

# La base de données toponymiques du CESIT Corsica

Stella Retali-Medori  
Maria Francescu Luneschi  
Jean-Louis Santini  
*France*

## Résumé

S'appuyant sur le constat d'une carence de données fiables pour mener des études de toponymie corse, et de la perte croissante du patrimoine onomastique traditionnel de l'île, le CESIT Corsica a créé une base de données mise en ligne en juillet 2011 (<http://www.cesitcorsica.org>). Cette base accueille en son sein le fruit d'enquêtes de terrains réalisées auprès de locuteurs natifs, et d'investigation en archives. Elle espère offrir le matériel d'analyses des toponymes recueillis. Si la structure interne de la base propose des fiches complexes, le grand public peut accéder aux toponymes géolocalisés mis en ligne. Le poster présente donc la base de données et ses principales fonctionnalités ainsi que des applications possibles pour la connaissance théorique et pratique de la toponymie insulaire.

## Abstract

Building on a constant of a lack of reliable data to conduct Corsican toponomastic studies, and of the increasing loss of traditional onomastic heritage of the island, the CESIT Corsica created a database available online since July 2011 (<http://www.cesitcorsica.org>). This database collects the results of field surveys conducted with native speakers, and the investigation of archives. Through its database the association is expecting to offer the material for analyses of collected toponyms. If the internal structure of the database offers complex records, the general public can access geolocation-based toponyms online. The poster presents the database and its main functionalities and some applications to the theoretical and practical knowledge of the toponymy on the island.

\* \* \*

## Contexte

Le *Comité d'Études Scientifiques et Informatiques de la Toponymie Corse* (CESIT Corsica), a été fondé en 2009. Ses membres fondateurs, chercheurs, universitaires de plusieurs champs disciplinaires et amateurs éclairés de toponymie corse, se sont rassemblés en association loi 1901, reconnue d'utilité publique, sur la base d'un double constat :

- la carence, pour les études de toponymie corse, d'un outil offrant des données fiables pour la connaissance pratique et scientifique de la toponymie corse (Medori 2009);
- la perte irrémédiable du patrimoine onomastique de l'île qui est en cours, perte liée au recul de la pratique linguistique du corse, phénomène accru par l'abandon des territoires ruraux de la Corse.

L'association, afin d'assurer la qualité de sa démarche, s'est dotée d'un Conseil Scientifique composé de chercheurs et d'universitaires français et italiens. L'association a œuvré, dès sa création, à la réalisation d'une base de données au service de la toponymie insulaire.

Répondant à un besoin identifié dans le cadre de la politique linguistique impulsée par la *Collectivité Territoriale de Corse* en faveur de la langue corse, le projet de création de base de données a reçu un soutien financier de cette collectivité. Conçue à partir de 2009, la base de données est ainsi en ligne depuis l'année 2011.

## La base de données: structure, contenu

### Les catégories de toponymes

Cinq catégories de toponymes ont été définies pour la base. Il y a tout d'abord les microtoponymes, qui sont les plus abondants, jalonnant particulièrement l'espace rural. Il y a ensuite les macrotoponymes pour l'ensemble des lieux habités : écarts, hameaux, villes. Ne sont bien entendu pas pris en considération dans cette catégorie les hameaux disparus ; il en est de même pour les zones d'urbanisation récente ne présentant pas une structure urbaine cohérente et stable. Les noms de communes ont des fiches spécifiques, ainsi que les choronymes. Nous avons décidé de déterminer une catégorie «petit choronyme» pour certaines unités géographiques relativement importante au niveau spatial sans atteindre le niveau de la région (choronymes); il peut s'agir par exemple d'une chaîne montagneuse ou d'une baie.

Pour des raisons pratiques, liées à l'outil informatique dont nous disposons, les macrotoponymes, les choronymes et les petits choronymes sont actuellement matérialisés par des polygones tandis que les microtoponymes sont matérialisés par des points. Les quartiers de village ou de villes sont aussi représentés par des points pour permettre une lecture de carte plus aisée.

