L'INFORMATISATION DES ATLAS LINGUISTIQUES DE LA FRANCE

par Jean LE DÛ
Université de Bretagne Occidentale

LABURPENA

Duela urte batzuk, CNRS-ek informatikara jotzea erabaki zuen bildumaren azken aleak (hogei bat) burutzeko. 1989an hartu zuen artikulu honen egileak lan horren ardura, GDR 9-ren zuzendaritza hartuz, horretarako kontratatu zen informatiko baten laguntzarekin.

Teknikoki, bi alderdi ezberdin dauzka proiektuak: ikertzaileek mikro-ordenagailutan egiten duten datu-bilketa eta mapak egitea. Hamaika arazo sortu dira fase guztietan, hala nola: leku ezberdinetako alfabeto fonetikoen arteko konpatibilitatea; mapa eta datu periferikoak aldi berean jasotzea, datu basea osatzeko; area bakoitzaren tratamendu informatikoa eta aurkibideen, sistema ezberdinen konpatibilitatearen, etabarren aurrikuspena.

SUMMARY

The CNRS (French National Scientific Research Centre) decided, a few years ago, to adopt computers systems to prepare the last volumes (about 20) of the collection. I was entrusted with this task in 1989 when I became director of the GDR 9, with the aid of a computer specialist who was employed for this purpose.

Technically, the project consists of two, very different parts: introduction of data from researchers into computers and preparation of maps. Numerous problems have arisen during all stages of the work, for example: compatibility between phonetic alphabets of different locations, simultaneous translation of the maps and of the peripheral data to establish data bases, computer processing of resulting areas and design of indexes and an estimation of the compatibility of the different systems, etc.

IKER 7, 299-317

1. Introduction

Le chantier des Atlas linguistiques de la France par régions est une entreprise de longue haleine, puisque c'est en 1939[1] qu'elle a été conçue par Albert Dauzat. Il s'agissait avant tout de compléter l'A.L.F. en constituant un questionnaire comprenant «une partie générale commune à toute la France et une partie régionale. Sa base sera le questionnaire de l'atlas Gilliéron, qui contient beaucoup de mots et formes bien choisis, mais qui manque de termes régionaux. La partie commune permettra la comparaison entre les deux atlas [2]».

L'énormité de la tâche, la quantité et la spécificité des données ne permettaient pas d'envisager l'élaboration d'un atlas unique, mais d'un ensemble d'atlas régionaux ayant chacun son propre directeur et ses enquêteurs. Dans la préface au premier volume de l'Atlas linguistique de la Gascogne de Séguy, publié en 1954, Dauzat nous apprend qu'«Un nouveau questionnaire a été élaboré pendant les années d'occupation et les suivantes dans une de mes conférences à l'Ecole Pratique des Hautes Etudes, avec le concours de futurs enquêteurs dialectologues originaires de divers points de la France».

Les premiers travaux sont réalisés à partir de 1946 par des chercheurs isolés -au sens du C.N.R.S.- c'est-à-dire non rattachés à une structure: Mgr Gardette à l'Institut catholique de Lyon, Jean Séguy à l'université de Toulouse, Nauton préparent leurs questionnaires, enquêtent et publient en obtenant des aides individuelles du C.N.R.S.

En 1962 est créée au sein du C.N.R.S. une Commission des atlas linguistiques ayant pour mission «de susciter la mise en chantier d'atlas linguistiques dans toutes les provinces de la France, et de coordonner ceux qui sont en cours de réalisation. Dans ce but elle réunit une fois par an les responsables des divers atlas»[3]. Présidée par Michel Lejeune, elle a pour secrétaire Mgr Gardette. Cette structure s'est maintenue jusqu'à nos jours sous diverses appellations: R.C.P. 160 jusqu'en 1977, G.R.E.C.O. 9 de 1977 à 1988 et enfin G.D.R. 9. Elle est dirigée tour à tour par Mgr Gardette, Gaston Tuaillon (de 1973 à 1979), Jean-Claude Bouvier (jusqu'en 1984), et Jean Lanher, auquel j'ai succédé au début de 1989.

Il semble que Dauzat envisageait initialement de ne couvrir que la France romane, comme l'avait fait Gilliéron: «La France romane sera divisée en dix régions qui auront chacune leur atlas, dont la préparation et l'enquête sont confiées à des spécialistes originaires de la contrée»[4]. Le chantier s'est depuis

1. Le programme de l'atlas linguistique régional de la France a été exposé dans *Le français moderne*, Paris, d'Artrey, avril 1939, 97 et octobre 1939, 289.

2. Dauzat, *Nouvel atlas linguistique de la France par régions*, Luçon, S. Pacteau, s.d. [1942], 5 (cité dans Sever Pop *La dialectologie*, 1950, p. 5).

3. Rapport de Mgr Gardette au Directeur du C.N.R.S., 17 septembre 1963.

4. Dauzat, *Le village et le paysan de France*, Gallimard, 1941, p. 186.

lors étendu à l'ensemble du territoire métropolitain et même au-delà pour couvrir, avec l'atlas de la Réunion, le nombre de 25 régions.

En dépit des difficultés de tous ordres inhérentes à la longue durée –abandons, départs à la retraite, décès– l'entreprise s'est poursuivie pour aboutir en 1991 à la publication de 64 atlas. Il en reste actuellement moins de vingt à publier.

2. Historique

2.1 Les premières étapes de l'informatisation

J'avais envisagé, dès 1980, de publier le nouvel atlas linguistique de la basse Bretagne par des moyens informatiques, en prévoyant une exploitation ultérieure des données. Le logiciel de saisie était opérationnel, des cartes provisoires avaient été tracées[5], mais des contretemps ont retardé la publication de cet atlas.

