

L'ORDINATEUR AU SERVICE DE LA GEOGRAPHIE DIALECTALE

Appliqués aux domaines les plus divers, les comptages ont l'avantage, sur les comparaisons qualitatives, de fournir des renseignements spécifiques aussi précis et utiles quoique différents, tout en étant d'une réalisation moins exigeante.

Sans aucun doute, le premier et le dernier mot - le choix et la préparation des données et l'interprétation des résultats - restent au spécialiste et ce sont les tâches essentielles; mais à condition de respecter certains principes élémentaires, un profane peut participer efficacement à des recherches par comptages sur des phénomènes qu'il ne connaît pas intimement; il peut les concevoir et, jusqu'à un certain point, en analyser les résultats. Si, en outre, les calculs sont automatisés, à l'économie de compétence s'ajoute celle d'innombrables heures de travail fastidieux et les risques d'erreur sont moindres.

Ces réflexions résultent d'une récente expérience personnelle. Malgré mon ignorance, j'ai pu me faufiler dans le domaine de la dialectologie dont la matière vivante et difficile n'est bien connue que de rares initiés, et avoir la chance de collaborer utilement à un travail entrepris par Louis Remacle, dialectologue wallon.

L'objet de mon exposé est d'expliquer et de discuter, sans traiter des faits dialectologiques mêmes, les méthodes que nous avons expérimentées et utilisées dans l'étude de la différenciation lexicale des parlers de Wallonie.

LES COMPTAGES ENTRE POINTS CONTIGUS

Beaucoup de linguistes refusent d'admettre l'existence des dialectes et donc celle des limites dialectales : le parler d'un endroit se distingue toujours de celui des endroits voisins, ne fût-ce que par d'infimes différences.

Mais si l'addition des phénomènes par lesquels deux lieux voisins diffèrent de façon sensible atteint un total relativement élevé, on a le sentiment d'une cassure qui peut d'ailleurs n'être qu'anomalie locale. Il se peut aussi que les différences restent nombreuses entre plusieurs points échelonnés en deux files plus ou moins parallèles; autrement dit, il se peut que la cassure se prolonge nettement et devienne un tracé de limite entre zones plus ou moins étendues, plus ou moins homogènes. Pour dresser la carte d'une région, il faut donc, pour un nombre de faits choisis, compter les différences entre chaque lieu repéré et ses voisins; la liste des comparaisons à faire est établie d'après un réseau qui s'inspire de la triangulation géodésique; les différences sont ensuite classées en une hiérarchie de valeurs qui détermine les tracés principaux et les lignes secondaires. C'est comme si on empilait autant de cartes que de phénomènes étudiés et que, par transparence, on vît apparaître des tracés plus ou moins sombres selon le nombre variable de coïncidences. Il ne peut être question d'inventorier toutes les variations dialectales; la valeur de la carte se limite donc strictement aux faits choisis qu'il s'agit, pour cela même, de bien préciser. Tel est le principe de la méthode que H. Guiter a très clairement expliquée dans un article important : *Atlas et frontières linguistiques* (Les Dialectes romans de France, Paris, Klincksieck, 1973, pp. 61-107).

J. Séguy, d'autre part, a basé sur le même procédé, sa méthode de dia-

lectométrie, longuement exposée dans *La dialectométrie dans l'A.L.G.* (Revue de Linguistique romane, tome 37, 1973, pp. 1-24). Avant eux, Atwood avait tracé, grâce à elle, pour les variantes phonétiques du langage, les limites sous-dialectales de la Wallonie (*The phonological divisions of Belgo-romance* dans Orbis, tome 4, pp. 367-389).

Elles ont été, dans une très large part, confirmées par une carte que L. Remacle a obtenue par un procédé plus rapide : le tracé sur une même carte des isoglosses de 22 faits phonétiques et de 10 faits morphologiques essentiels (*La géographie dialectale de la Belgique romane* dans Les Dialectes de France au moyen âge et aujourd'hui, Paris, Klincksieck, 1972, pp. 310-335). Il était tentant de voir si, dans le domaine lexicologique, plus fluctuant, il y aurait assez de coïncidences pour que se dessinent des tracés significatifs, et, dans l'affirmative, jusqu'à quel point ils pourraient correspondre à ceux de la carte phonétique d'Atwood.

L'enquête devait naturellement porter sur un nombre élevé de mots; seule était praticable la méthode par comptages dont j'ai pris en charge l'exécution automatique.

Avant de parler de notre travail, je voudrais faire comprendre que si les quatre enquêtes signalées reposent sur le même principe, les modalités d'application sont chaque fois différentes; elles dépendent notamment des matériaux dont on dispose et des objectifs qu'on a en vue; chaque étude doit être considérée en fonction de ces modalités bien particulières et les règles de l'une ne sont pas automatiquement vraies pour les autres.

Ainsi le réseau de triangulation paraît un jeu d'enfant, mais on déchanté bientôt. Guiter relie les points contigus en évitant que deux segments ne

se croisent, contrainte peut-être indispensable en géodésie, mais qu'il a le tort de ne pas justifier dans l'application présente. Or la triangulation est une opération qui laisse plus de liberté, donc plus de place à l'arbitraire, qu'on ne croirait, car les points voisins déterminent souvent des quadrilatères dont la configuration permet la scission aussi bien dans un sens que dans l'autre. Le dialectologue, d'après sa connaissance de la région, choisit le dessin qu'il présume devoir être le plus fructueux. Inventorier toutes les jonctions possibles entre les 305 points de l'Atlas linguistique de la Wallonie aurait mené à un gonflement impraticable de l'enquête; mais, à la liste des comparaisons établies d'après la triangulation simple, nous en avons ajouté auxquelles nous ne voulions pas renoncer, une quarantaine environ; ce qui porte à 955 le total de nos comparaisons. Séguy a, lui aussi, jugé bon d'admettre dans sa triangulation quelques segments croisés.

