Toponym recognition in custom-made map titles

Catherine Dominguès, Iris Eshkol-Taravella

Toponym recognition in custom-made map titles

Catherine Dominguès*, Iris Eshkol-Taravella**

* Université Paris-Est, IGN/SR, COGIT, Saint-Mandé, France
** LLL, UMR 7270, Orléans, France

Abstract. The titles of customized topographic maps constitute a specific corpus which is characterized by a very significant number of place names and spelling variations. This paper is about identifying toponyms in these titles. The toponym tracking is based on gazetteers as well as light parsing according to patterns. The method used broadens the definition of the toponym to include the nature of the corpus and the data in it. It consists of seven successive stages where both the extralinguistic context - in this case toponym georeferencing - and the linguistic context are taken into account. Mistakes in tagging are analyzed from the corpus characteristics and the results of each step tagging are evaluated (recall, precision, F-measure). Different conclusions can be suggested: i) toponym recognition in web corpora should take into account spelling changes, ii) toponym recognition cannot be limited to gazetteer proper nouns, iii) the notion of *subjective toponym* is relevant in this specific corpus, and could be considered with reference to the customization of maps.

Keywords: Toponym, Subjective place, Web corpus, Natural language processing, Gazetteer

1. Introduction

There has been an increasing demand for cartographic products which meet the specific needs of the user (for example, Loustau et al. 2009). Indeed, the French National Mapping Agency, IGN, launched a web service called "Carte à la carte" ("custom-made maps") in 2007. This service allows each Internet user to make a customized map by specifying the format, the size, the scale, where it is centered and its title. This requirement set is a useful source of information on why and how to customize a map which may be a holiday souvenir, a means of organizing or remembering an event, a gift to a

close friend, etc. One way of studying the users' needs is by tagging the titles in order to identify the different types of information they contain. This paper concerns the first tagging stage. It is the recognition of the toponyms which are employed by the web service users to name their customized maps. Consequently, this task deals with both cartographic knowledge and web users' habits.

Toponym definitions vary according to their fields of application.

For IGN, a toponym is the one or several word name of a place referring closely to a geographically located detail and to a group of people who use it. Toponomy distinguishes between inhabited and uninhabited places, places with relief, rivers, lanes and microtoponyms such as building names.

The National Commission on Toponymy of France (CNT)[1] defines a toponym as it refers to a determined geographical object.

In the natural language processing (NLP) field, toponyms are included in the named entities (NE). State-of-the-art NE identification systems are presented in (Chinchor 1998, Sekine 2004, Nadeau & Sekine 2009). NE is a linguistic expression which refers to a unique entity of the corpus in an autonomous way. Among toponyms, conventions of the Quaero[2] (2011) evaluation campaign distinguish between administrative places, physical places, lanes, buildings, addresses, etc. (Lesbegueries 2007) suggests distinguishing between absolute named entities which characterize NEs' own information (*Paris city*) and relative named entities which characterize spatial information associated to NE (*near Paris*).

Toponym definition causes a problem, therefore toponym identification too. According to the definitions, toponym classification is referential because it is based on the type of referent which is named by the toponym. Consequently, toponym definition cannot be limited to only proper nouns because common nouns can be used to name places in a neutral way: *the village*, in a personal way: *my village*, to refer to imaginary places: *the end of the world*, *my paradise*, etc. Lastly, toponym does not refer to a single referent because one place may be designated as several toponyms (referent ambiguity) and vice-versa, one toponym can be used to designate different geographical places (reference ambiguity) (Buscaldi 2009, Moncla et al. 2014).

---

1 CNI-CNIG (2010), Recommandations et observations grammaticales. http://www.cnig.gouv.fr/Front/docs/cms/cnt-grammaire-recommandation_126924688421947500.pdf [consulté le 18/12/2012]

2 http://www.quaero.org/media/files/bibliographie/quaero-guide-annotation-2011.pdf

Firstly, the title corpus is introduced in Section 2; it contains toponyms made from both proper nouns and common nouns. Because of the title corpus size, natural language processing (NLP) methods and tools are necessary to identify toponyms in the titles and annotate them. The process is made up of seven steps; they are described in section 5 with the results obtained. The process is assessed in Section 6. The discussion begins in Section 7, before the conclusions and perspectives presented in Section 8.

