

# **Entre toponymes et situation géographique : cartographie et autocorrélation spatiale des suffixes communaux suisses**

**Romain Loup**

Université de Lausanne

[romain.loup@unil.ch](mailto:romain.loup@unil.ch)

## **Résumé**

Cet article propose une analyse quantitative de la structure des noms de communes en Suisse, en s'intéressant plus particulièrement à la distance toponymiques entre les communes en analysant leur noms et leurs suffixes. En combinant des méthodes d'analyse de noyaux, de mesure de similarité entre noms, et d'autocorrélation spatiale (indice de Moran), il met en évidence une organisation territoriale marquée des suffixes selon les régions linguistiques. L'étude révèle que certains suffixes, comme *-ens*, *-wil* ou *-ano*, présentent une forte concentration géographique, traduisant des dynamiques historiques et culturelles régionales profondes. Cette approche interdisciplinaire, à la croisée de la géographie et des sciences du langage, contribue à une meilleure compréhension de la structuration spatiale des identités toponymiques suisses.

## **1 Introduction**

Les toponymes, bien plus que de simples désignations géographiques, constituent des témoins historiques, linguistiques et culturels des territoires. En Suisse, pays marqué par la coexistence de quatre langues nationales, les noms de communes reflètent une diversité phonétique et morphologique exceptionnelle. Chaque suffixe, chaque terminaison, porte la trace d'une stratification historique profonde, témoignant des migrations, des colonisations et des évolutions linguistiques successives.

La tradition philologique suisse a largement documenté cette richesse. [Gatschet \(1867\)](#) identifie des couches celtes, romaines,

germaniques et romanes dans l'onomastique nationale. [Kristol \(2023\)](#) dresse un historique qui montre que dès les années 1960, des historiens approfondissent cette analyse en mettant en évidence la persistance régionale de certains suffixes, tels que *-ens* en Suisse romande, *-wil* et *-dorf* en Suisse alémanique, ou *-ano* au Tessin. Plus récemment, des synthèses comme celle de [Jordan \(2012\)](#) ou [Cattin et al. \(2005\)](#) rappellent que les toponymes participent à la construction d'identités spatiales et culturelles.

Étudier les noms de lieux, c'est découvrir à quel point les différentes régions linguistiques de la Suisse sont mutuellement imbriquées, et interconnectées avec les régions européennes voisines ([Cattin et al., 2005](#), p. 9)

Cependant, malgré l'abondance d'études descriptives, peu de recherches ont quantifié la structure phonétique et morphologique des toponymes et leur organisation spatiale. Les avancées en statistique spatiale – notamment l'application des indices d'autocorrélation, du *Moran scatterplot* ([Anselin, 1996](#)) et des techniques de projection multidimensionnelle (MDS) ([Borg & Groenen, 1997](#)) – offrent de nouveaux outils pour explorer ces questions de manière rigoureuse (cf. [Bavaud, 2014, 2023](#)). Le présent article propose ainsi une approche quantitative originale de la toponymie suisse, en s'appuyant sur :

- l'extraction systématique des suffixes communaux,
- la construction d'une distance toponymique entre les communes,
- l'analyse de l'organisation spatiale de ces structures onomastiques.

En analysant la structure de chaque nom de commune ainsi que leur suffixe, nous visons à explorer deux axes principaux. (I) Les suffixes des noms de communes présentent-ils des structures régionales marquées en Suisse ? (II) Pouvons-nous, à partir d'une analyse phonétique ou onomastique, retrouver les dynamiques linguistiques et historiques connues du territoire ? Ainsi, en combinant méthodes de phonétique computationnelle, statistique spatiale, et cartographie analytique, cette étude souhaite contribuer à la compréhension de la géographie linguistique de la Suisse contemporaine.

## 2 Données et préparation

L'analyse repose sur l'ensemble des 2126 communes suisses recensées en 2024, couvrant les quatre principales aires linguistiques : la Suisse alémanique, la Suisse romande, Suisse italienne et la Suisse romanche. Les données utilisées ([Office Fédéral de la Statistique, 2024](#)) combinent des informations toponymiques, linguistiques et géographiques :

- noms officiels des communes et des localités,
- coordonnées géographiques (longitude, latitude) des centres communaux,
- langue principale de chaque commune (classée en quatre catégories : allemand, français, italien, romanche).

La préparation des données détaillées ci-après a suivi plusieurs étapes, visant à standardiser les formes linguistiques et à permettre une analyse toponymique robuste.

### 2.1 Nettoyage des noms

Les noms de communes ont été systématiquement nettoyés et standardisés afin d'homogénéiser leur représentation écrite et simplifier le traitement informatique :

- suppression des accents et caractères spéciaux (é, ä, ç, etc.),
- conversion en minuscules,
- suppression des éléments non alphabétiques (parenthèses, traits d'union, apostrophes).

### 2.2 Extraction des suffixes

Afin d'analyser la structure morphophonologique des noms, des suffixes de longueur variable ont été extraits :

- Suffixes de 3 lettres (par ex. *-ens*, *-wil*, *-(d)orf*, *-ano*),
- avec en complément des tests sur des suffixes de 1 à 4 lettres pour explorer d'autres régularités.

Chaque commune a ainsi été associée à un suffixe principal, défini comme les dernières lettres significatives du nom nettoyé.

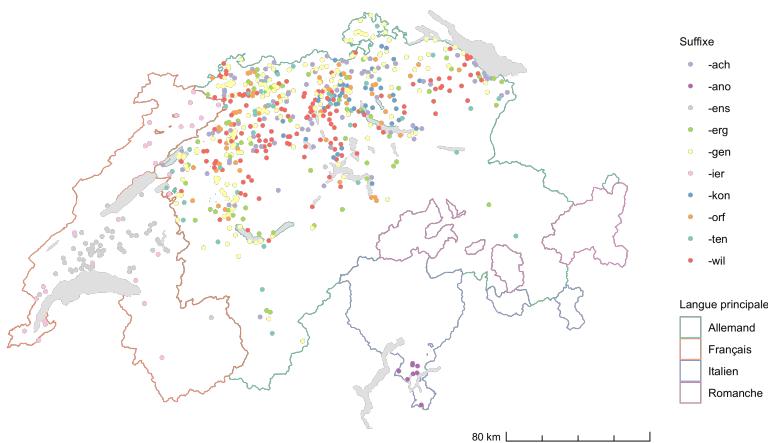


FIGURE 1 – Suffixes de trois lettres les plus fréquents en allemand, français et italien.