Peut-être Guiter n'a-t-il pas éprouvé le même besoin parce qu'il travaille sur le seul atlas linguistique de France qui soit exhaustif quant aux lieux inventoriés, celui des Pyrénées orientales. Pour 9000 km², il compte 382 points, soit une densité de 4,24 au myriamètre carré. Sans être exhaustif, l'A.L.W. offre un réseau relativement dense : 305 points pour 15.000 km², soit 2,03 au myriamètre carré contre les 0,36 de Séguy qui, dans l'Atlas linguistique de la Gascogne, ne dispose que de 154 points pour 42.250 km².

Une différence essentielle entre ces travaux provient du choix des phénomènes étudiés.

Guiter a pris les 200 premières cartes de l'A.L.P.O., telles qu'elles se suivent dans l'ordre alphabétique et comportant donc des faits linguistiques de natures diverses, mêlés et cumulés. Séguy procède à un inventaire plus ana-

lytique, cherche à équilibrer, dans le comptage global, les différents aspects du langage, en leur attribuant des pourcentages qu'il ne cherche pas à justifier rationnellement. Il ne le pourrait d'ailleurs pas. Je doute qu'on puisse jamais établir dans quelles proportions chiffrées interviennent la phonétique, la morphologie, le lexique... dans l'originalité d'un langage, ni savoir combien de faits lexicologiques "vaut" un phénomène phonétique aux occurrences bien plus nombreuses, ni combien de mots suffisent à représenter fidèlement un effectif lexical global dont on ignore le montant total exact complet ou restreint aux mots usuels.

C'est pourquoi la conduite plus élaborée de l'étude de Séguy permet des recherches plus fouillées, mais ne la rend pas plus rationnelle pour autant, et la carte d'ensemble n'a pas plus de valeur absolue que celle de Guiter car, je l'ai dit, toute carte de cette nature donne des informations nettement limitées par le choix et l'amalgame particulier des matériaux sur lesquels elle est fondée.

Cela dit, voici comment se répartissent les 426 faits étudiés par Séguy :

| | |
|-------------------------|-----|
| lexicologie | 170 |
| phonétique diachronique | 67 |
| phonologie | 77 |
| morphologie verbale | 44 |
| morpho-syntaxe | 68 |

Les 170 cartes lexicologiques ont été tirées au hasard, car l'A.L.G. groupe les cartes d'après un ordre logique et non alphabétique; elles représentent environ 10 % de l'effectif total. Séguy définit la distance linguistique entre deux

points en la calculant d'après les 5 paramètres dont les différences sont additionnées, et il baptise sa méthode, la dialectométrie. Outre les résultats globaux, il fournit séparément ceux qui concernent chaque paramètre, sans toutefois tracer les cartes particulières. Le chercheur peut comparer, mais au prix d'une patience qui peut décourager celui qui n'a pas lui-même manipulé ces données depuis le début. Au regard du labeur qu'a exigé cette étude monumentale qui force le respect, ce complément eût, semble-t-il, demandé un effort minime. Or de telles comparaisons sont intéressantes. Séguy le sait puisqu'il a justement trié les faits d'après leur nature. C'est le désir aussi de confronter les données de la lexicologie à celles de la phonétique qui nous a poussés à notre travail. Pour vérifier la valeur des deux méthodes, Séguy a appliqué la sienne à une partie du territoire étudié par Guiter et il reconnaît volontiers, sans pouvoir l'expliquer, que les résultats sont tellement proches qu'il estime les deux procédés également fiables. Guiter compte une différence par mot alors que ce mot peut être le support de plusieurs différences linguistiques de natures diverses. D'autre part, un fait phonétique compte chez Séguy pour une seule différence, alors que chez Guiter, grâce à leur répétition, les traits phonologiques et phonétiques à forte fréquence bénéficient d'une pondération naturelle : ainsi s'établissent sans doute les compensations qui expliquent la concordance des résultats.

Même si deux études portent sur des faits d'une seule et même nature, les principes qui guident leur choix peuvent différer au point que la comparaison des résultats ne peut se faire aveuglément. C'est seulement pour le prouver que je donne ici quelques précisions un peu fastidieuses qui, en elles-mêmes, n'intéressent que le spécialiste.

Comparons avec la nôtre la manière dont Séguy choisit ses cartes lexicologiques.

Il en a tiré 170 de façon tout à fait aléatoire. Je ne pense pas qu'il donne dans son article la raison de ce nombre en apparence donc arbitraire.

La variété lexicale, très grande, va de zéro (cartes monotypes) à 52. Il accepte les lacunes qu'il compte pour une différence. Beaucoup de lieux connaissent plusieurs types concurrents et si un seul de ces types est commun aux deux points comparés, aucune différence n'est comptée.

Voici, en regard, les caractéristiques de nos données.