2. Custom-made map requests and title corpus

Mapmakers can request custom-made maps through the interface of the IGN "Carte à la carte" website[3].

Every request contains technical specifications required to make the map and the map title; it has this following format:

normal;	*15000;*	*landscape;*	*719888;6274149;*	*La Clusaz – Have nice holiday!*
fomat	scale	orien-tation	centre coordinates	title

The title is made up from sentences or word groups; its length must not exceed 55 characters. All the request titles constitute the title corpus.

The corpus is made up by varied Internet users which use disparate spelling rules. In conceiving their map and naming it, Internet users are guided by their own needs and objectives which may be very different. They use unregulated language where typographic rules are not applied or are individually interpreted in a heterogeneous way. Therefore, the title text is not standardized and it shows different types of spelling variations: undifferentiated use of upper or lower cases or separators, abbreviations, typing errors, neologisms, etc.

There are place-names on continental French maps, but all the titles are not written in French. Mapmakers may use foreign languages (English, German, Spanish, etc.) or regional languages (Corsican, Basque, etc.) and they might mix different languages in one title.

Toponym identification in such a corpus is a difficult task. Most often, processes for NE identification involve either a symbolic approach based on local grammars (Bontcheva et al., 2002), (Friburger, 2002), (Poibeau, 2003) or a statistical approach based on automatic learning, or hybrid systems such as (Leidner 2007, Béchet et al. 2011). Our approach is guided by

[3] http://loisirs.ign.fr/carte-a-la-carte/carte-a-la-carte-randonnee-decouverte.html

corpus specifications: located information, spelling variations and a very large number of toponyms. So, it involves both lexical resources (gazetteers with toponyms and geographic coordinates) and patterns which detect toponyms made up from common nouns.

3. Toponyms made from proper nouns

3.1. Terminology resources for identifying toponyms

The process for identifying proper noun toponyms relies on terminological and ontological resources.

3.1.1. BDNyme

The French NMA offers a database, BDNyme, which lists continental France's toponyms with their coordinates. This toponym collection is based on field surveys and aims to remain as close as possible to present-day local use.

Quality criteria are guaranteed. For example, each toponym is submitted to a local expert, and then it is validated by the IGN toponomy office. The number of toponyms accepted meet cartographical criteria, which generally correspond to a density criterion. Lastly, its coverage is more than 1.7 million entries. Toponyms are classified according to their type. This typology is based on geographical and administrative criteria: the capital city of a province, an inhabited locality, an uninhabited locality, hydronym, mountain range, communication line, rail line and miscellaneous line and is used to tag the title corpus.

BDNyme toponyms are written in accented lower-case letters (utf-8 encoding); they are made up of one or several words and their separator may be a blank space, a hyphen or an apostrophe. Each toponym is followed by its geographical coordinates, for example:

fos;	*719888.90;6274149.00*
toponym	toponym coordinates

3.1.2. Geonames

In accordance with its specifications, BDNyme only contains names of places depicted as a point(mountain highest point: *Carlit peak*) or may come to their centre: *Paris*. Consequently, linear objects such as rivers or area objects such as administrative entities (for whom the notion of centre is not relevant) are not in BDNyme. Thus, another toponym database has been used to identify toponyms whose layout is linear or areal: GeoNames.

GeoNames is a crowd sourcing resource; each toponym can have alternative identifiers; BDNyme covers the whole world but does not have the same term density as BDNYme; for example, GeoNames offers 1 295 occurrences of term *Ardèche* (cities, rivers, hills, ...), while BDNyme offers 19 804 punctual toponyms in the *département* of the *Ardèche*,

Geonames uses the WGS94 standard, although the coordinates of BDNyme toponyms and custom-made map centres are based on the RGF93 standard. A conversion must be done: the Geonames toponym coordinates have been translated into the RGF93 standard by R software[4].