### 2.3 Que signifient ces suffixes ?

Les suffixes toponymiques suisses portent la trace de strates historiques et linguistiques profondes. Ils permettent de retracer les mouvements de peuplement, les dynamiques culturelles et les influences croisées entre langues germaniques et romanes. Plusieurs familles de suffixes sont particulièrement révélatrices.

Parmi les plus répandus en Suisse occidentale figurent les formes issues du suffixe germanique *-ingōs*, devenu *-ens*, *-in* ou *-an*, souvent combiné à un anthroponyme germanique. Très présent dans les cantons de Vaud et de Fribourg, ce suffixe indique l'appartenance à un clan ou un groupe familial : il signifie littéralement « chez les gens de », le premier élément désignant le fondateur de la localité (Cattin et al., 2005, p. 117). Une variante, *-ingun* (> *-ingen*, *-igen*), prolonge cette logique d'identification collective au travers de la notion de suite ou de clan (Cattin et al., 2005, p. 314).

Les formes romanes de suffixes dérivées du latin *-acum* ou *-akos*

donnent naissance à deux séries distinctes selon la région linguistique. En Suisse romande, ces formes aboutissent le plus souvent à *-y*, *-ier*, *-ey*, *-ay* ou *-ex* (comme dans Giez VD, Delley FR, Cossonay VD ou Thônex GE). Ces suffixes proviennent de gentilices ou de cognomina romains et ont progressivement évolué en fonction des dynamiques phonétiques locales. En Suisse alémanique, en revanche, la même racine latine se fossilise dans la forme germanisée *-ach* (Dornach SO, Bülach ZH), tandis qu'au Tessin elle devient *-ago* (Brissago TI) (Cattin et al., 2005, p. 208).

Autre exemple d'adaptation germanique, le suffixe *-(i)kofen* ou *-(i)kon*, dérivé de *-ing -hoven*, désigne des habitats aménagés, apparus entre les VII<sup>e</sup> et XVIII<sup>e</sup> siècles. On le retrouve sous la forme abrégée dans les cantons d'Argovie, de Lucerne et de Zurich, et sous la forme longue dans le canton de Thurgovie ou à l'ouest de la Suisse alémanique (notamment Berne et Soleure) (Cattin et al., 2005, p. 990).

Le cas du suffixe *-wil* est également emblématique. Il remonte au vieux haut allemand *wilāri*, issu du roman *villare*, lui-même formé sur l'adjectif latin *villaris* (dérivé de *villa*, désignant un domaine agricole). Ce suffixe renvoie à un petit village ou un hameau isolé. On le retrouve aussi bien en Suisse alémanique (Wil SG, Horgen ZH) que sous ses équivalents romans dans des noms comme Villars FR ou Villiers NE (Cattin et al., 2005, p. 673).

Enfin, dans les régions italophones, on trouve des formes en *-ano* ou *-an*, issues du latin *-anum*, *-ana*, *-ano*, qui servent à former des adjectifs de relation ou de possession (p. ex. proche de, appartenant à) (Cattin et al., 2005, p. 549; Vassere, 1996, p. 1443).

Il est à noter que ces suffixes restent en grande partie spécifiques à chaque région linguistique. On observe très peu de chevauchements entre les aires germanophone, francophone et italophone, ce qui renforce l'idée d'une forte structuration linguistique et culturelle des formes toponymiques, comme le montre la carte de la figure 1.

## 2.4 Construction de la distance toponymique

Deux approches principales ont été envisagées pour quantifier la similitude phonétique entre les noms de communes. La première, divisée

en deux, repose sur une comparaison directe des chaînes de caractères, soit par la **distance de Levenshtein** sur les noms nettoyés (Levenshtein, 1966), soit après **phonétisation** des noms à l'aide d'algorithmes comme Soundex ou NYSIIS (Odell & Strong, 1947). La seconde approche, plus structurelle, consiste à représenter chaque nom de commune sous forme d'un **vecteur de bigrammes** (séquences de deux lettres successives) et à mesurer leur proximité à l'aide d'une **distance cosinus** (Cohen, 2000; Cohen et al., 2003).

Après évaluation, la distance cosinus sur les vecteurs de bigrammes a été privilégiée car elle est indépendante des langues, elle capture efficacement les structures syllabiques et morphologiques typiques des toponymes suisses et est robuste aux variations orthographiques locales. Le résultat est une matrice de distances toponymique  $D_{\text{nom}}$  symétrique, de dimension  $2126 \times 2126$ .

## 2.5 Projections et voisinages spatiaux

En parallèle des distances toponymiques, deux structures de voisinage ont été construites : voisinage géographique basé sur les 10 plus proches voisins en coordonnées XY (k-NN), et le voisinage toponymique basé sur les 10 plus proches voisins dans l'espace MDS toponymique.

Ces structures permettent d'analyser la cohérence spatiale et toponymique des suffixes, via des mesures d'autocorrélation globale et locale ( $I$  de Moran).

## 3 Méthodes

L'objectif méthodologique est de quantifier la similarité toponymique entre les noms de communes suisses et d'analyser leur structuration spatiale. Les trois approches principales sont détaillées par la suite.

### 3.1 Distance de Levenshtein sur les noms nettoyés

La distance de **Levenshtein**  $d_L$  mesure le nombre minimal d'opérations élémentaires (insertion, suppression ou substitution de caractères) nécessaires pour transformer un mot en un autre (Levenshtein, 1966). Elle est définie par l'équation de récurrence :

$$d_L(i, j) = \begin{cases} d_L(0, 0) = 0 \\ \min \begin{cases} d_L(i - 1, j) + 1 \\ d_L(i, j - 1) + 1 \\ d_L(i - 1, j - 1) + \mathbf{1}_{s_i \neq t_j} \end{cases} & \text{sinon} \end{cases} \quad (1)$$

et la distance entre  $s$  et  $t$  est finalement  $d_L(|s|, |t|)$ .

Cette mesure – directement appliquée aux noms nettoyés – est sensible aux variations locales d’orthographe, mais peut être influencée par la longueur des noms.

### 3.2 Phonétisation suivie de comparaison

Pour atténuer les effets d’orthographe et se rapprocher de la réalité phonétique, les noms de communes ont été transformés à l’aide d’algorithmes de phonétisation : **Soundex**, conçu à l’origine pour l’anglais (Odell & Strong, 1947) et **NYSIIS**, légèrement mieux adapté aux langues romanes, car il groupe plusieurs lettres ensemble (Anon, 1968).

