

La « carte scolaire » des écoles élémentaires du Grand Nancy : traitement algorithmique d'un corpus hétérogène

Paul CLAUDE MICHEL*, Anne HECKER et Xavier ROCHEL**

La « carte scolaire » est un élément central du système éducatif français qui a de nombreuses implications sur le quotidien des enfants. C'est pour cette raison que nous nous y intéressons dans le cadre de nos travaux de recherche sur les mobilités enfantines. Ceux-ci visent à saisir l'impact des choix parentaux de scolarisation sur la mobilité des enfants et s'inscrivent donc dans l'axe 2 du laboratoire. Cet article vise à présenter les difficultés que posent sa mise à disposition, dans un premier temps, et son traitement, dans un second temps.

Une carte des périmètres scolaires

La scolarisation des élèves dans l'école publique se fait, en France selon l'adresse de résidence, ce qui n'est pas le cas dans d'autres pays (voir Barrault-Stella, 2022). Chaque adresse dispose de son école – primaire, collège ou lycée – de secteur dans laquelle l'enfant est amené à suivre sa scolarité¹. Pour les élèves de l'école primaire, l'affectation est décidée par le conseil municipal, c'est ce dernier qui fixe la « carte scolaire » pour ses écoles maternelles et élémentaires.

Si les zones d'affectation des élèves dans le collège sont accessibles de manière ouverte et centralisée², ce n'est pas le cas pour celles du primaire. Les communes produisent et exploitent des données relatives à l'affectation des enfants dans les écoles primaires, mais ne les rendent pas systématiquement disponibles au grand

public. Pour travailler sur les cartes scolaires, il est donc nécessaire de contacter chaque commune afin d'accéder à sa « carte scolaire ».

Notre étude porte sur les affectations scolaires en école élémentaire dans le Grand Nancy. Sur les vingt communes de la métropole, seules cinq³ mettent à disposition des informations sur les affectations scolaires sur leur site internet. Toutes les autres ont été contactées – et relancées – individuellement. Ce travail chronophage a permis de constituer un corpus de données de trois formats différents : des tableurs (10 communes)⁴ ; des cartes à proprement parler (7 communes) ; ou une simple information concernant l'affectation dans une école unique (1 commune). Reste que cinq communes⁵ n'ont pas répondu à nos sollicitations, ce qui ne pose pas problème pour quatre d'entre elles, qui ne disposent que d'une école élémentaire.

* - Étudiant en M2 Transition & Aménagement – TRAM, ** LOTERR, Université de Lorraine, F-54000 Nancy

¹ - Sauf dans le cas d'une dérogation ou d'une scolarisation dans le privé.

² - Elle est disponible sur le site de l'open-data du MÉN : <https://data.education.gouv.fr/explore/dataset/fr-en-carte-scolaire-colleges-publics>

³ - Laxou, Saint-Max, Malzéville, Laneuveville et Ludres.

⁴ - 3 communes proposent à la fois des tableaux et des cartes.

⁵ - Heillecourt, Seichamps, Pulnoy, Saulxures, Fléville, Houdemont, Art-sur-Meurthe et Dommartemont.

À partir de ce corpus varié, il s'agit de réfléchir à la manière de numériser et coder les données obtenues pour les rendre utilisables sur SIG.

L'utilisation de la Base adresse nationale

L'adresse joue un rôle central pour notre problématique. C'est à partir de celle-ci que les enfants sont affectés à une école, c'est donc à partir d'elle que nous allons travailler. Pour ce faire, nous utilisons la Base adresse nationale (BAN) qui associe une position géographique à chaque adresse. Cette base se présente sous la forme d'un tableau avec des points géoréférencés qui correspondent aux adresses postales. Elle s'appuie sur des Bases adresses locales (BAL) constituées par les communes ou sur d'autres sources, dans l'attente de la constitution de la BAL communale. La Métropole du Grand Nancy compte 58 814 adresses. 30 557 d'entre elles ont été validées par les communes, et 28 257 doivent encore l'être et sont ajoutées à la BAN selon d'autres sources.

