In Buchi, Éva/Chauveau, Jean-Paul/Greub, Yan/Pierrel, Jean-Marie (éd.), 2015 : *Actes du XXVIIe Congrès international de linguistique et de philologie romanes (Nancy, 15-20 juillet 2013). Allocutions d'ouverture, conférences plénières, tables rondes, conférences grand public.* Nancy : ATILF : http://www.atilf.fr/cilpr2013/actes/section-0.html

Allocution d'ouverture des représentants de l'ATILF

Au nom du comité d'organisation et de l'ensemble des membres du laboratoire ATILF, nous souhaitons remercier tous les congressistes qui sont venus à Nancy participer au XXVIIe Congrès international de linguistique et de philologie romanes[1]. Nous voudrions aussi remercier la Société de linguistique romane, son président et son bureau pour la confiance qu'ils nous ont témoignée en nous confiant l'organisation de ce prestigieux congrès.

În numele comitetului de organizare şi al întregului colectiv al institutului ATILF, vrem să mulţumim tuturor celor care au venit la Nancy pentru a participa la al douăzeci şi şaptelea Congres internaţional de lingvistică şi filologie romanică. Mulţumirile noastre se îndreaptă de asemenea către Societatea de lingvistică romanică, către preşedintele şi biroul său pentru încrederea pe care ne-au acordat-o, oferindu-ne posibilitatea de a organiza acest prestigios congres.

Le laboratoire ATILF (Analyse et Traitement Informatique de la Langue Française) est issu de trois anciennes unités : l'INaLF (Institut National de la Langue Française), une unité du CNRS, et deux équipes universitaires : le LanDisCo (Langue, Discours, Cognition) et le CRAPEL (Centre de Recherches et d'Applications Pédagogiques en Langues). Ses deux tutelles sont le CNRS et l'Université de Lorraine, et il réunit actuellement 138 membres : des enseignants-chercheurs et des chercheurs, des ingénieurs et des techniciens, du personnel administratif et des doctorants et post-doctorants. Le laboratoire est structuré en cinq équipes scientifiques : « Lexique », « Linguistique historique française et romane », « Discours, langue et cognition », « Didactique des langues et sociolinguistique » et « Ressources : normalisation, annotation et exploitation ».

Il laboratorio ATILF (Analyse et Traitement Informatique de la Langue Française) nasce dalla confluenza di tre precedenti unità: l'INaLF (Institut National de la Langue Française), un'unità del CNRS, e due gruppi di lavoro universitari, LanDisCo e CRAPEL. Le sue due istituzioni di riferimento sono il CNRS e l'Università della Lorena, e attualmente riunisce 138 membri tra docenti-ricercatori, ricercatori,

---

[1] Nos remerciements les plus vifs s'adressent à Xosé Afonso Álvarez Pérez (Lisbonne), Maria Reina Bastardas i Rufat (Barcelone), Myriam Benarroch (Paris), Giorgio Cadorini (Opava), Marco Maggiore (Florence), Mihaela-Mariana Morcov (Nancy), Simone Pisano (Sassari), Fernando Sánchez Miret (Salamanque) et Paul Videsott (Bolzano) pour les traductions en roumain, en italien, en sarde, en frioulan, en ladin, en catalan, en espagnol et en portugais qu'ils nous ont procurées.

ingegneri, tecnici, personale amministrativo, dottorandi e post-dottorandi. Il laboratorio è strutturato in cinque unità scientifiche: «Lessico», «Linguistica storica francese e romanza», «Discorso, linguaggio e cognizione», «Didattica delle lingue e sociolinguistica» e «Risorse: normalizzazione, annotazione e utilizzo».

Pour l'organisation de ce congrès, l'ATILF a non seulement pu compter sur l'infrastructure du laboratoire et l'engagement de ses membres, mais a pu bénéficier aussi du soutien de ses tutelles, le CNRS et l'Université de Lorraine, d'un ensemble de partenaires institutionnels : le Ministère de l'Enseignement Supérieur et de la Recherche, la Délégation Générale à la Langue Française et aux Langues de France, l'Agence Universitaire de la Francophonie, l'Institut de Linguistique Française, la Maison des Sciences de l'Homme Lorraine, et des collectivités territoriales avec lesquelles le laboratoire entretient depuis plusieurs années des partenariats importants : le Conseil Régional de Lorraine, le Conseil général de Meurthe-et-Moselle, la Communauté urbaine du Grand Nancy et la Ville de Nancy. Qu'eux tous reçoivent nos remerciements les plus chaleureux.

Pro s'ammàniu de su congressu, s'ATILF at àpiu sa fortuna manna de poder impreare sos mèdios de su laboratòriu e at retziu una manu manna de azudu dae sos chi bi trabbalant ma at fintzas pòtiu aer s'apozu de sos entes suos amparadores: Su CNRS e s'Universidade de Lorena; una cumarca de assòtzios istitutzionales: Su Ministèriu de Imparonzu Superiore e de sa Chirca, Sa Delegatzione Zenerale de sa limba frantzesa e de sas limbas de Franza, s'Agentzia de s'Universidade pro sa Francofonia, Istitutu de linguìstica frantzesa, sa Dommo de sas iscèntzias de s'òmine de Lorena; sas istitutziones de su territòriu chi traballant a costazu de su laboratòriu dae medas annos: Su Cossizu Regionale de Lorena, su Cossizu zenerale de Meurthe-et-Moselle, sa Comunidade de totu sas biddas a fùrriu de Nancy e sa Dommo Comunale de Nancy. A cada unu bos torramus milli gràssias sas prus sintzeras.

Nous voudrions aussi rendre hommage à l'action du comité de programme, présidé par Jean-Pierre Chambon en sa qualité de président de la Société de linguistique romane, lequel a sollicité différentes personnalités pour présenter des conférences plénières ou grand public ou pour présider des tables rondes, mais qui a surtout arrêté la thématique des seize principales sections du congrès, tout en en confiant la responsabilité à des présidents et vice-présidents. Une dix-septième section, restreinte à une seule journée, est par la suite venue compléter la gamme.

O volaressin ancje fâi onôr ae ativitât dal comitât programatic, che i à stât da cjâf Jean-Pierre Chambon inte sô cualitât di president da la Societât di linguistiche romaniche; cun di plui che al à interpeladis diviersis personalitâts par la presentazion des conferencis plenaris o di grant public e po ancje par la presidence des taulis tarondis, ma che soredut al à inmaneât la tematiche des sedis sezions primaris dal congrès, tal timp istès confidantjur la responsabilitât ai presidents e vice-presidents. Tal ultin e je rivade inmò une sezion disesiet, limitade a nome une zornade, che e complete la sielte.

Ce sont les présidents et vice-présidents qui ont eu à assumer la difficile charge de la sélection des communications à retenir parmi les 850 propositions qui nous sont parvenues. Ils ont eu à cœur d'assumer cette tâche dans un réel esprit d'ouverture, en faisant se côtoyer au sein de leur section, voire au sein d'une même session, un large éventail de langues analysées, de thématiques scientifiques et d'approches théoriques.

I presidenc y i vizepresidenc dles sezions à abù l'enciaria nia saurida da chirì fora i referac da porté dant anter les 850 propostes che é vegnudes manedes ite. Chest laour ài fat con n spirit davert, a na maniera che vigni sezion pò mostré su n gran numer de lingac adorés, de tematiches analisedes y de teories metudes a descuscion.

Afin de donner la possibilité à un plus grand nombre de romanistes de présenter leurs travaux, nous avons pris l'initiative, pour ce vingt-septième congrès de la Société, de mettre en place, au sein des différentes sections, des sessions de posters, qui, grâce au temps ensoleillé dont nous avons pu bénéficier durant toute la semaine, se sont déroulées en plein air, sous les arcades du beau bâtiment que la Faculté de droit, sciences économiques et gestion de l'Université de Lorraine a mis à notre disposition.

A fi de permetre que un nombre més gran de romanistes pogués presentar els seus treballs, hem pres la iniciativa d'organitzar, per aquest vint-i-setè congrés de la Société, i en el sí de les diferents seccions, unes sessions en format pòster. Aquestes comunicacions s'han presentat a l'aire lliure, gràcies al bon temps i al sol que ens han acompanyat durant tota la setmana, sota els porxos del bell edifici que ens ha cedit la Facultat de Dret, Ciències Econòmiques i Gestió de la Universitat de la Lorena.

En matière d'édition d'actes, nous avons choisi d'innover, avec l'objectif de garantir une lecture partagée et donc une appropriation collective des apports diversifiés du congrès. Les actes sont ainsi publiés dans leur intégralité sur internet et accessibles à partir des sites de l'ATILF et de la Société de linguistique romane et parallèlement, sous une forme partielle, sur papier. La publication internet répond aux meilleures exigences actuelles de cet outil de diffusion avec en particulier, pour chaque article, un DOI, identifiant d'objet numérique, équivalent pour les documents numériques du numéro ISBN des publications papier, ainsi qu'une pagination continue, permettant des citations précises. Par ailleurs, l'intégralité des actes sera versée dans les collections électroniques de la Bibliothèque nationale de France pour garantir leur longévité.