Chaque nom  $s$  est transformé en une représentation phonétique  $\phi(s)$ . La distance phonétique devient alors la distance de Levenshtein entre les représentations phonétiques :

$$d_\phi(s, t) = d_L(\phi(s), \phi(t)) \quad (2)$$

Cette approche standardise les formes similaires (p. ex. *Lausanne* vs *Lozan*) mais reste dépendante du choix de l’algorithme de phonétisation.

De plus, cette distance est sensible à la langue d’origine du toponyme : les algorithmes de phonétisation comme **Soundex** ou **NYSIIS** ont été conçus pour l’anglais et peuvent mal représenter les particularités phonologiques du français, de l’allemand ou de l’italien (Christian, 1998).

### 3.3 Distance cosinus sur vecteurs de bigrammes

Enfin, une approche plus structurelle a été utilisée. Elle consiste à représenter chaque nom de commune comme un **vecteur de bigrammes** de caractères, c'est-à-dire toutes les séquences de deux lettres successives apparaissant dans le mot. Cette représentation permet de capturer la forme morphographique des toponymes, en tenant compte des enchaînements de lettres typiques des langues concernées. L'approche est largement utilisée en traitement automatique des noms propres, en particulier dans les tâches de désambiguïsation ou de rapprochement d'entités textuelles similaires (Cohen et al., 2003). Plus formellement, à chaque mot  $s$  est associé un vecteur  $v(s) \in \mathbb{R}^d$ , où chaque composante correspond à la fréquence (ou la présence) d'un bigramme parmi les  $d$  bigrammes possibles (par exemple, *su*, *ul*, *ll*, *le*, *en*, etc.). La similarité entre deux noms  $s$  et  $t$  est alors mesurée par la **similarité cosinus** :

$$\cos(s, t) = \frac{v(s) \cdot v(t)}{\|v(s)\| \|v(t)\|} \quad (3)$$

Soient  $v(s)$  et  $v(t)$  les vecteurs de fréquences de bigrammes associés aux noms  $s$  et  $t$ . La **distance phonétique cosinus** dérivée est :

$$d_{\cos}(s, t) = 1 - \cos(s, t) \quad (4)$$

Cette approche est largement utilisée en traitement automatique des langues (TAL) pour mesurer la similarité entre mots, noms propres ou entités nommées, en particulier dans les contextes multilingues ou bruités où les distances basées sur l'orthographe stricte (comme Levenshtein) sont trop sensibles.

Les bigrammes permettent de modéliser de façon souple les motifs morphologiques (préfixes, suffixes, syllabes) sans recourir à une phonétisation complexe. Leur robustesse a été démontrée dans des tâches de désambiguïsation de noms propres, de reconnaissance d'entités nommées ou de recherche floue (Cohen et al., 2003; Kondrak, 2005). Cette mesure présente plusieurs avantages : elle est indépendante de la langue et de l'orthographe précise, elle capte les similarités

morphologiques globales (préfixes, suffixes, rythmes syllabiques) et elle est robuste aux différences mineures de longueur et d'orthographe entre noms.

Ces propriétés en font un outil particulièrement bien adapté à la diversité linguistique suisse, où les suffixes sont des marqueurs à la fois phonétiques, morphologiques et régionaux. Il a par conséquent été retenu pour la suite de l'analyse.

### 3.4 Construction de la matrice de distances toponymiques et vecteur de la distance locale

À partir de la méthode choisie, une **matrice de distances toponymiques**  $\mathbf{D}_{\text{nom}}$  de dimension  $2126 \times 2126$  a été construite, symétrique, carrée et de diagonale nulle. Cette matrice constitue le support principal pour les analyses de structuration spatiale et phonétique présentées dans la suite de l'article. Le vecteur de pondération  $\mathbf{f}$  où  $f_i = n_i/n_{\bullet}$  représente le poids spatial et  $n_i$  correspond au nombre d'habitants dans une commune  $i$ . Afin de quantifier l'hétérogénéité toponymique du voisinage immédiat de chaque commune, nous avons construit un indicateur de dissimilarité toponymique locale. Pour chaque commune  $i$ , nous avons identifié ses dix voisines géographiques les plus proches (en distance à vol d'oiseau), et calculé la moyenne des distances toponymique  $dl_i$  entre cette commune et ses voisines. Le vecteur de la **distance locale**  $dl_i$  repose sur une mesure préétablie de dissimilarité toponymique entre les noms des communes. Cette moyenne locale fournit un score de dissemblance contextuelle qui reflète dans quelle mesure une commune s'écarte phonétiquement de son environnement spatial immédiat. Cet indicateur constitue la variable d'entrée de l'analyse d'autocorrélation spatiale présentée ci-après (sec. 4.2.3),

$$dl_i = \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} \mathbf{D}_{ij}^{\text{nom}} \quad (5)$$

où  $\mathcal{N}_i$  est l'ensemble des  $k = 10$  plus proches voisines géographiques de la commune  $i$ , et  $\mathbf{D}_{ij}^{\text{nom}}$  désigne la distance toponymique entre les communes  $i$  et  $j$ .

## 4 Résultats : toponymie dans les communes

L’analyse s’appuie sur la matrice de distances toponymiques construite à partir des vecteurs de bigrammes des noms de communes. Les résultats sont présentés en trois volets : la structure toponymique générale, la distribution spatiale des suffixes et l’autocorrélation locale des formes toponymiques. Les analyses seront menées à la fois avec pondération  $f$ , selon le poids de communes, et avec des poids uniformes ( $f_{uni} = 1/n$ ). Cette double approche permet de distinguer deux dimensions complémentaires : d’une part, la structuration toponymique intrinsèque des noms, et d’autre part, l’influence potentielle de la taille ou du poids démographique des communes.

### 4.1 Structure toponymique générale

À partir de la matrice symétrique de distances toponymique  $D_{nom}$ , des **analyses multidimensionnelles classiques et pondérées (MDS)** ont été réalisées. La projection sur les deux premiers axes permet de visualiser la structuration toponymique globale.

Le MDS pondéré (ou de poids uniformes) utilisé est basé sur la matrice de centrage  $H$  et le noyau de caractéristiques (*feature kernel*)  $K_X$  (Bavaud, 2024), définis par :

$$H = I_n - \mathbf{1}_n \mathbf{f}^\top \in \mathbb{R}^{n \times n} \quad \text{et} \quad (6)$$

$$K_X = -\frac{1}{2} \boldsymbol{\Pi}^{\frac{1}{2}} H D^2 H^\top \boldsymbol{\Pi}^{\frac{1}{2}} \quad (7)$$

où  $f$  est le vecteur des poids relatifs des communes,  $\mathbf{1}$  est le vecteur colonne de taille  $n$  contenant des 1 et  $\boldsymbol{\Pi} = \text{diag}(f) \in \mathbb{R}^{n \times n}$  est la matrice diagonale de poids.