Grâce à une convention avec l'*Office de l'Environnement de la Corse*, les toponymes sont placés sur trois couches vectorielles : le Top 25 (IGN), la photographie aérienne de 2002, le cadastre rénové.

### Méthodologie d'enquête

Chaque commune de l'île fait l'objet d'une enquête menée selon une méthode semblable. La première étape consiste en un inventaire des toponymes sur les cahiers d'état de sections du cadastre napoléonien ainsi que des références cadastrales de chaque toponyme recensé. Les noms sont ensuite placés sur les cartes du cadastre.<sup>1</sup> Une enquête orale est ensuite réalisée, à l'aide du cadastre napoléonien et de la carte IGN, auprès de locuteurs natifs, dépositaires du patrimoine toponymique de leur région.

L'enquête enregistrée fait ensuite l'objet d'un dépouillement des toponymes recueillis, transcrits en forme phonique et en forme graphique, la forme graphique tâchant de représenter fidèlement la forme phonique, prenant en considération le plan phonologique et la

---

<sup>1</sup> Plans accessibles en archives ou sur le site du Conseil Général de Haute-Corse à l'adresse: <http://www.haute-corse.fr/site/index.php?page=le-cadastre-en-ligne>.

variation phonétique des localités enquêtées. Des contrôles (forme et positionnement des toponymes) sont réalisés si nécessaire auprès des témoins.

## Structure des fiches

### *Accès public*

Pour l'heure, le grand public a accès à des fiches simplifiées. N'y sont en effet visibles que la forme graphique corse, le nom de lieu et de la commune INSEE ainsi que le nom du cadastre napoléonien. La forme graphique corse, pour des raisons d'ordre pédagogique, note l'accent tonique lorsque le nom est accentué sur la dernière syllabe ou sur l'antépénultième, par exemple pour les toponymes *Aculò* ou *Petra Fràbbica* [p'edʒa vr'abbiga]. Lorsqu'un toponyme n'est présent que sur le cadastre napoléonien, sans forme orale correspondante, c'est celui-ci qui est représenté sur la carte; il est alors précédé du symbole \* (par exemple: *\*Costa di Piano*)

Les macrotoponymes, et les microtoponymes qui correspondent à des quartiers ainsi que les communes ont aussi des champs réservés aux noms des habitants: ethnonymes.

Il est prévu à terme de développer l'accès de données au public, notamment de la forme phonique, mais aussi de modules de recherches comme nous l'évoquerons *infra*.

### *Structure interne*

Pour chaque toponyme saisi, en accès interne,<sup>2</sup> les informations consignées sur les fiches sont les suivantes:

- a. forme phonique et orthographique ;
- b. ethnonyme corse pour les lieux habités (régions, communes, hameaux, quartiers) ;
- c. informations INSEE (nom de commune et référence) ;
- d. métadonnées : noms de l'enquêteur, de l'informateur, et du collecteur pour les sources écrites (cadastres, archives), dates ;
- e. données cadastrales et du Plan terrier de la Corse (formes et références dont numéros de parcelles ou de rouleau) ;
- f. données d'archives (publiées et inédites) ;
- g. analyse linguistique (classification, étymologie, reconstruction du lemme et du signifié originels) ;
- h. référent ;
- i. documents sonores, textuels, iconographiques.

Les enquêteurs et collecteurs et agents de saisie remplissent les champs a, b, c, d, e, f, h, i de la liste. L'analyse linguistique est réservée à la direction scientifique de la base de données (S. Retali-Medori). En effet, l'analyse des toponymes requiert des compétences spécifiques et elle n'est pas sans poser problèmes. Ainsi, la classification fait toujours l'objet de réflexions. Elle est conçue hiérarchiquement, selon cinq niveaux possibles:

<sup>2</sup> L'accès interne offre trois niveaux de confidentialité : public, agents de saisie et membres, administrateurs.

- a. classification générale
- b. thème général
- c. thème
- d. sous-thème
- e. type de \*signifié.