De son côté, Marthe Philipp développait, avec l'aide des informaticiens du centre de Cronenbourg, un système extrêmement performant de saisie et de traçage de cartes qui a permis la publication du volume II de l'atlas de l'Alsace en 1984. Il n'a malheureusement pas été possible d'adapter ce logiciel (qui ne fonctionnait que sur gros système) à l'ensemble des équipes. Dès 1981, le comité directeur du G.R.E.C.O. 9 avait évoqué la nécessité de concevoir des index informatisés de formes typisées qui auraient été saisis dans quatre centres puis regroupés dans une banque de données à Strasbourg. Ce projet n'a pas eu de suite, et il a fallu attendre le 15 mai 1987 pour que se tienne à Paris une réunion de réflexion sur l'informatisation des atlas, au cours de laquelle furent envisagées les techniques de saisie des données et la manière de les traiter. Une nouvelle rencontre sur le même sujet eut lieu le mois suivant à Strasbourg, au cours de laquelle Madame Salomon-Bayet, directeur-adjoint des S.H.S.[6], évoqua la création d'une nouvelle série d'atlas qui sortirait de l'adoption des techniques nouvelles. La liste des derniers atlas restant à publier selon l'ancienne formule était dressée: le volume V de l'*Atlas de la Franche-Comté* de M. Dondaine vient de paraître; c'est le volume VI de l'*Atlas du Languedoc Occidental* de Xavier Ravier qui va clore la série. Tous les autres volumes restant, quel que soit leur degré d'avancement, devront désormais se conformer à la nouvelle procédure.

5. En raison de mes tâches d'enseignement, je n'avais pu finir à temps les enquêtes et surtout les dépouillements. De plus, B. Tanguy, alors I.T.A., qui avait commencé la saisie, demandait à passer dans le corps des chercheurs et abandonnait le projet.

6. Sciences Humaines et Sociales, une des subdivisions du C.N.R.S.

A la fin du mandat de Monsieur Lanher, le C.N.R.S. me proposa de prendre en charge la mise au point des procédures d'élaboration de la nouvelle série des atlas linguistiques, tâche qui devenait la mission prioritaire du G.D.R. 9.

2.2. Conception du projet

Le 3 février 1989, je présentai à l'Assemblée générale la conclusion des travaux d'une commission[7] sur l'informatisation des atlas, réunie à Paris les 8 et 9 décembre 1988 à la demande du Comité scientifique. La discussion proprement dite a été précédée d'une réflexion sur le problème de l'exploitation ultérieure des atlas. Il est clair, en effet, que si le problème immédiat est la publication des derniers volumes, l'un des objectifs fondamentaux de la recherche au sein du G.D.R. 9 va être désormais de tirer tout le profit scientifique possible de ce formidable ensemble de données que constitue la collection. On a pu, dès lors, aborder l'objectif premier de la réunion qui était la fourniture à l'imprimeur de cartes prêtes à clicher.

L'un des problèmes majeurs était le choix du format des ouvrages. L'adoption du format A4[8], qui entraînait *ipso facto* l'adoption de cartes à symboles sur le modèle de l'atlas linguistique de la Wallonie, n'a pas été retenu. Ce changement aurait en effet constitué la transformation radicale d'une collection en cours de publication et une rupture épistémologique avec la tradition française: depuis Gilliéron, on le sait, les atlas français de première génération fournissent des données brutes en alphabet phonétique, l'interprétation ne devant venir que dans un deuxième temps. C'est pour respecter cette conception que la commission a opté pour le format A3, le minimum indispensable à la lisibilité des cartes, qui présente en outre l'avantage de leur conserver l'aspect qu'elles avaient dans l'ancienne formule.

La discussion a alors tourné autour des modalités pratiques de l'opération, qui peuvent se résumer à deux questions majeures:

-qui effectuera la saisie?

-comment et où seront tracées les cartes destinées à être remises à l'imprimeur?

Dans ses conclusions, la commission décide de traiter séparément les deux opérations:

a) la saisie devra, dans la mesure du possible, être faite par les chercheurs eux-mêmes, seuls à même de vérifier les données. Cette solution, étant donné

7. Les participants à cette réunion étaient: Mme Salomon-Bayet (C.N.R.S.), Mme M.J. Dalbéra, MM. M. Contini, Ph. Dalbéra, J. Lanher, Y. Le Berre, J. Le Dû, (enseignants-chercheurs), MM. R. Calvez et D. Hénaff (Centre de Calcul de l'Université de Brest), Mme García-Mouton et M. Antonio Quilis, collègues espagnols.

8. 21 x 21,7 cm.

la dispersion des diverses équipes, implique l'adoption du micro-ordinateur. Tous les chercheurs n'étant pas experts en informatique, le logiciel devra être de maniement simple, permettant par exemple de saisir au choix –ou en alternance– toutes les données d'une même carte ou l'ensemble des données d'un seul point;

b) pour le traitement, on préférera au contraire centraliser les opérations. Seul en effet un centre de calcul peut assurer le stockage et le traitement des millions d'informations recueillies par les divers chantiers surtout si on envisage une mise en commun ultérieure des données. Un gros système calculant à très grande vitesse permet l'unification des procédures, donne accès à des tables traçantes performantes, et dispose de la capacité de mémoire nécessaire à la constitution de bases de données.

2.3. Organisation du travail

Pour des raisons de commodité, le travail de conception et de mise en route du projet s'est fait à Brest. Un poste d'informaticien sur contrat à durée déterminée, créé spécialement pour cette tâche, n'ayant pu être pourvu avant la fin de l'année 1989, ce fut d'abord Didier Hénaff, technicien au Centre Universitaire de Calcul, qui commença le travail. Après le recrutement d'André Cherblanc, on a décidé d'un partage des tâches entre les deux informaticiens, Didier Hénaff continuant à assumer la responsabilité du logiciel de saisie sur micro-ordinateur tandis qu'André Cherblanc mettait au point sur gros système le logiciel de traçage des cartes. Des réunions de travail régulières nous ont permis de résoudre les nombreux problèmes qui apparaissent au fur et à mesure de l'avancement des travaux. C'est à Jean-Claude Potte, qui travaille à Thiers, qu'est revenue la tâche redoutable de saisir le premier volume de la collection, et donc d'expérimenter le logiciel qui s'est, pour une bonne part, inspiré de ses critiques et remarques. Une liaison téléphonique par MODEM permet d'intervenir de Brest sur son ordinateur et d'échanger directement des données. Ce procédé est maintenant étendu aux autres chercheurs.