La projection toponymique finale  $\mathbf{X}$  du MDS (Bavaud, 2023) est obtenue par :

$$\mathbf{X} = \boldsymbol{\Pi}^{-1/2} \mathbf{U} \Lambda^{1/2}, \quad x_{i\alpha} = \frac{1}{\sqrt{f_i}} u_{i\alpha} \sqrt{\lambda_\alpha} \quad (8)$$

où  $\mathbf{U}$  et  $\Lambda$  sont respectivement les vecteurs et valeurs propres de  $K_X$ .

## 4.2 Tester et visualiser l'autocorrélation

Cette partie a pour but de comprendre dans quelle mesure les noms des communes sont autocorrélés, s'il y a des groupes de noms qui se forment et comment la pondération  $f$  joue un rôle dans les représentations graphiques. Les scores seront aussi comparés avec d'autres indicateurs.

### 4.2.1 Scores d'autocorrélation spatiale

Comme défini par Loup & Bavaud (2025, p. 5, Eq. 7), le test

$$z = \frac{\delta - \mathbb{E}_0(\delta)}{\sqrt{\text{Var}(\delta)}} > u_{1-\alpha} \quad (9)$$

où  $u_{1-\alpha}$  représente le quantile standard normal permet de rejeter l'hypothèse  $H_0$  au niveau  $\alpha$ . Le tableau 1 présente les **valeurs standardisées des scores**  $z$ , mesurant l'autocorrélation spatiale entre le noyau toponymique  $\mathbf{K}_{\text{nom}}$  (calculé à partir d'une distance cosinus sur les bigrammes) et six noyaux d'interprétation spatiale : politique ( $\mathbb{K}_{\text{pol}}$ ), temps de trajet ( $\mathbb{K}_{\text{time}}$ ), taille des communes ( $\mathbb{K}_{\text{size}}$ ), langue ( $\mathbb{K}_{\text{ling}}$ ), richesse ( $\mathbb{K}_{\text{rich}}$ ), et transport optimal des richesses ( $\mathbb{K}_{\text{TO}}$ ). On observe que l'autocorrélation est particulièrement importante pour plusieurs de ces dimensions, notamment pour le noyau politique, mais aussi pour celui reflétant la taille des communes. Ces valeurs sont bien au-dessus de scores usuels d'une loi normale. Il est remarquable que les noyaux politique et linguistique présentent une autocorrélation plus forte que le noyau fondé sur le temps de trajet, suggérant que les clivages symboliques et institutionnels structurent l'espace aussi fortement, voire davantage, que les contraintes de mobilité. Cela appuie l'idée que les noms de lieux en Suisse portent la trace d'un enracinement local durable, en lien avec les dynamiques territoriales historiques, culturelles et politiques. En revanche, les noyaux liés à la richesse et au transport optimal montrent des niveaux d'autocorrélation plus faibles, traduisant une relation moins directe entre les toponymes et les dimensions socio-économiques contemporaines.

	$\mathbb{K}_{\text{pol}}$	$\mathbb{K}_{\text{time}}$	$\mathbb{K}_{\text{size}}$	$\mathbb{K}_{\text{ling}}$	$\mathbb{K}_{\text{rich}}$	$\mathbb{K}_{\text{TO}}$
$\mathbb{K}_{\text{nom}}$	200.8	118.3	189.4	132.7	40.6	95.1

TABLEAU 1 – Scores d'autocorrélation spatiale standardisés ( $z$ ) mesurant l'alignement entre le noyau toponymique  $\mathbb{K}_{\text{nom}}$  (ligne) et six noyaux spatiaux (colonnes) : politique, temps de trajet, taille des communes, langue, richesse, et transport optimal des richesses.

#### 4.2.2 MDS : Des groupements bien marqués

La **carte toponymique** obtenue révèle plusieurs regroupements remarquables : les communes de Suisse romande se rassemblent autour de suffixes en *-ens*, *-ier*, *-res*, *-gny* alors que les communes alémaniques forment des ensembles distincts autour de *-gen*, *-wil*, *-ach*, *-(d)orf*. Enfin les communes tessinoises et grisonnes se regroupent dans une zone associée aux suffixes en *-ano*, *-one*.

Les figures 2 et 3 présentent les deux premiers axes de la décomposition multidimensionnelle classique (MDS), respectivement dans les versions pondérée et non pondérée. Dans les deux cas, on observe une structure en *clusters* bien marquée, où certains suffixes apparaissent fortement regroupés : c'est notamment le cas pour les suffixes *-gen*, *-wil*, *-ach*, *-ens* et *-ano*. La pondération par la taille des communes (Fig. 2, *scree graph*, annexes Fig. 7) a pour effet d'étaler la projection et d'éloigner certaines communes plus influentes démographiquement. Cela rend visibles des effets de masse : les communes de grande taille occupent des positions plus centrales ou plus isolées selon leur profil toponymique. La figure 3 – pondération uniforme – regroupe encore plus les suffixes et les effets de taille disparaissent.

On note par exemple que la ville de Berne est projetée légèrement à l'écart du centre de gravité des autres points, ce qui pourrait s'expliquer par la particularité phonétique de son suffixe et par son profil linguistique hybride ou neutre dans un espace fortement structuré par les oppositions linguistiques. En effet, *Berne* est certainement un mot celtique donné par le nom d'une rivière (Cattin et al., 2005, p. 143; Müller, 2002, p. 87). Cette hypothèse est renforcée par la présence de regroupements linguistiques nets dans les deux configurations : les

suffixes germanophones, francophones et italophones tendent à occuper des zones distinctes dans l'espace factoriel. Il est particulièrement intéressant de constater que cette structuration linguistique est aussi marquée que celle induite par les suffixes eux-mêmes, illustrant la profondeur culturelle et territoriale des régularités toponymiques en Suisse.