Par exemple, le toponyme *Castagnetu* peut être analysé ainsi :

- a. classification générale: agrotoponyme
- b. thème général: agriculture
- c. thème: arboriculture
- d. sous-thème: castanéiculture
- e. \*signifié: châtaigneraie.

Le fait que le signifié soit reconstruit est mentionné par l'astérisque. Pour l'heure, seul le premier élément des syntagmes fait l'objet d'une classification. Il faut souligner que lorsque le nom commun correspondant au nom propre a disparu mais que l'appellatif originel est identifiable, la notation du signifié se fait en doublant l'astérisque. On opposera ainsi le signifié 'forêt' des toponymes de type *Valdu* noté «\*signifié: forêt» de celui des toponymes de type *Selva* noté «\*signifié: \*forêt». En effet, si *Valdu* est toujours en usage dans certains dialectes insulaires, il n'en est pas de même pour *Selva* qui n'est conservé que dans la toponymie (Dalbera-Stefanaggi *et* Retali-Medori 2013 *et* Retali-Medori *et* Luneschi, à paraître).

Notons que la classification donne aussi sa part aux idéotoponymes, qui s'avèrent, à l'épreuve de l'enquête, assez nombreux et aux métaphores qui sont parfois révélées par la confrontation avec la nature du référent. Nous pouvons citer l'exemple de *Manghja Cane* qui pourrait être traduit littéralement par «mange-chien» et que nous avons choisi de classer parmi les idéotoponymes, le nom composé évoquant un trait de comportement. Si l'on pourrait penser à un surnom ayant donné naissance au toponyme, la zone très escarpée dans laquelle celui-ci se trouve, laisse envisager qu'il s'agit d'une métaphore, qui dans le cas présent se trouve au niveau trois de la classification.

## Résultats et perspectives

### Quelques chiffres

Depuis la mise en ligne de la base de données en 2011, plus 9000 toponymes ont été saisis, ce qui représente une moyenne de 3000 toponymes saisis par an, chiffre qui doit être majoré pour le recueil lui-même. La mise en ligne est progressive, chaque toponyme étant soumis à une correction des transcriptions par la responsable scientifique de la base de données. Le CESIT Corsica a couvert, depuis sa création, presque un tiers des communes corses.

## Outil pratique

La base de données du CESIT Corsica répond, comme nous l'avons dit précédemment, à un besoin pour les collectivités corse et pour le grand public. Des applications pratiques sont en effet possibles dans le domaine de l'adressage, des révisions cartographiques, ou encore des documents d'urbanisme. Des communes montrent déjà un intérêt pour les relevés effectués sur leur territoire

## Outil scientifique

D'un point de vue scientifique, la base de données du CESIT est déjà exploitée dans des travaux linguistiques faisant référence à la toponymie corse.<sup>3</sup> En effet, cette base de données est conçue pour devenir un outil scientifique.

### *Formes lexicales figées dans la toponymie*

Bien entendu, l'interrogation simple de recherche de toponymes montre déjà des résultats intéressants pour la recherche d'éléments lexicaux disparus de l'usage mais figés dans la toponymie, comme nous avons pu le mettre en évidence par ailleurs,<sup>4</sup> par exemple pour le type lexical *\*salettu*, qui a été substitué, au Moyen-Âge, par *salicetu* 'saulaie', en Corse comme dans le reste de l'Italoromania, mais qui s'est figé dans la toponymie (Aebischer 1963, Medori 2012 et Retali-Medori à paraître).

L'interrogation depuis le signifié reconstruit (seulement en accès privé pour l'instant), permet également de mettre en évidence des formes lexicales dont seule la toponymie peut témoigner. Ainsi en est-il par exemple de 'châtaigneraie' pour lequel on obtient les réponses *Castagnetu*, *Castagniccia* ou *Pulluniccia* qui répondent à des noms communs circulant toujours dans la langue, mais aussi *Castineta*, semblant ainsi attester de la présence en Corse, en des temps reculés, de *\*castina* 'châtaigne' < CASTINA, emprunt tardif du latin au grec, attesté dans l'Oribase Latin et dont ne sont conservés que quelques continuateurs de la variante CASTINEA dans certains dialectes gallo-italiens.<sup>5</sup>