Le poste d'André Cherblanc vient d'être pérennisé (IR2), ce qui fait que le G.D.R. 9 dispose désormais des services d'un informaticien à temps complet. Grâce à cette nomination et aux crédits importants d'équipement et de fonctionnement dont il est doté par le C.N.R.S. le groupe peut, me semble-t-il, envisager l'avenir avec confiance.

3. Description du logiciel de saisie

Le logiciel est conçu pour micro-ordinateurs IBM et compatibles. L'atlas de la Corse dispose de son propre logiciel fonctionnant sur Macintosh[9].

9. Cet atlas constitue une partie d'un vaste projet de banque de données subventionné par la Région Corse.

L'atlas du Pays Basque a également son propre logiciel, fonctionnant sur IBM. Il fait exception puisque, dans le cadre d'un accord passé entre l'Académie Basque et le C.N.R.S., il sera publié à part et couvrira l'ensemble du territoire basque. De nombreux échanges ont eu lieu entre l'équipe du G.D.R.9 et celle des atlas corse, basque et piémontais occidental afin d'harmoniser les conceptions.

Il n'est pas question ici de retracer les métamorphoses du logiciel de saisie que nous décrivons ici: d'abord dérivé de Base4 et tournant sous Windows, il fonctionne actuellement sous DOS (version 4) et comporte plusieurs modules dont les programmes sont écrits en langage Turbo C++.

Ces modules, constitués en fonction des données à saisir, se répartissent comme suit[10]:

- *le module nomcarte* permet la saisie du nom de la carte[11].

- *le module saisie* est destiné à la partie la plus délicate de l'opération: la saisie des entrées en caractères phonétiques ainsi que des formes doubles dotées d'un numéro de renvoi sur la carte. La disparité des notations phonétiques a amené à la création pour chacun des atlas de son propre alphabet phonétique avec codage particulier. Une comparaison attentive de plusieurs atlas a en effet révélé que chaque auteur a apporté pour son propre usage des modifications au système Rousselot, non seulement en fonction des particularités de son terrain d'enquête, mais aussi à partir de ses propres conceptions de la notation et donc, en dernière analyse, de sa personnalité: l'un utilise une transcription large et tend vers une notation plus phonologique, quand l'autre s'attache plus volontiers à une transcription impressionniste. La conception même du projet, entraînant l'indépendance de chaque chantier, a sans aucun doute favorisé cette «dialectalisation» du système Rousselot: Dauzat n'imaginait pas la difficulté qu'il pouvait y avoir pour des chercheurs dispersés aux quatre coins de la France à collaborer de façon suivie à une époque où n'existaient pas les moyens de communication actuels.

La somme des caractères utilisés par les divers atlas constitue un nombre impressionnant de symboles et de diacritiques. Ajoutons à cela que l'atlas de

10. Au démarrage du module le numéro de carte *1* est proposé automatiquement par défaut, le curseur étant dans la zone de saisie du numéro de point. La touche *z* permet de passer à une autre carte dont il suffit d'entrer le numéro. Le nom de la carte apparaît à l'écran quand on a validé un numéro de carte (Voir *NOMCARTE* pour la saisie des noms de carte). Une fois saisi un numéro de point, s'il existe déjà une donnée pour la carte et le numéro de point indiqués, celle-ci est alors affichée. Les données sont vides en l'absence de saisie.

11. Dès le lancement du module, le programme affiche le numéro de carte 1, qu'on peut valider ou changer. Le programme affiche alors le nom de la carte, que l'on peut modifier. Tant qu'aucun nom n'est saisi, le nom proposé est *carte n°xxx*, xxx étant le numéro de carte saisi. Pour sortir du programme on tape 000 à la place du numéro de carte.

basse Bretagne utilise l'A.P.I.[12], et que les atlas germaniques ont leur propre système. Comme il n'est pas question à ce stade d'unifier les transcriptions, on imagine la complexité extraordinaire de la tâche que les informaticiens doivent affronter.

Le principe de codage est cependant le même pour tous les atlas, ce qui permettra par la suite, on l'espère, un transcodage facilitant les passerelles entre les divers chantiers. Il consiste à distinguer les lettres simples -que nous appellerons formes de base ou simplement formes- des diacritiques ou mises en forme. On parlera donc de code d'une forme et de code d'une mise en forme. Les caractères phonétiques simples ou formes de bases sont réalisés à l'aide d'une seule touche, les diacritiques ou mises en forme à l'aide d'une ou de plusieurs touches supplémentaires -selon leur nombre- en respectant, dans la mesure du possible, la localisation des lettres sur le clavier azerty.

Chaque signe phonétique complet -forme de base et diacritiques- est codé par trois caractères ASCII. Les deux premiers régissent les diacritiques ou mises en forme et le troisième la lettre simple ou forme de base[13]. Les caractères s'affichent sur l'écran au fur et à mesure qu'ils sont entrés au clavier. Notons que le dessin du caractère, dessiné spécialement pour cette opération, n'est pas parfaitement identique à celui du caractère tel qu'il sera tracé sur la carte.

Sur l'écran figurent le numéro de la carte, le numéro du point et une zone réservée à la saisie de la donnée. Il est toujours possible d'insérer un caractère, de le supprimer ou de le remplacer par un autre[14]. A chaque étape, une

12. Alphabet Phonétique International.

13. Pour l'obtention de leur affichage à l'écran lors de leur saisie, comme pour l'obtention des cartes ces deux types de formes sont l'objet de traitements différents. On peut combiner plusieurs mises en forme, par exemple en tapant *F1*, suivi de *F3*. La série des touches *F1* à *F4* et *Shift+F1* à *Shift+F4* et leurs combinaisons détermine le premier caractère de code et la série des touches *F5* à *F8* et *Shift+F5* à *Shift+F8* le deuxième. En ce qui concerne le recours à la touche *Shift*, dans la pratique, seule les associations *Shift+F1* à *Shift+F2* ont un effet de mise en forme. La représentation des diverses combinaisons de touches s'opère par l'utilisation d'une structure de huit bits où l'état de chaque bit traduit la frappe ou non de la touche associée à ce bit. La structure ainsi obtenue est transcrite dans le fichier par un caractère ASCII dont le code hexadécimal est l'addition du nombre 30 en représentation hexadécimale et du nombre représenté par la structure de bits en représentation binaire. L'addition de 30 hexadécimal correspond à la volonté d'éviter d'inclure dans les fichiers des caractères non identifiables sous DBASE. Ainsi, cela correspond, pour les combinaisons possibles des touches de *F1* à *F4* aux caractères *0* à *?* et à la colonne 30 à 3F de la table des codes ASCII hexadécimaux. La même transcription est reprise pour les deux séries de touches de fonction définies plus haut. La touche *F9* permet d'effacer une mise en forme. La touche *Entrée* permet de valider les entrées saisies et de passer au numéro de point suivant.