En cuanto a la edición de las actas, hemos decidido innovar, con el objetivo de garantizar una lectura compartida que facilite una apropiación colectiva de las variadas aportaciones del congreso. Por este motivo las actas se han publicado íntegramente en internet y, paralelamente, una versión reducida ha aparecido en papel. La publicación en internet responde a las máximas exigencias actuales de este medio de difusión: los artículos se han etiquetado con un formato XML, se han paginado de manera que se puedan hacer referencias bibliográficas precisas y cuentan con un número DOI (identificador de objeto digital), que equivale al ISBN de las publicaciones en papel. Los textos de cada sección se pueden descargar desde los sitios del

ATILF y de la Société de linguistique romane. Además, a fin de garantizar su longevidad, las actas completas se volcarán en las colecciones electrónicas de la Biblioteca Nacional de Francia.

Pour ce qui est de la publication sur papier, elle regroupe le texte des événements communs – les allocutions prononcées lors de la cérémonie d'ouverture, les conférences plénières et les tables rondes – ainsi qu'un aperçu synthétique du travail de chacune des sections, assorti d'un choix représentatif de communications.

Quanto à publicação em papel, reúne o texto dos acontecimentos comuns – as alocuções pronunciadas na cerimónia de abertura, as conferências plenárias e as mesas redondas – assim como uma amostra sintética do trabalho de cada uma das seções, acompanhada por uma seleção representativa de comunicações.

Secrétaire général du CILPR 2013 — Jean-Marie PIERREL

Secrétaire adjointe du CILPR 2013 — Éva BUCHI

Allocution d'ouverture du représentant de l'Université de Lorraine

Au nom de Pierre Mutzenhardt, Président de l'Université de Lorraine, je vous souhaite la bienvenue en Lorraine, à Nancy et dans les locaux de l'UL.

Comme vous le savez sans doute, l'Université de Lorraine est une jeune et en même temps une ancienne université, puisque créée en janvier 2012 par la fusion de quatre universités quadragénaires, et compte près de 55 000 étudiants dont environ 2 000 doctorants. Mais l'histoire remonte en fait encore plus loin, à 1572, année de création de la première université de Lorraine, soit 440 ans d'âge…

Le site lorrain associe l'Université de Lorraine à quatre Établissements Publics à caractère Scientifique et Technologique (CNRS, Inra, Inria et Inserm) et à un Centre Hospitalier Universitaire. Ces six acteurs sont engagés dans un partenariat académique fort depuis de nombreuses années et partagent des volontés politiques et stratégiques en recherche déclinées en six axes thématiques prioritaires sur des enjeux sociétaux et résolument orientées vers la pluridisciplinarité dans une articulation affirmée entre recherche fondamentale, finalisée et appliquée, tout ceci en partenariat étroit avec le monde socio-économique.

Ce site lorrain comprend plus de 4 000 personnels permanents dans les structures de recherche, soit 61 laboratoires répartis en dix pôles scientifiques qui couvrent presque toutes les disciplines scientifiques. Parmi ces dix pôles, trois couvrent les sciences humaines et sociales, juridiques et de gestion et l'un d'entre eux (dénommé *Connaissance, Langage, Communication, Sociétés*) abrite le laboratoire ATILF, coorganisateur de l'évènement d'aujourd'hui avec la Société de linguistique romane.

L'ATILF est l'un des laboratoires SHS les plus conséquents de l'Université de Lorraine et est parfaitement en phase avec la politique de l'établissement. En effet, il participe en particulier à l'axe prioritaire « Ingénierie des langues et des connaissances », qui représente une thématique fortement ancrée en Lorraine depuis l'installation du *Trésor de la Langue Française* (TLF) en 1962.

Aujourd'hui l'ATILF est acteur des programmes d'investissement d'avenir au travers de l'Equipex ORTOLANG (Outils et Ressources pour un Traitement Optimisé de la LANgue) et est fortement impliqué dans des projets de recherche internationaux d'envergure.

D'ailleurs, si ce XXVIIe Congrès international de linguistique et de philologie romanes a choisi la Lorraine pour se réunir, après une absence de 30 ans en France,

c'est sans aucun doute significatif de la place des travaux menés à l'ATILF dans ce domaine, et ses deux tutelles, l'Université de Lorraine et le CNRS, ne peuvent que s'en féliciter (tout en félicitant une fois de plus ses directeurs successifs).

Voilà, au nom de l'Université de Lorraine et de son Président, je vous renouvelle nos mots de bienvenue et, 1 200 ans après le Concile de Tours, en cette année anniversaire pour les langues romanes, je vous souhaite que vos travaux d'aujourd'hui aient un même retentissement suite à ce qui pourrait devenir le Congrès de Lorraine, ne voyant point dans cet amphithéâtre les marques élémentaires d'un concile.

Chargé de mission « Partenariat académique avec les EPST »
de l'Université de Lorraine Pierre ARCHAMBAULT

Allocution d'ouverture du représentant du CNRS

C'est un honneur de participer à l'ouverture du XXVIIe Congrès international de linguistique et de philologie romanes, fort de près de 900 participants venus d'une quarantaine de pays différents. Depuis sa création, c'est la première fois que ce congrès est accueilli en Lorraine et la sixième fois en France, où il ne s'est pas tenu depuis trente ans. La linguistique, le lexique, les corpus, la littérature, la philologie, le multilinguisme, le traitement automatique de données textuelles, la création de nouvelles ressources informatisées et de nouveaux outils pour l'étude et l'utilisation des langues seront au cœur de vos échanges.

Ce congrès est un événement majeur de la Société de linguistique romane, qui a comme vocation de fédérer au niveau international des chercheurs et des institutions qui souhaitent développer les études de linguistique et philologie romanes. Fondée en 1924, elle compte aujourd'hui un millier de membres de 45 pays différents.

Le congrès est organisé sur place par l'Unité Mixte de Recherche ATILF (Analyse et Traitement Informatique de la Langue Française), qui associe l'Université de Lorraine et le CNRS.

Le Centre National de la Recherche Scientifique est un organisme public de recherche. Avec près de 34 000 personnes (dont 25 000 statutaires), un budget pour 2013 autour de 3 milliards et demi d'euros, le CNRS exerce son activité dans tous les champs de la connaissance, en s'appuyant sur plus de 1 100 unités de recherche et de service. Des chercheurs éminents ont travaillé et travaillent dans des laboratoires du CNRS, qui, avec dix-huit lauréats du prix Nobel et onze de la Médaille Fields[1], est fort d'une longue tradition d'excellence.

L'ATILF est une unité importante du dispositif de recherche du CNRS en linguistique, avec un nombre d'effectifs CNRS (ingénieurs, techniciens et chercheurs) qui dépasse la trentaine. Créé en 2001, ce laboratoire demeure une référence sur les plans national et international en linguistique historique, en didactique des langues, en analyse du discours ainsi que dans le domaine du lexique, une thématique à caractère transversal.

L'ATILF est porteur de l'EQUIPEX ORTOLANG (Open Resources and TOols for LANGuage), membre de la fédération CNRS Institut de Linguistique Française (ILF) et du consortium international Text Encoding Initiative (TEI). Il est aussi labo-

[1] Depuis la date de la conférence, le CNRS a vu augmenter en 2014 la liste de ses prix par attribution du prix Nobel d'économie et d'une médaille Fields de mathématiques.

ratoire support du Centre National de Ressources Textuelles et Lexicales (CNRTL), mis en place par le CNRS et impliqué au sein de l'infrastructure européenne Common Language Resources and Technology Infrastructure (CLARIN).

L'architecture de l'ATILF est articulée en cinq équipes : Lexique ; Linguistique historique française et romane ; Discours, langue et cognition ; Didactique des langues et sociolinguistique ; Ressources : normalisation, annotation et exploitation.

L'ATILF est une unité très dynamique sur le front des coopérations nationales et internationales et sur celui de la valorisation, notamment en ce qui concerne l'analyse de la langue française et son traitement informatique.

Pour conclure, je tiens à remercier les organisateurs d'un congrès qui demeure la référence internationale des études de linguistique romane. En particulier, ma gratitude va à Jean-Marie Pierrel, l'ancien directeur de l'ATILF, et Éva Buchi, l'actuelle directrice, pour avoir su relever un défi si important et mobiliser les effectifs de l'unité. Un remerciement spécial va au Président de la Société de linguistique romane, Jean-Pierre Chambon, qui a su impulser une dynamique remarquable à la Société et fédérer des forces internationales considérables.

Dedico infine una parola di riguardo alla comunità linguistica italiana così fortemente rappresentata per quantità e qualità di participanti a questo congresso, ricordando ai miei connazionali quanto fu generosa in tempi recenti la terra di Lorena nei confronti dei nostri emigranti che partirono d'Italia per ragioni prima politiche e poi economiche.

Je souhaite ainsi à tous les participants des échanges fructueux sur le plan scientifique et un agréable séjour dans cette ville culturellement très riche.

Directeur adjoint scientifique, en charge de l'économie, de la gestion et
de la linguistique, de l'Institut des sciences humaines et sociales
du Centre National de la Recherche Scientifique Stefano BOSI

Allocution d'ouverture du président de la Société de linguistique romane

Mesdames et Messieurs les vice-présidents, Monsieur le doyen, Monsieur le directeur adjoint, Monsieur le chargé de mission, chers consœurs et confrères, chers amis,

Quatre-vingt-cinq ans après le premier congrès de la Société de linguistique romane tenu à Dijon en 1928, les « pèlerins de la latinité » (la formule est d'Albert Henry) et/ou de la romanité que nous sommes reviennent en France d'oïl. Après de longues pérégrinations, nous devons ce retour à Jean-Marie Pierrel et Éva Buchi, que je voudrais remercier tout d'abord, ainsi que toutes les personnalités et institutions qui ont permis la réalisation de leur projet.