Nous avons travaillé d'abord sur 50 mots; puis, encouragés par les résultats, nous avons fait porter nos investigations sur 100 mots. Ce qu'ont apporté les nouveaux calculs était assez différent pour que nous nous réjouissons d'avoir pris cette précaution. (1) D'ailleurs, la quantité de nos matériaux rejoint ainsi celle d'Atwood qui utilise les 100 cartes du premier volume de l'A.L.W., toutes relatives à des phénomènes phonétiques.

Guiter qui, comme nous, avait d'abord expérimenté l'étude de 50 mots, en a inventorié 100, puis 200, ensuite consigné et analysé les trois séries de résultats. Il prouve que 100 mots forment un lot fidèlement représentatif. Sa démonstration a malheureusement la faiblesse d'être empirique et basée sur sa seule expérience. Mais d'autre part, nos 100 mots qui n'ont pas la charge de tous les faits linguistiques mais seulement de la lexicologie, qui en outre ont été choisis, comme je vais l'expliquer, parmi les plus significatifs ont, me semble-t-il, une valeur représentative plus sûre que les 100 mots de Guiter fournis par le hasard de l'ordre alphabétique. Voilà pour le nombre; reste le choix.

La distinction des types est fondée, comme en général pour Séguéy, sur

l'étymologie. Puisque le but était de faire apparaître d'éventuelles lignes de dissemblance, on a rejeté les cartes monotypes ou trop confuses pour retenir celles où des aires se dessinent avec une certaine netteté. Beaucoup de mots sont représentés par 3 ou 4 types; deux en ont 9; un seul, tout à fait exceptionnel, en a 21. Dans le même souci et aussi pour ne pas compliquer l'exécution automatique des comparaisons, on a évité les cartes où trop de lieux comportent des formes concurrentes et quand le cas s'est présenté isolément dans une carte retenue, les spécialistes n'ont fait figurer que le type dominant ou le plus remarquable; ils ont enfin jugé opportun de combler quelques lacunes inévitables. On comprend tout de suite que cette politique dans le choix des mots tend à accentuer les tracés sur la carte.

C'est, parmi d'autres, une raison pour laquelle la hiérarchie que Guiter établit selon le pourcentage des différences, du parler à la langue en passant par le sous-dialecte et le dialecte, n'est sans doute pas applicable à nos résultats.

Tout principe de comptage qui répond à un dessein particulier est légitime. L'essentiel est de bien savoir ce que l'on compte et de ne pas l'oublier.

L'EXECUTION DE NOS COMPTAGES

L. Remacle et ses assistants, après avoir choisi les mots, ont confectionné les fiches d'après lesquelles j'ai mis les matériaux sur cartes perforées. Comme il s'agissait d'un emploi occasionnel, il n'y avait aucune raison de les enregistrer sur bande ou sur disque.

J'ai constitué 3 fichiers en économisant le nombre de cartes pour la facilité de la manipulation.

L'un comprend la liste des sigles des 305 lieux, chaque sigle comptant 5 caractères : 2 lettres (ou 1 lettre et 1 blanc) et 3 chiffres (A 025, NA110...); chaque carte porte donc 16 sigles. Comme repérer les lieux parmi cette liste est une des fonctions constantes du programme, ils sont rangés par ordre alphabétique pour permettre le repérage par dichotomie, plus économique.

Un autre fichier empile des lots de 5 cartes, un par mot étudié : une carte comporte ce terme et n'est, en fait, utile que pour les vérifications car elle ne joue aucun rôle dans l'exécution même du programme, et les suivantes alignent, rangés dans l'ordre alphabétique des lieux, les sigles représentant le type que le mot revêt dans chacun des 305 points; chaque sigle, une lettre, occupe une seule position.

Le troisième fichier comporte la liste des paires de lieux à comparer; chacune prend 10 positions, puisque le sigle d'un lieu en exige 5; j'aurais pu en mettre 8 par carte, mais par une habitude de clarté graphique que devraient désapprendre les usagers de l'informatique, j'ai souhaité un intervalle entre chaque groupe; pour obtenir des intervalles égaux, je n'avais pas le choix : j'ai dû me borner à caser 5 groupes par carte. La programmation, qui ne comporte que comparaisons et comptages n'offre aucune particularité (2).

Les résultats, triés par fourchettes, ont été reportés sur une carte, d'abord segment par segment, puis sur une seconde carte où, réunissant les valeurs différentielles relativement élevées, des traits plus ou moins épais selon

l'importance de celles-ci, dessinent de belles limites franches, telle celle qui cerne le domaine liégeois, étonnamment homogène, ou morcèlent d'autres régions comme l'Ardenne cruellement tailladée.

La carte, confrontée avec celle d'Atwood, montre de longs tracés concordants, mais aussi des dissemblances.

Ces résultats sont exposés et commentés par L. Remacle dans l'article qu'il publie dans *Les dialectes de Wallonie*, tome 4 (1976). J'ai tenu cependant à fournir ici cette carte lexicologique de la Wallonie, aboutissement essentiel du travail. On trouvera cette carte en annexe.

L'ANALYSE DES RESULTATS

On aimerait être persuadé que les valeurs différentielles entre points contigus quand elles atteignent un certain niveau ne mesurent pas seulement la dissemblance locale entre ces deux lieux, mais sont le signe d'une limite entre zones.

Or il est évident que si deux endroits voisins connaissent des formes différentes, l'une peut être isolée ou appartenir à un îlot minuscule, ou bien n'être qu'un point quelconque au sein d'une aire fort confuse; or dans le comptage automatique, cette différence peu significative compte évidemment comme une autre. Il ne pouvait être question de vérifier la valeur des 95.500 comparaisons exécutées par l'ordinateur, sinon mieux valait tout faire à la main dès le départ. On ne voit pas bien non plus comment évaluer la fréquence des cas litigieux, ni trouver un critère simple qui les dénoncerait.