14. Pour déplacer le curseur on utilise les flèches de direction *flèche à gauche* et *flèche à droite*. La touche *home* permet de placer le curseur sur le premier caractère de la donnée et la touche *Fin* de le placer en fin de donnée.

ligne de message apparaît en bas de l'écran pour indiquer ce que le programme attend de l'opérateur[15]. Le nombre d'entrées qu'il est possible de saisir pour chaque carte, correspondant au nombre de points d'enquête, est un paramètre propre à chaque atlas.

- Le module *commentaire de marges*, destiné à la notation du texte des commentaires en caractères ordinaires, peut être saisi au moyen d'un traitement de texte quelconque. Chaque carte est accompagnée de son propre fichier de marge. L'insertion des mots en phonétique est effectuée au moyen de renvois[16]. On peut utiliser pour cela l'éditeur du logiciel résident *sidekick* qui a l'avantage de permettre la saisie alternée des marges en phonétique et du texte en caractères ordinaires. Pour basculer d'un programme à l'autre, il suffit d'appuyer en même temps sur les deux touches *Shift*.

- Le module *marges* est destiné à la saisie des formes phonétiques à insérer dans le texte des marges. Le nombre maximum de données phonétiques susceptibles d'être reporté dans la marge de chaque carte est un paramètre propre à chaque atlas qui dépend pour beaucoup de la configuration géographique du territoire concerné[17].

Tous ces fichiers, outre leur finalité première, constituent, rappelons-le, des bases de données qui trouveront d'autres applications, notamment la confection d'index.

- Le module *impressi* permet d'éditer à l'écran ou d'imprimer l'ensemble des entrées de plusieurs cartes (dix au maximum) pour contrôle de la saisie. L'auteur peut ainsi faire ses corrections sur papier à partir de tirages réalisés sur des imprimantes simples[18]. Il faut se souvenir que, contrairement à ce qui se passait dans l'ancien système, on a la possibilité d'apporter des corrections ou des modifications aux cartes jusqu'au moment du tirage définitif.

- Le module *sortie* permet la sortie des fichiers d'une carte sur disquette à des fins de sauvegarde[19]. Les chercheurs, nous l'avons dit plus haut, sont en règle générale équipés de modems qui permettent un transfert des données

15. Pour sortir du programme, il suffit de taper 000 à la place du numéro de carte.

16. Les renvois ont la forme suivante: deux parenthèses ouvrantes, le numéro de renvoi, deux parenthèses fermantes, par exemple *((5))*.

17. En l'absence de saisie, toutes les données de ce type sont considérées comme vides.

18. On entre successivement les numéros des cartes désirées, puis le numéro 000 si l'on désire obtenir moins de dix cartes. Le programme demande si l'on veut imprimer. Une réponse négative permet de faire défiler les entrées de la carte à l'écran sans les imprimer. Il est toujours possible d'imprimer l'image d'un écran en appuyant sur la touche i. Pour sortir du module on entre 000, comme premier numéro de carte.

19. On entre le numéro de carte, on met une disquette formatée dans le lecteur A:. Le programme crée un fichier sur A: qui a pour nom le numéro de carte. Pour sortir on tape 000 à la place du numéro de carte.

vers le centre de calcul de Brest où sont confectionnées les cartes au fur et à mesure qu'avance la saisie.

- *autres fichiers.*

Selon les besoins de chacun on peut ajouter à la demande des fichiers de données diverses: nom de l'enquêteur, nom du transcripteur, nom du ou des informateurs, date de l'enquête, date de la transcription, formes lemmatisées, étymons, morphologie, ethnotextes, etc.

4. Le tracé des cartes

Nous avons dit que la deuxième grande étape avant la publication est la fourniture des cartes prêtes à clicher à l'imprimeur. Quand les fichiers que nous venons de décrire ont été saisis par les chercheurs sur micro-ordinateur, ils sont transférés par modem sur l'IBM 9370 du centre de calcul de l'Université de Brest. Les cartes de contrôle destinées aux corrections et les épreuves des cartes servant à la réalisation des clichés destinés à l'impression sont réalisées sur un traceur de marque Benson grâce à des programmes écrits en langage Fortran.

Ces programmes ont pour base ceux qui avaient été élaborés pour l'atlas de la basse Bretagne. Ce point d'appui a été fort utile à la réflexion qui a mené à son amélioration, sa modernisation et son adaptation aux nouvelles exigences du projet global.

Une première opération consiste à définir les divers types de tracés composant une carte ainsi que les fichiers nécessaires à leur mise en oeuvre. Du point de vue technique, on considère que les cartes représentent, textes et encadrements compris, les paramètres de divers types de tracés dont la réalisation est commune à tous les atlas, étant bien entendu que chacun d'entre eux possède ses variantes propres. Seuls les paramètres de position et de taille sont modifiés d'un atlas à l'autre en fonction de la mise en page.

4.1. Les fichiers des formes de tracés

Ces divers paramètres sont répartis de la façon suivante:

-un ou plusieurs fichiers rassemblent les formes des contours géographiques, selon que l'on souhaite distinguer, par exemple, les limites départementales des limites nationales ou linguistiques[20].

- un fichier contient les emplacements des numéros identifiant les points d'enquête.

20. Ces fichiers contiennent autant d'enregistrements que de points définissant le contour qu'ils décrivent, chaque enregistrement comportant les deux coordonnées du point ainsi que la position de la plume du traceur.

-un fichier réunit les formes d'encadrement des cartes, des titres et des emplacements de marge.

-un fichier rassemble les formes des caractères phonétiques de base, comportant pour chacun son code et le nombre de points qui constituent son dessin, chaque point étant défini par ses coordonnées relatives et la position de la plume du traceur.