La ville lorraine qui nous accueille est devenue depuis quelques décennies, grâce à l'INaLF, puis à l'ATILF (CNRS/Université de Lorraine), à leurs chercheurs et aux projets que vous connaissez tous – le *Trésor de la langue française*, le *Dictionnaire du moyen français*, le *Französiches Etymologisches Wörterbuch* et le *Dictionnaire Étymologique Roman* –, l'une des capitales de la linguistique et de la lexicographie françaises, puis romanes. Si bien que beaucoup d'entre nous se sentent déjà un peu chez eux à Nancy, au plan intellectuel en tout cas.

Chacune de nos réunions triennales est pour les organisateurs l'occasion d'apporter des retouches à l'architecture de nos congrès et d'esquisser une certaine idée de nos deux disciplines, la linguistique romane et la philologie romane : des disciplines bien distinctes, mais non irrémédiablement opposées et dont chacun de nos congrès proclame au contraire la nécessaire complémentarité.

À son tour, le congrès de Nancy apporte certaines modifications à ce qui est devenu, au fil des éditions, notre schéma de base.

Une première innovation réside dans l'organisation de quelques conférences « en ville ». Cette initiative vise à faire partager à un plus grand nombre certains aspects de nos travaux. Une première conférence a déjà eu lieu, permettant d'annoncer le congrès ; d'autres suivront, consacrées à l'histoire du français, aux apports des savants allemands à la linguistique/philologie romane et aux noms de famille en Lorraine. Je remercie les confrères qui ont bien voulu accepter de prononcer ces conférences « grand public », qui seront suivies aussi, j'en suis sûr, par beaucoup d'entre nous.

Le congrès de Nancy met d'autre part l'accent, dans ses deux premières sections, sur une articulation plus nette et plus ferme de la linguistique romane avec les branches de la linguistique super-ordonnées : la linguistique latine et la linguis-

tique générale. On peut souhaiter, je crois, que cette articulation soit pérennisée dans nos prochains congrès. C'est là, d'ailleurs, un retour aux sources : en 1925, l'organigramme de la Société commençait par la linguistique générale, dont le responsable était Joseph Vendryes, et par le latin, représenté notamment par Alfred Ernout.

Par ailleurs, les littératures romanes médiévales, qui ont déjà fait des apparitions disons épisodiques dans nos congrès, sous des intitulés divers, reviennent à Nancy avec une section qui leur est explicitement consacrée. Il me semble désirable que cette expérience puisse être prolongée.

La répartition des communications en sections tente, autant que possible, de mettre de l'ordre dans la richesse proliférante des propositions, grâce à la collaboration – essentielle – des présidents et des vice-présidents. À cet égard, la section « Syntaxe » a atteint à une belle ordonnance thématique de ses sessions.

Les conférences plénières et les table rondes nous permettront de lutter contre les tendances centrifuges et de faire le point sur quelques questions centrales : philologie textuelle, philologie et syntaxe historique, géolinguistique, sémantique, unité de la romanistique. Max Pfister nous rappellera, *in fine*, que la Lorraine est de longue date une terre de contact entre le monde roman et le monde germanique.

La très large majorité des grands ensembles de variétés romanes feront l'objet de communications, y compris le macédoroumain, le dalmate, l'istroroman, le ladin dolomitique, le francoprovençal, l'occitan, l'asturien, le galicien ou le français exporté. En feuilletant le programme, il m'a semblé que l'italien méridional et le portugais du Brésil étaient particulièrement bien représentés. Les influences romanes sur diverses langues non romanes (le polonais, le basque, le guarani et le chinois) seront aussi envisagées.

Sept langues romanes, sauf erreur, seront employées comme métalangues : le roumain, l'italien, le français, le catalan, l'espagnol, le portugais, mais aussi le frioulan.

Cette double diversité des langues et des métalangues est un bien extrêmement précieux, qui est constitutif à tous égards de l'identité de notre Société.

La linguistique romane n'est pas l'addition de la linguistique française, de la linguistique espagnole, de la linguistique italienne etc., mais une discipline *sui generis* ayant constamment pour fondement et comme horizon la totalité des variétés romanes, dans leur diversité et dans leur unité, dans le présent comme dans le passé. Pour la linguistique romane, toutes les variétés romanes, standardisées ou non standardisées, standard ou sub-standard, possèdent un poids épistémique identique.

Ne serait-ce qu'en raison de leur nombre, les variétés primaires, c'est-à-dire les variétés non standardisées (dont les variétés anciennes pré-standard), revêtent une importance fondamentale pour la linguistique romane. Ces variétés sont partout menacées, si bien que la linguistique romane se trouve confrontée à la disparition progressive d'une partie essentielle de son objet. Cette disparition est une catastrophe scientifique et une raison supplémentaire pour souhaiter que la dialectologie romane conserve le rôle central qu'elle a tenu dans notre discipline. Certes, la notion de dia-

lectologie est notoirement mal formée : faire de la dialectologie, ce n'est pas autre chose que faire de la linguistique (et même au sens éminent de ce terme !), mais l'existence d'une dialectologie romane vigoureuse constitue un rempart protecteur contre le poids social écrasant des grandes langues standardisées et contre le *standardocentrisme* dans nos études.

En ce qui concerne à présent les métalangues, vous savez que dans nos congrès, dans la vie de notre Société et dans la *Revue de linguistique romane*, toutes les langues romanes sont admises, mais rien que les langues romanes. Ce dispositif de défense de la *latinità* est dû aux refondateurs de notre Société. Nous savons gré à nos confrères non romanophones natifs qui veulent bien s'exprimer dans un idiome roman lors de nos congrès et dans notre *Revue*.

Si, dans le grand processus glottophagique en cours, les variétés romanes primaires sont tout simplement menacées de mort, les variétés romanes standardisées sont, quant à elles, menacées d'un début de déstandardisation, à travers la 'démaximalisation' de leurs fonctions, en particulier dans la communication scientifique, et ce en faveur de la langue de la puissance mondiale dominante. Certes, disait Georges Straka, ce ne sont pas les linguistes et les philologues qui font l'histoire des langues. Ils ont cependant, comme tout un chacun, leur petite part de responsabilité individuelle et collective en tant que parleurs et que scripteurs. On peut toujours choisir de résister à l'hégémonisme de la langue – et de la pensée – uniques.

Cette question est clairement liée aux bouleversements qui affectent les universités européennes. Sur ce point, je me borne à redire ce que j'ai eu l'occasion de déclarer au congrès de Valence : « la Société de linguistique romane est une compagnie de savants, dotée de sa logique propre, distincte de la logique strictement universitaire ». J'espère qu'en dépit des inévitables pressions, notre Société saura maintenir ce cap d'autonomie et de liberté.

Je signale ici que la Société de linguistique romane et le congrès se réjouissent de donner l'hospitalité à la réunion qui jettera les fondations d'un Conseil international d'études romanes regroupant les associations scientifiques actives dans notre champ de travail.

Vous savez sans doute que les romanistes aiment à caresser l'idée qu'ils forment une grande « famille ». Je préfère néanmoins reprendre, en terminant, une formule d'Albert Henry, formule que je mets à l'impératif : efforçons-nous donc de faire vivre, pour la durée de ce congrès, « une démocratie savante, représentant tout ce qui est Rome et ce qui est amour de la latinité » !

Je déclare ouvert le XXVIIe Congrès international de linguistique et de philologie romanes.

Professeur de linguistique romane et
langue et littérature d'oc à l'Université de Paris-Sorbonne,
Président de la Société de Linguistique Romane Jean-Pierre CHAMBON

Lachmann et Bédier. La guerre est finie

Opposer Joseph Bédier et Karl Lachmann, en parlant de l'édition des textes, c'est comme inventer un duel entre une personne réelle et un fantôme. Sebastiano Timpanaro[1] et surtout Giovanni Fiesoli[2] ont montré que Lachmann n'est pas le responsable de la méthode dite lachmanienne. Il n'empêche que cette méthode, largement adoptée par les spécialistes, est un point de repère, grâce aussi à l'importance de la filière Lachmann-Paris-Bédier.

Comme dans le cas de la provenance des fabliaux ou celui des origines de l'épopée, en ce qui concerne l'édition de textes Bédier renverse totalement la position de son maître, Gaston Paris. Bédier a saisi avant les autres le changement épistémique entre la culture romantique et les nouvelles tendances historiographiques. Par exemple, au lieu de partir à la recherche des origines reculées et mystérieuses de l'épopée, Bédier focalise son attention sur les chansons de geste telles que nous les connaissons, et évite de s'égarer dans le haut Moyen Âge, sur des pistes trop problématiques. La critique philologique, élaborée par les philologues anglais et allemands sur des textes d'abord grecs et latins et ensuite allemands et romans aussi, avait l'ambition de restituer et reconstituer les œuvres qui sont arrivées jusqu'à nous, généralement grâce à des manuscrits. Une tâche d'envergure, si l'on remarque que ces philologues, pour la plupart protestants, ont affronté également le texte de la Bible. Depuis le milieu du XIXe siècle, la philologie allemande a assuré un travail gigantesque de récupération et reconstitution. Son idéal était de trouver des principes logiques de validation et de restitution ; la perspective de la méthode des sciences naturelles était dominante. Et dans ces mêmes années, Eugène Viollet-Le-Duc refaisait les principaux monuments médiévaux selon la doctrine exposée dans son très savant Dictionnaire raisonné de l'architecture française (1854-1868).

