Assistance informatique à l'interprétation des données en cartographie linguistique. Informatisation anthropocentrée du Nouvel Atlas Linguistique de la Basse-Bretagne

Ioannis Kanellos, Jean Le Dû, François Legras, Ludovic Tanguy

▶ **To cite this version:**

Ioannis Kanellos, Jean Le Dû, François Legras, Ludovic Tanguy. Assistance informatique à l'interprétation des données en cartographie linguistique. Informatisation anthropocentrée du Nouvel Atlas Linguistique de la Basse-Bretagne. Géolinguistique, 1999, 8, pp.181-196. halshs-01322261

HAL Id: halshs-01322261
https://shs.hal.science/halshs-01322261

Submitted on 26 May 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Assistance informatique à l'interprétation des données en cartographie linguistique. Informatisation anthropocentrée du Nouvel Atlas Linguistique de la Basse-Bretagne

Ioannis Kanellos[1], Jean Le Dû[2], François Legras[1] et Ludovic Tanguy[3]

Résumé

Nous discutons les motivations et les caractéristiques fondamentales d'un projet de recherche qui vise à mettre sur pied un système informatique d'assistance à l'interprétation pour les besoins de la recherche en cartographie linguistique. Dans la première partie nous exposons rapidement les traits fondateurs de la démarche de la cartographie linguistique sur la trace d'un morceau de son histoire récente. Notre objectif est de montrer combien ses aspirations tissent, dès le début, avec le traitement des données, et, par voie de conséquence, combien elle est tributaire des moyens de ces traitements. Dans la seconde, nous discutons le thème de l'introduction de l'informatique à son secours. Cependant, nous cherchons à défendre parallèlement l'idée suivant laquelle l'appel à l'informatique peut être plus qu'un simple service de calcul, et qu'il peut être pensé comme facteur de jonction intersémiotique, entre les différents niveaux de données qui intéressent la cartographie linguistique. L'objectif de notre argument est coordonné par une vision interprétative de la discipline. Enfin, dans la troisième partie, nous montrons les premières réalisations informatiques issues d'une telle vision.

Mots clés : Cartographie linguistique, systèmes anthropocentrés, traitement de données, assistance informatique à l'interprétation, atlas linguistique.

1. Le problème de la constitution et de l'interprétation des cartes linguistiques comme problème sémiotique

Les concepts de langue, dialecte, patois restent flous malgré les avancées de la linguistique. Les parlers populaires ruraux, vecteurs d'une civilisation rurale multiséculaire aujourd'hui en voie d'extinction, sont d'une variété extrême, au point que certains ont pu parler à leur propos « d'émiettement dialectal ». Depuis le dix-neuvième siècle, pour des raisons idéologiques, les chercheurs de l'école allemande se sont appliqués à démontrer l'existence de frontières dialectales tranchées, tandis que les chercheurs français penchaient plutôt pour l'existence d'un continuum. L'Atlas Linguistique de la France de Gilliéron et Edmont, dont les points d'enquête couvrent les régions de tradition gallo-romane, y compris ceux qui sont situés en Belgique et en Suisse – tout en laissant de côté les territoires allophones, dont la Basse-Bretagne – fournit un matériau brut à partir duquel il est possible de rechercher une réponse au dilemme à partir de données objectives. Publié au cours de la première décennie du vingtième siècle, cet ouvrage est un recueil de cartes fournissant sur un réseau serré des formes brutes, en écriture phonétique, constituant les réponses, recueillies sur place, à des questions simples portant sur la

[1] Ecole Nationale Supérieure des Télécommunications de Bretagne, IASC, BP 832, 29285 Brest Cedex.
E-mail : ioannis.kanellos@enst-bretagne.fr, francois.legras@enst-bretagne.fr.
[2] Université de Bretagne Occidentale, Faculé des Lettres, Centre de Recherche Bretonne et Celtique, BP 814, 29285 BREST CEDEX. E-mail : ledu@univ-brest.fr.
[3] ERSS - Université de Toulouse le Mirail, 5 allées Antonio Machado, 31058 Toulouse Cedex 1.
E-mail : Ludovic.Tanguy@univ-tlse2.fr.

vie quotidienne et la civilisation rurale. Cet ouvrage monumental n'a pas reçu l'accueil qu'il méritait dans une France en voie d'unification linguistique. Il n'a toujours pas, en cette fin du vingtième siècle, fait l'objet d'une interprétation globale. Les études qui lui ont été consacrées sont plus positivistes que spéculatives.

Le premier ouvrage de géographie linguistique bretonne, réalisé sur le modèle de l'ALF, est l'Atlas Linguistique de la Basse-Bretagne (ALBB) de Pierre Le Roux. Ses données, recueillies dans 77 points d'enquête, ont été rassemblées entre 1913 et 1920. La publication des six fascicules de 100 cartes s'est étalée sur une période d'une quarantaine d'années (1924, 1927, 1937, 1943, 1953 et 1963). François Falc'hun s'est appuyé sur les quatre premiers fascicules pour sa thèse *L'Histoire de la langue bretonne d'après la géographie linguistique* (1951), qui a révolutionné les connaissances sur la langue bretonne par l'application d'une méthode d'analyse originale des cartes. Son étude a permis de mieux comprendre l'évolution historique de la langue, la manière dont le vocabulaire français s'est introduit le long des routes du Sud, et a même abouti à la conclusion, controversée à l'époque, que le breton constituait, du moins en partie, la continuation de la langue gauloise. Il explique la découverte de sa méthode dans l'introduction de son ouvrage *Perspectives nouvelles sur l'histoire de la langue bretonne* :