-un fichier des signes diacritiques (ou mises en forme) est constitué selon les mêmes principes que celui des caractères de base. La position des signes diacritiques relativement à celle du caractère de base qu'ils viennent modifier peut varier selon la taille et la position du caractère. Les mises en forme sont donc saisies dans une position fixe, leur position définitive n'étant calculée qu'au moment de la confection des cartes.

4.2. Création des formes des dessins des limites géographiques, des cadres et des caractères phonétiques

Les formes des dessins autres que les textes et les nombres sont obtenues sur micro-ordinateur à l'aide du logiciel Autocad[21] fonctionnant sous DOS. Des indications reportées dans les fichiers rendent ces dessins exploitables par le traceur.

Le dessin des divers contours géographiques composant le fond de carte, ainsi que les coordonnées des numéros de points d'enquête sont établis dans un même repère en veillant à ce que leurs leurs dimensions soient compatibles avec la taille des cartes à dessiner.

Les caractères phonétiques et les signes diacritiques (mises en forme) sont dessinés un à un: c'est à ce niveau qu'est réalisée la concordance entre les codes de ces formes et les codes du programme *saisie* propres à chaque atlas.

Pour chaque dessin, un nouveau plan est défini dans le fichier de dessin. Un caractère simple (ou forme) est désigné par la chaîne de caractères représentant le numéro de code ASCII décimal associé à la touche du clavier utilisée pour l'obtenir. Un signe diacritique (mise en forme), en revanche, est désigné par un chiffre de *1* à *9*, et au-delà par le caractère *A*. Chaque nom de ce type correspondant au numéro d'une des touches de fonction, de *F1* à *F8*, puis aux combinaisons de *Shift* et de *F1*, et de *Shift* et de *F2*, une concordance

21. Pour dessiner les fonds de cartes et les caractères phonétiques, la version de base d'Autocad a été enrichie grâce à Autolisp, pour disposer de commandes spécifiques à cette application, permettant d'automatiser au maximum la confection des divers dessins, puis, ensuite, pour chaque point qui les compose, l'écriture dans des fichiers, de paramètres reconnaissables pour le pilotage du traceur. L'ensemble de ces commandes peut être mis en oeuvre grâce à un menu, chargé lors du lancement de l'application par la commande Atlas lancée à partir du DOS. Ainsi l'utilisation d'Atlas ne nécessite pas la connaissance préalable du logiciel Autocad, d'autant plus qu'une aide peut être appelée à partir du menu.

est ainsi établie avec l'utilisation des touches du clavier qui, dans le programme *saisie*, permettent d'obtenir les diverses formes phonétiques. Le nom de ce plan constitue le code interne de la forme phonétique qui permet de le ranger dans la bibliothèque attachée à l'atlas.

4.3. Tracé des textes et des nombres

Le traceur ne pilote selon ses procédures propres que les tracés composant les textes des titres de cartes, des commentaires non phonétiques et des numéros de points d'enquête. Il reste cependant à traiter séparément certains caractères français comme le ç et les signes accentués.

4.4. Autres fichiers communs à toutes les cartes

Plusieurs indications sont indispensables à la conduite du traceur: échelle et position des cartes sur la page, nombre de cartes par page, nombre de points d'enquête par carte, taille des divers types de caractères (titre de page, titres de carte, références, marges, entrées phonétiques...) et leur angle d'italique, emplacement des titres et des textes de marges etc. Ces indications sont rangées dans des fichiers qui viennent ainsi s'ajouter à ceux qui définissent la forme des divers types de dessins particuliers à chaque atlas.

4.5. Fichiers de données propres à chaque carte

Les fichiers de données proprement linguistiques occupent un espace mémoire très important. Afin d'accroître la vitesse d'accès, on les décompose en autant de fichiers que l'atlas comporte de cartes:

-un fichier des entrées phonétiques à reporter à chaque emplacement de lieu d'enquête.

-un fichier de texte des commentaires de marge.

-un fichier des entrées phonétiques à reporter dans le texte des commentaires de marge, chacune étant référencée par un numéro défini lors de la saisie du texte de ces commentaires.

-un fichier de texte des références à d'autres atlas éventuellement adjointes au titre de carte.

4.6. Tracé des dessins sur la carte

Le tracé d'un dessin peut être obtenu par des procédures de divers niveaux de complexité géométrique qui se réduisent toutes au tracé de vecteurs élémentaires. Le logiciel fait appel à la procédure *Traa*, qui positionne la plume du traceur en un point donné[22]. Ce choix permet de définir des para-

22. Un vecteur élémentaire est tracé par deux appels successifs de cette procédure. Lors du premier appel, sont assignées comme paramètres les coordonnées de l'origine du vecteur et

mètres identiques pour tout le tracé sans considération de la structure géométrique des formes à obtenir. Il permet aussi de diminuer le nombre de vecteurs, donc le nombre de déplacements de la plume, au prix d'un traitement préalable plus complexe des paramètres requis; il facilite aussi les modifications d'angles d'italique.

De la calligraphie et de la finesse des tracés au moment du tirage dépend la qualité proprement esthétique des épreuves de cartes réalisées.

5. Description du tirage des cartes sur Benson

Le commentaire concerne surtout le tirage des cartes de données phonétiques. Rappelons cependant que d'autres types de cartes peuvent aussi être inclus dans un atlas, dont deux sont actuellement programmés: la carte des noms des points d'enquête et la carte du réseau, qui est un graphe reliant les points d'enquête de proche en proche par ordre de numéro croissant.

5.1. Mise en place des numéros de points d'enquête

Pour le tracé des numéros des points d'enquête aux emplacements prévus, la bibliothèque de pilotage du traceur dispose d'une procédure qui permet d'écrire un nombre de taille et de position définies sur la feuille de papier, en utilisant une plume -et partant une couleur d'encre- données. Les coordonnées, obtenues par lecture du fichier correspondant, deviennent alors accessibles comme éléments d'un tableau; elles sont indexées par le numéro du point et par la carte concernée s'il y en a plusieurs par page. Le programme fera par la suite appel de nombreuses fois à cette indexation pour le calcul des positions des entrées phonétiques et des caractéristiques de leur encombrement.