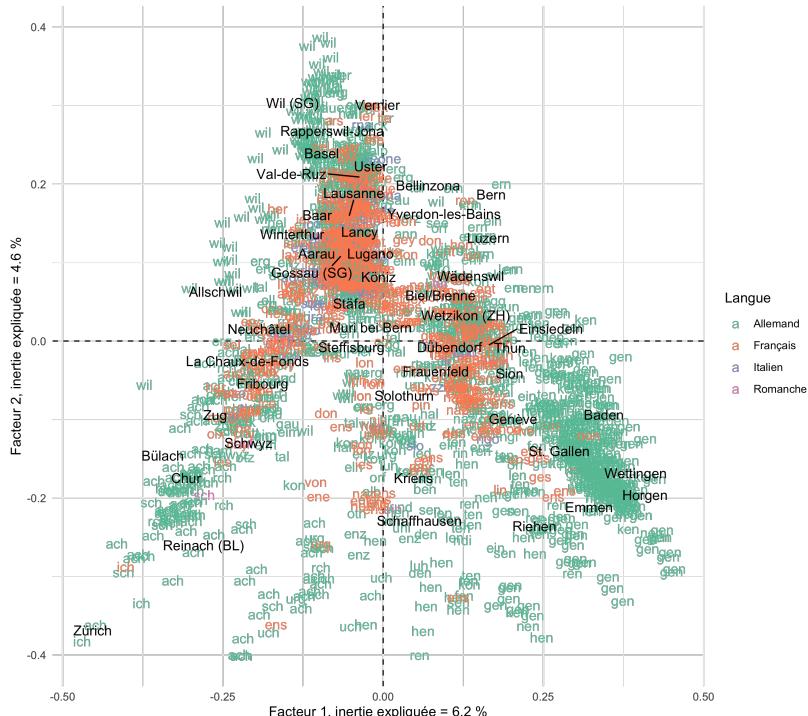


FIGURE 2 – MDS sur les toponymes, pondéré par le poids des communes et les 50 plus grandes communes nommées.

#### 4.2.3 Autocorrélation spatiale des dissimilarités toponymiques : effets d'échelle et de pondération

Afin d'évaluer dans quelle mesure les dissimilarités toponymiques locales entre les communes suisses présentent une organisation spatiale, nous avons mobilisé l'indice de Moran pondéré (Bavaud, 2024). Cette

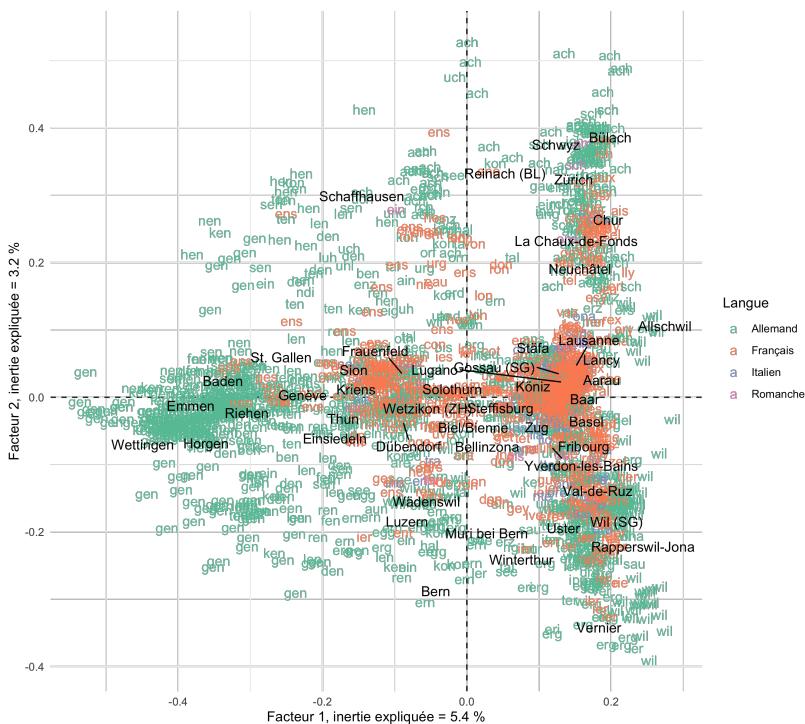


FIGURE 3 – MDS sur les toponymes, pondération uniforme et les 50 plus grandes communes nommées.

mesure, adaptée à nos données par l'intégration d'un noyau de chaleur (*heat kernel*), permet d'apprécier l'autocorrélation spatiale à différentes échelles, modulées par le paramètre de diffusion  $t$  (Loup & Bavaud, 2025, Eq. 23, p. 13). Deux types de pondération ont été testés : une pondération uniforme (chaque commune ayant le même poids dans l'ensemble) et une pondération régionale, fondée sur un vecteur  $f$  reflétant l'importance démographique des communes.

Les *scatterplots* de Moran-Anselin (Fig. 4) obtenus pour une faible valeur de  $t = 1$ , donc pour une interaction spatiale fortement locale, révèlent une autocorrélation toponymique marquée, avec une pente de 0.66 dans le cas d'une pondération uniforme, et de 0.5 sous pon-

dération régionale. Ces résultats indiquent que, localement, les communes voisines ont tendance à présenter des structures toponymiques similaires, suggérant une certaine continuité onomastique à l'échelle micro-régionale.

Lorsque le paramètre  $t$  est augmenté à 50 (Fig. 5), étendant l'influence spatiale à des voisinages beaucoup plus larges, l'intensité de l'autocorrélation diminue significativement et des groupes de noms romands et alémaniques se séparent verticalement. La pente du *scatterplot* tombe alors à 0.13 (pondération uniforme) et 0.07 (pondération régionale), révélant une structure spatiale plus diffuse et une perte de cohérence phonétique à large échelle. Autrement dit, les similitudes phonétiques observées à petite échelle tendent à s'estomper lorsqu'on considère des relations intercommunales plus distantes, ce qui atteste du caractère principalement local des régularités onomastiques.

L'effet de la pondération est également instructif. Dans tous les cas, la pondération régionale ( $f$ ) tend à réduire l'estimation de l'autocorrélation par rapport à la pondération uniforme. Cela suggère que les grandes communes – qui reçoivent un poids plus important – affichent une dissimilarité toponymique plus marquée par rapport à leur voisinage immédiat. Ce phénomène peut être interprété comme le reflet d'une plus grande hétérogénéité interne ou d'une situation linguistique plus complexe dans les centres urbains, souvent situés aux interfaces entre régions linguistiques ou dotés d'une toponymie historiquement composite. De plus, la variance expliquée de deux premiers scores factoriels est un peu plus élevée dans ce cas.

Ces résultats plaident ainsi pour une lecture multi-échelle de l'organisation spatiale des noms de communes, dans laquelle la cohérence phonétique apparaît fortement structurée à l'échelle locale, mais perd en intensité lorsque l'on s'éloigne géographiquement ou lorsque l'on pondère par des poids démographiques. L'approche par noyau de chaleur permet de moduler finement cette échelle d'analyse et de mieux cerner les dynamiques linguistiques et historiques qui sous-tendent la géographie des toponymes suisses.