Geonames metadata (*feature codes*) can be used to determine toponym type: country, region, stream, lake, mountain, etc. Granularity of information in Geonames is different from BDnyme's, and the feature codes hard to transpose into BDNyme tags. Nevertheless, since Geonames is especially used for non punctual places, its city topology is not taken up. On the other hand, feature code A and its sub-types can be used to tag regions and departments. New tags are added: *forest*, *pointOfInterest*, *park*.

Lastly, because of reference ambiguity (Smith and Mann 2003): a single string can refer to different places, Geonames toponyms are searched in two different steps. Firstly, after BDNyme toponym identification (Steps 1, 2, 3 and 4 in Table 1), Geonames toponyms which are located in France are searched and tagged in the title maps (Steps 5 and 7). Once toponyms based on common nouns have been tagged, Geonames toponyms, wherever they are located in France, are searched in the titles.

3.2. Variations in toponym spelling

For each language, toponym spelling rules apply to every toponym, whatever the language of the region where the place is located. Toponym spelling rules are complex because they depend on linguistic and extra-linguistic knowledge. Names of geographical objects are not standardized and often come from oral tradition. Moreover, toponym spelling rules differ according to use; for example, in France, street signs are in capitals, but street names are in small-letters on their addresses. Bioud (2006) notes that it is more and more usual to find a word spelled in different ways, in the same text. On the web, spelling rules change according to the Internet user and the new forms of written texts strongly affect toponym spelling.

Nevertheless, spelling rules exist, but they are complicated, subtle and heterogeneous. In particular, two typographic marks make compound toponyms difficult to write: capital and hyphen. For example, the spelling rules

---

http://www.r-project.org/

lead to writing: *le massif du Mont-Blanc* and *le mont Blanc*. In the first case, the compound *Mont-Blanc* is a complement of a geographical generic noun: *massif*; in the second one *Blanc* is an adjective, complement of a geographical noun: *mont*.

In the majority of cases, even if spelling rules gain consensus on the toponym spelling, they rely on different grammatical concepts and notions. Mathieu-Colas (1998) underlines that, even if each author shows his/her rules on an imperious form, divergences are numerous and there is no universal standard. Writers cannot memorize typographical nuances which assume deep linguistic knowledge and rely on preliminary syntactic and semantic analysis. Mathieu-Colas (1998) adds that for automatic processing it would be detrimental to cling to arbitrary and fussy norms. That is why this work is descriptive, not prescriptive, and takes into account freedom in spelling toponyms.

3.3. Spelling changes

Title text is not standardized and it shows different types of spelling variations which are taken into account for toponym identification:

- capitals, separators, diacritic marks, prepositions, determiners may either be present or not: for example, *FRESNES-LES-MONTAUBAN*, *FRESNES lès-Montauban* and *Fresnes-Lès Montauban* equal *fresnes-lès-montauban*;
- abbreviations: / for *sous* (*under*) or *st*(*e*) for *saint*(*e*)*:* for example, *rosny/bois* equals *rosny-sous-bois* and *ste-victoire*, *sainte-victoire*;
- empty words (determiner or preposition) may be omitted: for example, *fresnes-montauban* equals *fresnes-lès-montauban*.

3.4. Shortened toponyms

Toponym identification takes into account that compound toponyms may also be shortened on certain conditions. For example, *Bouc* is recognized as the abbreviated form of the toponym: *Bouc-Bel-Air* but *Sainte-Victoire* cannot be shortened to *Sainte* or *Le Mans* to *Le* (determiner) or *Mans*.

4. Toponyms made from common nouns

The title corpus shows that places are not named only by proper nouns. For example:

Where I run	(translated from: *Là où je cours*)
Our home, our sweet home	(in English in the corpus)
Good places for mushrooms	(translated from: *Bons coins à champignons*)

4.1. Identifying common noun toponyms with patterns

Obviously, terminal resources such as BDnyme or GeoNames are useless in order to detect these place-names because they are made from a common noun. Consequently, local grammars (see Fig. 1) have been designed on the Unitex platform (Paumier 2003). Local grammars are automata used to locate specific sequences and to tag them. In this corpus:

- they detect generic nouns of place. Generic nouns may be alone: *lake*, *forest*, *plains*, *hotel*, *mountain*, *house*, *home*, *place* or accompanied by a complement: *country house*, *St Cucufa forest;*
- they may be based on verbs, locative nouns and locative prepositions: *to leave*, *departure*, *arrival*, *beside*, *alongside*, *close to,* etc.;
- deictic adverbs: *here*, *there*, *where* are marked with the tag *DeicticPlace;*
- the tag *AddressPlace* identifies addresses: *221B Baker Street*;
- the tag *SubjectivePlace* marks places which have been appropriated or customized by the user: *mon paradis* (in English: *my paradise)*, *far east* or imaginary places: *Tamalou-Land*.