Chaque numéro de point se voit également indexé par les caractéristiques définissant son encombrement, terme qui désigne ici le rectangle assorti de marges au sein duquel il s'inscrit.

5.2. Mise à blanc des contours

Le fichier des paramètres nous permet de moduler le tirage des divers éléments qui composent la carte. Pour que l'auteur puisse contrôler la carte, on fait un premier tirage rassemblant l'ensemble des éléments, fond de carte compris. Afin d'améliorer la lisibilité de ces tirages, une procédure permet le gommage des contours géographiques lorsque les numéros de points d'enquête

un entier égal à 0, indiquant que la plume doit être levée pour atteindre ce point à partir de sa position antérieure. Lors du second appel immédiatement consécutif, les paramètres sont alors les coordonnées de l'extrémité du vecteur et un entier égal à 1, pour que le second parcours, ainsi commandé, s'effectue plume baissée.

les chevauchent[23]. On fait aussi appel à cette procédure après l'écriture des entrées phonétiques.

On peut également tirer une liste des entrées phonétiques d'une carte, par exemple dans le cas où la longueur des entrées nécessite un contrôle plus minutieux.

L'imprimeur ne reçoit pour chaque atlas qu'une épreuve des contours et des numéros de points à partir de laquelle sera réalisée l'impression en couleur du fond de carte. Les données variables (phonétique et textes) sont tirées en noir sur des épreuves séparées, ce qui diminue le prix de revient. Les entrées étant imprimées en noir sur la couleur plus claire des contours restent toujours parfaitement lisibles, et il n'y a donc plus lieu à ce stade de gommer les contours lorsque les données les chevauchent.

5.3. Recherche dichotomique des paramètres de formes phonétiques

Les traitements réglant la position et la taille des entrées phonétiques et pilotant leur écriture nécessitent de fréquents recours aux paramètres régissant la forme et l'encombrement des caractères qui les composent. Moyennant l'attribution à la machine virtuelle disponible sur l'IBM d'un espace mémoire suffisant, ces paramètres sont chargés dès le lancement du programme. Leur accès rapide est obtenu par une recherche dichotomique de l'index des formes concernées dans les tableaux où ils sont ordonnés selon leurs codes. Cela permet de réduire à quelques minutes le traitement de chaque carte.

5.4. Encombrements des entrées phonétiques

Pour éviter de possibles chevauchements des entrées phonétiques de points voisins, chacune d'entre elles peut occuper huit positions relativement au numéro de point: à gauche ou à droite, au dessus ou au dessous, centrée ou non sur le numéro.

Lors du chargement des paramètres de forme, la taille de chacun des caractères est calculée pour une échelle de référence donnée. L'encombrement[24] est ensuite déterminé pour chaque position possible à partir de la taille des caractères qui composent l'entrée, en tenant compte également de l'encombrement du numéro de point.

23. Nous obtenons ce résultat par la mise à l'état 0, du paramètre qui commande la position de la plume, pour les zones concernées. Pour ce faire, une procédure *misablanc* est mise en oeuvre après chaque écriture de numéro, le tracé définitif des contours n'intervenant qu'après les autres tracés. Cela présente dans certain cas un intérêt, compte tenu du fait que la couleur choisie pour imprimer les numéros est la même que celle des contours.

24. Nous entendons ici par encombrement un rectangle assorti de marges dans lequel s'inscrit la donnée pour une taille de caractères prise comme référence.

5.5. Chevauchements d'écriture et débordements des limites de carte

La position définitive de toutes les entrées phonétiques relativement au numéro du point d'enquête et la taille des caractères ne sont fixées qu'à la suite de l'examen des chevauchements possibles avec les entrées d'autres points et des débordements éventuels des limites de la carte. C'est là l'objet de la procédure *misenplace,* qui intervient avant l'écriture proprement dite.

Pour chaque numéro de point, le programme vérifie l'éventualité de chevauchements ou de débordements à partir d'une position initiale de l'entrée fixée à sa droite et au même niveau que lui. Dans l'affirmative, la position de l'entrée est modifiée tant que cela est logiquement possible. Si le changement de position ne suffit pas à éliminer l'interférence, la taille des caractères est progressivement réduite jusqu'au moment où l'anomalie a disparu. Ce traitement de mise en place garantit que les entrées transcrites sur la carte ne déborderont pas du cadre et ne se chevaucheront pas.

Dans la majeure partie des cas, les entrées réduites restent lisibles. Si cependant l'absence de chevauchement n'est obtenue qu'au prix d'une réduction trop importante de la taille des caractères, il incombe à l'auteur de décider d'une modification éventuelle des entrées, par exemple en confectionnant deux cartes au lieu d'une.

5.6. Traitements requis pour l'écriture des entrées phonétiques

La position de la ligne de base et du début du texte de l'entrée phonétique ayant été fixées par rapport à celle du numéro de point, il suffit pour écrire le texte phonétique de réitérer la procédure d'écriture d'un caractère autant de fois que nécessaire. La mise en forme de chaque caractère doit être examinée afin de déterminer sa position par rapport à la ligne de base et l'éventualité de sa réduction, et de chercher si des diacritiques doivent éventuellement lui être adjoints; dans le cas où plusieurs diacritiques viennent affecter le même caractère, il faut tenir compte de leur position relative selon les paramètres propres à chaque atlas[25]. A la fin de l'opération la plume doit être convenablement positionnée sur le début du caractère suivant pour éviter d'éventuels chevauchements.

5.7. Tracé de caractères alphabétiques

Le dessin de caractères alphabétiques pour l'écriture des titres, des références et des commentaires de marge utilise, nous l'avons dit plus haut, une police interne disponible sur le traceur et des commandes de la bibliothèque

25. C'est une procédure *misenform* qui interprète les deux codes de mises en forme et fait les appels nécessaires à la procédure *dessin* qui trace le dessin correspondant à chaque mise en forme, à l'emplacement et selon la taille et l'angle d'italique appropriés, et détermine la taille et la position de la forme de base. Le dessin du caractère de base doté des caractéristiques de taille, de position et d'italique requises, n'intervient qu'ensuite.

de pilotage permettant d'écrire des textes selon un format défini, à l'emplacement désiré.