En 1872, Gaston Paris met au point une édition du célèbre poème hagiographique sur *Saint Alexis* selon les principes de l'école allemande, qu'il avait acquis en tant qu'élève de Friedrich Diez, considéré comme le fondateur de la philologie romane[3]. L'édition de Paris, qui est la première édition lachmannienne d'un texte roman et un véritable monument de maîtrise philologique, reconstituait le texte de ce poème en utilisant tous les manuscrits, y compris ceux qui transmettaient des remaniements du texte original, datable de la fin du XIe siècle. Paris résumait les résultats de sa

---

1 Timpanaro (1963).

2 Fiesoli (2000).

3 Paris / Pannier (1872).

recherche sur la tradition textuelle dans un arbre généalogique, ou *stemma codicum*. Cet arbre était fondé sur le principe des fautes communes. En pratique, si deux ou plusieurs manuscrits contiennent les mêmes fautes, on considère qu'ils remontent sans aucun doute à un seul ancêtre commun, puisqu'il est invraisemblable que des manuscrits qui ne sont pas apparentés présentent les mêmes fautes aux mêmes endroits.

Une fois l'arbre reconstruit, l'accord de deux ou plusieurs traditions indique les leçons originaires, qu'il faudra considérer comme correctes, sur la base d'un raisonnement inverse : comme il est impossible que deux manuscrits contiennent les mêmes innovations au même endroit, l'accord de deux contre un indique la leçon correcte. Pour donner un exemple, dans le cas d'une œuvre transmise par trois manuscrits, ou trois sous-archétypes, *ABC*, on peut considérer probablement comme correctes les leçons communes à *A* et *B*, ou à *A* et *C*, ou encore à *B* et *C*. Étant donné, en revanche, un arbre *AB* ; *C*, l'accord de *A* avec *C* ou de *B* avec *C* indique encore une fois la leçon correcte, tandis que dans le cas d'une opposition *AB* contre *C*, il n'existe aucun critère logique pour donner raison à l'un ou à l'autre. C'est un contre un.

C'est justement sur cela que Bédier prend appui dans son article révolutionnaire de 1928[4]. Mais il faut préciser qu'entre 1872 et 1928, la France, comme d'ailleurs l'Italie, avait fait son entrée dans le concert ecdotique en produisant des éditions critiques lachmanniennes d'un niveau excellent, par exemple celles du *Roman de Troie* par Léopold Constans (1904-1912) ou du *Couronnement de Louis* (1888) et du *Roman de la Rose* par Ernest Langlois (1914-1924), ou encore les œuvres de Machaut éditées par Hoepffner (1908-21), que l'on peut comparer aux produits du lachmannisme allemand, comme Chrétien de Troyes par Wendelin Foerster (1884, 1887, 1890, 1899, etc.), ou les *Fables* de Marie de France par K. Warnke (1898) ou les magnifiques éditions, toujours allemandes, des troubadours occitans, parmi lesquelles il suffira de citer celle de Bernart de Ventadorn (1925), par Carl Appel.

À vrai dire, le premier travail lachmannien sur des textes romans est l'étude, qui pourtant n'aboutira pas à une édition, de la tradition du *Fierabras* (1869) menée par Gustav Gröber, qui par la suite, en 1877, nous a donné sa recherche magistrale sur les *Liedersammlungen der Troubadours*. En Italie, après l'édition très précoce d'Arnaut Daniel par Ugo Angelo Canello (1883), pour trouver des éditions significatives il faut arriver aux *Reali di Francia* de Andrea da Barberino, édités par Giuseppe Vandelli (1892-1900), ou à la *Divina Commedia* et aux autres ouvrages de Dante (*Vita nuova* et *Rime* éditées par Michele Barbi, en 1907 et 1932) : nous sommes là autour du sixième centenaire de la mort du grand poète (1921).

L'importance attribuée à l'établissement critique du texte est telle que Léon Gautier, après avoir publié son commentaire classique de la *Chanson de Roland*, en 1872, sur la base de l'édition lachmannienne de Theodor Müller (1863), modifie radicalement son commentaire à partir de 1880 (VIIe édition), parce qu'il apporte des retouches au texte de Müller en le conformant aux résultats des premiers travaux d'Edmund

[4] Bédier (1928). Voir à ce propos Leonardi (2009).

Stengel (1877), et en anticipant ainsi l'édition que Stengel ne publiera qu'en 1900. Une décision, faut-il ajouter, déplorable, car Stengel avait conçu son texte sur la base d'un arbre qui était lachmannien, bien sûr, mais décidément erroné, tandis que l'arbre précédent de Müller est encore valable, même selon Bédier.

À cette activité d'édition d'inspiration rigoureusement lachmannienne participe aussi, et en première ligne, Joseph Bédier, notamment avec ses deux éditions du *Lai de l'ombre* (1890 et 1913), dont la seconde est conçue comme une réponse à un compte-rendu de Gaston Paris, qui à son tour avait proposé son propre arbre.

Dans son article de 1928, Bédier change sa perspective, et semble se poser pour la première fois le problème de la facilité avec laquelle on peut remplacer un arbre par un autre. Il s'amuse même à multiplier les combinaisons possibles, en proposant au moins 14 arbres. C'est un grand jeu de l'intelligence, mais pas si surprenant, si l'on envisage posément les faits. Évidemment, les manuscrits du *Lai de l'ombre* contiennent très peu de fautes indiscutables, et les philologues sont obligés de chercher des fautes là où il n'y en a pas : la combinatoire devient infinie. En général, on a remarqué que les copistes consacrent beaucoup plus d'attention et de soin aux textes d'un niveau stylistique plus élevé : dans les poèmes troubadouresques aussi, il est très difficile de trouver des fautes certaines. Je signale, en tout cas, que quelqu'un s'est chargé de contrôler tous les arbres conçus par Bédier, et a conclu que celui qui demeure valable est le premier[5].

Mais Bédier avait fait une autre remarque, qui nous laisse franchement rêveurs. En recueillant tous les arbres proposés par les romanistes pour leurs textes, Bédier remarquait que sur 110 arbres, 105 sont bipartis. Arrigo Castellani, qui a essayé de refaire les comptes de Bédier[6], est arrivé à établir une liste de 120 arbres, dont 82-83% à deux branches. Bref, cette loi admet peu d'exceptions, même si dans le calcul de Castellani leur nombre est un peu plus élevé. Mais pourquoi tous les textes conservés auraient-ils donné lieu à deux copies, d'où descendraient toutes les autres? Pourquoi faut-il que les arbres à trois, quatre ou plusieurs branches soient si rares ? Les philologues, à partir de Sebastiano Timpanaro, ont réfléchi là dessus, et ont trouvé bon nombre de raisons qui justifient cette prédominance des arbres bipartis. Leurs arguments sont pour la plupart statistiques. Et chacun peut les contrôler. Ensuite, Castellani a expliqué la prédominance des arbres à deux branches par les modalités du processus de copie, qui doit être exécuté le plus vite possible. D'où sa notion ingénieuse de « production maximum ».

Le résultat de la recherche de Bédier était en tout cas intéressant, même s'il a peut-être exagéré son importance. Mais il faut souligner que, pour Bédier, la prédominance des arbres bipartis dépendait, plutôt que de la réalité de la tradition manuscrite, de raisons qui nous échappent, mais qui pourtant jettent l'ombre du doute sur

5 Whitehead / Pickford (1973). Pour une nouvelle proposition stemmatique (toujours bifide) au sujet du *lai* de Jean Renart voir Trovato (2013).

6 Castellani (1957).

le mécanisme d'élaboration des arbres mêmes. Parmi les explications avancées par Bédier, en effet, il y a le soupçon que l'éditeur vise spontanément un arbre biparti parce que celui-ci lui laisse cette liberté de choix que l'arbre à trois branches anéantirait. Le travail de collation des manuscrits étant de nature essentiellement binaire (leçon considérée comme correcte *vs* leçon considérée comme fautive), on peut penser que le système de reconstruction de la recension amène lui aussi inévitablement à une solution binaire.

Pour comprendre le succès de sa thèse, sans aucun doute révolutionnaire, il faut tenir compte du fait que Bédier était désormais le chef de file de la romanistique française, qu'il était célèbre aussi pour sa traduction-remaniement très élégante du *Roman de Tristan* (1900), et presque un héros national pour sa contribution à la documentation des crimes allemands pendant la première guerre mondiale[7]. Incontestablement, dans l'histoire que je suis en train d'évoquer, l'antagonisme franco-allemand a joué un rôle considérable. Rabaisser la gloire philologique des Allemands, voilà ce qui pouvait représenter une belle satisfaction. Et Bédier était doué d'une capacité rhétorique exceptionnelle, quitte à frôler le raisonnement captieux.

Certes, l'opération de Bédier est d'une envergure considérable. Si son article consacré au *Lai de l'ombre* date de 1928, l'année précédente il avait publié les *Commentaires* à son édition de la *Chanson de Roland*, dont la première parution date de 1922[8]. Dans ses *Commentaires*, Bédier expose les critères qu'il a suivis pour l'édition de la *Chanson*, des critères qui à première vue apparaissent déjà, pour ainsi dire, bédiériens.