> *« Ma première préoccupation fut de préciser la véritable situation des limites dialectales. A cette fin, je séparai par des traits au crayon, sur les cartes de l'Atlas, les variantes phonétiques, morphologiques ou lexicologiques, puis je reportai ces isoglosses sur une même carte. De plusieurs cartes ainsi obtenues, il ressortait avec évidence que, sauf sur le pourtour du pays de Vannes, les principaux faisceaux d'isoglosses ne s'alignaient ni sur les limites de diocèses, ni sur les rivières, ni sur les montagnes. A quels principes obéissaient donc leurs courbes capricieuses ? » (p. 17).*

Il a alors l'idée de superposer aux cartes des transparents des « grands chemins » figurant sur l'Atlas itinéraire de Bretagne, d'Ogée (1769). C'est alors l'illumination :

> *« Aussitôt, j'eus comme l'impression de voir les mots courir sur les routes, spécialement de Carhaix à Tréguier. L'influence des routes révélait du même coup l'action immense et totalement insoupçonnée de Carhaix, point de convergence du réseau des voies romaines en Basse-Bretagne ». (p. 18).*

Cette nouvelle méthode allie donc à l'étude des cartes en elles-mêmes celle de tout ce qui conditionne l'évolution linguistique : histoire des voies de communication, de la vie économique, chronologie des faits de civilisation, témoignage des textes, de la toponymie etc.

Déjà, la cartographie linguistique comme secteur d'exercice d'une science explicite ce qui restera toujours son hypothèse fondamentale de travail. Elle repose sur la valeur qu'on accorde aux régularités phonématiques, recherchées à travers un corpus relativement représentatif. Ces régularités définissent des classes d'équivalences, dont la forme la plus remarquable en matière d'exploitation est celle de « l'isoglosse ». Schématiquement, une isoglosse est une attestation d'une identité locale (lexicale, morphologique, phonématique, syntaxique…) qui redistribue l'espace géographique, en le transcrivant sous forme de régularités linguistiques. Mais, surtout, elle fait apparaître des stabilités formelles qu'on tente d'interpréter comme incidences de phénomènes de nature socio-économique.

Le rapport de la mise en relief de telles stabilités avec les techniques d'élaboration des données est évident. La retranscription de la carte d'une région devient, en quelque sorte, un

« épiphénomène » macroscopique, où les éléments atomiques doivent céder devant l'émergence de structures qui les intègrent massivement et de manière pluraliste.

Faite *manuellement*, cette recherche a des limites évidentes. On doit avec patience tracer une multitude d'isoglosses par carte selon les points de vue (lexicologique, phonétique) ; trouver des analogies, tenter des superpositions, etc. Poser une hypothèse devient alors le début d'un travail long et fastidieux, fort ingrat dans la mesure où on ne peut vérifier qu'un petit nombre d'hypothèses. La finalité restant, bien évidemment, la recherche des lois de la variation, de ses causes, des grands mouvements de transformation, le rêve de tout chercheur serait de pouvoir superposer à loisir des cartes à même échelle comportant des faits linguistiques, historiques, anthropologiques etc. afin de pouvoir multiplier les hypothèses. Mais ceci demande de disposer d'importantes ressources…

Cependant, le recours à la machine et aux techniques avancées du traitement des données ne doit pas occulter, ni même déplacer au second plan le rôle de l'acteur humain. Le travail du cartographe est et restera pertinemment un travail d'interprétation. Deux experts du domaine ne trouverons pas nécessairement les mêmes cartes, puisqu'ils ne valoriseront pas obligatoirement la même forme d'isoglosse.

Ainsi, le problème devient d'emblée non pas d'évacuer l'instance humaine dans la production cartographique mais de la seconder dans la pluralité de ses objectifs interprétatifs. Cette perspective nous oblige à réfléchir sur le thème de l'interprétation des données de la cartographie linguistique en tant que thème fondateur dans l'utilisation de l'informatique.

2. Deux ou trois réflexions sur le statut interprétable des objets en Cartographie Linguistique

Cette rapide présentation de la tradition récente en matière de méthodes et objectifs de la cartographie linguistique, volontairement schématique, n'avait pas de prétention. Elle visait seulement à poser un premier canevas pour réfléchir sur les moyens et les techniques que la cartographie linguistique emploie dans un cadre plus formel, où le calcul peut être légitimement appelé. Non pas pour récupérer une tradition dans un projet positiviste. Non plus pour dévier ses objectifs vers une technologisation désormais de mode. Mais seulement pour nous donner l'occasion de discuter sa méthodologie comme une tentative de créer des ponts intersémiotiques. Et, se fondant sur cet argument, pour montrer que l'informatique peut avoir un rôle plus important que celui qu'on lui réserve habituellement, rôle dévolu généralement à des calculs intensifs. En effet, dans cette étude, nous souhaiterions développer l'idée que l'informatique peut revendiquer un rôle authentique en tant que *médiateur sémiotique*. Mais commençons par le début.