On souhaite associer l'école élémentaire de secteur à chacune de ces adresses et cela se fait de manière différente selon le format de la donnée. D'abord, si l'ensemble des enfants d'une commune est affecté à une école unique, on indique manuellement cette information pour l'ensemble des adresses de la commune. Ensuite, si l'on dispose d'une image, on procède à son géoréférencement, puis l'on sélectionne manuellement les adresses selon la zone d'affectation. Enfin, si l'on dispose d'un tableau, on procède à un traitement automatisé avec Python.

Une première étape pour ce traitement consiste à préparer les tableaux d'affectation scolaire. Ces derniers sont réorganisés avec le nom de l'école d'affectation et le nom de la rue concernée. Dans la majorité des cas l'ensemble de la rue est affectée à une seule école, mais dans un plus petit nombre de cas, seuls certains numéros sont concernés. Dans les tableaux fournis, cette information est indiquée sous diverses formes : « Du numéro 3 au numéro 47 » ; « Pairs de 2 à 24 et impairs de 3 à 35 » ; « Du 4 au 25 P/I » ; « À partir de la rue Alfred Krug », etc. En prévision du traitement automatisé, on formate manuellement les numéros de rues, compte tenu de la variété des formats pour l'indiquer et de l'inconsistance de cette notation au sein d'un même document.

De l'adresse à l'école : le script Python

L'objectif de notre script est d'ajouter une information concernant la sectorisation scolaire au tableau des adresses de la BAN. Il commence par nettoyer l'ensemble des noms de rue des tableaux d'affectation scolaire et

de la BAN. Précisément, les caractères spéciaux sont supprimés, le contenu des parenthèses placé au-devant du nom de rue et le tout est mis en minuscule ; ceci pour rapprocher au maximum les adresses issues des deux bases, en prévision de l'étape suivante.

Pour chaque adresse de la BAN on va ensuite, comme le montre la figure 1, (1) calculer la distance de Levensh-

