aerial images from that time.

#### 4.2.2 Enriching display of place name spatial target

Creating maps about Npr and Nc including Npr place names calls for clarification about their spatial target corresponding to extent and location. Consequences on their display may be important. A defined semiology was proposed based on point geometry. It has yet to be evaluated (Harrie et al., 2015; Fabrikant et al., 2012). A difficulty is still to symbolize one location whereas there are several quotes corresponding to different spatial targets about this location. Besides, place names could be displayed with more relevant extent and location than point geometry given by the built gazetteer. Natural Earth and IGN databases could be used to allocate line or area geometries, e.g. for transport network or large regions.

Consideration of modified Npr and Nc as well as of spatial relations would complete the definition of the targeted places as formulated by the interviewees. It would enrich and precise the five categories of spatial target listed in this article. If successfully located, Nc forms which do not include Npr would be interesting to map. Locations would probably be imprecise and an adapted semiology would have to be determined. Visualizing uncertainty is a challenge: uncertainty concerns the NLP results, spatial analysis conclusions and the geographical databases used for mapping results or as background. It is an additional piece of information to display (MacEachren et al., 2005).

#### 4.2.3 Adding time to places to map individual stories

Counts of annotated pieces of information and synthetic maps have been produced to provide an overview on the whole corpus. Detailed counts by life stories were carried out so that to capture their variety of content. Maps could be designed by life story based on already extracted Npr, Nc and sentiment words. Time could be taken into account so that to be able to enrich mapping of individual stories such as in maps published on the RAHMI website<sup>7</sup>. Corresponding dates for mentioned places would allow displaying movement patterns and itineraries (Morcrette, 2016) and also studying changes of perception over time regarding places.

## 5. Conclusion

Mapping and studying spatial distributions of place names in a life story corpus allowed an overview of its geography at an individual level and at a shared level among the interviewees and regarding other corpus.

Exploratory analyses of quoted places, associated features and perceptions may enrich further hypotheses on how places are structural components of life courses and what additional aspects are worth considering so that to define those places. Aspect like perception can be integrated in a similar or contrasted spatial context, and can potentially evolve in time, during the told events and due to the present memory of the past events. Analysis summaries and maps may support works and researches in geography and history. They finally contribute to graphically enhance Spanish Republican oral life stories.

## 6. References

- Boccagni, P. and Baldassar, L. (2015). Emotions on the move: Mapping the emergent field of emotion and migration. *Emotion, Space and Society*, Volume 16, p. 73-80.
- Caquard, S. (2017). Cartography I: Mapping narrative cartography. *Progress in Human Geography*, Volume 37 (1), p. 135-144.
- Clementini, E., Di Felice, P. and Van Oosterom, P. (1993). A small set of formal topological relationships suitable for end-user interaction. *Advances in Spatial databases*, Volume 692, p. 277-295.
- Cunningham, H. et al (2013). Getting more out of biomedical documents with GATE's full lifecycle open source text analytics. *PLoS computational biology*, Volume 9 (2), online.
- Dominguès, C., Jolivet, L., Brando, C. and Cargill, M. (2019). Place and Sentiment-based Life story Analysis. *Revue française des sciences de l'information et de la communication*, Volume 17, online.
- Egenhofer, M. and Franzosa, R. (1991). Point-Set Topological Spatial Relations. *International Journal of Geographical Information Systems*, Volume 5 (2), p. 161-174.
- Fabrikant, S. I., Christophe, S., Papastefanou, G. and Maggi, S. (2012). Emotional response to map design aesthetics. *Proceedings of GIScience*, September 18-21, 2012, Columbus, Ohio, USA.
- Finkel, J. R., Grenager, T. and Manning, C. (2005). Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling. *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL 2005)*, p. 363-370.
- Griffin, A.L. and McQuoid, J. (2012). At the Intersection of Maps and Emotion: The Challenge of Spatially Representing Experience. *Thematische Kartographie*, Volume 62 (6), p. 291-299.
- Harrie, L., Stigmar, H. and Djordjevic, M. (2015). Analytical Estimation of Map Readability. *ISPRS International Journal of Geo-Information*, Volume 4 (2), p. 418-446.
- Hoffart, J., Yosef M. A., Bordino, I., Fürstenau, H., Pinkal, M., Spaniol, M., Taneva, B., Thater, S. and Weikum, G. (2011). Robust Disambiguation of Named Entities in Text. *Proceedings of the 2011 Conference on*

<sup>7</sup> <https://www.rahmi.fr/collecte-memoire-orale/republicains-espagnols/mariano-alcala-serrano>

- Empirical Methods in Natural Language Processing, p. 782-792.
- Huyghe, R. (2012). Événements et localisation spatiale : quand le lieu est un événement. In *Langue, espace, cognition*, Fagard, B. and Stosic, D. (eds), *Corela* numéro thématique.
- Leroy, S. (2004). *Le nom propre en français*. L'essentiel Français, Ophrys.
- MacEachren, A. M., Robinson, A., Hopper, S., Gardner, S., Murray, R., Gahegan, M. and Hetzler, E. (2005). Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know. *Cartography and Geographic Information Science*, Volume 32 (3), p. 139-160.
- Morcrette, Q. (2016). L'espace récité de la carte d'itinéraire: exemples français et américains. *Mappemonde*, n° 118.
- Nguyen, D. B., Theobald, M. and Weikum, G. (2016). J-NERD: Joint Named Entity Recognition and Disambiguation with Rich Linguistic Features. *Transactions of the Association for Computational Linguistics*, Volume 4, p. 215–229.
- Olmedo, É (2016). Pour une cartographie affective des récits des femmes de Sidi Youssef Ben Ali (Marrakech, Maroc). In Fournier, M. (dir.), *Cartographier les récits*, Clermont-Ferrand, Presses universitaires Blaise Pascal, Collection CERAMAC, p. 179-196.
- Purves, R. S., Winter, S. and Kuhn, W. (2019). Places in Information Science. *Journal of the association for information science and technology*, Volume 70 (11), p. 1173-1182.
- Ramshaw, L. A. and Marcus, M. P. (1995). Text Chunking using Transformation-Based Learning. In *Natural language processing using very large corpora*, Springer, Dordrecht, 1999, p 157-176.
- Tuan, Y.-F. (1977). *Space and place: The perspective of experience*. University of Minnesota Press, Minneapolis, Minnesota, USA.
- Zeile, P., Resch, B., Exner, J.-P. and Sag, G. (2015). Urban Emotions: Benefits and Risks in Using Human Sensory Assessment for the Extraction of Contextual Emotion Information in Urban Planning. In *Planning Support Systems and Smart Cities*, Geertman, S. et al. (eds), Springer International, p. 209-225.