Le texte de base (le « bon manuscrit ») est, comme dans la quasi-totalité des éditions de la *Chanson de Roland*, celui d'Oxford, qu'il faut suivre presque toujours parce que, selon Bédier, « le texte d'Oxford a autant d'autorité à lui seul que tous les autres textes réunis » (p. 93). Et pourtant, on pourrait affirmer – au grand scandale des bédiériens – que ces *Commentaires* laissent apercevoir un établissement du texte de type lachmannien, pour les raisons suivantes. Premièrement : Bédier souligne, avec une grande évidence, sa fidélité à l'arbre de Müller, établi grâce à des critères lachmanniens. Deuxièmement : celui de Müller est l'un des ces arbres bipartis que Bédier condamne dans son article de 1928. Lorsqu'il s'agit de la *Chanson de Roland*, Bédier n'insiste jamais sur la fiabilité douteuse de ce type d'arbre. Troisièmement : comme on vient de le dire, Bédier affirme, à raison, que « le texte d'Oxford a autant d'autorité à lui seul que tous les autres textes réunis ». Mais dans l'enthousiasme de son plaidoyer, il ne s'aperçoit pas que cette affirmation signifie également, en bonne logique : « [le témoignage de] tous les autres textes réunis a autant d'autorité que celui d'Oxford ». Quoi qu'il en soit, dans la *Section V* (p. 178-195), Bédier examine tous les cas où les versions du groupe *b*[9] « peuvent être de bon secours », et sa réponse est souvent favo-

[7] Voir sur ce sujet Corbellari (1997, chap. IX).

[8] Bédier (1927).

[9] Que Bédier dénomme *a*.

rable. De plus, il reviendra encore sur son texte, avec de nouvelles retouches à O[10]. Que Bédier souhaite, de toute façon, réduire au maximum le recours aux manuscrits *b* est un choix personnel, que tout lachmannien trouverait acceptable. Remarquons, en outre, que l'étude d'ensemble de la recension de la *Chanson de Roland* est extrêmement approfondie, ce qui serait même trop dans une édition de l'école bédiérienne.

Que l'on puisse vouer au mépris public la fréquence des arbres avec une conformation bipartie et, en même temps, baser l'édition de la plus grande *chanson de geste* française du Moyen Âge sur un arbre à deux branches, est une question psychologique, ou peut-être rhétorique. Chaque fois, les lecteurs sont emportés par le feu de la démonstration de Bédier : certes, dans ses *Commentaires* il écrit : « on ne doit pas s'étonner, ni s'inquiéter, si l'on aboutit, dans le cas de la *Chanson de Roland*, à un classement à deux branches. En faveur de celui-ci, j'ai pu invoquer, j'invoquerai encore, comme on a vu et comme on verra, nombre de faits » (*Com*. 134). Il n'y avait donc pas de raison de s'inquiéter face à la majorité des arbres bipartis.

Je parlais de logique. Eh bien, toujours dans cette année fatale 1928, et toujours dans son article consacré au *Lai de l'ombre*, Bédier à l'occasion de discuter une tentative révolutionnaire de bâtir des arbres, celle de dom Henri Quentin; une tentative aujourd'hui oubliée, sauf pour ressemblance avec quelques expériences modernes de construction mécanique des arbres. À la fin de sa revue des suggestions critiques de dom Quentin, Bédier, en les repoussant, accuse le savant bénédictin d'avoir renoncé à « tracer les lignes de faîte du schéma, celles qui doivent relier les manuscrits réels à l'archétype », de ne s'être pas soucié de « peser les variantes » pour bâtir des groupes et des sous-groupes. C'est un résumé parfait du travail d'un éditeur lachmannien, même si, dans ce cas aussi, Bédier refuse la notion lachmannienne de « faute ». Et alors ? Alors, il faudrait penser que Bédier ne s'est jamais débarrassé tout à fait des principes de Lachmann.

Quoi qu'il en soit, après 1928 s'est affirmée une sorte de vulgate bédiérienne, selon laquelle il ne serait pas indispensable d'étudier systématiquement la tradition manuscrite d'un texte : il suffit de choisir « le bon manuscrit », ou « le meilleur manuscrit », et l'éditer de la façon la plus fidèle possible. Les éditions se sont succédé, toujours selon le même schéma, toutes pareilles, même si les meilleurs disciples de Bédier condamnaient le choix de la facilité et de la paresse : en effet, des philologues tels qu'Albert Henry (Adenet le Roi, 1951-1971) ou Jean Rychner (Marie de France, 1966) ont basé leurs éditions sur des recherches approfondies et systématiques, en discutant et en résolvant au fur et à mesure des problèmes que les lachmanniens résolvent préliminairement dans la discussion de l'arbre.

La notion de « bon manuscrit » est l'une des plus vagues et contradictoires dans le domaine de l'édition des textes. Le grand Maurice Delbouille dénonçait le fait qu'avant comme après Bédier « le "meilleur manuscrit" a continué d'imposer ses

[10] Bédier (1937-1938).

leçons les moins heureuses »[11]. Si l'on considère dans l'ensemble les usages des divers philologues, le « bon manuscrit » est le plus complet, ou celui qui offre la leçon la plus correcte ; ou encore, il s'agit du manuscrit avec les traits linguistiques les plus proches de ceux de la région d'appartenance de l'auteur. En outre, le « bon manuscrit » est, préférablement, l'un des plus anciens, parce que de cette façon il a moins subi la détérioration du temps et le cumul des altérations dues au travail de copie.

Je me permets de remplacer ces définitions, qui souvent se contredisent mutuellement, par une seule : le « bon manuscrit » est un manuscrit qui nous transmet le texte dans les *meilleures conditions* : moins de fautes et moins d'altérations. Mais comment le reconnaître ? Le seul moyen est d'approfondir la recension, ce qui nous ramène en pleine méthode lachmanienne.

J'ai eu l'occasion de réfléchir sur cette question lors de la découverte du manuscrit T du *Bestiaire d'Amours*[12]. Un manuscrit que Xenia Muratova, qui l'a découvert, aurait tendance à définir « bon », car il est complet, apparemment correct et relativement ancien. Malheureusement, l'examen de la tradition, que je peux faire aisément comme auteur de l'édition critique de ce texte[13], dont les apparats offrent toutes les données utiles pour une telle appréciation, m'a révélé immédiatement que *T* remonte à une mauvaise copie, dont il a hérité forcément les fautes et les banalisations[14]. Malgré son aspect apparemment correct, il ne s'agit donc pas d'un « bon manuscrit ». Lorsque Bédier célèbre le manuscrit *O* de la *Chanson de Roland* comme le « bon manuscrit », il a beau jeu, parce que la supériorité de *O* est universellement reconnue. Mais la qualité de *T* n'a rien à voir à celle de *O*. J'ajouterai que le manuscrit *T* du *Bestiaire* s'insère dans l'arbre d'une façon tellement précise que l'on peut prévoir presque constamment quelles seront ses leçons.

En tout cas, le succès de la théorie de Bédier a eu également des conséquences sur la géographie philologique. Car cette théorie a été accueillie presque sans exception en France, tandis que dans les autres pays elle a été moins suivie. Et il faut reconnaître que, souvent, la multiplication d'éditions de textes français basées sur un « bon manuscrit », sans une étude systématique de la tradition, n'assure pas un véritable progrès de nos connaissances. Un grand philologue bédiérien, Jacques Monfrin, le reconnaît, lorsqu'il déclare : « on ne peut s'étonner du discrédit où est tombée cette méthode qui a justifié tous les abandons, mais il est étrangement simplificateur d'en attribuer, surtout sous les formes caricaturales qu'elle a parfois prises, la paternité à Joseph Bédier »[15].

Je passe, sans pour autant l'absoudre, sur le niveau souvent avilissant qu'ont atteint les débats entre bédiériens et lachmanniens. Mais je souligne qu'il est utile,

---

11 Delbouille (1976).

12 Muratova (2005).

13 Segre (1957).

14 Pour la démonstration, voir Segre (2010).

15 Monfrin (1984-1986, 354).

et fécond, de prendre en compte, d'un côté, les changements d'avis des bédiériens, de l'autre, plus généralement, la conscience, présente depuis longtemps chez les néo-lachmanniens les plus avertis, que l'édition critique n'est pas une idole, mais l'un des traitements du texte suggérés par la pratique philologique.

Car il y a des éditions qui excluent à juste titre toute retouche trop lourde, comme les éditions diplomatiques et interprétatives ; il y a des approches dialectologiques des textes, qui seraient faussées par les conjectures. Et il y a des éditions synoptiques du même texte selon différents manuscrits. Ni les répertoires des rimes d'un auteur ni les concordances n'admettent aucune intervention du philologue.

L'édition critique n'est fondamentale que quand on cherche à récupérer quelque chose qui puisse représenter vraisemblablement et au niveau le plus élevé de probabilité l'aspect original du texte[16]. J'ajoute que les hypertextes, s'ils sont réalisés avec intelligence, nous permettent de parcourir l'ensemble des variantes beaucoup plus aisément qu'une édition critique, et de multiplier et comparer les différentes possibilités, en les mettant à l'épreuve dans des hypothèses de reconstitution.

L'expérience de la philologie romane italienne montre, mieux que d'autres expériences nationales, comment l'ecdotique a abouti à une heureuse intégration des principes les plus valables du bédiérisme et du lachmannisme : pour cette raison (et je m'excuse de cette simplification forcée), presque tous les exemples que je vais citer seront tirés de cette expérience.