La démarche de la cartographie linguistique n'a décidément pas pour objectif d'instituer des objets en soi, voués à figer, sous forme de carte achevée et pérenne, l'identité d'une langue. Encore moins, de proposer un modèle de langue stable et univoque. Elle vise, au contraire, à construire des entités de format linguistique qui retracent un moment remarquable d'une mémoire partagée. A travers elles, elle vise certes, *in fine*, une interprétation. Cependant, l'interprétation qui coordonne son projet s'engage toujours dans une conception de la langue nécessairement dynamique, qui se pose comme une donnée à établir descriptivement sur plusieurs axes. Celui de la *diatopie*, par exemple, qui concerne la variabilité de ses unités en fonction de l'espace, celui de la *diachronie*, qui projette un segment de langue sur l'écran de son histoire, celui de la *diastratie*, qui retrace les rapports entre les niveaux de langue et les strates sociales qui les utilisent, celui de la *diapraxie*, qui convoque la dialectique entre le genre

linguistique et la pratique au sein de laquelle il s'inscrit… Sous un tel objectif, la cartographie linguistique abandonne volontiers l'hypothèse d'une langue unifiée, puisque non opérationalisable (sauf bien sûr idéologiquement). Mal définie, mais constituée par une argumentation normative ou par un souci d'objectivation, la langue devient plutôt un objet-cadre, un objet idéel. Elle est plus une hypothèse qu'une forme en soi scientifiquement explorable. La cartographie linguistique lui préfère l'élément attestable qu'elle peut exploiter comme *donnée*. Mais, en réalité, ce sont là des mutations épistémologiques dues à des différences de motivations : la langue n'occupe pas la cartographie linguistique comme finalité mais comme moyen, comme espace d'observables à travers laquelle elle cherche des indices d'incidence de la réalité linguistique à la réalité socioculturelle. La possibilité d'une corrélation entre parlers et faits socioculturels reste certes une hypothèse. Cependant, elle bénéficie d'une certaine naturalité et se voit corroborée de remarques comme celle de Falc'hun. Peut-être de manière médiate, mais toujours sous un régime opérant. Tout le reste est interprétation.

Mais, on vient de le voir, pour atteindre le niveau interprétable, la cartographie linguistique doit s'engager dans une voie de traitement de données. Un traitement de facture plutôt standard. Non pas par un effet de mode, ni par la volonté de s'approprier les règles, et par là même la promotion, d'une scientificité exacte. Simplement par tradition méthodologique et pratique.

Certes, la constitution de son espace de données, ses méthodes de recueil et le tri initial de ses unités sont, et resteront probablement à jamais problématiques, dans la mesure où plusieurs paramètres sont condamnés à rester en dehors de la sphère de la rationalité – éprouvée ou encore recherchée – qui est la sienne. On ne saurait prétendre à l'exhaustivité des données dans le temps, l'espace et les strates sociales convoquées, ni à leur équilibre qualitatif au sein des différentes pratiques. On ne saurait défendre un argument objectif concernant les conditions de leur recueil, l'équité des protocoles qui sous-tendent leur conformité. On ne pourrait que consentir, confrontés aux critiques sur l'unité des moyens employés pour leur collecte, et leur premier traitement.

Ces critiques portent cependant sur des traits nécessaires, et finalement inhérents à toute approche empirique. Techniquement, elles ne concernent pas plus la cartographie linguistique que tout autre travail qui, avant de passer au traitement « neutre, juste et désintéressé » de ses données, fait intervenir diverses formes de subjectivité. Elles se fondent le plus souvent sur un postulat d'objectivité, calqué sur le paradigme exact. Mais il ne saurait être admis comme universellement valide. Le genre auquel se rattache la cartographie linguistique est plus visible par sa tradition en matière de pratiques que par son mouvement à prétendre à une science de type véricconditionnel. Elle s'affirme comme projet culturel et non pas comme défi épistémologique. Car, précisément, le problème premier de la cartographie linguistique est l'interprétation, et non pas la modélisation d'une réalité censée exister *a priori*. Aussi, son objet est-il constitué *dans* et *par* la succession des opérations qu'elle mène, entendues toujours comme des opérations d'un sujet interprétant, et animées par un seul esprit, au sein du même projet interprétatif. Comme telles, ces opérations, mais aussi les objets qu'elles constituent, doivent être envisagés en tant qu'étapes de l'interprétation. Et ceci dès la constitution du corpus.

Discipline frontière, faisant jonction entre deux domaines aux pratiques et aux caractéristiques distinctes, la cartographie linguistique se doit par conséquent de définir son identité sans pour autant renier les deux mondes qu'elle articule. Si son unité opérationnelle peut être cernée (comme un énoncé, ou une unité langagière projetée dans un espace géographique), elle doit hériter de ces deux facettes des particularités structurales, dont la définition semble essentielle dans les manipulations interprétatives envisagées, à court ou à long terme, de manière immédiate ou médiate. L'unification interprétative recherchée se situe ici dans l'intersémioticité

entre la linguistique et une forme de représentation de l'espace à un moment historique privilégié. Dans chaque plan, un ensemble de relations formelles, relevant de traditions et de considérations théoriques établies, peuvent être dégagées. Or, ces deux plans convoquent des représentations qui font d'eux, déjà, des supports interprétatifs pouvant être abordés indépendamment. Il nous faut ainsi préciser leur influence sur notre problématique, qui est celle de la réappropriation du rôle de l'informatique.