J'ai procédé à plusieurs sondages, cherchant si, entre 2 ou 3 points choisis, le nombre des différences qui pourraient avoir réelle valeur de limite entre zones était assez proche du résultat des calculs aveugles. J'en ai fait plus tard pour comparer la valeur des comptages entre points contigus et entre points distants. L'expérience prouve qu'on se trouve presque toujours devant un cas d'espèce ne permettant pas de conclusion sûre. Procéder à ces inventaires est une tâche lente et fastidieuse puisque chaque mot, pour chacun des lieux comparés, doit être examiné individuellement. Comme de toute manière on ne pourrait aller au-delà d'un échantillonnage très partiel, il était inutile de la prolonger.

On est vite convaincu qu'on ne peut se fier aveuglément aux chiffres. Même si l'on borne l'examen à des lieux très voisins, on voit aussitôt que la différence entre 2 points, A et C, entre lesquels se trouve B, est loin de valoir automatiquement la somme des 2 différences AB et BC. Il y a sans cesse des retours en arrière, des formes isolées, des imbrications complexes qui font qu'un écart important marque toujours, bien sûr, une grande dissemblance entre deux points contigus, mais n'est pas le signe certain d'une limite de portée moins restreinte.

Les chiffres sont d'autant plus trompeurs que si C reprend une forme de A, pour ce mot la différence entre A et C est nulle alors que l'addition des différences partielles mène à 2; il faut ajouter comme correctif que la valeur différentielle de chaque segment n'est cependant affectée que d'une unité.

Nous avons repéré ainsi quelques cas de valeurs différentielles surfaites, mais il ne faut pas en exagérer les conséquences.

D'abord ces cas se signalent à nous comme exceptionnels; nous les avons

examinés parce qu'étonnés des chiffres fournis ou sachant que l'endroit pouvait faire attendre une anomalie.

D'autres sondages, en des endroits également soupçonnés de turbulence, ou choisis un peu au hasard, ont prouvé, au contraire, la fiabilité de nos calculs.

Pour que la valeur de l'écart soit vraiment mal évaluée, il faut d'ailleurs que, par une coïncidence qu'on n'a aucune raison de croire fréquente (mais qui est un faisceau de facteurs non élucidés par le calcul automatique), la conjoncture que je viens d'expliquer se rencontre entre deux points voisins pour un nombre relativement élevé de mots.

D'autre part, les accrocs à la progression géographique régulière de la différenciation du langage sont très fréquents; la plupart des segments prennent ainsi une valeur plus ou moins surfaite comme départageurs de zones, et on peut penser que, dans l'ensemble, les différences relatives ne sont pas gravement affectées. Puisque ce sont des différences relatives et non absolues que nous ambitionnons seulement de connaître, nous pouvons donc, jusqu'à un certain point, être rassurés sur la validité de nos comptages.

LES COMPTAGES ENTRE POINTS DISTANTS

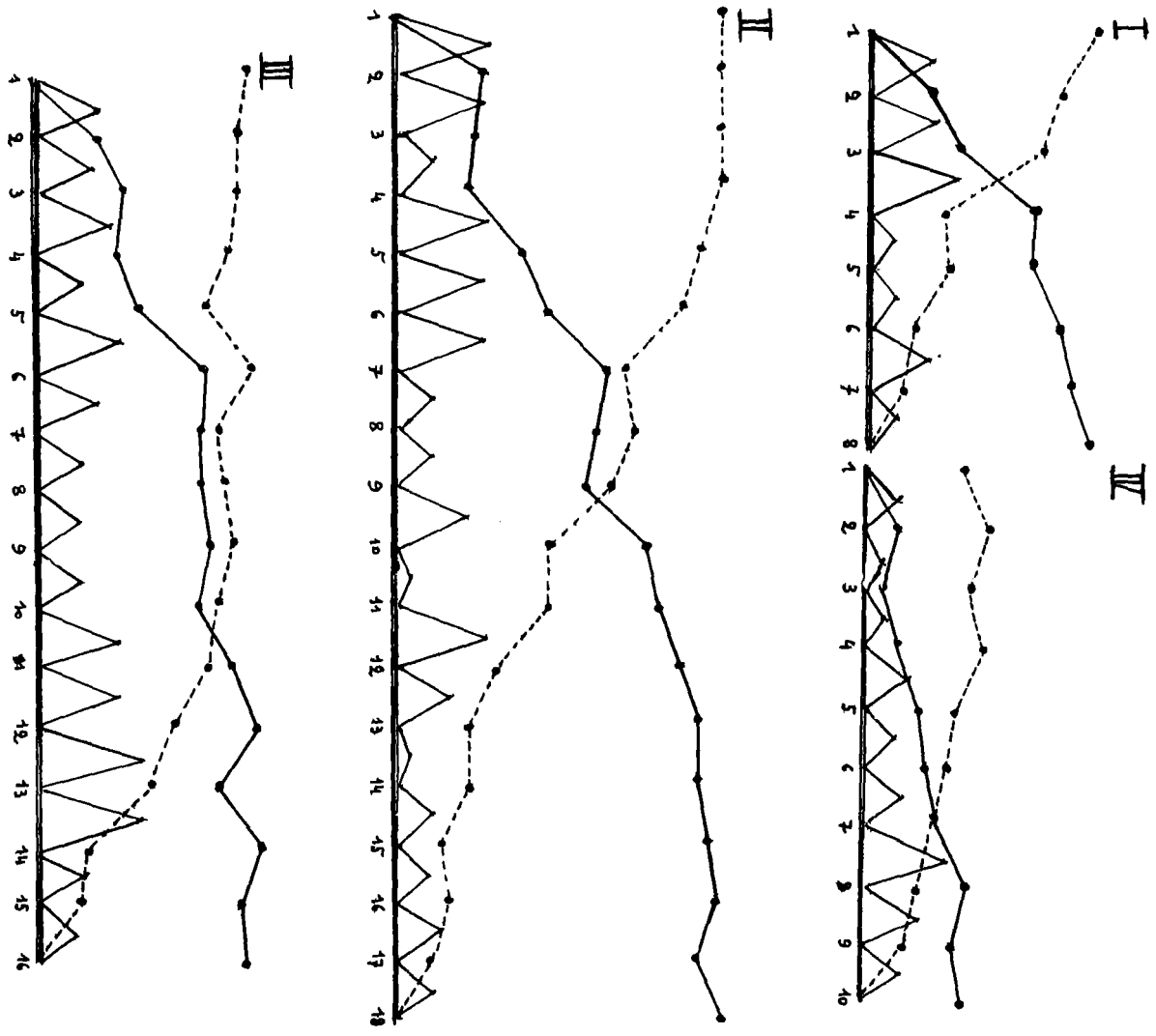
Les comptages entre points distants fournissent des mesures qui ne font peut-être que confirmer des faits connus, mais qui permettent de les apprécier de façon objective et plus précise, puisque chiffrée. Comme Séguéy l'avait fait pour la Gascogne, nous avons balafré la Wallonie de nombreuses droites (en tout 23 : 8 horizontales, 7 verticales et 8 obliques).