Cette police ne contenant pas tous les caractères imprimables, il a fallu prévoir des traitements spécifiques pour les caractères manquants: une procédure *caract* permet d'écrire n'importe quel texte imprimable, à un emplacement et selon un format de caractères définis.

5.8. Mise en place des textes de marge

Les textes de marges sont composés de mots en écriture alphabétique dont certains sont en gras, en italique ou soulignés tandis que des données phonétiques peuvent être intercalées au sein des commentaires en caractères courants.

Lors du tirage, ces textes se présentent sous la forme de lignes de 80 caractères. Comme le traceur ne fonctionne pas comme un traitement de texte, il est indispensable de placer les références aux modifications souhaitées juste avant les parties de textes concernées sous la forme de caractères qui seront interprétés au moment du tracé[26]. Les données phonétiques à inclure sont référencées par un numéro qui leur est attribué au moment de leur saisie. On peut aussi, si on le souhaite, inclure dans la marge d'une carte l'ensemble des réponses à une question qu'on ne souhaite pas cartographier[27].

Outre l'interprétation préalable des caractères gouvernant le format des mots qui composent chaque ligne de texte, il faut veiller à ce qu'aucun d'entre eux ne déborde de la marge. Si un chevauchement se produit, il faudra reporter le mot à la ligne suivante. Si la nouvelle ligne déborde à son tour, le tirage est interrompu et un message signale qu'il faut modifier les paramètres gouvernant la taille des caractères.

L'écriture des données phonétiques incluses dans le texte et identifiées par leur numéro de référence fait appel aux mêmes traitements que ceux des entrées placées près des numéros de points. Une césure est prévue pour les mots longs arrivant en fin de ligne.

5.9. Mise en place des textes de titre et des références

Leur position, la taille de leurs caractères et l'épaisseur de tracé étant modifiables en tant que paramètres de la carte, leur écriture à l'emplacement voulu ne pose aucun problème. En cas de débordement, une nouvelle ligne est écrite.

26. *G* ou *N*, *I* ou *D*, *S* ou *NS*. Ces indications seront placées par convention au sein d'un jeu de deux parenthèses ouvrantes «((» et de deux parenthèses fermantes «))».

27. On écrit alors, par exemple, *((5))* pour indiquer qu'on souhaite noter dans la marge d'une carte l'ensemble des données des réponses à la question 5.

5.10. Tracé du cadre de page

Le tracé d'un cadre de page, incluant un numéro et une référence aux conditions techniques de réalisation, peut éventuellement être adjoint au dessin des cartes.

5.11. Confection d'une série de cartes

Grâce au langage de commande interprété disponible sur l'IBM, le programme prévoit le lancement automatique de plusieurs tirages de carte successifs dont les tracés sont stockés sur disque. Chaque carte ainsi traitée est alors prête dans un fichier de dessin dont il suffit de commander l'exécution sur le traceur.

Le programme prévoit aussi le lancement sur l'IBM du tirage des épreuves de cartes en tâche de fond par un passage automatique des paramètres de désignation des fichiers nécessaires à la confection de chaque carte.

6. Etat d'avancement des travaux en octobre 1991

Le volume III de l'atlas Auvergne-Limousin de Jean-Claude Potte sera le premier volume de la nouvelle série. Il comportera 600 cartes, à raison de 2 cartes par page, et ressemblera beaucoup aux volumes précédents.

Le volume II de l'atlas de Picardie dont la saisie est faite par Fernand Carton est également en bonne voie.

Jacques Tourrel, à Aix-en-Provence, qui a déjà assuré la saisie de l'atlas de la Réunion selon d'autres principes, a terminé la saisie des formes phonétiques du volume III de l'atlas de la Normandie de Patrice Brasseur, dont les marges ont déjà été saisies par l'auteur sous Macintosh. Il va maintenant se charger du volume III de l'Ile-de-France de Marie-Rose Simoni, puis du volume IV de l'atlas de Provence de Jean-Claude Bouvier et Claude Martel.

Jean-Paul Chauveau, à Angers, a commencé la saisie du volume III de l'Albram; Brigitte Horiot, à Lyon, va commencer la saisie du volume IV de l'Ouest et Jacques Boisgontier, à Toulouse, celle de son volume IV du Languedoc oriental. Arlette Bothorel commencera la saisie du volume III de l'Alsace pendant l'été 1992, Guy Levieuge et Marthe Philippe du volume II de la Lorraine germanophone. M. Bourcelot s'occupe, bien qu'en retraite, de son dernier volume de l'atlas de la Champagne et de la Brie: la saisie en sera effectuée à Aix-en-Provence. Enfin, j'assure personnellement la saisie des données de la basse Bretagne.

La Corse, nous l'avons dit, dispose de son propre programme. Afin que les cartes soient réalisées à Brest, un logiciel de transfert des données de Macintosh sur IBM doit être réalisé. Un premier volume de vocabulaire fondamental présentant un réseau plus dense que le projet de plus longue haleine sera réalisé en priorité dans un an.

Le perfectionnement actuel du logiciel de saisie permet désormais d'envisager de faire appel pour cette tâche à une secrétaire, le chercheur n'ayant plus à assurer que le contrôle.

7. Les travaux connexes

Parallèlement, une commission dirigée par Gérard Taverdet établit les index généraux informatisés des atlas, qui prennent bien entendu en compte les volumes parus. Des réunions de travail ont eu lieu à Nancy en juin 1991 et à Dijon en septembre. Cette tâche concerne plus particulièrement les auteurs ayant achevé la publication de leurs atlas. Toute cette opération devra déboucher sur la constitution d'une banque de données.

Le chantier de recherche sur les ethnotextes ou textes oraux a commencé en 1977. Les ethnotextes recueillis au cours des enquêtes peuvent être saisis parallèlement aux autres données, ce qui est de loin la meilleure solution.

Le lien avec les ethnologues, qui devait être une des composantes fortes du projet initial, va être ravivé par la création dans un premier temps d'un index des données de type ethnologique contenues dans les atlas.

