

L'Atlas linguistique normand : aperçu et perspectives

Dans le cadre large de l'avenir des atlas linguistiques galloromans, nous proposons un état des lieux du chantier de l'Atlas linguistique normand. Notre contribution s'articule en deux volets, l'un consacré à la publication physique des matériaux normands et l'autre à la possibilité d'une transition numérique de cet ouvrage. Cette seconde partie, prospective, s'appuie entre autres sur la réflexion du projet APPI (Baiwir/Kaisin 2020).

1. Le chantier de l'ALN physique

La publication, en 2019, du 5^e et dernier volume marque le terme du chantier de l'ALN, dont les enquêtes principales se sont déroulées de 1970 à 1976. Si les deux premiers volumes ont été publiés rapidement (1980 et 1984), le troisième a été longtemps retardé en raison de la complexité de la saisie informatique, puis de la réalisation des cartes, qui se voulait automatisée, mais a subi tous les inconvénients de la nouveauté. Et c'est peu dire. Entre temps, le groupement de chercheurs du CNRS (GDR 9) auquel plusieurs dialectologues appartenaient avait été dissous, ce qui avait amené P. Brasseur à quitter l'institution pour rejoindre l'Université, après un passage à l'Institut d'études créoles d'Aix-en-Provence. D'autres tâches l'ont alors occupé et il n'a pas pu consacrer aux parlers normands le temps qu'il aurait souhaité. En outre, le financement n'était plus assuré par le CNRS et il a fallu le secours de plusieurs institutions et organismes normands pour réunir les fonds nécessaires à la publication des deux derniers volumes.

Dans ce bref aperçu, nous souhaitons donner une vue d'ensemble de l'ALN en montrant sa spécificité par rapport aux autres atlas du domaine d'oïl. En effet, outre les données des cartes et des marges, l'ALN comporte plusieurs particularités. Par exemple, le 3^e tome fournit des renseignements utiles sur les 697 informateurs, sauf leurs noms, qui figurent cependant dans les cahiers d'enquêtes déposés aux Archives de la Manche à Saint-Lô. Le 5^e tome inclut le questionnaire, un index des sens français des mots dialectaux de l'*Atlas* ainsi que l'équivalent en API des notations de l'alphabet de Rousselot-Gilliéron qui est employé dans l'ALN.

Des *addenda* et *corrigenda* des volumes précédents se trouvent dans les tomes 4 et 5. En effet, malgré tout le soin qui a été mis au classement des données, les dépouillements révélaient, à mesure de l'avancement du travail, un grand nombre de formes isolées dans les notes marginales produites hors questionnaire qui auraient trouvé une place plus « naturelle » dans l'un des volumes précédents. Dans une telle masse de données, des oublis et quelques erreurs sont inévitables. Dans la mesure du possible, la place que les *addenda* auraient dû occuper dans l'atlas est indiquée, ce qui permettra de les intégrer dans une version ultérieure numérisée.

1.1. Le réseau d'enquêtes

Le domaine normand a été envisagé sous l'aspect historique. De ce fait, le réseau de l'ALN s'étend sur trois points situés sur le territoire du Royaume-Uni, mais ne comprend pas le Perche (sud-est de l'Orne). D'autre part, de manière variable quantitativement, les Îles anglo-normandes de Jersey et Guernesey ont fait l'objet de plusieurs enquêtes distinctes, les parlers de ces îles n'étant pas considérés dans leur globalité, mais en fonction des différentes « paroisses » (unités administratives, de superficie variable, comparables aux communes françaises). Des données ont donc été recueillies dans 9 des 10 paroisses de Guernesey¹ et dans 9 des 12 paroisses de Jersey².

Les 111 points restants sont répartis dans les cinq départements de la région Normandie (Manche, Calvados, Orne, Eure et Seine-Maritime). À Guernesey et Jersey, les données sont localisées avec précision dans les « paroisses ». Par comparaison, l'ALF possède 4 points dans les Îles (les mêmes plus Aurigny) et 31 sur le continent.