Schématiquement parlant, nous aurons à préciser l'interaction entre trois registres : le linguistique, le « cartographique » ou géographique, et le médium qui permet d'établir, par les élaborations qu'il permet, précisément ce type de jonction recherchée et spécifiquement interprétative. La cartographie linguistique va donc devoir mêler intimement ces deux registres interprétatifs, en élaborant des unifications de ces deux types distincts de données. L'exploration de cet espace hybride se fait dès lors par la projection simultanée de ces deux plans. Or, chacun de ces trois registres peut être l'objet d'une modélisation. La complexité émerge bien entendu de la sélection des modes d'organisation propres à chacun des deux premiers registres (linguistique et géographique). Mais aussi de leur caractère massif, techniquement envisageable désormais par la médiation du calcul et des techniques informatiques, « déviation » naturelle des objectifs initiaux de la cartographie linguistique.

Cette interaction est bien entendu orientée : les deux registres initiaux, le linguistique et le géographique, ne peuvent pas être équivalents. Bien qu'ils se mêlent étroitement, l'un l'emporte sur l'autre, pour chaque technique employée. Tracer des isoglosses revient à projeter des régularités linguistiques sur un espace géographique. Repérer la liste des variations phonétiques inter-lexicales de deux villages voisins sépares par une rivière est le mouvement inverse. Certes, le linguistique réécrit le géographique plus que l'inverse. La deuxième composante apparaît sans doute plus réduite. Mais elle reste opérationnellement aussi importante. Car, sur le plan des élaborations et des considérations connexes qu'elle permet, elle se révèle plus ouverte. Le régime de l'espace, de la proximité, de la distribution sur deux dimensions, est bien sûr pris en compte, même lorsque les lieux géographiques ne sont plus que des lieux d'élocution. Au fond, si le concept de l'espace est marqué par le médium linguistique, il ne perd pas ses rapports avec une authentique spatialité géographique. Le médium final est *a priori* toujours unitaire : une carte ou une liste de mots ou de phonèmes. Cependant, les passages de l'un à l'autre ne doivent pas se limiter à un aller ou un retour simples. Et chaque nouvelle projection est exactement le lieu d'une décision herméneutique qui peut se traduire par la création ou le repérage de relations structurelles entre des unités qui auront toujours un correspondant à l'arrivée. C'est peut-être cette difficulté des passages successifs qui peut être dépassée par l'informatisation : les projections successives, avec, à chaque demi-tour, des résultats à analyser et des décisions à prendre.

La complexité des relations spatiales qui est ainsi générée, précisément par des réélaborations, systématiques et surtout systématisées des unités (qu'elles soient premières ou dérivées), s'étale sur plusieurs niveaux d'organisation. Toutefois, en s'autorisant différents régimes de structuration des paramètres géolinguistiques, cette possibilité ne risque pas de devenir chaotique, puisque déjà normée par la légitimité du projet d'interprétation. En réalité, cette complexité n'est autre que le pendant herméneutique des structures intermédiaires produites. Elle annonce les possibilités d'interprétation offertes à chaque instant et à chaque niveau. L'informatisation peut être ainsi envisagée comme plus qu'une simple commodité pour les calculs : son appel peut en effet traduire le besoin constant d'une assistance à l'interprétation. Son intervention peut être aperçue comme le médiateur qui institue, opérant sur des unités sélectionnées par un chercheur, d'authentiques niveaux d'interprétation, précisément en faisant apparaître (mais aussi en permettant la régulation de leur structure herméneutique) des espaces originaux du point de vue de l'interprétation. Ainsi, le calcul peut devenir un catalyseur de

l'interprétation en matière de cartographie linguistique tant qu'il élabore des unités d'un niveau de base pour les intégrer à un niveau supérieur, où les nouvelles unités et les nouvelles formes structurales peuvent avoir une fonction interprétative qui contribue à avancer notre compréhension. Dans d'autres contextes, on parlerait d'interprétants : l'informatisation, sous une orientation toujours anthropocentrée de l'architecture globale, peut se poser comme une technique de génération d'interprétants, i.e. d'unités facilitatrices voire productives de l'interprétation. De ce point de vue, elle devient un véritable médiateur sémiotique.

Cette remarque peut être investie dans la nature même de la cartographie linguistique, pour y situer de nouvelles méthodologies de type coopératif entre différentes formes de sémiose, mais toutes communicantes par le biais de l'informatisation. Récupérée, une telle idée imprime une inflexion sur les objets de la cartographie linguistique qui se multiplient en nature, fonction et structure, et deviennent l'héritage d'élaborations normées toutes régies par le calcul. Comme tels, ils commandent l'agencement de différents paliers d'un même « espace interprétable ». On le disait, ils ne sont pas générés pour révéler une vérité linguistique généralement admise ou supposée préexistante. Encore moins pour pénétrer ses secrets. Mais simplement, pour instituer les conditions d'une lecture opératoire. De ce point de vue, le recours à la médiation informatique exprime la volonté de rechercher cette forme de représentativité qui débouche sur une valeur interprétable. Pour ce faire, une telle médiation adopte un régime qui n'est aucunement celui de l'expression de la vérité, mais celui de la contribution à la plausibilité d'une interprétation. Sans doute, il sera encore possible d'interpréter les mêmes unités ou les mêmes phénomènes différemment, même si les opérations informatiques qui les sous-tendent, voire les génèrent restent les mêmes. Car si l'informatisation augmente l'opportunité interprétative, elle ne la détermine pas complètement.