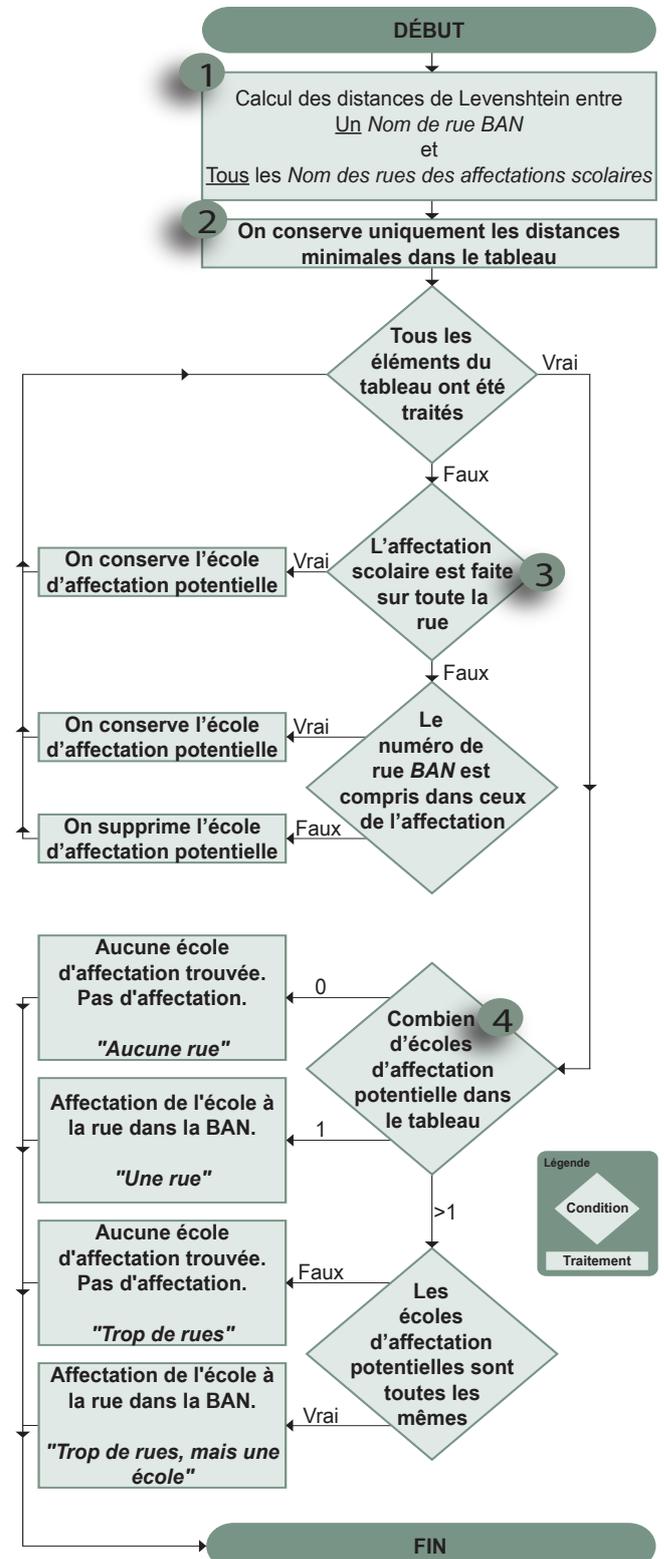


Figure 1 : Logigramme du script Python d'affectation scolaire par adresses

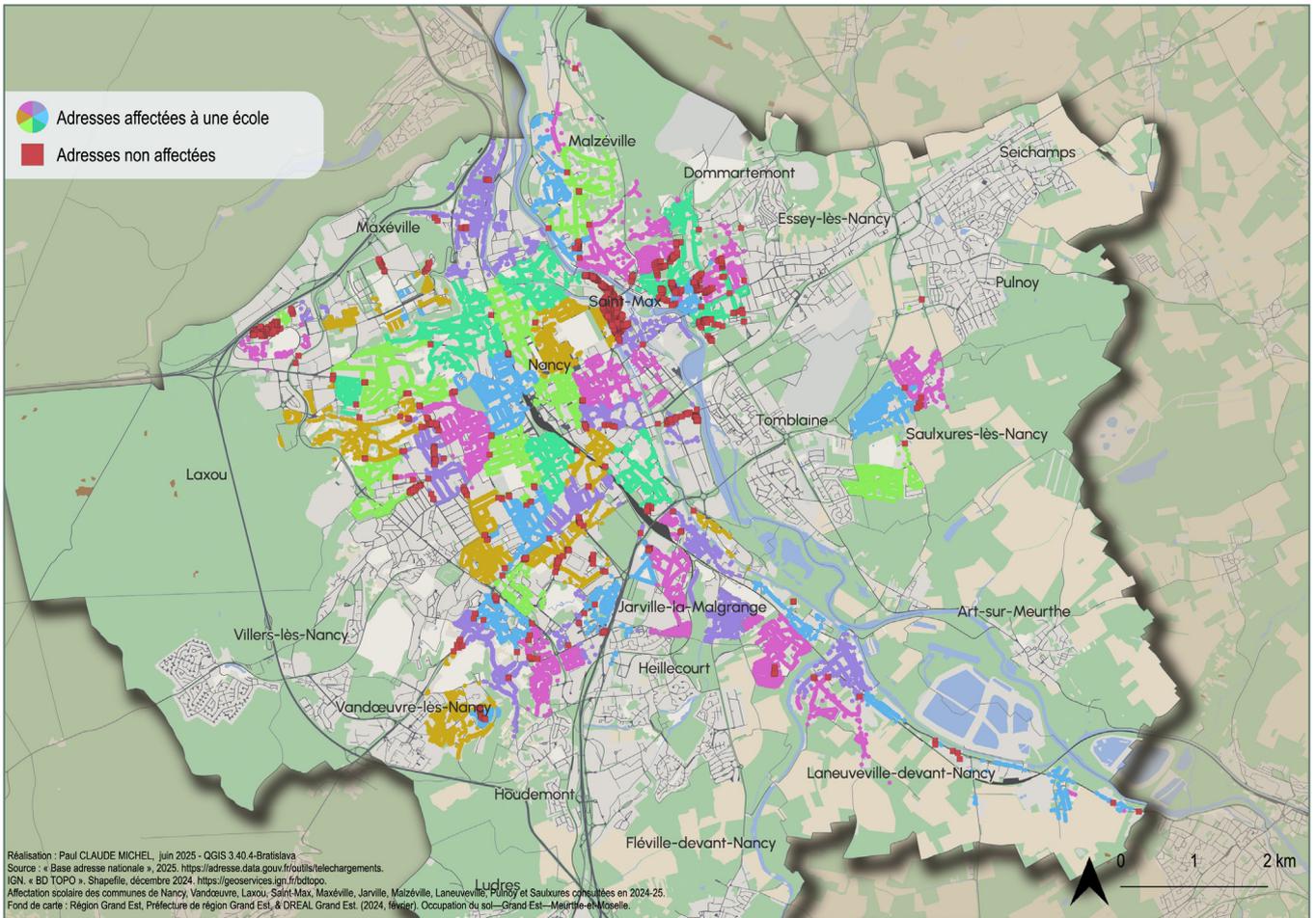


Figure 2 : Résultat géoréférencé du script Python d'affectation scolaire par adresses

tein avec toutes les rues du tableur d'affectation scolaire. La distance de Levenshtein désigne le plus petit nombre de caractères à modifier pour passer d'un nom de rue à l'autre, comme le montre le tableau 1. (2) On sélectionne ensuite le(s) nom(s) de rue avec la plus petite distance dans le tableau d'affectation scolaire avant de (3) vérifier que l'affectation porte sur l'ensemble de la rue ou à défaut, que les numéros de rue correspondent. Après quoi (4) on associe l'école à l'adresse.

Il est possible que plusieurs rues aient la même distance de Levenshtein, ou que le numéro de rue issu de la BAN ne corresponde pas à ceux du tableau d'affectation, dans ce cas, on n'effectue pas d'association.

Rue 1	Rue 2	Distance de Lv
Rue de Lorraine	Rue de Lorraine	0
Rue de Loraine	Rue de Lorraine	1
<u>Rue de Lorraine</u>	Lorraine	7
<u>Rue de Lorraine</u>	Alsace	13

Tableau 1 : exemples de distances de Levenshtein

Le choix d'utiliser la distance de Levenshtein est lié à l'absence de codage des noms de rue dans les tableurs d'affectation scolaire transmis par les communes. Si cette méthode permet d'associer des noms de rues agencés différemment (par exemple « Bd Voltaire » et « Voltaire »), elle associe aussi des rues qui ne devraient pas l'être. Par exemple, « rue de Lorraine » est plus proche de « rue de la Moraine » que de « Lorraine » ; ce qui implique de rester vigilant.

Globalement, une large majorité (88%) des adresses de la BAN ont vu leur affectation se faire sur la base d'une seule école potentielle. Moins de 2% disposaient de plusieurs écoles potentielles, mais qui en représentaient en réalité une unique. En effet, les tableurs d'affectation scolaire peuvent présenter plusieurs fois la même école. Sur l'ensemble des adresses traitées par le script, moins de 10% n'ont pas été affectées à une école, et moins de 4% ne l'ont pas été si l'on exclut Nancy.