Toujours sur le plan lexical, le seul qui nous occupe, nous avons calculé les différences, non seulement entre les extrémités, mais aussi entre chacune d'elles et un assez grand nombre de points voisins et plus ou moins équidistants échelonnés le long de ces lignes, parcourues ainsi dans les deux sens.

Rien de tel pour apprécier facilement des informations chiffrées que de les traduire en diagrammes. Sur un même graphique, par ligne, nous avons tracé les courbes de croissance, une dans chaque sens, et relié entre eux les points du parcours par des angles qui figurent la différence calculée entre points contigus, prise absolument. Pour la facilité, je l'appellerai *l'écart*, et je nommerai *saut* l'accident de la courbe qui mesure de combien tel point est, plus ou moins que son voisin, distant par la lexicologie de l'extrémité du parcours prise comme point de référence.

Des graphiques, dont je joins 4 spécimens où par raison pratique j'ai remplacé les sigles des lieux par des chiffres, découle une constatation générale, déjà faite par Séguy et que le bon sens pouvait faire prévoir : la courbe de croissance de la valeur différentielle a tendance à s'aplanir, ou même s'aplanit très nettement à partir d'un certain point; la différence maximale est atteinte ou peu s'en faut, et seul subsiste le capital commun aux deux points extrêmes. Même quand ce phénomène est moins sensible, la différence maximale apparaît souvent avant la fin du parcours; il n'est même pas rare qu'elle dépasse légèrement celle de l'extrémité.

On remarque vite aussi que l'écart entre deux points contigus détermine parfois des sauts d'importance inégale selon le sens du parcours et qu'il arrive qu'entre l'écart et ces sauts, la différence soit frappante.



Ces faits sont la conséquence naturelle de la complexité bien connue avec laquelle le dialecte s'est diversifié au sein d'une aire; j'y reviendrai.

Les comptages entre points distants permettent de comparer les différences maximales, très variables selon les lignes puisqu'elles passent de 30 à plus de 80 pour 100 mots; même pour deux distances égales, les chiffres peuvent être de 30 et 50, ou 50 et 84; ce dernier cas, par exemple, est celui des deux droites les plus longues : la seconde, horizontale, longe à court intervalle la frontière nord de la Wallonie (Tournai-Verviers); l'autre, oblique, est l'hypoténuse du triangle rectangle que dessine le domaine (Tournai-Virton).

Si l'on tient compte de la longueur très variable du parcours, on peut calculer le pourcentage de différenciation par unité de distance; dans l'exemple choisi, il est respectivement de 2, 6 et 4, 4, et pour l'ensemble des 23 droites, il peut aller du simple (Tournai-Virton) au triple (sur la droite qui, partant de Hannut, coupe verticalement la Wallonie) ou même plus (sur l'oblique reliant Corbion sur la Semois à Vielsalm). Si à présent, on observe le rythme de croissance des valeurs différentielles, on trouve, cela va de soi, des parcours qui montent à pas réguliers et d'autres qui restent au ras du sol, puis s'élèvent en quelques bonds jusqu'à leur plafond.

Dans l'ensemble, ces courbes de croissance, ces pourcentages de différenciation correspondent entre eux, et également avec le caractère plus ou moins homogène des régions traversées telles que les décrit la carte lexicologique dressée d'après le réseau de triangulation.