Ainsi justifié, le projet de la cartographie linguistique ne diffère guère d'une science herméneutique, même si pour son articulation elle doit recourir à un appareil soumis à des exigences de rigueur de type calculatoire. En tant que telle, sa norme restera pertinemment l'égard : non pas à une objectivité établie de manière rationnelle et suivant le mode exact, mais à une époque, à une mémoire, à un moment privilégié d'une culture, en deux mots, à une identité voulue ou simplement recherchée.

Cette remarque canalise et même force, en réalité, le rapport de la démarche de la cartographie linguistique avec le traitement des données. Cependant, même contraint par des principes herméneutiques, un tel rapport n'est pas à l'abri des difficultés. En effet, si la cartographie linguistique semble depuis toujours s'être approprié un cadre d'analyse robuste pour transformer ses données brutes en informations exploitables, elle s'est aussi heurtée aux difficultés bien connues de toute analyse de données : si la combinatoire est une chance pour fonder un espace interprétable, elle est rédhibitoire dans la mesure où elle devient rapidement difficile à maîtriser, puisqu'elle finit par permettre la génération de tels phénomènes de façon à la fois mécanique et massive. Et donc suggérer des parcours de lecture et d'interprétation concurrents difficilement pondérables. Ainsi, la prolifération informationnelle devient une contrainte supplémentaire, qui amende la qualité du matériau interprétable, précisément en le noyant dans une quantité de structures qui revendiquent, parfois de manière autonome et inattendue, les mêmes droits face à l'interprétation.

Ainsi, si l'outil informatique devient aujourd'hui, plus qu'un service, une nécessité, il ne peut pas se débarrasser des risques de surinterprétation. Et il n'est pas légitime d'espérer que les techniques du calcul et la puissance des machines puissent indéfiniment pallier la surabondance interprétative. Il y a deux raisons à cela.

La première est inhérente à la théorie même de la calculabilité, et concerne la complexité de la tâche : il est assez aisé de montrer que le rapport du corpus aux structures interprétables possibles qu'on peut y découvrir suit un ordre exponentiel. Les structures émergentes, et pouvant être le thème des interprétations visées, constituent généralement des unités de niveau supérieur et peuvent faire l'objet de traitements étagés et consécutifs, pratiquement sans fin. Un tel constat se reformulerait aussi sous les traits de la prudence calculatoire : avant même la formulation des premières promesses par l'introduction de l'informatique dans la manière de faire de la cartographie linguistique, on doit abandonner le rêve de l'exhaustivité.

La seconde fait appel à l'indécidabilité de la démarche même : en effet, on le disait tout à l'heure, le véritable problème de la cartographie linguistique n'est pas structurel mais interprétatif. Il ne concerne pas les régularités en tant que figures de structuration de l'espace des données, mais les ressources globalement pertinentes convoquées par ces dernières ainsi que l'évaluation de leur rôle dans l'échafaudage d'une interprétation, étape indispensable de la compréhension. Mais une interprétation est généralement hors des capacités de toute machine, et dépend d'un ensemble de facteurs subjectifs (et intersubjectifs), et de nombreuses incidences contextuelles, qui ne sauraient se décider à l'avance. Ni suivre toujours des chemins déterministes. Le rôle de la machine et de l'automatisme doit pouvoir (mais aussi savoir) se suspendre.

Ces deux remarques, si elles ne sont pas nécessairement positives, ne sont pas non plus négatives. Elles formulent en réalité, et avec clarté, les conditions d'une contribution légitime de l'informatique au sein de l'approche de la cartographie linguistique. En effet, abandonnant le rêve d'une automatisation large et inconciliable avec les pratiques herméneutiques, qui remplacerait globalement l'homme à l'instance interprétative même, il convient de penser leur rencontre sous le signe de l'assistance. Une assistance, cependant, anthropocentrée, qui ne signifierait pas partage équitable mais subordination de la machine, et donc du mode de calcul, aux priorités et choix interprétatifs fixés par l'homme. Ainsi, le calcul serait précisément convoqué pour assurer ce que l'homme peine à assurer (la conservation, le recouvrement aisé des informations, la constitution de structures suivant des critères choisis par l'homme et respectant ses projets interprétatifs, les comparaisons formelles, le maintien de la cohérence logique entre attributs, voire, parfois, la suggestion de voies alternatives...), et réaliserait, suivant de visées interprétatives différentes, des ponts intersémiotiques aux termes de notre discussion ci-dessus. Mais, dans tous les cas, il se limiterait au rôle de l'assistance. La primauté interprétative serait réservée à l'homme. L'indépendance de ses choix aussi. Une vision anthropocentrée signifierait, précisément, le respect de la compétence interprétative de l'homme, mais aussi un mode de collaboration entre l'homme et la machine qui institue une sorte d'instance métissée, combinant à la fois des ressources calculatoires et des compétences sémantiques en vue d'un même objectif interprétatif, mais subordonnant toujours les premières aux secondes.