Après la seconde guerre mondiale, il y a eu en Italie une reprise très particulière de la méthode lachmanienne, qui par ailleurs avait déjà obtenu des résultats excellents avec Michele Barbi et Pio Rajna. Le protagoniste de cette reprise a été Gianfranco Contini. Contini a donné naissance à une grande école, qui s'affirme d'abord avec l'édition (1960) des poètes italiens du XIIIe siècle[17], établie sous sa direction par un groupe de jeunes et même très jeunes chercheurs, parmi lesquels Franca Ageno, D'Arco S. Avalle, Romano Broggini, Domenico De Robertis, Ezio Raimondi et Cesare Segre. Quelques années plus tard, Contini crée, avec Alfredo Schiaffini, une collection de textes critiques, les « Documenti di Filologia », qui ont renouvelé le panorama philologique et littéraire de la Romania.

Il faut remarquer que tous les auteurs choisis par les chercheurs italiens appartiennent en fait au monde roman: France du Nord et France du Sud, Italie, Espagne. Cela élargissait le domaine de l'expérience de l'édition des textes, et en même temps renforçait l'aspect comparatiste de la tradition italienne dans le domaine de la philologie romane, qui dans les autres pays ne travaille le plus souvent que sur des textes dans la langue nationale correspondante.

---

16 Voir, par exemple, le débat constructif entre Roncaglia / McMillan / Segre / Lecoy (1978).

17 Contini (1960).

Le jeune Contini, après avoir étudié la méthode lachmannienne à l'école de Santorre Debenedetti et avoir réfléchi sur l'ouvrage capital de Giorgio Pasquali[18], avait fait un séjour de deux ans à Paris (1935-1936), à l'école de Bédier, auquel il a consacré un superbe portrait[19]. Bédier a été constamment un point de repère pour Contini, qui n'a pas cessé pour autant d'être lachmannien ou plutôt néo-lachmannien[20]. En effet, il a aisément accueilli plusieurs points décisifs de la philologie de Bédier : éviter les reconstitutions en mosaïque des textes, et même suivre, pour l'aspect linguistique, un seul texte (qui n'est pas toujours le « bon manuscrit ») ; éviter des restaurations traumatiques des textes ; éviter d'être influencé par les grammaires historiques, mais partir de données documentaires et localisables.

Contini soumet à une analyse méthodologique, je dirais presque scientifique, le travail de Bédier. Son principe fondamental est la conception de l'édition en tant qu'« hypothèse de travail » : une formule où ceux qui ne sont pas experts dans la méthodologie scientifique voient un aveu de faiblesse (hypothèse), en oubliant que, dans le domaine scientifique, cette expression indique l'un des rapprochements progressifs de la vérité. « Hypothèse de travail » signifie : la meilleure solution que n'importe quelle science a pu fournir pour un problème. Si nous faisons confiance au philologue, nous pouvons considérer comme valable son travail, jusqu'au moment où un autre philologue, sur la base de données plus nombreuses et plus sûres, propose une solution plus satisfaisante. C'est ainsi que la science avance dans son chemin. Nous sommes là aux antipodes du texte critique conçu comme une vérité ultime.

L'un des résultats principaux de la réflexion de Contini est la notion de « diffraction des leçons », une forme de la *lectio difficilior* pourvue d'une considérable rentabilité opérationnelle. Mais l'opération la plus importante de Contini, du point de vue symbolique aussi, est son retour au problème des arbres bipartis, en prenant en considération justement l'élaboration la plus ancienne du domaine lachmannien : l'arbre du *Saint Alexis* proposé par Gaston Paris. En effet, selon Contini, l'arbre correct du *Saint Alexis* serait un arbre à trois branches, et non à deux branches, comme le croyait Paris[21]. Le poème hagiographique le plus illustre aurait échappé à la force mystérieuse des arbres bipartis, tandis que Gaston Paris serait l'une des nombreuses victimes du binarisme[22].

---

18 Pasquali (1934).

19 Contini (1939).

20 Tout en prônant une reprise de la méthode lachmannienne, Contini antéposait le préfixe néo (néo-lachmannienne) pour souligner d'emblée que sa propre méthode était bien différente de celle utilisée jusqu'à 1928. Dans des pages malheureusement très difficiles, Contini a discuté chaque point de son opération d'édition des textes, voir Contini (1986).

21 Contini (1970).

22 Il est significatif qu'après la prise de position de Bédier nombre de philologues soient allés à la recherche d'arbres tripartis. Contini (1978) avait déjà signalé la probabilité d'un arbre triparti, mais voir aussi Crespo (2012).

Le nombre des expériences des néo-lachmanniens dans l'édition de textes est tel que le problème des arbres bipartis se situe dans des lointains brumeux. Les arbres à deux ou trois branches sont les cas les plus simples d'une forêt où l'on peut trouver des descendants de deux archétypes distincts, des rédactions doubles et plurielles, et où on se voit obligé d'avoir recours à des représentations horizontales de la tradition, ou à des représentations stéréoscopiques, ou encore à la notion linguistique de diasystème[23].

Il faut accepter également que dans plusieurs cas, la méthode lachmannienne n'est pas utilisable. Il s'agit d'une phénoménologie très vaste, dont il faut se rendre compte et comprendre les raisons. Quant à moi, si la méthode néo-lachmannienne a été fondamentale pour mon édition de la *Chanson de Roland* et pour le *Bestiaire d'Amour*, ainsi que pour plusieurs textes italiens, je la trouve inutilisable pour les *lais* de Marie de France, comme Jean Rychner et moi-même l'avons constaté lors d'un débat amical[24].

Je ne chercherai pas à décrire la situation d'ensemble des études actuelles d'ecdotique, mais je peux affirmer, en synthèse, qu'à mon avis, la guerre entre lachmanniens et bédiériens est désormais un vestige archéologique. On a compris qu'entre les deux méthodes il n'y a pas d'opposition, mais une différence de finalité.

La reconstitution de l'archétype n'est pas le seul but du travail d'édition ; les éditions interprétatives sont souvent indispensables, et peuvent représenter la base de nombreuses recherches. Les recherches sur la tradition des textes ne peuvent pas toujours utiliser la méthode des fautes, qui pourtant reste fondamentale dans les cas où elle est applicable.

Il faudrait en tout cas une certaine prudence lorsqu'on définit comme éditions critiques celles qui n'utilisent pas tous les procédés de la critique textuelle. Certes, pour moi, souvent désigné comme un dangereux lachmannien, il a été surprenant de constater que, dans le manuel pour l'édition des textes médiévaux publié en 2002 par deux illustres chartistes telles que Pascale Bourgain et Françoise Vielliard, on affirme que la méthode lachmannienne est

---

23 L'existence de deux archétypes du *Bestiaire d'Amours* est démontrée dans mon édition, Segre (1957), qui fournit également une méthode pour la recherche des contaminations ; Chiarini (1964) a résolu le problème de la prétendue double rédaction du *Libro de Buen Amor* ; Contini (1984) cherche à démontrer la paternité dantesque du *Fiore* grâce à la notion de « mémoire interne » ; Avalle (1965) démontre l'existence d'une koinè poitevine en partant du *Sponsus* provençal ; les chansons d'Arnaut Daniel, selon Perugi (1978), révèlent, à travers la « théorie du hiatus », certains traits conservateurs des textes des *canzonieri*. Pour une réflexion sur le poids stemmatique des variantes adiaphores voir Squillacioti (1999). Mon édition italienne de la *Chanson de Roland* est une tentative d'édition stéréoscopique, car elle aide le lecteur, grâce à des caractères spéciaux, à suivre le texte de α ou celui, plus récent, de β. Malheureusement, je n'ai pas réussi à me faire comprendre par mes collègues, et dans l'édition française, Segre (1989), j'ai dû me rabattre sur une présentation plus traditionnelle. J'ai proposé aussi une nouvelle notion d'apparat, en considérant la stratification dialectale des textes comme un cas de « diasystème ».

24 Segre (1959) ; Rychner (1966) ; Segre (1968). Ce qui est ressorti de cet échange d'opinions, c'est que l'arbre que je proposais n'était pas plus sûr que celui proposé par Jean Rychner. Les deux étaient équivalents.

> « un merveilleux outil de logique, très rigoureux, extrêmement satisfaisant pour l'esprit, dans des conditions généalogiques normales. On ne peut échapper à ces principes, qui découlent du fait même de la reproduction par copie. Les éléments essentiels de cette méthode restent en place jusqu'à nos jours, avec les différentes étapes de la recension des témoins, de l'examen de la tradition et de l'établissement du texte »[25].

Un autre signe intéressant est le retour à l'étude des grandes éditions lachmanniennes des Langlois et des Foerster : non seulement pour leur rendre le prestige qu'ils méritent, mais pour approfondir le précieux travail déjà accompli, qu'il ne faut pas gaspiller mais poursuivre.

Nous savons bien que le domaine de la critique textuelle n'est jamais inerte. L'un des plus grands lachmanniens italiens, D'Arco Silvio Avalle, auteur d'une édition exemplaire de Peire Vidal[26], a produit un changement considérable. Avalle a préparé des *Concordanze della lingua poetica italiana delle origini* (CLPIO), dont le premier volume, qui pour l'instant est le seul publié, a paru en 1992 chez la maison d'édition Ricciardi. Ce volume monumental recueille tous les textes poétiques italiens attestés par des manuscrits d'avant l'an 1300. Bien entendu, ces poèmes ne sont pas transcrits d'après leurs éditions critiques, remplies de leçons conjecturales qui représenteraient des entrées étrangères aux textes, mais d'après les manuscrits, en respectant fidèlement leurs particularités graphiques.