1 — Nous n'avions aucun informateur dans la paroisse urbaine de St-Pierre-Port. St-Martin, où quelques relevés ont été effectués, a été oublié sur la carte n° 7 'noms officiels des localités enquêtées'.

2 — Nous n'avions pas d'informateur dans la paroisse urbaine de St-Hélier, ni dans celles de St-Sauveur et St-Clément. Mais les secteurs de la Moie (au sud de la paroisse de St-Brelade) et des Landes (à l'extrême ouest de St-Ouen) sont mentionnés spécifiquement.

La densité du réseau d'enquêtes de l'ALN est sensiblement identique sur l'ensemble du territoire de la Normandie continentale (d'un point d'enquête pour 213 km² dans le Calvados à un point pour 296 km² dans l'Orne).

1.2. La fabrication des différents tomes

Les cartes des deux premiers volumes ont été réalisées de manière artisanale au moyen de lettres à décalquer, appartenant aujourd'hui à un autre âge. Chaque signe, figurant sur des planches fabriquées à cet effet, était reporté individuellement ; le tout était maintenu tant bien que mal sur le support par un film de fixateur. Les risques de dégradation des signes en raison de leur fragilité et d'une mauvaise fixation sur le support étaient importants. Certains blocs de signes pouvaient aussi se déplacer pour réapparaître ici ou là, obligeant à des vérifications jusqu'à l'étape du cliché off-set. Les données figurant en marge étaient composées avec les moyens relativement classiques d'une machine à écrire pourvue d'une mémoire permettant un traitement du texte sommaire et surtout des corrections aisées. Les deux éléments étaient ensuite assemblés. Une seule taille de police était disponible : l'imposant format des deux premiers tomes est à mettre au compte de cet inconvénient.

Le troisième volume a été réalisé informatiquement, mais avec les moyens du moment nécessitant une saisie complexe, particulièrement lourde, même si elle a été effectuée avec une compétence, une assiduité et une patience infinies par Jacques Tourrel, à Aix-en-Provence. Mais la production des cartes, opérée à Brest, a été une longue et pénible épreuve, le procédé mis en place semblant particulièrement rigide : les coordonnées de chaque point étaient fixées d'avance, ce qui ne permettait pas un placement facile des réponses multiples, par exemple. La modulation de la taille des caractères n'a pu être obtenue que pour quelques cartes (ex. : 790 « marchand de bestiaux mesquin ») et souvent de manière globale (ex. : 836 « diarrhée »), ce qui rend certaines données difficilement lisibles tant la taille de la police est inutilement réduite. Le résultat n'est certainement pas en rapport avec l'énergie consacrée.

Le quatrième volume, composé de manière très professionnelle par l'éditeur à partir du fichier informatique de P. Brasseur a évité tous ces inconvénients. Il a cependant été nécessaire de transcoder la police de caractères dans une police compatible avec le logiciel de l'imprimeur. Cette étape a été délicate, étant donné le grand nombre de combinaisons de signes. Le logiciel de dessin utilisé n'exige cependant ni compétence en informatique ni habileté particulière et permet de placer les données de manière précise, tout en modulant finement la taille des caractères. Cette technique

avait l'avantage de permettre une cartographie directe, la vérification des cartes se faisant ainsi en une seule étape.

Le procédé de fabrication du cinquième volume a été encore plus simple, puisque la police de caractères originale a été conservée sans modification. Le résultat est particulièrement satisfaisant, la lisibilité parfaite malgré une réduction du format initial.