Le projet que nous menons doit, par conséquent, être compris dans cette perspective. Il aurait comme définition la mise sur pied de ce que l'on appellerait volontiers « un poste d'interprétation », suivant une architecture précisément anthropocentrée pour le compte de la cartographie linguistique. Et donc, pour le compte des hommes qui la pratiquent. Les éléments que nous allons exposer dans la suite doivent être interprétés comme moments d'une telle démarche. Même si déjà relativement avancés, ils sont encore, dans notre esprit, loin de la vision globale que nous nous faisons du projet, et dont le signe de clôture, certes jamais définitif et toujours à travailler, s'annoncerait par l'intersection ergonomique de trois axes : l'intégration de ressources intertextuelles larges (et non seulement linguistiques et géographiques), le façonnage d'interfaces conviviales et répondant à divers régimes

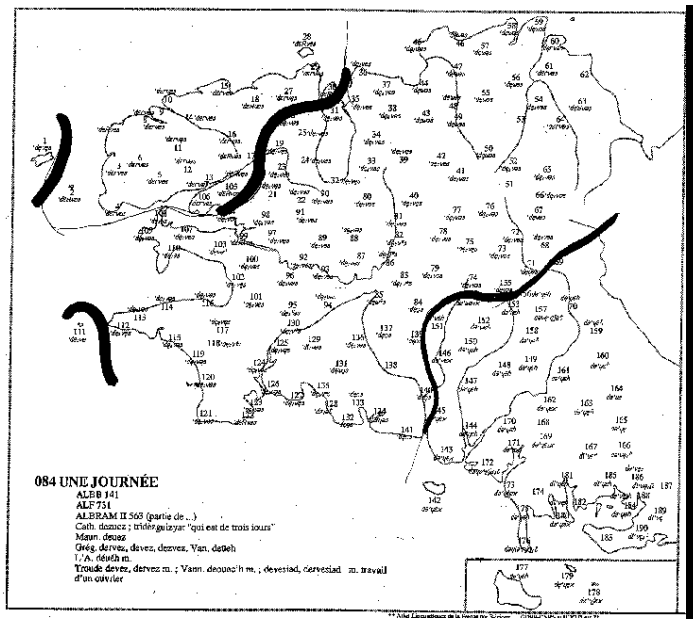
interprétatifs propres aux normes d'usage en vigueur dans la tradition de la cartographie linguistique, et l'affinement des possibilités d'assistance interprétative.

Dans la suite, nous décrivons succinctement les résultats, encore exploratoires, de la première rencontre, telle que nous l'imaginons, entre cartographie linguistique et informatique. Le cadre est d'un côté l'achèvement récent (1999) d'un Nouvel Atlas Linguistique de la Basse-Bretagne élaboré au Centre de Recherches Bretonne et Celtique, à l'Université de Bretagne Occidentale, comprenant 600 cartes réparties sur un réseau de 187 points d'enquêtes (presque trois fois plus que l'ALBB) ; et de l'autre, les recherches menées autour du thème des systèmes anthropocentrés en matière d'assistance en interprétation au sein du Laboratoire d'Informatique Avancée et Sciences de la Complexité de l'Ecole Nationale Supérieure des Télécommunications de Bretagne.

3. Technique informatique et interprétation. Les isoglosses comme objet informatique : génération, traçage, comparaisons…

On ne saurait éviter le concept d'isoglosse. Omniprésent, sa manipulation pose le point de départ de tout recours à l'informatisation. On comprend les raisons, dès lors que l'on voit dans le concept d'isoglosse le terme premier de tout échafaudage interprétatif en cartographie linguistique. Techniquement, cette interrogation vise, en premier lieu, le tracé des cartes.

La représentation la plus simple des données consiste, pour une unité linguistique donnée (mot ou expression, que l'on a demandé aux personnes interrogées de prononcer), à placer sur un fond de carte la forme phonétique locale correspondant à chaque point d'enquête. Cette mise en place vise à opérer une première topologie géolinguistique susceptible d'être exploitée différentiellement (comme exposé dans la partie précédente), sur des critères d'ordre lexical, morphologique, phonologique ou prosodique. Sa validité est relative à l'unité choisie et au mode d'établissement des différences, et sa mesure toujours locale.

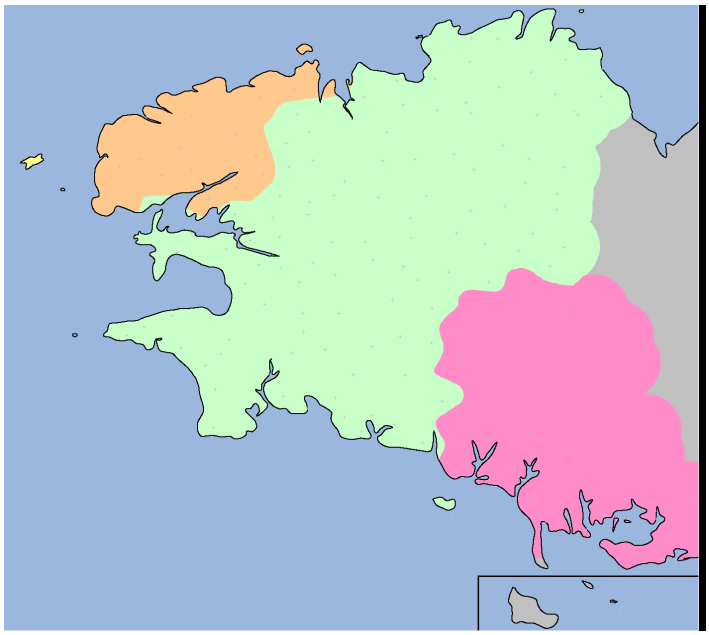


- 0 - Tracé manuel d'isoglosses pour "une journée".