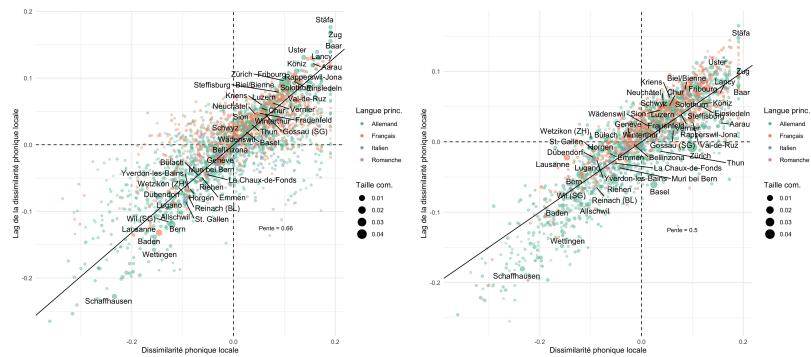


FIGURE 4 – Scatterplots de Moran-Anselin,  $t = 1$ . Gauche : poids f. Droite : poids uniformes. 50 plus grandes communes nommées.

### 4.3 Distribution spatiale des suffixes

Pour approfondir la structuration régionale, les suffixes de trois lettres les plus fréquents ont été sélectionnés. Des cartes ont été réalisées pour visualiser leur répartition (Fig. 1 et 8) ainsi qu'un tableau (2) regroupant les suffixes les plus fréquents avec leur nombre d'occurrences.

Ces cartes mettent en évidence une forte concentration des suffixes *-ens* dans les cantons de Vaud et de Fribourg, la prédominance des suffixes *-wil* et *-gen* dans la Suisse orientale, la localisation des suffixes *-ano*, *-one* exclusivement dans le Tessin. La structure phonétique des suffixes suit donc clairement la répartition historique des langues nationales et de leurs variantes dialectales.

Allemand (1373)	Français (615)	Italien (122)
<i>-gen</i> (168)	<i>-ens</i> (61)	<i>-ano</i> (8)
<i>-wil</i> (134)	<i>-ier</i> (25)	<i>-one</i> (7)
<i>-ach</i> (84)	<i>-res</i> (18)	<i>-gio</i> (4)
<i>-orf</i> (55)	<i>-gny</i> (17)	<i>-gno</i> (4)
<i>-eng</i> (54)	<i>-lle</i> (17)	<i>-ino</i> (4)

TABLEAU 2 – Occurrences des suffixes les plus fréquents pour les trois langues principales.

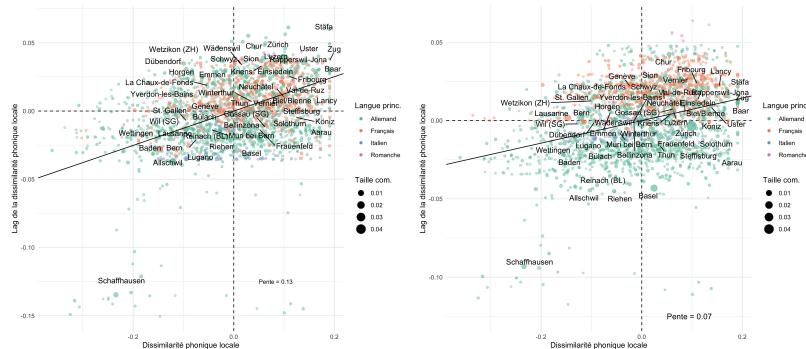


FIGURE 5 – Scatterplots de Moran-Anselin,  $t = 50$ . Gauche : poids f. Droite : poids uniformes. 50 plus grandes communes nommées.

Afin de comprendre de façon quantitative ces observations, le tableau 3 présente les indices de Moran (non pondérés) calculés pour une sélection de suffixes toponymiques parmi les plus fréquents en Suisse, ainsi que leur langue d'attache et la significativité statistique de leur distribution spatiale. L'indice de Moran  $I$  mesure ici le degré d'auto-corrélation spatiale d'un suffixe donné : une valeur élevée indique que les communes partageant ce suffixe ont tendance à être spatialement proches les unes des autres, signalant une concentration régionale.

Suffixe	Langue	Moran $I$	Valeur p
<i>-ens</i>	Fra	0.23	$1.5 \times 10^{-135}$
<i>-ins</i>	Fra	0.16	$1.8 \times 10^{-73}$
<i>-wil</i>	All	0.15	$1.7 \times 10^{-63}$
<i>-gen</i>	All	0.14	$4.1 \times 10^{-49}$
<i>-kon</i>	All	0.10	$3.1 \times 10^{-30}$
<i>-see</i>	All	0.07	$6.2 \times 10^{-16}$
<i>-orf</i>	All	0.07	$2.9 \times 10^{-14}$
<i>-ach</i>	All	0.06	$5.8 \times 10^{-13}$
<i>-ges</i>	All	0.06	$1.6 \times 10^{-13}$
<i>-ken</i>	All	0.05	$1.2 \times 10^{-9}$

TABLEAU 3 – Dix indices de Moran les plus forts pour des suffixes communaux selon la langue.

Comme le montre le tableau 3, le suffixe *-ens*, particulièrement romand, présente la plus forte autocorrélation ( $I = 0.23, p < 10^{-130}$ ), suivi par *-ins*, également en français. « En Suisse occidentale, il existe trois régions qui connaissent une densité significative de noms de lieux en *-ānum* (<*ins*) : la région de Nyon, colonie romaine, celle d’Avenches, ancienne capitale de l’Helvétie romaine, et celle de Studen (la Petinesca romaine, au bout du lac de Bienne) » (Cattin et al., 2005, p. 715). Cette forte structuration spatiale suggère que certains suffixes romands sont fortement enracinés dans des zones géographiques précises, vraisemblablement en lien avec l’histoire linguistique et la toponymie locale de la Suisse romande (cf. annexes, Fig. 8).

Du côté alémanique, plusieurs suffixes présentent également une autocorrélation marquée, tels que *-wil* ( $I = 0.15$ ), *-gen* ( $I = 0.14$ ) ou *-kon* ( $I = 0.10$ ) (Fig. 6). Ces résultats indiquent que certains éléments lexicaux propres au suisse allemand ont également une répartition géographiquement structurée, même si l’intensité de leur autocorrélation est légèrement inférieure à celle observée pour les suffixes romands.