La principale limitation de ce script réside dans l'utilisation de la distance de Levenshtein pour associer les rues. Paradoxalement, ce sont ainsi les données numériques, c'est-à-dire les tableurs d'affectation scolaire, qui sont les plus imprécises. On comprend aisément que plus une commune a de rues, plus il est possible que diffé-

rentes rues aient une distance de Levenshtein identique et plus le risque d'associer des mauvaises rues augmente alors. C'est notamment le cas à Nancy. En revanche, pour les autres communes de la métropole, l'impact reste acceptable. Un contrôle visuel à l'aide de la figure 2 montre des zones homogènes autour ou à proximité des écoles. Dès lors, on conserve ces affectations avec le script, sauf dans le cas de Nancy où nous privilégions la méthode par géoréférencement, puisque nous disposons à la fois du tableur d'affectation scolaire et d'une carte.

Croiser les bases de données

Les adresses issues des trois types de traitement sont rassemblées dans un fichier unique. À ce stade, toutes les adresses ont été affectées à une école, y compris celles ne faisant pas référence à un logement. L'information sur l'usage d'un bâtiment peut être trouvée dans la BD TOPO de l'IGN, mais cette dernière ne dispose pas de relation directe avec la BAN.



Figure 3 : Double jointure avec la BDNB

Pour faire le lien entre ces deux bases, on passe par la Base de données nationale des bâtiments (BDNB) du CSTB. Cette dernière rassemble de nombreuses informations concernant les bâtiments comme : le DPE, le numéro de parcelle, l'adresse et le lien avec la BD TOPO. Une simple jointure attributive entre, d'un côté, la BD TOPO et la BDNB et, de l'autre, la BDNB et la BAN permet d'attribuer un identifiant de bâtiment BD TOPO à une adresse. La figure 3 permet de visualiser cette double jointure.

Sur toute la zone d'étude (c'est-à-dire 58 814 adresses), on dénombre 1 592 adresses pour lesquelles aucun bâtiment issu de la BD TOPO n'a été associé. Pour-



Figure 4 : Un bâtiment avec plusieurs adresses

tant, si la BD TOPO dénombre 128 225 logements sur la métropole, on en compte 178 155 avec cette méthode par double jointure. Cette différence s'explique par le fait qu'un même bâtiment puisse disposer de plusieurs adresses et ainsi être compté plusieurs fois, comme le montre la figure 4. Pour notre usage cette inexactitude ne pose pas de difficulté puisque l'on souhaite uniquement savoir si une adresse dispose d'au moins un logement, pour l'intégrer à la carte d'affectation scolaire. En revanche, pour un usage statistique, cette méthode ne pourrait être employée.

Une méthode à perfectionner ?

Comme nous l'avons expliqué, le recours à la distance de Levenshtein pour associer les rues de la BAN à celles des affectations scolaires soulève plusieurs difficultés. Bien que d'autres méthodes puissent permettre de rapprocher les noms de rues de manière plus efficace (voir Negre, 2013), les limites inhérentes à notre approche subsistent : dès lors que l'on tente de faire correspondre des libellés de rues différents, des imprécisions surviennent.

Le choix de cette approche tient uniquement à l'absence de codification des données transmises par les communes. Pourtant, l'affectation scolaire se fait, nous l'avons dit, sur la base de l'adresse. Or, les communes sont aussi chargées de construire les BAL. Elles pourraient dès lors utiliser ce même codage pour les affectations scolaires. En procédant ainsi, il serait bien plus facile de travailler la « carte scolaire » et de la faire interagir avec d'autres bases de données, facilitant donc sa construction et sa réutilisation.

Remerciements

Les auteurs remercient le programme ORION pour sa contribution au financement du stage de recherche de Paul Claude Michel. Ce travail a bénéficié d'une aide de l'État gérée par l'Agence Nationale de la Recherche au titre du Programme d'Investissements d'Avenir portant la référence ANR-20-SFRI-0009.

Références

- Barrault-Stella, L. (2022). Que faire de la carte scolaire ? Pour une politique étatique d'égalisation des conditions d'accès à l'école. *Germinal*, 5(2), 286299. <https://doi.org/10.3917/ger.005.0286>
- Negre, E. (2013). Comparaison de textes: quelques approches... Manuscrit soumis pour publication. Repéré à <https://hal.science/hal-00874280>