Lorsque ce travail doit être effectué à la main, la solution généralement utilisée consiste à imprimer un fond de carte comportant les formes phonétiques locales, et à délimiter directement les zones au crayon. Hormis la lenteur, on imagine vite les inconvénients d'une telle procédure. Tout d'abord, les risques de confusion et d'erreur devant la masse d'informations à traiter (pour les travaux sur la Bretagne, les 187 formes linguistiques, de longueur variable, doivent êtres réparties sur une carte réalisée sur un papier de format A3 maximum). Ensuite, dessiner une carte plus aboutie qu'une simple esquisse est une tâche longue et laborieuse dont le résultat est souvent inégal.

'outil informatique propose déjà des solutions à ces différents problèmes, tout en améliorant le corélat interprétatif leur étant associé.

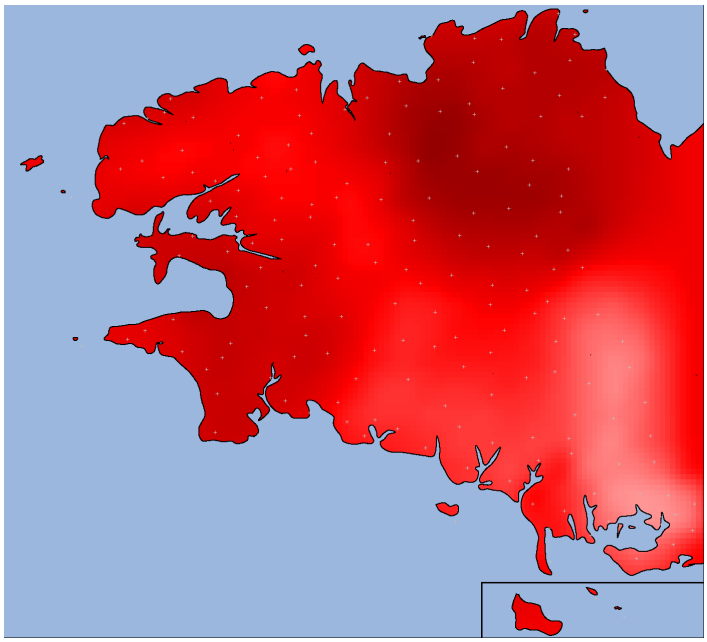
Tout d'abord, plutôt que de fournir les données brutes réparties spatialement, sont présentées à l'utilisateur uniquement une liste des différentes formes phonétiques existant sur la carte. D'une par, cela réduit la charge cognitive de l'utilisateur en réduisant la taille des données et en les présentant sous une forme plus facilement appréhendable. Et d'autre part, cela permet d'éliminer un biais interprétatif précaire en dissociant dans un premier temps le point de vue linguistique du point de vue géographique. Cette séparation est capitale : elle fonde les deux premiers registres discutés dans la partie précédente et prépare la médiation interprétative des traitements envisageables. L'outil permet à l'utilisateur de répartir lcs formes phonétiques en plusieurs classes comme il l'entend. Pour l'aider dans cette tâche, ln lui propose d'utiliser des expressions régulières (combinaisons de « jokers ») pour les répartir ln classes d'équivalence lorsque le nombre de formes différentes est important. Au fond, chacune de ces classes d'équivalence est la traduction logique - donc calculatoire - d'une isoglosse. Les critères utilisés pour séparer les formes en classes sont souvent suffisamment simples pour êtres exprimés sous cette forme ou comme combinaison logique de telles expressions. Une fois que l'utilisateur a réparti ces différentes formes en plusieurs classes qu'il juge pertinentes suivant une présomption interprétative, le programme tire de ces classes linguistiques des classes géographiques par l'association qui existe entre les notions de « point d'enquête » et « forme phonétique ». Connaissant les coordonnées de ces points d'enquête, il est alors possible de tracer les régions correspondantes grâce à un algorithme de dessin qui trace des frontières définies mathématiquement (l'utilisation d'une fonction ayant de bonnes propriétés rend la figure naturelle avec un tracé des frontières sans rigidité ni cassure). Dès lors il devient également très facile de corriger les éventuelles erreurs : si par exemple, une forme phonétique a été mal classée par l'utilisateur ou s'il change d'avis, il lui suffit de la changer de classe et de relancer le calcul de la carte. Cette facilité n'est cependant pas un gadget. Car, grâce à elle, s'établit entre le système et l'utilisateur un authentique travail de rectification interprétative, une sorte de commerce entre deux formes de sémiose susceptible de valider ou d'invalider la présomption interprétative qui fonde le tracé des cartes.



- 0 - Dessin des mêmes isoglosses par l'outil.