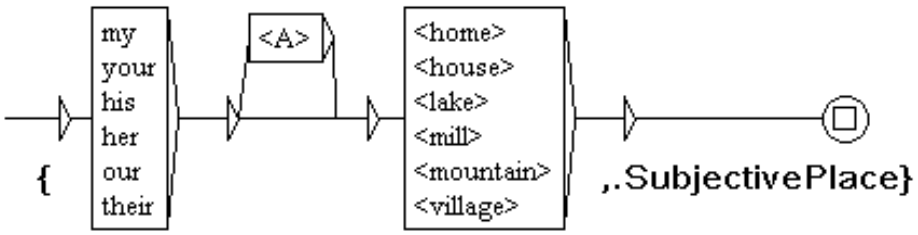

Figure 1. In Unitex's syntax, the symbol <*A*> represents any adjective. For example, this pattern identifies the sequences: *our first home* and *my mill* (which are considered as subjective places) and replaces them with the tagged sequences: {*our first home*,*.SubjctivePlace}* and *{my mill*,*.SubjectivePlace}*.

4.2. Subjective toponyms

We define a subjective toponym as a toponym used to name places in a personal way. This personal way may be obtained by:

- appropriation or customization of the place. Different cases have been observed:

  i)the place becomes subjective thanks to the contexts which may be linguistic: a possessive (*our territory*, *our fief*) or a qualifier (*magic place*) are added to the place-name; or extralinguistic: by using

South Africa for naming a place in France, the mapmaker associates a singular customization to the French place;

ii)- the noun or the compound do not name a place but they are employed as a name place as in the following examples: *Hiking around our love* (translated from *Randonnées autour de notre amour*)*; Walks around Mary* (from *Promenades autour de Marie*)*; Marseilles and Peter* (from *Marseille et Pierre*);

iii) coordination links two toponyms whose extent is different but which acquire the same status because of the coordination property; in the following example, *Thomery* is a town: *THOMERY or at Peter and Mary's* (from: *THOMERY ou chez Pierre et Marie*);

- reference to an imaginary place: *the end of the world* (*le bout du monde*), *my paradise* (*mon paradis*), *my Atlantis* (*mon Atlantide*);
- assessment and taste of the place which may be preceded or followed by an expression of emotion about the place: *my demanding Alps* (*mes Alpes exigeantes*) or *LE CAROUX mountain of light* (*LE CAROUX montagne de lumière*).

5. Method for toponym tagging in the map titles

Toponym identification relies on a comparison between character strings which are in the lexical resources (BDNyme and Geonames) or can be recognized with the patterns, and those in the titles.

Toponym tagging is separated into seven steps (see Table 1). In every step, the toponyms which are identified in the input corpus are tagged and typed. This output version of the corpus becomes the input corpus of the next step. The process relies on the context; there are two types of context: location context with the geolocation of the map centre (map area in Steps 1, 4 and 5, enlarged area in Steps 2 and 3) and linguistic context for the surface analysis of the title (Step 6).

The process combines different options: area size, lexical and ontological resources, spelling changes, shortened toponyms and patterns.

	Step 1	**Step 2**	**Step 3**	**Step 4**	**Step 5**	**Step 6**	**Step 7**
Area							
- map area	X			X	X		
- enlarged area		X	X				

- no area constraint							X
Resources							
- BDNyme	X	X	X	X			
- Geonames					X		X
Spelling changes			X				
Shortened toponyms				X			
Patterns						X	

Table 1. Toponym recognition steps

5.1. Map area

Toponym identification is firstly based on an extra linguistic context: geolocation. In Step 1, the only resource toponyms which are examined are in the area shown by the map (rectangle[5] a_1xb_1 in Figure 2). In Step 2, the resource toponyms looked for in the titles, are located in an enlarged area (rectangle a_2xb_2), centered on the map centre. In case of ambiguity, i.e. when different analyses may be proposed for one string, the only tag, corresponding to the longest string, is placed. In the example: *La Sainte-Baume* where *La Sainte* and *La Sainte-Baume* are both place-names, but only the longest string is tagged.