1.3. *L'ordre des mots dans les réponses multiples*

Dans le cas de réponses multiples, l'ordre des réponses est soumis aux contraintes de place disponible. En aucun cas, il ne s'agit d'un ordre de fréquence, impossible à établir dans le cadre d'une telle enquête. La mention « rare » signale cependant des mots peu présents à la mémoire des informateurs et qui ont nécessité plusieurs investigations. La mention « vieux » porte le plus souvent sur des mots qui correspondent à des techniques obsolètes et qui n'ont pas été affectés à d'autres emplois. Le traitement de nombreuses cartes selon le principe des aires à dominantes dégagées fera certainement difficulté pour une future exploitation numérisée des cartes, à moins peut-être de rétablir préalablement les formes en question dans chaque point de l'aire concernée. Il en est de même pour l'utilisation du sigle ~ qui sert à noter des réponses identiques ou partiellement semblables, notamment dans les locutions.

1.4. *Les données phraséologiques*

En plus des commentaires d'ordre linguistique, les marges des tomes 1 à 4 contiennent de nombreuses notes à caractère ethnographique. Un dépouillement exhaustif a permis de dénombrer les rubriques suivantes :

- Proverbes, dictons et expressions proverbiales : 156.

Ex. : *Guéret mottu, poulain bourru, homme n'a jamais perdu* (123) ; *rîn qui valle qué nous counaît vaut mux que rîn qui valle qué nous ne counaît pas* (1279) ; *l'auge est faite pour les cochons* : qui se ressemble s'assemble (788) ; *un bon dégel n'est jamais chaud* (568*) ; *en juillet tout coucou devient émouchet* (672*).

- Formulettes et comptines : 79 (+ celles sur les doigts de la main).

Ex. : *Sapré rignolet, jamais je ne te pardonnerai* (469*) ; *arc-en-ciel de Saint-Martin, si tu passes dans mon jardin, je te coupe les boudins* (539*).

- Devinettes : 6.

Ex. *qui qu'a mille urs qui n'y vét goutte : ch'est le daille* » (4) : qui a mille yeux et qui n'y voit goutte ? C'est le dé.

- Autres rubriques marginales : des onomatopées, appellations plaisantes, expressions métaphoriques, mots employés par dérision, vocabulaire enfantin ou argotique sont parfois répertoriés comme tels, même si leur classement peut être discutable.

1.5. Les données ethnographiques

Certaines font l'objet de relevés systématiques. Ex. : « l'assolement » (108), « ce qu'on voit dans la lune » (534*), « animal imaginaire » (les farces faites aux naïfs) (618). Ces observations peuvent être générales comme sous « râtelier » (788*), mais elles sont le plus souvent localisées. La forme des objets, leur utilisation, les matériaux employés, la réalisation des travaux quotidiens, leur périodicité, etc. sont souvent décrits. De plus, les trois premiers volumes comportent aussi de nombreuses rubriques marginales de type ethnographique concernant divers aspects :

- Médecine (populaire) et remèdes : 128.

Ex. : Infusions de racines de bardane pour soigner les furoncles (374), graisse de poule pour soigner les engelures (564*).

- Prières de guérisseurs : 8.

- Croyances : 98. (Elles ne sont cependant pas toujours répertoriées ou même identifiées comme telles, mais souvent données comme des paroles rapportées. Certaines font, par ailleurs, l'objet de questions). Ex. : La foudre ne tombe jamais sur un érable (441*) ; si on laisse le trépied vide dans la cheminée, on peut faire tarir les vaches (973).

- Coutumes et usages : très nombreux. Cela concerne, par exemple, l'usage de l'arrière-faix ou du colostrum, au vêlage (820, 821).

- Plaisanteries et farces traditionnelles : très nombreuses.

Ex. : (quand le temps s'assombrit) : *le soleil tombe par morceaux* (553*). (D'un cheval rétif) : *il est franc comme 19 sous* (884*) ; *il éclé à mouques comme un bourdon qu'est au tiers* (1195*) : il contemple les abeilles comme un bourdon au piquet (de quelqu'un qui reste à ne rien faire).

- Jeux : 8 (outre les cartes qui leur sont consacrées).