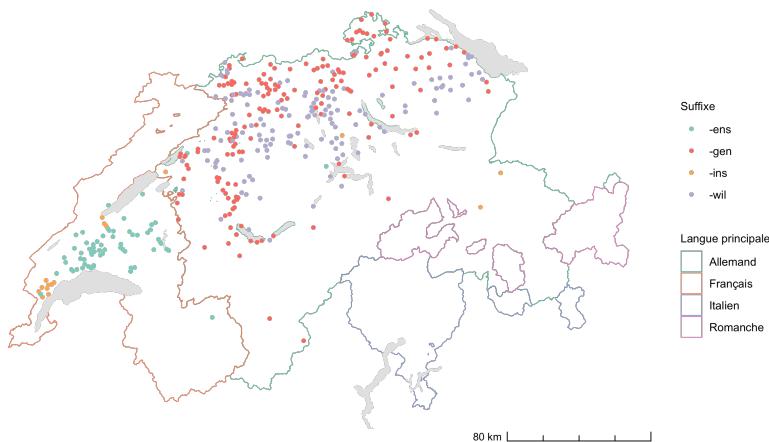


FIGURE 6 – Les quatre suffixes les plus autocorrélés localisés.

L'ensemble des suffixes retenus présentent une significativité statistique très forte ( $p < 10^{-8}$ ), ce qui confirme que la distribution spatiale des éléments phonétiques des noms de communes en Suisse n'est pas aléatoire, même si les scores du  $I$  restent contenus. Elle témoigne au contraire de dynamiques régionales cohérentes, reflétant à la fois les frontières linguistiques, les histoires de peuplement et les influences morphologiques propres à chaque aire culturelle.

#### 4.4 De commune à localité

À partir des 2126 communes suisses, les mêmes analyses peuvent être étendues aux 5764 localités définies par l'[Office Fédéral de la Topographie \(2024\)](#), dont 3973 portent un nom unique. L'analyse à cette échelle plus fine renforce les résultats obtenus : pour le test du  $I$  de Moran, les suffixes *-wil* (0.24), *-ens* (0.24), *-see* (0.20) ou *-gen* (0.20) présentent des valeurs d'autocorrélation encore plus élevées, accompagnées de  $p$ -valeurs encore plus significatives que celles observées au niveau communal. Cela s'explique en partie par le fait qu'en Suisse alémanique – plus encore qu'en Romandie – les fusions de localités en communes ont souvent atténué les effets de proximité entre noms partageant des suffixes similaires.

### 5 Conclusion

Cette contribution visait à mettre en avant l'interdisciplinarité de travaux à la croisée de la géographie et des lettres, en mêlant la distance spatiale entre des communes et leurs toponymes. Comme nous avons pu le voir tout au long de cet article, la géographie se mêle à l'histoire et à la formation des noms. Cela est mis en évidence grâce à plusieurs méthodes quantitatives, telles que la transformation de distances en noyaux qui a permis une visualisation en MDS montrant déjà une spatialisation claire des noms. L'autocorrélation de toponyme a aussi pu être vérifiée de plusieurs manières, par l'étude du noyau, comparé à plusieurs autres distances, mais aussi à l'échelle des suffixes. L'analyse de la répartition des toponymes dans les noms de communes suisses ainsi que de leurs suffixes révèlent une organisation spatiale loin d'être aléatoire. Plusieurs suffixes – notamment *-ens*, *-wil*,

*-gen* ou *-ach* – présentent une autocorrélation significative, indiquant une concentration régionale nette. Cette structuration n'est pas homogène : elle est plus marquée pour les suffixes francophones que pour les suffixes alémaniques, ce qui suggère des dynamiques historiques et linguistiques distinctes selon les aires culturelles. Les résultats obtenus à l'aide de l'indice de Moran pondéré montrent que l'autocorrélation toponymique est avant tout locale. Les régularités s'estompent rapidement à mesure que l'on élargit l'échelle spatiale d'interaction, comme en témoignent les courbes décroissantes obtenues avec le noyau de chaleur. Ce comportement met en lumière le rôle structurant des micro-régions linguistiques, souvent invisibles dans les découpages administratifs, mais bien ancrées dans l'onomastique. L'introduction d'un vecteur de pondération régional dans le calcul de l'indice de Moran nuance encore cette lecture : les grandes communes ou celles ayant un poids démographique plus important tendent à présenter une dissimilarité toponymique plus élevée par rapport à leur environnement immédiat. Ce phénomène pourrait s'interpréter comme le reflet d'une plus grande hétérogénéité interne, ou d'une position liminaire à l'interface entre espaces linguistiques.

Ces résultats s'inscrivent dans le sillage des travaux de François Bavaud sur la formalisation de la proximité spatiale et la modélisation des structures morpho-linguistiques régionales. L'analyse toponymique proposée ici rejoint, dans une perspective morphogénétique, des approches similaires appliquées au champs politique ou économique. Les suffixes toponymiques, au-delà de leur portée linguistique, apparaissent ainsi comme des marqueurs spatiaux d'appartenances régionales cohérentes, et constituent une voie féconde pour étudier la formation, la persistance et la diffusion des identités territoriales.

## Références

- Anon (1968). *NYSIIS (New York State Identification and Intelligence System)*.  
Office of Justice Programs, New York, NY, 2ème édition.
- Anselin, L. (1996). The Moran scatterplot as an ESDA Tool to assess local instability in spatial association. In Fischer, M. (éd.), *Spatial Analytical*,