Il est normal qu'un long parcours à travers des aires homogènes provoque un pourcentage de différenciation et même une différence maximale absolue

plus faibles qu'un autre trajet, même court, parmi une zone de turbulence linguistique, et que l'homogénéité d'un territoire succédant à un autre très morcelé entraîne, par compensation, un pourcentage moyen. Outre ces particularités locales, il est logique qu'en raison du phénomène d'aplanissement des fins de parcours, le pourcentage des lignes les plus longues, où naturellement il est le plus sensible, soit relativement faible; logique aussi qu'aux limites que la carte trace entre deux aires correspondent sur les courbes des sauts importants.

Certains résultats pourtant peuvent intriguer, comme les pourcentages différents de deux parcours parallèles traversant des régions assez également mouvementées, ou, entre deux lignes, une inégalité de pourcentage que l'on comprendrait mieux inversée.

Mais aucune anomalie vraiment inquiétante n'a été révélée. Les sujets d'étonnement ou en tout cas de réflexion sont plus nombreux quand on cherche de plus près jusqu'où va le parallélisme entre les deux sauts en sens inverses qui caractérisent un segment, entre ces sauts et l'écart.

Si la majorité des tracés répondent à ce qu'on attend, les anomalies locales ne sont pas rares.

Théoriquement, les courbes des deux parcours d'une même ligne en sens opposés devraient être symétriques et cela leur arrive souvent (1).

On connaît assez la complexité de la réalité linguistique pour ne pas espérer trouver constamment pareille ordonnance; les dérogations seront justement la marque d'une plus grande diversification dans la région qu'elles affectent.

Il arrive que l'un des parcours soit plan quand l'autre effectue un saut, ou qu'ils dessinent deux pentes d'inclinaison très inégale (I entre 2 et 3), ou même, ce qui est beaucoup plus anormal, deux pentes parallèles et non plus opposées. Un tracé de ce genre résulte souvent de l'existence d'un point qui, nettement plus ou moins que ses voisins, diffère de l'extrémité de référence, ce qui explique que le trait formant un angle inattendu doit, pour reprendre un niveau plus normal, devenir momentanément parallèle au parcours inverse; le graphique III en offre des exemples clairs entre 5 et 7, 11 et 13, 13 et 15.

J'ai déjà dit que les investigations approfondies montrent que chaque situation a ses particularités; on ne peut conclure qu'à l'action conjuguée de facteurs qui s'additionnent ou s'annulent.

LA CONFRONTATION DES DEUX PROCÉDES DE COMPTAGE

Très souvent les résultats des comptages entre points distants confirment les données des comptages entre points contigus, mais parfois ils dévoilent des divergences dont il est bon de connaître l'existence, même si on ne peut les élucider entièrement ou si on ne cherche pas à le faire,

Les anomalies qui frappent sans doute le plus au premier regard que l'on porte sur les graphiques, ce sont des discordances entre l'importance des écarts et celle des sauts.

Elles résultent d'abord du phénomène d'aplanissement des fins de parcours dont il a été question. Le graphique II le représente typiquement et l'analyse détaillée des matériaux, effectuée sur 50 mots, l'explique parfaitement.

Les écarts entre les points 1, 2, 3 et 4 sont importants, mais la courbe qui a son origine à l'extrémité opposée, en 18, est parfaitement plane; or il n'y a que 6 formes communes aux points 18 et 4, formes qui se retrouvent en 3, en 2 et en 1 alors que ne réapparaît en ces points aucune autre forme de 18; donc 1, 2, 3 et 4 peuvent rivaliser d'indépendance et se disputer autant qu'ils le veulent les diverses formes des 44 autres mots inconnues en 18, le tracé du parcours n'en sera pas le moins du monde affecté.

En dehors de cette situation, disons classique, apparaissent d'autres discordances entre écarts et sauts.

Ainsi entre deux lieux contigus du Namurois, l'écart a une valeur très modérée; or le tronçon figure dans deux lignes (une horizontale et une oblique) et les quatre courbes (deux dans chaque sens) dessinent à cet endroit des sauts proportionnellement très importants. La situation inverse se rencontre également.

Quel parti tirer de cette comparaison entre les écarts et les sauts ? S'ils ne s'accordent pas trop mal (I, II), les deux procédés de mesure prouvent évidemment une aptitude égale à révéler la réalité. La stagnation de la courbe et les écarts faibles caractérisent une certaine uniformité du langage tandis que sauts et écarts moins modérés montrent une réelle progression.

Mais si stagnation du parcours et grands écarts se démentent (III, IV), il y a des retours en arrière, des formes isolées, des imbrications, soit de nombreuses particularités locales. Cette constatation parle sûrement en faveur du comptage entre points contigus qui rend mieux compte du caractère qu'a pris le dialecte dans l'aire considérée.

Le grand inconvénient d'ailleurs de l'autre méthode est que, contrairement aux écarts, les sauts n'ont pas tous même valeur de signification.

Dans les extrémités du parcours, surtout s'il a une certaine longueur, les écarts peuvent traduire des différences locales ayant force de limites, ne fût-ce que secondaires, et que le comptage en lignes rabote à l'excès dans un des deux sens. On ne peut non plus donner la même valeur au premier saut du parcours : les différences entre les deux premiers points ont le caractère absolu de l'écart, n'ont pas encore eu l'occasion d'être neutralisées comme ce sera le cas pour les suivantes; or ce point de départ n'en est pas un réellement, c'est un point choisi arbitrairement au sein d'une réalité continue.