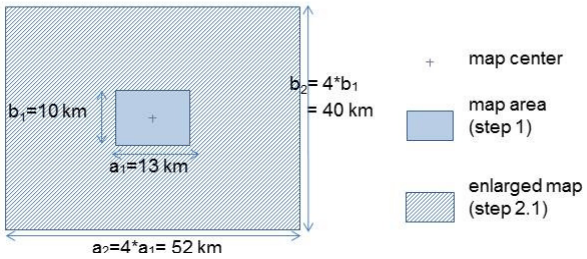

Figure 2. Dimensions of map areas: the map area (Steps 1, 4 and 5) and the enlarged area (Steps 2 and 3) for a map whose format is: *normal* (i.e. 91.5 cm x 69 cm), scale: *1/15 000* and orientation: *landscape*.

In Step 4, the area where shortened toponyms are looked for is the only map area in order to minimize the wrong tagging of words which are am-

[5] Different map sizes are offered by the web service "Carte à la carte".

biguous. For example, in French, *Bouc* may be a part of *Bouc-Bel-Air* (toponym) or an animal (billy goat).

5.2. Example of tagging

After all the process steps, proper and common noun toponyms are tagged in the titles. For example, the title:

CHOLET Forest of Nuaillé At Oliver and Mary's

is tagged as follows:

{CHOLET,.<AdministrativeCentreStep2>} {Forest of Nuailé,.<MisceanellousToponymStep1>},{At Oliver and Mary's,.<SomebodyPlaceStep6>}

In this example, *Forest of Nuaillé* is recognized in Step 1 because it is located in the map area. *CHOLET* is tagged in Step 2 because it is located in the enlarged area. *At Oliver and Mary's* is tagged in Step 6 by means of the *SomebodyPlace* pattern.

6. Evaluation

The toponym tagging method has to be assessed. The corpus has been split into two parts: the work corpus which allowed to identify the necessary lexical resources and to make patterns, and the reference corpus on which the method has been tested. Evaluation is based on a comparison between automatically obtained tags in the reference corpus and those which are placed manually into the same corpus.

6.1. Reference corpus

Specific features of the corpus have made it difficult to constitute the reference corpus:

- lack of context leads to cases of ambiguity, for example: *STE AGATHE* may be either a person's name, a nickname, an abbreviated place name (*Ste-Agathe en Donzy*) or a celebration. In the reference corpus, these sequences have always been tagged as places;
- some toponyms have not been tagged because, in the map titles, they are not used to name a place:
  * a toponym contributes to describe the map author or addressee: *the cousins of Pietrosella* (translated from *Les cousines de Pietrosella*);
  * a toponym contributes to describe another entity, here it is a cycling club: *TEAM U MARSEILLE*;

* generic nouns are ambiguous because, in French, they may describe either a place or an activity: *LA MONTAGNE* (which can be translated by either *MOUNTAINS* or *MOUNTAINEERING*).

Lastly, the reference corpus contains 1 749 lines, 15 403 words where 4 124 toponyms have been identified and tagged.

6.2. Result evaluation

Every step has been assessed (cf. Table 2) with recall, precision and F-measure. These measures are based on the counts of the true positives (i.e. the strings correctly tagged as toponyms by the process), and the false positives (the strings incorrectly tagged as toponyms) which are compared with the total number of strings that are actually toponyms, i.e. the sum of true positives and false negatives (the strings which are not tagged although they are toponyms).

recall = number of true positives/ (number of true positives + number of false negatives)

precision= number of true positives/ (number of true positives + number of false positives)

F-measure = 2 x (recall x precision) / (recall + precision)

In the first step, toponyms which exactly match those in the BDNyme (spelling and geo-location) are recognized. Consequently, the precision is excellent (1) but the recall is poor (0,18). The mistake analysis shows that i) many changes in toponym spelling are used by map users and ii) toponyms are not necessary in the map area.