- Illustrations : elles sont abondantes dans les volumes 1 et 2, mais absentes des volumes suivants, malgré une riche iconographie recueillie sur le terrain. Ce sont des dessins d'après les photographies prises sur le terrain.

Enfin, nous ne revenons pas sur la méthodologie de l'enquête, le rapport enquêteur / enquêtés, le choix des points, le recueil

des données en général et leur présentation cartographique, tous points déjà présentés en détail dans quelques articles (Brasseur 1973 et 2005a, par exemple) ainsi que dans les introductions des volumes 4 et 5.

Pour conclure ce bilan, nous aimerions souligner que l'exploitation des cahiers d'enquêtes et des enregistrements réalisés au cours des sept années d'enquêtes sur le terrain a été menée systématiquement et que la totalité des données recueillies aura finalement trouvé place dans l'un des cinq volumes de l'*ALN*.

2. Les pistes pour une numérisation

2.1. Plusieurs possibilités, un seul objectif

La numérisation de l'*ALN*, comme celle de tout atlas linguistique, peut s'envisager de différentes manières. La première, la plus simple, est la numérisation des cartes en mode image. Cette option, qui résout à elle seule les problèmes d'accessibilité physique des ressources, est à l'étude au sein de la région Normandie, qui pourrait assurer à la ressource un large public. Dans cette optique, une indexation des notions traitées peut utilement compléter les scans.

Toutefois, nous pouvons également envisager une numérisation individuelle de l'ensemble des matériaux, qui devraient être réencodés manuellement. C'est le parti pris par l'équipe du projet APPI à Lille, dont le fonctionnement pourrait être exporté pour les matériaux normands.

2.2. Examen d'une intégration dans la nomenclature du projet APPI

Le premier réseau auquel il conviendrait d'intégrer les données de l'*ALN* est évidemment celui des autres atlas linguistiques gallo-romans, mais la comparabilité des données n'est pas une évidence. « Cela tient, bien évidemment, à la nature de cette entreprise [des atlas par région], qui s'est déroulée sur près de 50 ans, à l'absence d'un véritable plan de publication, chaque chantier avançant au gré des disponibilités des chercheurs, parfois bénévoles, les dates du début des enquêtes variant beaucoup d'une région à l'autre. L'atlas du Massif Central, par exemple, était terminé depuis longtemps quand les enquêtes ont commencé dans le domaine normand » (Brasseur 2005b : 163). L'outil informatique est nécessaire pour lever une part de ces disparités, mais il n'est pas suffisant ; rien qu'au niveau macrostructurel, une analyse sémantique doit être effectuée afin d'évaluer le contenu précis des cartes, dont le titre ne rend pas toujours compte. C'est à cet exercice que nous nous sommes livrés pour un échantillon des matériaux.

L'objectif était d'examiner la possibilité d'un rapprochement avec les matériaux contigus du domaine picard, dont le traitement informatique est en cours dans le cadre de l'APPI. Notre corpus de test est donc composé des 205 notions actuellement disponibles dans la macrostructure du projet picard. Cette macrostructure est le résultat d'une analyse sémantique des matériaux de l'ALPic, de l'ALW et de l'ALF (voir ce volume, Baiwir/Kaisin). Les points d'enquête de ce dernier actuellement intégrés dans la base sont évidemment ceux du domaine picard, mais l'intégration des points normands ne représenterait pas de difficulté majeure.

Un premier tri montre que dans 128 cas, les notions telles qu'elles sont définies sont également traitées dans l'ALN, que ce soit sous un titre identique ou sous un intitulé différent (le 'porte-seau' est intitulé *joug*, la 'sieste' est traitée sous la notice *méridienne*). Dans ces cas, le contenu sémantique est identique, les matériaux ont été recueillis dans tous les points et pourraient facilement être introduits dans une base de données commune.