Si on a le sentiment qu'il est dangereux de bâtir sur les débuts et fins de lignes, à partir de quel point les tracés seront-ils fiables ? La réponse est d'autant moins possible que la configuration triangulaire de la Wallonie leur donne forcément des longueurs très inégales.

Pour pallier cet inconvénient, on a envisagé de choisir d'autres points de départ, en allant par exemple du point médian des lignes vers les deux extrémités, ou en comptant les différences le long d'axes en étoile à partir d'un seul centre, ou de plusieurs points choisis comme centres d'aires présumées assez homogènes.

Convaincu que les deux méthodes de comptage se corrigent mutuellement, on a même cherché si, en travaillant sur les graphiques, on ne pourrait inventer une formule où se combineraient, mais dans quelles proportions ?, les résultats qu'apporte chacune d'elles.

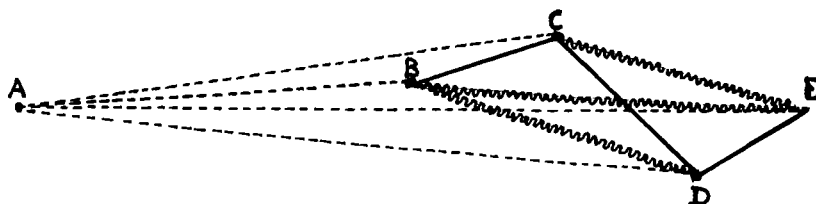
Après quelques essais et toutes réflexions faites, on a jugé ces procédés plus défectueux qu'efficaces et on y a renoncé. D'après les résultats des 23 lignes, les calculs portant sur 50 mots, on a ébauché une carte. Bien sûr les lignes maîtresses rejoignent celles de la carte par triangulation, mais il y avait tout de même des particularités saillantes qui n'étaient pas imaginaires. On ne pouvait renoncer à de telles informations. On se souvenait aussi que les écarts entre points contigus pouvaient donner trop d'importance à des fluctuations étroitement locales.

Qu'a-t-on décidé ?

D'établir sur le comptage entre points contigus la carte qui était le but même du travail et qui, pour être valablement comparée à la carte phonétique d'Atwood devait reposer sur des principes semblables. Mais d'appliquer à nos matériaux les comptages entre points distants à titre de vérification et d'information complémentaire.

L'EXPERIMENTATION SYSTEMATIQUE DU SECOND PROCÉDE

On a donc pratiqué un traitement plus systématique et portant sur les 100 mots. Les lignes horizontales et verticales ont été rapprochées et les points retenus beaucoup plus nombreux; malgré ce resserrement, on s'est vite aperçu qu'entre les lignes forcément sinueuses qui reliaient les points, il demeurait d'autres points qu'on n'avait pas plus de raison d'abandonner que les autres, et aussi que certains lieux devaient figurer dans deux séries contiguës, leurs rapports avec chacune ne pouvant être négligés; plutôt que de lignes, il s'agit donc maintenant de bandes de largeur irrégulière de sorte qu'il arrive qu'on profite souvent d'évaluations par recoupement.



Ainsi en calculant les différences entre une extrémité A et les points B, C, D, E, en plus des sauts BC, CD et DE, on obtient à volonté la valeur des sauts BE, BD et CE. Il arrive aussi qu'un même segment ait été retenu dans deux bandes perpendiculaires et soit donc examiné par rapport à 4 points de référence différents.

On a vérifié en outre si, dans cet inventaire, figuraient tous les segments qui, par le comptage entre points contigus, s'étaient révélés significatifs. Par contre, puisque cette enquête n'avait plus qu'un but de vérification et pour éviter les calculs inutiles, on a tenu compte de tout ce que l'on savait déjà et traité plus légèrement quelques coins où la langue est indubitablement très uniforme, de même que les extrémités de parcours.

En fin de compte, on arrive à un réseau très serré, plus serré même en certains endroits que le réseau de triangulation, non qu'on l'ait nécessairement voulu, mais par le jeu de la disposition des points d'enquête, du croisement des bandes perpendiculaires et des résultats qui s'offrent indirectement, parfois cherchés, parfois non. Il était superflu d'ajouter des bandes obliques à ce quadrillage. On a seulement conservé la grande hypoténuse parce que les points qu'elle supporte ne peuvent figurer, dans les autres bandes, qu'en début ou en fin de parcours. Par le gros millier de comparaisons effectuées, qui sont en fait 102.000 comparaisons de mot à mot, 510 segments sont ainsi appréciés (une fois dans chaque sens), auxquels

s'ajoutent, si on le veut, plus de 120 autres segments dont la double estimation est obtenue par voie indirecte.

Le fouillis de cette enquête en fait la force : tout le terrain est ratissé, ou peu s'en faut; mais aussi, sans doute, la faiblesse : ces innombrables comparaisons qui se croisent, qui se font à partir de points de référence si divers deviennent une matière plus hétéroclite et peu aisée à bien utiliser. Ainsi, si l'on veut chiffrer un saut entre deux points, laquelle choisir des deux valeurs que donnent les parcours inverses ? ou bien faut-il les additionner ? Le manque de rigueur de la méthode malgré nos soins et nos recherches nous rend perplexes et, avouons-le, un peu découragés.