### *Toponymie et reconstruction étymologique*

D'un point de vue diachronique, les recueils toponymiques ont aussi permis de conforter des reconstructions étymologiques et d'attester de strates d'évolution que les dialectes insulaires ne documentent plus. Ainsi en est-il de l'étymologie du corse *piopa* ou *piova* 'canal qui amène l'eau au moulin', qui a été rattaché à l'étymon *\*PLŨBĨCUS*, variante méthatétique de *PŨBLĨCUS* REW 6805 dont Aebischer (1937) a inventorié des attestations italo-romanes entre le IX<sup>e</sup> et le XIII<sup>e</sup> siècle. La toponymie corse atteste de variantes de *Piubbica* et de formes apocopées de type *Piobba*, permettant de reconstruire une strate intermédiaire, médiévale, conduisant de *\*PLŨBĨCUS*, ou plus exactement de *\*(AQUA) PLŨBĨCA* à *piopa*. Les référents

<sup>3</sup> Cf. notamment Medori (2013).

<sup>4</sup> (Dalbera-Stefanaggi et Retali-Medori 2013) ou encore Medori (à paraître).

<sup>5</sup> Cf. DÉRom s.v. *\*/kas'tania/* ~ *\*/kas'tinia/*, LEI s.v. CASTANEA, \*CASTENEA, CASTINEA, et pour *Castineta* en Corse (Dalbera-Stefanaggi et Retali-Medori 2013).

des toponymes mettent en outre en évidence le signifié originel '(ouvrage) public' du type lexical \**piùbbica* puisque les lieux désignés par *Piùbbica* sont des canaux et des voies (voir Medori, à paraître).

### *Toponymie et géolinguistique*

La géolocalisation des données et le développement à venir de modules de symbolisation devraient permettre d'appréhender la base de données comme un outil géolinguistique. Il devrait être possible, ainsi, d'établir par les traces toponymiques,<sup>6</sup> l'ancienne extension d'une unité lexicale par comparaison avec des cartes d'atlas linguistiques.<sup>7</sup> Ainsi en est-il des variantes et dérivés de *Valdu*, que nous avons évoqué supra, et qui est aujourd'hui restreint, dans l'usage, à une aire dialectale située dans le centre-ouest de la Corse, alors que la toponymie et la documentation médiévale permettent d'en attester l'extension sur l'ensemble du territoire insulaire.<sup>8</sup>

### *Polysémie*

La polysémie des appellatifs que les toponymes est une problématique intéressante, bien que parfois ardue à résoudre. Pour certaines formes, il est évident que la possibilité d'accéder au référent, soit par la case renseignée à cet effet, soit par le géoréférencement du toponyme, permet de reconstruire certains signifiés pour des noms communs devenus monosémiques. Le cas des toponymes de type *Lama* et son dérivé *Lamaghju* ainsi que *Prunu*, sont, à ce titre, significatifs. En effet, les noms communs correspondants *lamaghju*, ainsi que *prunu* désignent, dans les parlers corses, des plantes à épine, soit la 'ronce' pour les deux, le 'prunellier' et l' 'aubépine' pour *prunu* (BDLC, NALC 3). Or, les référents des toponymes *Prunu* et *Lama* ainsi que *Lamaghju* sont régulièrement des sources, des zones où l'on trouve des ruisseaux ou à proximité des fleuves. Cela peut laisser sous entendre, comme nous l'avons déjà mis en évidence pour le toponyme *Prunu* en lien avec l'appellatif *ochju* (Medori 2008), que *prunu* et *lama* ont été antérieurement polysémiques.

## **Conclusion**

La base de données du CESIT Corsica, qui devrait s'enrichir et connaître de nouveaux développements dans les années à venir, est déjà un outil opérationnel pour la connaissance de la toponymie insulaire et démontre aussi son utilité dans le cadre de la recherche scientifique, particulièrement linguistique.

---

<sup>6</sup> Dans cette perspective voir Pfister (1999).