Therefore, in Step 2, toponyms are searched in an enlarged area, and the recall is improved. The hypothesis is - designating a place which is little known, the author chooses to add the name of a much more popular place or which has a higher status (county seat or named place) even if the place is not in the map area. This step improves recall (0,24 i.e. +33 %) without damaging precision (0,98).
Similarly, the spelling changes (Steps 3 and 4) affect recall without reducing precision.

In Step 5, Geonames essentially identifies toponyms whose layout is linear or areal. Because toponyms are searched in the map area, the recall is not much improved, but the precision is kept.

The patters in Step 6 aim to identify toponyms made from a common noun. They significantly improve the recall: from 40 % to 66 %. It means that they are required to complete gazetteers.

The last step aims to identify all the toponyms which have not been tagged before. Therefore, Geonames is used without areal limit. This last step improves the F-measure, although it reduces the precision. This step matches the best result.

	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7
Tags	467	576	680	800	873	1306	1619
False positives	0	8	7	15	17	70	215
True positives	285	362	447	525	580	935	1077
False negatives	1260	1160	1055	943	871	491	321
Recall	0,18	0,24	0,30	0,36	0,40	0,66	**0,77**
Precision	1,00	0,98	0,98	0,97	0,97	0,93	**0,83**
F-Measure	0,31	0,38	0,46	0,52	0,57	0,77	**0,80**

Table 2. Step comparison based on recall and precision ratios.

6.3. Mistake analysis

Mistakes in automatic tagging have been analyzed and some of them result from the cases explained in paragraphs 3 and 4. They lead to tagging as a place the sequences which do not name places or which are ambiguous and untagged in the reference corpus. This affects the precision.

Other place names have not been recognized because:

- it is difficult to foresee any spelling change:
  - unusual abbreviations as *ch* for *chemin* (*lane*): *Ch-St Hilaire* for *Chemin St-Hilaire*;
  - lack of separator: *VillardBonnot* for *Villard-Bonnot*;
  - typing errors: *St Aygul* for *St Aygulf*;
  - phonetic transcription of regional accent: *NOT'BARAQUE ché par ichi* for *NOTRE BARAQUE c'est par ici* (whose traduction may be: *OUR PAD it's around here*);
- some places cannot be tagged automatically, lexical creations: *Tamalouland* (where *Tamalou* neither is a toponym proper noun nor a common noun) or *ARAMOUN* where *Aramoun*, which is a Lebanese village, metaphorically names a place in France.

7. Discussion

Firstly, the tagging process has been tested on the reference corpus. In order to broaden the observations, the process has been applied to a larger corpus: the work corpus. Table 3 recapitulates the characteristics of the two corpora.

Reference corpus description		
Line number	1 526	
Word number	7963	
Words per line	4,29	
Lines without title	20	1,3 %
Titles without toponym	246	16 %
Toponym number	1 619	
Toponymes per title	1,08	
Toponymes per word	0,20	

Work corpus description		
Line number	3 388	
Word number	18 791	
Words per line	5,55	
Lines without title	38	1,6 %
Titles without toponym	**517**	15 %
Toponym number	**4124**	100 %
Toponymes per title	1,23	
Toponymes per word	0,22	
BDNyme tags	2310	56 %
Geonames tags	773	19 %
Toponyms made on common nouns:		25 %
- *Subjective toponyms*	210	
- *Others*	831	

Table 3. Corpus description of the two corpora.

The titles are short: 4,29 words per title in the reference corpus.

Few maps have not any title (1,3 %). In the titles, toponyms are numerous: 0,20 toponym per word. That means that about one word out of five is a toponym or is one of a compound toponym. Nevertheless, 16 % of the titles do not have any toponyms: it means that, in 16 % of the cases, toponyms are not necessary to designate the custom-made map.

The recall and precision ratios (Table 2, Steps 3 and 4) show that it is necessary to take spelling changes, even on proper nouns, into account to identify toponyms in this corpus.