8. Conclusion: les projets d'avenir

8.1. Cartes à symboles

Les entrées réunies pour la publication des volumes informatisés permettront de réaliser automatiquement des cartes à symboles (cartes de la deuxième génération) du type de celles publiées dans le volume II de l'atlas de l'Alsace. Cette opération ne pourra être effectuée de la même manière pour les volumes déjà publiés qu'après l'achèvement des opérations décrites ci-dessous.

8.2. Informatisation des volumes déjà parus

Il est semble-t-il utopique de prétendre récupérer les entrées des volumes parus dans des fichiers informatisés grâce à des logiciels de reconnaissance de caractères. Il faut cependant y penser, en espérant que les progrès de l'informatique le permettront. Mais une telle opération, qui nécessiterait la mobilisation d'un espace mémoire gigantesque est-elle vraiment utile? On ne saurait en douter. D'abord par la possibilité de rapprocher les données par la constitution d'atlas «à cheval» sur plusieurs régions. Ensuite et surtout pour un retour vers l'A.L.F.

8.3. Rassemblement des données

La conception même de l'entreprise des atlas linguistiques de la France par régions induisait inévitablement, nous l'avons déjà souligné, un repli toujours plus grand des chercheurs sur eux-mêmes, en raison de l'étalement de

l'entreprise sur une période de près de cinquante ans; du renouvellement des hommes; des statuts différents des personnes engagées dans le projet - universitaires, chercheurs, ingénieurs; de la dispersion des terrains d'enquête; de la variété des situations sociolinguistiques - on pense au contraste entre l'Ile-de-France, où Marie-Rose Simoni fait de l'archéologie linguistique, et le Languedoc où les parlers sont toujours vivants.

Cette autonomie a conduit à un éloignement progressif des conceptions. S'il est vrai que les questionnaires devaient être différents afin de coller au terrain, ils devaient aussi initialement comprendre une partie commune destinée à la comparaison. Or, on s'aperçoit que chaque atlas a conservé ou éliminé des cartes selon leur «*rendement*», c'est-à-dire selon la variété plus ou moins grande des réponses obtenues, la variété étant assimilée à une richesse. Ce procédé était d'ailleurs déjà prévu par Dauzat, mais il devait affecter la partie générale du questionnaire: «Un certain nombre de mots qui, à l'expérience, se sont révélés de moindre intérêt, seront supprimés et remplacés par un choix de termes caractéristiques de chaque région»[28]. Dans ces conditions, une forme uniformément semblable ou à peu près sur l'ensemble de la zone était laissée de côté: Jean-Claude Potte a pu ainsi montrer dans un travail inédit que dans le domaine des parties du corps, peu susceptible de variations, seul l'oeil était commun aux différents questionnaires. Pour l'Atlas linguistique du Lyonnais, conçu avant même le projet de Dauzat, l'abbé Gardette, «rédigea lui-même "un questionnaire très proche des pensées du paysan lyonnais" en laissant de côté tous les termes généraux» (Pop 1950, 223).

Nous avons aussi montré que les alphabets phonétiques, tous basés chez les romanistes sur celui de l'abbé Rousselot, se sont également éloignés par l'adjonction ponctuelle de signes nouveaux, par l'interprétation différente de signes existants, par l'usage plus ou moins grand des diacritiques.

Enfin, et c'est encore plus important, des conceptions divergentes sont apparues dans la représentation cartographique des faits: on trouve d'un côté des cartes respectant la méthode de Gilliéron -à chaque point sa donnée- et de l'autre des cartes dites «*à aires dégagées*» qui sont en réalité une étape intermédiaire entre la méthode de Gilliéron et les cartes interprétatives[29].

L'informatisation des atlas a induit par la force des choses un retour vers la centralisation des données et provoqué une réflexion sur la comparabilité des divers atlas. Maintenant que l'urgence née des contraintes matérielles

28. Dauzat, *Nouvel atlas linguistique de la France par régions*, Luçon, S. Pacteau, s.d. [1942], 5 (cité dans Pop 1950, 5).

29. «L'atlas ne devait pas être l'oeuvre d'un lin eguiste; s'il eût été l'oeuvre d'un linguiste ou de nous-même, il ne présenterait point les mêmes garanties de désintéressement et serait un bloc déjà dégrossi en vue d'une oeuvre encore indéfinie» Gilliéron in *Généalogie des mots qui désignent l'abeille*, Paris, Champion, 1918, 1-13.

liées à l'achèvement des derniers volumes est moindre, il faut passer au travail de réflexion commune. Y sont conviés, cela va de soi, les auteurs d'atlas, hommes de terrains, mais aussi les utilisateurs effectifs ou potentiels des atlas. Que cherchent-ils, quels sont leurs souhaits? Des chantiers rassemblent des données internationales en vue de l'interprétation, soit, comme l'*Atlas Linguarum Europae*, sur le modèle de la comparaison entre langues d'origines diverses, soit, comme l'*Atlas Linguistique Roman*, sur le modèle de la comparaison entre langues génétiquement apparentées.

Il reste, je crois, à revenir, à partir des données dont nous disposons désormais, vers un *Atlas Linguistique de la République française*, étendu aux zones alloglottes afin de couvrir l'ensemble du territoire. L'*Atlas des Côtes de France*, pour lequel les enquêtes sont en voie d'achèvement, va ouvrir la voie. En ces années de construction européenne, n'est-il pas scientifiquement concevable d'observer les courants qui ont circulé et circulent au sein du territoire national de la France?

Je suis convaincu que, si les problèmes sont bien posés, des solutions techniques seront toujours possibles, et qu'en tous les cas seul un débat entre utilisateurs et informaticiens permettra d'avancer. La collaboration internationale, qui tient une place importante au sein du G.D.R. 9, doit se maintenir et se renforcer: le comité de direction de l'A.L.E. comprend trois membres du groupe; l'A.Li.R.[30] est centralisé à Grenoble, tandis que des échanges réguliers ont lieu avec les collègues chargés des atlas du Piémont occidental, du Portugal, du Pays basque (ce dernier dans le cadre du G.D.R. 9).

Nous sommes dans une problématique qui dépasse de loin la France, comme en témoigne le débat qui se déroule actuellement dans les *Quaderni di Semantica* sur le problème **Où va la dialectologie?**

30. Atlas Linguistique Roman.