Pour notre information personnelle, nous avons tout de même voulu traduire les calculs sur une carte, On a représenté par un trait les limites qu'indiquent les sauts considérés selon une certaine orientation, par exemple W-E; puis, par des traits de trois couleurs différentes, celles que déterminent les sauts selon les trois autres directions. Pour garder à la carte sa clarté, on n'a pas fait varier l'épaisseur du trait selon l'importance des sauts, se contentant de ne pas descendre sous un certain seuil. Mais comme, selon les endroits, des 4 traits, 2, 3 ou même tous coïncident, on trouve une évaluation que l'on pourra juger valable ou non, de la force des limites tracées.

Confrontant les deux cartes, nous avons eu enfin la satisfaction, et pour tout dire le soulagement, de constater que, malgré notre scepticisme, la très grande majorité des tracés concordent même si la force de la limite n'est pas toujours identique, ce qui ne peut étonner si on pense à la différence des systèmes d'évaluation. Il y a pourtant divergence en quelques

endroits. Le plus déconcertant c'est, dans le Namurois, une limite fortement tracée dans la seconde carte et à peine perceptible ou inexistante dans la première, limite qui isole la bande de Bierwart, Andenne, Ohey, Florée de ce qui l'avoisine à l'ouest et au sud. Elle est tellement nette qu'elle doit être significative. Un dialectologue curieux pourra peut-être élucider cette anomalie et quelques autres de moindre extension. La confirmation d'ensemble qui prouve la validité du procédé donne du prix à la révélation de ces divergences locales et l'appoint de cette seconde méthode n'est donc pas négatif. Ces expériences que nous n'aurions jamais eu la patience de poursuivre sans le secours de l'informatique, ces tâtonnements en vue d'améliorer ou de compléter une méthode connue n'ont cependant pas mené à la solution sûre et rationnelle que nous espérions; l'on pourra penser que je me suis trop étendue à les décrire; mais si un chercheur veut un jour mettre ses pas dans les nôtres, il le fera sur un terrain déjà débroussaillé.

CONCLUSION

Peu habituée à une matière aussi constamment et imprévisiblement fluctuante que le dialecte, j'ai eu souvent, au cours de notre travail, le sentiment que c'était naïveté et leurre que de lui appliquer une méthode qui, basée sur le calcul et agrémentée de graphiques, s'annonçait rigoureuse, pour ne pas dire scientifique.

J'ai été rassurée cependant, et peu à peu convaincue par l'unanimité avec laquelle nos différents procédés d'évaluation décidaient de la très grande majorité des tracés que nous cherchions, même si en certains points se marquaient des divergences, voire des contradictions très localisées. La valeur d'aucun de ces procédés n'est scientifiquement démontrable, mais ils

ne peuvent nous tromper avec autant de connivence, et la vérité que nous voulions cerner - des lignes de dissemblance - existe donc et les tracés auxquels nous avons abouti traduisent, dans leur ensemble, d'une façon approximative mais assez fidèle, des faits linguistiques certains.

Si on songe à la complexité du matériau, au nombre de facteurs occasionnels qui, depuis l'enquête sur le terrain, entrent en jeu à tous les niveaux, à la dissémination plus ou moins irrégulière des points inventoriés, on se console mieux de ne pouvoir trouver une méthode qui mène à la vérité absolue, puisqu'on est persuadé, dès le départ, que dans ce domaine, moins encore que dans bien d'autres, il n'y a pas de vérité absolue.

L'attitude scientifique est dès lors de bien savoir ce que représente ce qu'on trouve et d'en avertir ses lecteurs.

M. DUCHESNE-DEGEY

NOTES

- (1) Nous avons réparti les valeurs différentielles des segments en 6 fourchettes. On constate des différences importantes entre les résultats fournis par chacun des 2 lots de 50 cartes; ainsi parmi les 11 ou 12 comparaisons révélant le plus de différences, donc situées dans la fourchette 6, une seule est commune aux deux lots, les autres rétrogradant dans la série parallèle parfois jusqu'à la fourchette 2. Comme il fallait s'y attendre, en passant de 50 à 100 mots, ce sont les catégories extrêmes qui voient leur effectif réduit au profit des valeurs plus modérées; mais il y a, en gros, constance dans la répartition proportionnelle des valeurs entre les fourchettes.

Voici, à titre indicatif, la répartition des 955 comparaisons effectuées sur 100 mots; le nombre de différences le plus élevé est de 38.

| <u>De 0 à 6</u> | <u>De 7 à 12</u> | <u>De 13 à 18</u> | <u>De 19 à 24</u> | <u>De 25 à 30</u> |
|-----------------|------------------|-------------------|-------------------|-------------------|
| 155 | 390 | 263 | 115 | 28 |

Plus de 30

4

- (2) Pour les curieux peu familiarisés avec la programmation, qui aimeraient cependant comprendre comment s'organise pour l'ordinateur la coordination des 3 fichiers, j'en explique le circuit logique.

L'ordinateur lit successivement les deux premiers fichiers et enregistre les données dans deux tableaux indépendants, puis en abordant le troisième fichier, il commence l'exécution. Après la lecture de la première paire de lieux, il repère dans le premier tableau le numéro d'ordre de ces lieux. On lui ordonne de comparer ensuite sigle à sigle, c'est-à-dire

mot à mot, les lettres du second tableau casées aux deux numéros d'ordre qui viennent d'être repérés, et d'additionner les différences rencontrées; la liste des mots épuisée, le résultat est classé selon le total de ces différences dans un tableau récoltant les résultats répartis par fourchettes. L'ordinateur lit la paire suivante, recommence l'opération, et quand le troisième fichier a été entièrement parcouru, le tableau des résultats est prêt à sortir.