<sup>7</sup> Pour la Corse, cartes du NALC et de la BDLC.

<sup>8</sup> Voir Dalbera-Stefanaggi *et* Retali-Medori (2013).



Stella Retali-Medori  
Francescu Maria Luneschi  
Jean-Louis Santini  
Association CESIT Corsica  
France  
cesitcorsica4@gmail.com

## Références

- Aebischer, P. (1937) 'Les formes métathétiques romanes \**plubicus* et \**pulbicus* pour *publicus*'. *Zeitschrift für romanische Philologie* 57. 57-68.
- Aebischer, P. (1963) *Miscelànea Paul AEBISCHER*. Barcelona: Istituto de Cultura Romanica de la Excma Diputacion Provincial de Barcelona, Biblioteca Filologica Historica.
- BDLC = Dalbera-Stefanaggi, M.J. (éd.) (1986-) *Banque de Données Langue Corse*. Corte: CNRS – Università Pasquale Paoli. Disponible sur: <http://bdlc.univ-corse.fr>.
- Dalbera-Stefanaggi, M.J. et Retali-Medori, S. et al. (2013) *Castagni è puddoni, le lexique corse de la castanéiculture*, coll. *Detti è usi di paesi, Matériaux et analyses extraits de la Banque de Données Langue Corse*. Bastia: stamperia Sammarcelli.
- DÉROM = Buchi, É. et Schweickard W. (éds.) (2008-) *Dictionnaire Étymologique Roman*. Disponible sur: <http://www.atilf.fr/DERom>.
- LEI = Pfister, M. et Schweickard, W. (éds.) (1979-) *Lessico Etimologico Italiano*. Wiesbaden: L. Reichert.
- Medori, S. (2008) 'Barabattula, barabuledda... essai sur les noms corses du 'papillon' et de la 'coccinelle''. *Géolinguistique* 10. 19-69.
- Medori, S. (2009) 'Toponymie corse: études et matériaux'. *Rivista Italiana di Onomastica* (RION) 15.1. 71-88.
- Medori, S. (2012) 'Studio di un campione toponomastico della Pieve di Lota e del Capo Corso (Corsica)'. Dans : Arcamone, M.G., Bremer, D., de Camilli, D. and Porcelli, B. (éds.) *Atti del XXII Congresso Internazionale di Scienze Onomastiche, Pisa 28 Agosto – 4 Settembre 2005*. Vol. 5. Pisa: Edizioni ETS, Collana Nominatio. 265-276.
- Medori, S. (2013) 'Éléments gallo-italiens et gallo-romans dans les parlers corses'. *Revue de Linguistique Romane* 305-306, tome 77. 121-138.
- Medori, S., (à paraître, a) 'Présence du corse dans des textes en latin médiéval et en vulgaire toscan (XIe-XVIe siècles): quelques éléments lexicaux'. Dans : Buchi, É., Chauveau, J.P. et Pierrel, J.M. (éds.) *Actes du XXVIIe Congrès international de linguistique et de philologie romane (Nancy, 15-20 juillet 2013)*. Strasbourg: Société de Linguistique Romane/ÉliPhi.
- NALC = Dalbera-Stefanaggi M.J. (1995-) *Nouvel Atlas Linguistique et Ethnographique de la Corse*. Paris/Ajaccio: Éditions du CTHS/Éditions Alain Piazzola.
- Pfister, M. (1999) 'L'importanza della toponomastica per la storia della lingua nella Galloromania e nell'Italoromania'. *RION* V/2. 449-464.
- Retali-Medori, S. (à paraître, b) 'Notes sur l'onomastique corse médiévale'. Dans : *Écrits sur la variation*.

Retali-Medori, S. et Luneschi, F.M. (à paraître) 'Une base de données au service de la toponymie corse'. *XXVIIe Congrès de Linguistique et de Philologie Romanes, Nancy 15-20 Juillet 2013*.

REW = Meyer-Lübke, W. (1935) *Romanisches Etymologisches Wörterbuch*. Heidelberg: Carl Winter Universitäts.