- pages 111–125. Taylor and Francis, an imprint of Routledge, Boca Raton, FL, 1ère édition.
- Bavaud, F. (2014). Spatial weights: Constructing weight-compatible exchange matrices from proximity matrices. In Duckham, M., Pebesma, E., Stewart, K., & Frank, A. U. (éd.), *Geographic Information Science*, volume 8728, pages 81–96. Springer International Publishing, Cham.
- Bavaud, F. (2023). Exact first moments of the RV coefficient by invariant orthogonal integration. *Journal of Multivariate Analysis*, 198:105227.
- Bavaud, F. (2024). Measuring and testing multivariate spatial autocorrelation in a weighted setting: A kernel approach. *Geographical Analysis*, 56(3):573–599.
- Borg, I. & Groenen, P. (1997). *Modern multidimensional scaling: theory and applications*. Springer, New York, NY, USA, 1ère édition.
- Cattin, F., Aquino-Weber, D., Kristol, A. M., & Université de Neuchâtel (éd.) (2005). *Dictionnaire toponymique des communes suisses : DTS = Lexikon der schweizerischen Gemeindenamen*. Huber [u.a.], Frauenfeld.
- Christian, P. (1998). Soundex - can it be improved? *Computers in Genealogy*, 6(5):215–221.
- Cohen, W., Ravikumar, P., & Fienberg, S. (2003). A comparison of string distance metrics for name-matching tasks. *IIWeb*, 3:73–78.
- Cohen, W. W. (2000). Data integration using similarity joins and a word-based information representation language. *ACM Transactions on Information Systems*, 18(3):288–321.
- Gatschet, A. S. (1867). *Ortszymologische Forschungen als Beiträge zu einer Toponomastik der Schweiz*. 1, volume 10, 325 S. Haller, Bern.
- Jordan, P. (2012). Place names as ingredients of space-related identity. *Oslo Studies in Language*, 4(2):117–131.
- Kondrak, G. (2005). N-Gram Similarity and distance. In Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J. M., Mattern, F., Mitchell, J. C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M. Y., Weikum, G., Consens, M., & Navarro, G. (éd.), *String Processing and Information Retrieval*, volume 3772, pages 115–126. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Kristol (2023). *Histoire linguistique de la Suisse romande volume 1 : De la préhistoire au Moyen Age*. Editions Alphil-Presses universitaires suisses, Neuchâtel.
- Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady*, 10(8):707–710.
- Loup, R. & Bavaud, F. (2025). Spatial autocorrelation of political opinions: A kernel approach. *Journal of Geographical Systems*, pages 1–27.

- Müller, W. (2002). Siedlungsgeschichte und Ortsnamen in der Suisse romande. In *Ortsnamen Und Siedlungsgeschichte*, pages 83–94. Regesta Imperii, Heidelberg.
- Odell, M. K. & Strong, E. P. (1947). *Records management and filing operations*, volume 9. McGraw-Hill Book Co., New York, NY.
- Office Fédéral de la Statistique (2024). Application des communes suisses. <https://www.agvchapp.bfs.admin.ch/fr>.
- Office Fédéral de la Topographie (2024). Répertoire officiel des localités. <https://www.swisstopo.admin.ch/fr/repertoire-officiel-des-localites>.
- Vassere, S. (1996). 222. Morphologie et formation des microtoponymes: Domaine roman. In *Halbband+Registerband*, chapitre 2. Halbband+Registerband, pages 1442–1447. De Gruyter Mouton, Berlin, Boston, 2ème édition.

## Annexes

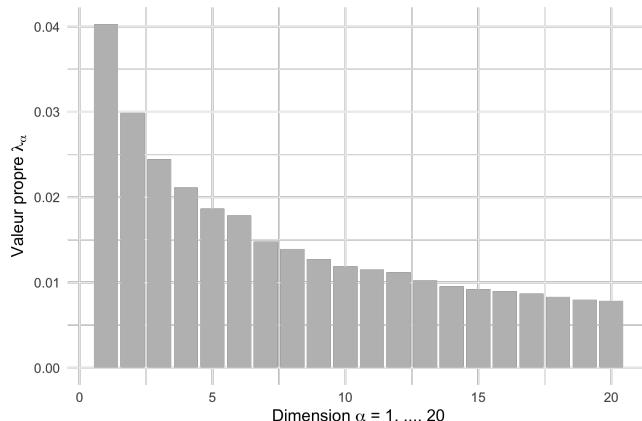


FIGURE 7 – *Scree graph* des 20 premières valeurs propres de  $\mathbf{K}_{\text{nom}}$ , pondéré par la taille des communes.

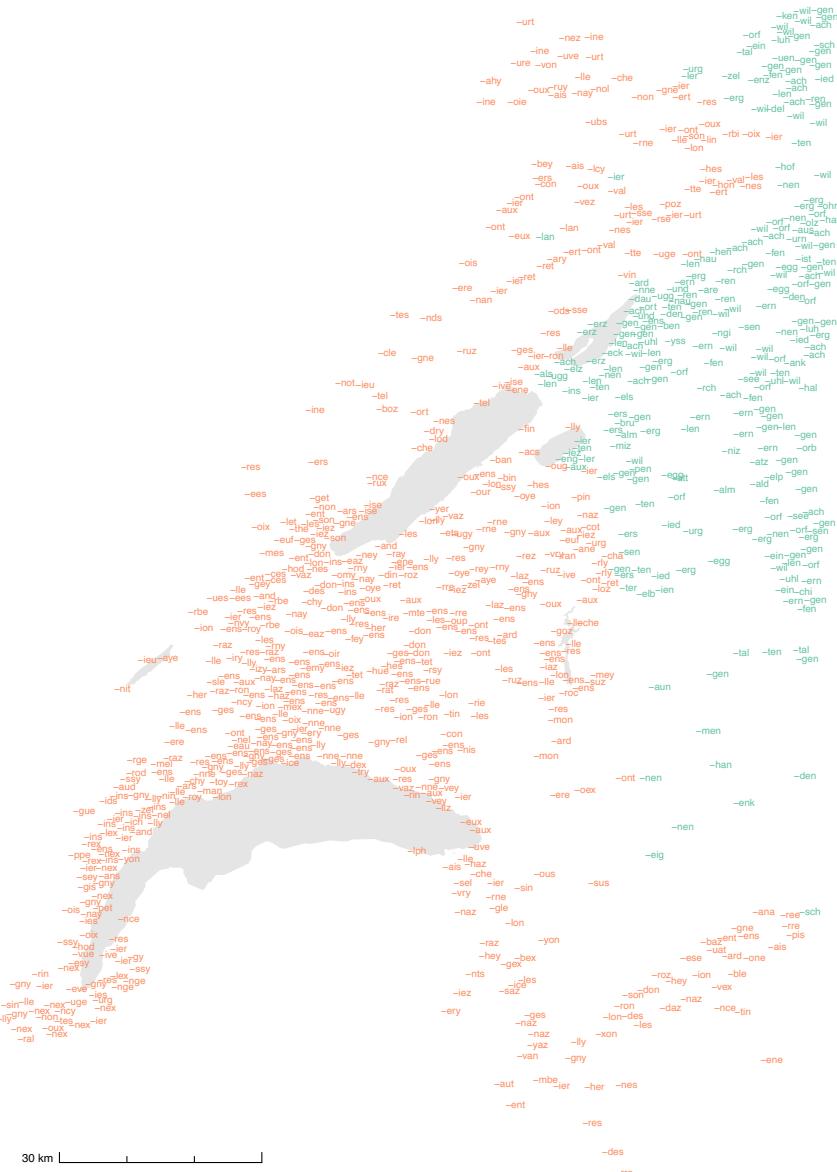


FIGURE 8 – Suffixes de la Romandie.