Table 3 shows how the title toponyms spread into the proper noun category and common noun category. One toponym out of four is made from a common noun and more than 5 % are obviously appropriated or customized by mapmakers.

8. Conclusions and future perspectives

The method proposed is determined by the features of the title corpus. Our toponym recognition is based on gazetteers (BDNyme and GeoNames), and patterns which use the linguistic context of the toponyms. The results show that:

- toponym recognition in corpora which come from the web should take into account spelling changes, currently not taken into account by search engines;
- toponym recognition cannot be limited to proper nouns because toponyms derived from generic geographical nouns (common nouns) designate a good number of places;
- the notion of *subjective toponym* is relevant in this specific corpus. This notion could be considered with reference to the customization of maps.

This work shows that the definition of the toponym poses a problem. Particularly, it is inappropriate to restrict the definition only to proper nouns, because common nouns can designate places too. In addition, a new class of subjective toponyms has been added to group toponyms which we consider as subjective into one category.

Toponym identification constitutes a preliminary and necessary step for tagging the entire title corpus which may help to better understand the users' needs, for example: the addressees (*Grandpa's map*), encouraging (*let's go!*), time elements (*Summer 2007*), events (*20th wedding anniversary*). Then, one of the objectives would be to adapt scales, data selection, map legends, typographies, cover illustrations, etc. to the needs of numerous cartographic web service users. This perspective would occur in the larger context of textual spatial information identification and exploitation.

Toponym location constitutes another relevant context for this study. Indeed, on the Internet, the majority of requests contain localized information. The use of computing and natural language processing tools in reply, increases the importance of toponym spelling norms. The title corpus comes from the Internet and its analysis gives relevant clues for retrieving toponyms in the web corpus.

References

Béchet F, Charton E (2010) Unsupervised knowledge acquisition, for extracting named entities from speech. In *IEEE International Conference on Acoustics Speech and Signal Processing*

Bontcheva K, Dimitrov M, Maynard D, Tablan V, Cunningham H (2002),Shallow Methods for Named Entity Coreference Resolution. *TALN 2002*.

Buscaldi D (2009) Toponym ambiguity in geographical information retrieval, In: Proceedings of the 32nd international ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR'09. ACM, New York, NY, 847-847

Chinchor N (1998) MUC-7 Named Entity Task Definition, version 3.5. Proc. of the Seventh Message Understanding Conference, http://www.aclweb.org/anthology/M98-1028. Accessed 15 October 2014

Friburger N, Maurel D (2004) Finite-state transducer cascade to extract named entities in texts, Theoretical Computer Science, vol. 313, 94-104

Leidner J L (2007) Toponym Resolution in Text: Annotation, Evaluation and Applications of Spatial Grounding of Place Names. PhD Thesis in Philosophy, University of Edinburgh

Lesbegueries J (2007) Plate-forme pour l'indexation spatiale multi-niveaux d'un corpus territorialisé, Thèse de doctorat, université de Pau et des Pays de l'Adour

Loustau P, Nodenot T, Gaio M (2009) Design principles and first educational experiments of PIIR, a platform to infer geo-referenced itineraries from travel stories, *Interactive Technology and Smart Education*, 2009, 6 (1), pp. 23 - 39

Mathieu-Colas M (1998) La majuscule flottante. Remarques sur l'orthographe des noms propres composés (type *N Adj*). *BULAG* n° 23, Centre Lucien Tesnière, Université de Franche-Comté, Besançon, 1998, pages 123-144

Moncla L, Renteria-Agualimpia W, Gaio M, Nogueras-Iso J (2014) Geocoding for texts with fine-grain toponyms : an experiment on a geoparsed hiking descriptions corpus

Nadeau N, Sekine S (2009) A survey of named entity recognition and classification. S. Sekine and E Ranchhod, ed. John Benjamins publishing company, pages 3-28

Paumier S (2003) De la reconnaissance de formes linguistiques à l'analyse syntaxique, Thèse de doctorat, Université de Marne-la-Vallée

Sekine S (2004) Named Entity; History and Future. http://www.cs.nyu.edu/~sekine/papers/NEsurvey200402.pdf. Accessed 10 November 2014