D'autre part, dans son étude la plus ancienne, consacrée à la *Letteratura medievale in lingua d'oc nella sua tradizione manoscritta* (1961)[27], Avalle renouait inévitablement avec les vieilles recherches de Gröber. En étudiant les *canzonieri* provençaux, Gröber les avait classés, et avait étudié leur généalogie, non pas sur la base des fautes, trop rares pour offrir au philologue un fondement solide, mais par rapport à la succession des poèmes, qui peut être considérée comme un indice du rapport entre les manuscrits (*Kritik der Sammlungen*). C'était un type inhabituel de recherche, qui a été adopté essentiellement par ceux qui ont étudié les grandes anthologies lyriques médiévales, provençales, françaises, italiennes. Tout d'abord, Napoleone Caix. En outre, Gröber cherchait à reconstruire les phases fondamentales de la diffusion des poèmes : feuilles volantes, recueils d'un seul auteur, recueils anthologiques, etc. Son but était de retracer les différentes phases de la tradition de chaque texte.

Après sa récupération par Avalle, la recherche fondée sur les critères de Gröber a eu une impulsion nouvelle, se répandant très rapidement. Aujourd'hui, par exemple, l'attention est centrée sur les recueils d'auteur ou leurs traces[28], sur la présence éventuelle de variantes d'auteur, sur la reconstitution des grands collecteurs des recueils, sur la définition des différentes couches linguistiques, sur la reconstruction de l'ordre

---

25 Bourgain / Vieilliard (2002, 16).

26 Avalle (1960).

27 Nouvelle édition augmentée : Avalle (1993).

28 Voir la sinthèse de Meliga (2006).

originel éventuel des poèmes[29]. Il s'agit d'une série d'opérations génétiques, qui, au lieu de viser le modèle, nous amènent à reconstruire la micro-histoire des textes. Il faut tenir compte du fait que, surtout dans les usages de la poésie italienne des origines, la distance chronologique entre la rédaction et la transcription n'est souvent que de quelques années, au plus de quelques décennies : notre regard se pose presque sur des copies de manuscrits autographes. Dans ce cas, on est tenté – et Avalle lui même l'a fait – d'opposer deux méthodes, de mettre face à face deux vérités, dans le sillage d'Averroès[30]. Toutefois, je voudrais le préciser, il ne s'agit pas de comparer deux entités ontologiques (dans notre cas, la « vérité » du texte et la « vérité » du témoin)[31], mais de choisir entre deux approches différentes, dont chacune apporte un enrichissement de connaissance.

Università di Pavia / Accademia Nazionale dei Lincei Cesare SEGRE †

[29] Voir, par exemple, les recherches rassemblées par Leonardi (2001), avec des contributions de R. Antonelli, A. Petrucci, M. Palma, P. Larson, C. Bologna, S. Zamponi, G. Frosini, G. Savino, T. De Robertis, Valentina Pollidori, M. L. Meneghetti.

[30] Avalle (1985), où l'auteur parle d'une double vérité : « la verità dei protagonisti, che è quella cui aspirano le edizioni critiche di singoli autori, e la verità dei testimoni per il momento affidata […] alle edizioni diplomatiche o semi-diplomatiche » (p. 275). Avalle (1972) avait déjà exprimé un certain penchant pour Bédier.

[31] La ligne de travail qui privilégie l'analyse de la matérialité des témoins manuscrits ne concerne pas que les chercheurs italiens ; mais il faut attribuer aux Italiens une priorité, à partir même de l'étiquette de 'filologia materiale', avec laquelle on désigne d'habitude cette approche, et qui a été formulée par Antonelli (1985, 207): voir, par exemple, Tyssens (1991) et Cadioli / Meneghetti (2008).

Réferences bibliographiques

Antonelli, Roberto, 1985. «Interpretazione e critica del testo», in: Asor Rosa, Alberto (ed.), *Letteratura italiana*, IV. *L'interpretazione*, Turin, Einaudi, 141-243.

Avalle, D'Arco Silvio, 1960. Peire Vidal, *Poesie*, édition établie et commentée par D'A.S.A., Milan-Naples, R. Ricciardi, 1960, 2 vol. (Documenti di Filologia).

Avalle, D'Arco Silvio, 1965. *Sponsus. Dramma delle vergini prudenti e delle vergini stolte. T*esto letterario a cura di D'A.S.A., testo musicale a cura di Raffaello Monterosso, Milan-Naples, R. Ricciardi.

Avalle, D'Arco Silvio, 1972. «La critica testuale», in: Jauss, Hans Robert / Köhler Erich (ed.), *Grundriss der romanischen Literaturen des Mittelalters*, I, *Généralités*, Heidelberg, Winter, 538-558 ; pubblicato in seguito con il titolo «Fenomenologia ecdotica del Medioevo romanzo», in : Id., *La doppia verità. Fenomenologia, ecdotica e lingua letteraria del Medioevo romanzo*, Florence, Edizioni del Galluzzo, 2002, 125-153.

Avalle, D'Arco Silvio, 1985. «I canzonieri : definizione di genere e problemi di edizione», in: *La critica del testo, Problemi di metodo ed esperienze di lavoro*, Atti del Convegno di Lecce (22-26 ottobre, 1984), Rome, Salerno, 363-382, Id. *La doppia verità. Fenomenologia, ecdotica e lingua letteraria del Medioevo romanzo*, Florence, Edizioni del Galluzzo, 2002, 155-173.

Avalle, D'Arco Silvio, 1993. *I manoscritti della letteratura in lingua d'oc*. Nuova edizione a cura di L. Leonardi, Turin, Einaudi, 1993.

Bédier, Joseph, 1900. *Le Roman de Tristan reconstitué d'après les poèmes français du XIIe siècle*, Paris, Piazza.

Bédier, Joseph, 1927. *La* Chanson de Roland (*Commentaires*), Paris, Piazza.

Bédier, Joseph, 1928. «La tradition manuscrite du *Lai de l'Ombre* : réflexions sur l'art d'éditer les anciens textes», R 54, 161-196, 321-356.

Bédier, Joseph, 1937-1938. «De l'édition princeps de la *Chanson de Roland* aux éditions les plus récentes. Nouvelles remarques sur l'art d'établir les anciens textes», R 63, 433-469; R 64, 145-244, 489-521.

Bourgain, Pascale / Vielliard, Françoise, 2002. *Conseils pour l'édition des textes médiévaux*, fascicule III : *Textes Littéraires*, Paris, École Nationale des Chartes, 2002.

Cadioli, Alberto / Meneghetti, Maria Luisa, 2008. *La materialità nella filologia*, a cura di A.C. e M.L.M. = «Moderna» 10.

Castellani, Arrigo, 1957. *Bédier avait-il raison? La méthode de Lachmann dans les textes du Moyen Âge*, Fribourg, Editions universitaires.

Chiarini, Giorgio, 1964. Juan Ruiz, Arcipreste de Hita, *Libro de buen amor*. Edizione critica a cura di G.C., Milan-Naples, R. Ricciardi.

Contini, Gianfranco, 1939. «Ricordo di Joseph Bédier», *Letteratura* 3, 145-152.

Contini, Gianfranco, 1960. *Poeti del Duecento*, a cura di G.C., Milan-Naples, Ricciardi.

Contini, Gianfranco, 1970. «La *Vita* francese *di Sant'Alessio* e l'arte di pubblicare i testi antichi» (1970), Id., *Frammenti di filologia romanza. Scritti di ecdotica e linguistica (1932-1989)*, textes recueillis par Giancarlo Breschi, Florence, Edizioni del Galluzzo, 2007, II, 957-986.

Contini, Gianfranco, 1978. «Fragments inconnus d'un ancien chansonnier français à Einsiedeln», in Güntert, Georges / Jung, Marc-René / Ringger, Kurt (ed.), *Orbis Mediaevalis. Mélanges de langue et de littératuture médiévales offerts à Reto Raduolf Bezzola à l'occasion de son quatre-vingtième anniversaire*, Bern, Francke, 29-59, Id., *Frammenti di filologia romanza*, Florence, Edizioni del Galluzzo, 2007, II, p. 1025-60.

Contini, Gianfranco, 1984. *Il* Fiore *e il* Detto d'amore *attribuibili a Dante Alighieri*, a cura di G.C., Milan, A. Mondadori.

Contini, Gianfranco, 1986. *Breviario di ecdotica*, Milan-Naples, R. Ricciardi.

Corbellari, Alain, 1997. *Joseph Bédier, écrivain et philologue*, Genève, Droz.

Crespo, Roberto, 2012. *Il primo episodio del* Couronnement Louis, Modène, Mucchi.

Delbouille, Maurice, 1976. «La philologie médiévale et la critique textuelle», in Boudreault, Marcel / Möhren, Frankwalt (ed.), *Actes du XIIIe Congrès International de Linguistique et Philologie Romanes tenu à l'Université Laval (Québec, Canada) du 29 août au 5 septembre 1971*, Québec, Université Laval, 57-73.

Fiesoli, Giovanni, 2000. *La genesi del lachmnannismo*, Florence, Sismel.

Gröber, Gustav, 1869. *Die handschriftlichen Gestaltungen der chanson de geste « Fierabras » und ihre Vorstufen*, Leipzig.

Gröber, Gustav, 1877. *Die Liedersammlungen der Troubadours*, RSt 2, 337-360.

Leonardi, Lino, 2001. *Canzonieri della lirica italiana delle origini. Saggi critici*, Florence, Edizioni del Galluzzo.