Une fois plongés dans cet espace de reconfiguration aisée des données (premières et secondaires), il nous est possible de prolonger le processus interprétatif : trouver, par exemple, des cartes similaires (et même suivant des critères de similitude paramètrables par l'utilisateur), rechercher des régions ou des frontières similaires à celles de notre carte, fusionner plusieurs cartes, demander des résultats statistiques, etc. Toutes ces opérations reconnaissent leur point de départ dans le registre linguistique. Mais pourquoi ne pas également suivre le mouvement inverse ? En partant précisément d'un point de vue géographique, la plate-forme informatique permet de plonger dans le « monde linguistique » en identifiant les phénomènes linguistiques qui se « cachent derrière » les phénomènes géographiques pris comme données de départ, puis trouver des phénomènes linguistiques similaires, les unifier, en extraire des formes typiques, etc. Par exemple, à partir de la frontière entre deux diocèses de Bretagne, l'outil pourrait identifier la variation morphologique d'un terme tel que « bonjour » qui aurait lieu sur cette

frontière, puis pourrait rechercher dans l'ensemble des données si des variations similaires existent.



- 0 - Utilisation du phonème "O".

Les analyses statistiques sont l'exemple le plus parlant de ce type d'application. Dans ce type de traitement, la base est interrogée transversalement. La requête (une expression régulière) est adressée sur tout le corpus, et on enregistre pour chaque point le nombre de formes phonétiques qui y ont vérifié le schéma d'interrogation. Les résultats peuvent alors être représentés sous la forme d'une carte en nuances de couleur. Ce type d'analyse est typiquement utilisé pour déterminer les zones dans lesquelles un phonème ou un ensemble de phonèmes est plus ou moins usité. Bien sûr, on ne peut envisager ce type d'approche qu'avec des données suffisamment nombreuses pour êtres considérées représentatives. Avec un corpus de plus de 600 mots et près de 200 points d'enquête, on voit bien qu'il est illusoire d'essayer de réaliser ce genre d'étude sans utiliser un ordinateur. Ce type d'approche, difficile à réaliser auparavant permet d'obtenir de très intéressants résultats. Et en particulier en inférer, par exemple, des attracteurs de structure, des régulations de dynamiques linguistiques.

L'outil, même sous sa forme actuelle, et donc exploratoire, propose déjà d'autres fonctionnalités intéressantes. Par exemple, en utilisant des techniques classiques d'analyse de données, il est possible de proposer à l'utilisateur de répartir les formes phonétiques automatiquement en autant de classes qu'il le désire. Ceci est réalisé en utilisant une distance d'édition (une distance numérique définie entre mots, considérés comme de simples suites de symboles), et un algorithme de classification. Cela revient à rassembler les formes par « paquets » à l'intérieur desquels les différentes formes se ressemblent. L'utilisateur spécifie donc un nombre de classes, et la machine se charge automatiquement des regroupements. Cette méthode, bien que se basant sur des critères relativement grossiers (puisque non sémantiques) permet d'alléger la tâche de l'utilisateur en dégrossissant le travail de classification lorsque l'on travaille avec un grand nombre de formes phonétiques différentes. Grâce à des techniques similaires, on peut faire émerger des zones stables à partir d'un ensemble de cartes déjà générées par l'utilisateur, ces zones correspondant alors probablement à des régions homogènes d'un point de vue linguistique. La technique pour cela revient à repérer des groupes de points d'enquête qui ont tendance à appartenir aux mêmes zones sur plusieurs cartes. On obtient alors une sorte de « fusion » des cartes de départ en une autre, qui en possède les caractéristiques communes.

A terme, le système devrait permettre d'assister et de guider l'utilisateur dans cette sorte de « spirale herméneutique », ce va-et-vient interprétatif permanent entre les deux registres linguistiques et géographiques, en composant librement les orientations des traitements, et

même permettre la prise en compte d'interventions « manuelles » de l'utilisateur. Nous ne sommes qu'au début. Nous ne pourrons probablement être qu'au début, dans la mesure où la fonction herméneutique a pour objectif non pas d'achever, mais d'initialiser constamment un processus de compréhension. Quoi qu'il en soit, dans cette initialisation, il y aura toujours des paliers de complexité à affronter. Et la médiation d'un traitement informatique anthropocentré, entendue comme médiation sémiotique au sein même d'une vision d'assistance à l'interprétation, si elle ne saurait prétendre à la clôture de la compréhension, peut sans conteste faire valoir l'argument et l'intérêt de formes et de chemins alternatifs, parfois mêmes originaux pour comprendre.

Bibliographie

Eco, U. : 1965 *L'œuvre ouverte*. Points. Seuil.

Eco, U. : 1996 *Interprétation et surinterprétation*. PUF.

Gadamer, H.G. : 1982/91 *L'art de comprendre* (vol. 1 et 2). Aubier.

Kanellos, I. : 1998 Exégèse néoplatonicienne et sémantique de l'interprétation. *Protée* 26(1) : 83-94.

Le Dû, J. : 1999 Nouvel Atlas Linguistique de la Basse-Bretagne (sous presse)

Rastier, F. : 1987 *Sémantique Interprétative*. PUF.

Rastier, F. : 1989 *Sens et Textualité*. Hachette.

Rastier, F. : 1995 Communication ou transmission ? *Césure* 8 : 151-195.

Ricœur, P. : 1995 *Histoire et vérité*. Seuil.

Ricœur, P. : 1995 *De l'interprétation*. Seuil.

Tanguy, L. : 1997 Traitement automatique de la langue naturelle et interprétation : contribution à l'élaboration d'un modèle informatique de la sémantique interprétative. Thèse de doctorat, Université de Rennes 1.

Thlivitis, Th. : 1998 Sémantique Interprétative Intertextuelle : assistance informatique anthropocentrée à la compréhension des textes. Thèse de doctorat, Université de Rennes 1.