Pour 41 notions, des matériaux existent également, mais ne concernent qu'une part plus ou moins importante de points de l'enquête normande ; ce sont les enquêtes partielles, dont les matériaux sont édités en marge des cartes. Dans certains cas, ces marges comprennent plusieurs dizaines de réponses ; ainsi de la notion 'babeurre', avec 48 formes sur les 114 points d'enquête que comporte l'ALN. Ces enquêtes partielles, publiées autour des cartes, interrogent la façon dont l'archivage des scans doit être effectué : doit-on conserver l'intégrité des pages des volumes physiques, avec le risque d'une mauvaise interprétation si deux notions renvoient au même scan ? Ou est-il préférable d'isoler les matériaux de chaque notion ? On le voit, les questions techniques et scientifiques ne doivent pas être traitées séparément.

Si 8 notions sont totalement absentes de la nomenclature de l'ALN, dans les 28 derniers cas, des matériaux existent, mais présentent des divergences à envisager avant de se prononcer sur la validité d'un rapprochement. Quatre catégories ont été identifiées :

- les différences de nombre des substantifs (9 cas) : ce cas ne génère généralement pas d'incidence sur les formes et partant, sur leur équivalence avec les matériaux de l'APPI ;

- les différences flexionnelles des verbes (2 cas de formes infinitives dans l'APPI et conjuguées dans l'ALN) : il convient d'analyser l'impact de la flexion sur le radical et de vérifier si l'infinitif n'est pas traité en enquête partielle ;

- les découpages sémantiques différents (8 cas) : une sélection des matériaux doit être opérée ; c'est le cas par exemple de la notion 'limace', répartie sur deux cartes dans l'ALN, intitulées (*une*

petite) limace grise et *(une grosse) limace rouge* ; dans celles-ci, les hyponymes éventuels doivent être écartés ;

- les approximations formelles (9 cas de catégories grammaticales différentes pour un contenu sémantique proche) : ici aussi, un traitement au cas par cas doit être effectué ; ainsi de l'entrée APPI 'à bouillons', qui peut être rapprochée de la carte ALN intitulée *(des) bulles (sur l'eau, quand il pleut) / il y a des bulles sur l'eau*.

On le voit, un rapprochement est donc largement possible sur la base de la seule nomenclature de l'ALPic. Celle-ci devrait ensuite être complétée par toutes les notions spécifiques à l'ALN, particulièrement riche dans certains champs sémantiques : ainsi, le vocabulaire de l'élevage des porcs est représenté par 15 cartes (sans compter les cartes liées à l'abattage et à la transformation de la viande), là où l'ALPic en possède 3 ! Le vocabulaire de la pêche et des poissons est également d'une grande richesse. On touche là à la raison d'être des atlas régionaux : ils enrichissent notre connaissance des spécificités locales, même si parfois, cette pratique participe à la fragmentation du panorama dialectal de la Galloromania de façon artificielle.

2.3. Enrichissement des matériaux : étymologisation et étiquetage

Différents enrichissements ouvriraient la possibilité d'interrogations variées. La première couche serait une lemmatisation de toutes les données, qui permettrait de faire dialoguer les formes avec leurs cognats d'autres atlas. Ainsi, les formes *gòg* ('noix', carte 361 de l'ALN) seraient reliées aux formes picardes *goke* (APPI 'noix'), mais aussi à *gaye* (idem), dont la parenté est pourtant moins évidente. L'analyse et la lemmatisation pourraient également porter sur des morphèmes, comme les préfixes ou les suffixes.

Dans un second temps, l'étymologisation autoriserait les renvois vers le FEW, à l'instar de ce que prévoit l'équipe de l'APPI (voir ce volume, Baiwir/Kaisin). Le recours à un « étiquetage FEW » permettrait d'ancrer les données dans un contexte galloroman, mais aussi de naviguer entre les différents domaines linguistiques. Les potentialités d'un tel réseau ne feront que s'élargir au fur et à mesure de la mise en ligne de la version dynamique du FEW et de l'ouverture des fonctionnalités d'interrogation de cet ouvrage incontournable (voir ce volume, Robecchi ou Renders 2015).

3. Conclusion

Il convient de souligner que la mise en ligne des matériaux d'enquêtes dialectologiques nécessite l'accessibilité de ces matériaux. À ce titre, l'ALN fait figure de modèle, eu égard à l'achèvement

du programme d'édition des données. Les cinq volumes physiques contiennent en effet l'ensemble des matériaux recueillis dans le cadre de la campagne des enquêtes. Numérisation, enrichissement et mise en réseau sont les trois niveaux d'une mise en valeur des riches matériaux de l'ALN, dans un but unique : que ceux-ci soient davantage exploités et qu'ils figurent, avec ceux des autres atlas, dans la boîte à outils de tout romaniste, à côté des ressources lexicographiques traditionnelles. Il s'agit pour l'instant d'un espoir, mais les conditions pour que celui-ci devienne réalité semblent accessibles.

Patrice BRASSEUR,
Avignon Université
Esther BAIWIR,
Université de Lille³

Références

- ALF = J. GILLIÉRON & E. EDMOND, *Atlas linguistique de la France*, 1920 cartes, Paris, Champion, 1902-1910.
- ALN = P. BRASSEUR, *Atlas linguistique et ethnographique normand*, 5 vol., Paris, Éd. du C.N.R.S. / Caen, Presses universitaires de Caen / Strasbourg, ELiPhi, 1980-2019.
- ALPic = F. CARTON & M. LEBÈGUE, *Atlas linguistique et ethnographique picard*, 2 vol. Éditions du CNRS, 1989-1998.
- ALW = L. REMACLE, E. LEGROS, J. LECHANTEUR, M.-T. COUNET, M.-G. BOUTIER, E. BAIWIR, *Atlas linguistique de la Wallonie*, Liège, Université de Liège (10 vol.), 1953.
- E. BAIWIR & C. KAISIN, « L'ALPic en ligne et le projet d'un Atlas pan-picard informatisé (APPI) », *Bien dire et bien apprendre*, t. 35, 2020, p. 47- 60.
- P. BRASSEUR, « L'atlas linguistique normand après trois années d'enquêtes », in *Annales de Normandie* XXIII-3, 1973, p. 253-263.
- P. BRASSEUR, « L'enquête dialectologique. Les atlas linguistiques », dir. L.-J. Calvet et P. Dumont, *L'enquête sociolinguistique*, Paris, L'Harmattan, 1999, p. 25-59.
- P. BRASSEUR, « Enquêteur et enquêtés l'enquête dialectologique en Normandie », dans *Mélanges Lothar Wolf je parle donc je suis... de quelque part*, dir. B. Horiot, E. Schafroth et M.-R. Simoni-Aurembou, Lyon, Centre d'études linguistiques de Lyon 3, 2005a, p. 67-77.

3 — Univ. Lille, ULR 1061 - ALITHILA - Analyses Littéraires et Histoire de la Langue, F-59000 Lille, France.

- P. BRASSEUR, « L'indexation de l'Atlas linguistique et ethnographique normand ; pour un trésor des parlers dialectaux », dir. J. KABATEK, Cl. D. PUSCH et W. RAIBLE, *Romanistische Korpuslinguistik II Korpora und diachrone Sprachwissenschaft / Romance Corpus Linguistics II Corpora and Diachronic Linguistics. Freiburg-in-Breisgau, 11-13 septembre 2003*], Tübingen, Narr, 2005b, p. 163-175.
- FEW = W. von WARTBURG, *Französisches etymologisches Wörterbuch*, Bonn, Tübingen, Bâle, Nancy, 1922-2002.
- P. RENDERS, *L'informatisation du Französisches Etymologisches Wörterbuch*, TraLiRo - Linguistique de corpus et philologie informatique, Strasbourg, 2015.
- M. ROBECCHI, « La place des *Atlas* dans la rétro-conversion du FEW », *Bien dire et bien apprendre*, t. 35, 2020, p. 147-168.