Leonardi, Lino, 2009. «L'art d'éditer les anciens textes (1872-1928). Les stratégies d'un débat aux origines de la philologie romane», R 127, 273-302.

Meliga, Walter, 2006. «Le raccolte d'autore nella tradizione trobadorica», in: Lo Monaco, Francesco / Rossi, Luca Carlo / Scaffai, Niccolò, *«Liber», «fragmenta», «libellus» prima e dopo Petrarca. In ricordo di D'Arco Silvio Avalle. Seminario Internazionale di Studi, Bergamo 23-25 ottobre 2003*, Florence, SISMEL – Edizioni del Galluzzo, 81-91.

Monfrin, Jaques, 1984-1986. «Problèmes d'éditions de textes», in Actes du XVIIe Congrès International de Linguistique et Philologie Romanes (Aix-en-Provence, 29 août-3 septembre 1983), Aix-en-Provence, Université de Provence, vol. IX, 351-364.

Muratova, Xenia, 2005. «Un nouveau manuscrit du Bestiaire d'Amours de Richard de Fournival », in: van den Abeele / Baudouin (ed.), *Bestiaires médiévaux. Nouvelles perspectives sur les manuscrits et les traditions textuelles. Communications présentées au XVe Colloque de la Société Internationale Renardienne*, Louvain-la-Neuve, Institut d'études médiévales de l'Université Catholique de Louvain, 265-281.

Paris, Gaston / Pannier, Léopold, 1872. *La Vie de Saint Alexis. Poème du XIe siècle et renouvellements du XIIe, XIIIe et XIVe siècles*, Paris, Frank.

Pasquali, Giorgio, 1934. *Storia della tradizione e critica del testo*, Florence, Le Monnier.

Perugi, Maurizio, 1978. *Le canzoni di Arnaut Daniel. Edizione critica*, a cura di M.P., Milan-Naples, R. Ricciardi, 2 voll.

Roncaglia, Aurelio / McMillan, Duncan / Segre, Cesare / Lecoy, Félix, 1978. «Premesse ideologiche della critica testuale», in: Vàrvaro, Alberto (ed.), *XIV Congresso internazionale di Linguistica e Filologia Romanza. Atti. Napoli, 15-20 aprile 1974*, Naples-Amsterdam, Macchiaroli-Benjamins, I, 479-514.

Rychner, Jean, 1966. *Les lais de Marie de France*, publiés par J.R., Paris, Champion.

Segre, Cesare, 1957. *Li bestiaires d'Amours de Maistre Richart de Fornival e Li Response du Bestiaire*, a cura di C.S., Milan-Naples, R. Ricciardi.

Segre, Cesare, 1959. «Per l'edizione critica dei *lai* di Maria di Francia», *CN* 19, 215-237.

Segre, Cesare, 1968. Compte-rendu de Rychner, Jean, 1966, *CCM* 11, 243-246.

Segre, Cesare, 1971. *La Chanson de Roland*. Edizione critica a cura di C.S., Milan-Naples, R. Ricciardi.

Segre, Cesare, 1989. *La Chanson de Roland*. Édition critique par C.S.. Nouvelle édition revue traduite de l'italien par Madeleine Tyssens, Genève, Droz, 2 voll.

Segre, Cesare, 2010. «Due interventi sul *Bestiaire d'Amours* di Richard de Fournival», *MedRom* 34, 396-405.

Squillacioti, Paolo, 1999. *Le poesie di Folchetto di Marsiglia*. Edizione critica a cura di P.S., Ospedaletto, Pacini.

Timpanaro, Sebastiano, 1963. *La genesi del metodo del Lachmann*, Florence, Le Monnier (édition augmentée : Padoue, Liviana, 1981).

Trovato, Paolo, 2013. «La tradizione manoscritta del *Lai de l'ombre*. Riflessioni sulla tecnica d'edizione protonovecentesca», *R* 131, 338-380.

Tyssens, Madeleine, 1991. *Lyrique romane médiévale. La tradition des chansonniers*, Actes du Colloque de Liège, 1989, édités par M.T., Liège, Université de Liège.

Whitehead, Frederick / Pickford, Cedric E., 1973. «The Introduction of the *Lai de l'ombre*. Sixty Years Later», *R* 94, 145-156.

Du chemin parcouru entre Coquebert de Montbret père et fils et la plus récente dialectométrie : une reconsidération critique

1. De Coquebert de Montbret père et fils à Jules Gilliéron : deux points de repère de la genèse de l'analyse empirique de l'espace de la France

Depuis un certain temps je défends l'idée que l'ALF et l'œuvre entière de Jules Gilliéron (1854-1926) dépendent d'un substrat intellectuel « typiquement français » et qu'ils doivent donc être considérés dans une perspective historique particulière. La même remarque vaut pour la genèse de la dialectométrie qui, de par son souci enveloppant de généralisation spatiale, a navigué dans le même sillage[1].

Rappelons brièvement les jalons les plus importants de ces implications. L'ALF est caractérisé, entre autres, par trois propriétés qui ont grandement favorisé sa réussite fulgurante :

- un maillage d'enquête « économique » et « rationnel », basé sur le réseau départemental de la France et partageant avec lui, de ce fait, tous les avantages y ayant trait, comme l'équidistance entre les mailles de la grille et le nombre « opérationnel » (= limité) des points d'enquête[2].
- un questionnaire d'accessibilité « pan-galloromane », c'est-à-dire contenant des items universellement compréhensibles et partant – métrologiquement parlant – pourvus d'un haut degré de reliabilité et de validité.
- les consignes très strictes données à l'enquêteur Edmond Edmont qui consistaient avant tout à n'enregistrer que la première réponse des sujets interviewés et d'éviter toute sorte d'extorsion.

Avec la curiosité de l'historien, l'on peut se demander deux choses :

- Tout ceci a-t-il été inventé de toutes pièces par Gilliéron lui-même,
- ou l'a-t-il conçu sous le coup d'une tradition spirituelle ou intellectuelle antérieure dans laquelle il avait baigné dès 1876, l'année de son arrivée en France?

---

1 Voir aussi nos contributions de 2002(b) et 2006.

2 Rappelons que la limitation programmatique des travaux de sondage à quelque 600 points d'enquête par rapport au total des 36 000 communes de la France équivaut, en dernière analyse, au tirage d'un échantillon aléatoire et représentatif.

Une autre certitude est son indépendance intellectuelle complète de l'entreprise géolinguistique allemande « Deutscher Sprachatlas » de Georg Wenker (1852-1911), pourtant antérieure à la sienne de quelque 20 ans[3].

Mais d'où avait-il pris l'idée heureuse de construire le réseau d'enquête de l'ALF suivant le compartimentage départemental de la France? Toujours est-il que cette décision s'est révélée, après coup, comme étant vraiment *heureuse*, vu son excellente maniabilité, qui se détache considérablement de la contrepartie allemande lourde – et non pas riche, en fin de compte – de plus 50 000 points d'enquête.

Or, je pense que les idées linguistiques de Gilliéron se nourrissaient – au cours de la période de gestation de l'ALF – d'un substrat intellectuel préexistant dont on parle peu en matière de linguistique ou de géolinguistique.

L'esprit de la départementalisation de la France – qui, elle, était une des premières mesures prises par la Révolution française, devenue par la suite très efficace – a favorisé, entre autres, une nouvelle considération de l'espace de la Nation, tout en la définissant non pas comme un continuum amorphe, mais bien plutôt comme la somme d'un nombre relativement petit de vecteurs, entre eux égaux et équidistants[4]. En bref, c'était un acte de mensuration visant à une discrétisation rationnelle de l'espace de la France, lourd de conséquences. Précisons que les préambules intellectuels de cette réorganisation de l'espace de la France se situaient déjà aux XVIIe et XVIIIe siècles.

Mais, ce qui compte le plus en l'affaire, c'est la dissémination à tous azimuts de l'usage de la nouvelle grille départementale de la France pour toutes sortes de relevés et d'études empiriques qui ont été réalisés au cours du XIXe siècle. Il en est né ce qu'on appelle la « Statistique Départementale ».

Dans ce contexte, marqué par une curiosité empirique toujours portant sur l'ensemble de la France, brillent – en ce qui concerne nos propos de linguistes – les figures et l'œuvre de Coquebert de Montbret père[5] et fils[6], devenus les auteurs, entre autres, de la première enquête linguistique du territoire entier de la France (et aussi bien au-delà), réalisée entre 1806 et 1812[7].

Une autre des émanations les plus connues de cette nouvelle veine d'investigation est la « Carte figurative de l'instruction populaire de la France », dressée en 1827 par le mathématicien Charles Dupin (1784-1873) : voir la figure 1[8].

---

3 Cf. Herrgen (2001).

4 Cf. Ozouf-Marignier (1989).

5 Il s'agit de Charles-Etienne Coquebert de Montbret (1755-1831).

6 Il est question d'Eugène-Barthélémy Coquebert de Montbret (1785-1847), un des cinq enfants de Charles-Etienne.

7 Au sujet de Coquebert de Montbret père et fils voir surtout la somme magistrale d'Isabelle Laboulais-Lesage de 1999, la thèse récente de Sven Ködel (2014) soutenue en 2013 aux Universités de Bamberg et Paris 7 (Diderot) et les indications fournies par S. Pop (1950, 19-26).

8 Pour la personne et l'œuvre de Ch. Dupin, cf. Palsky (1996, 62-67).

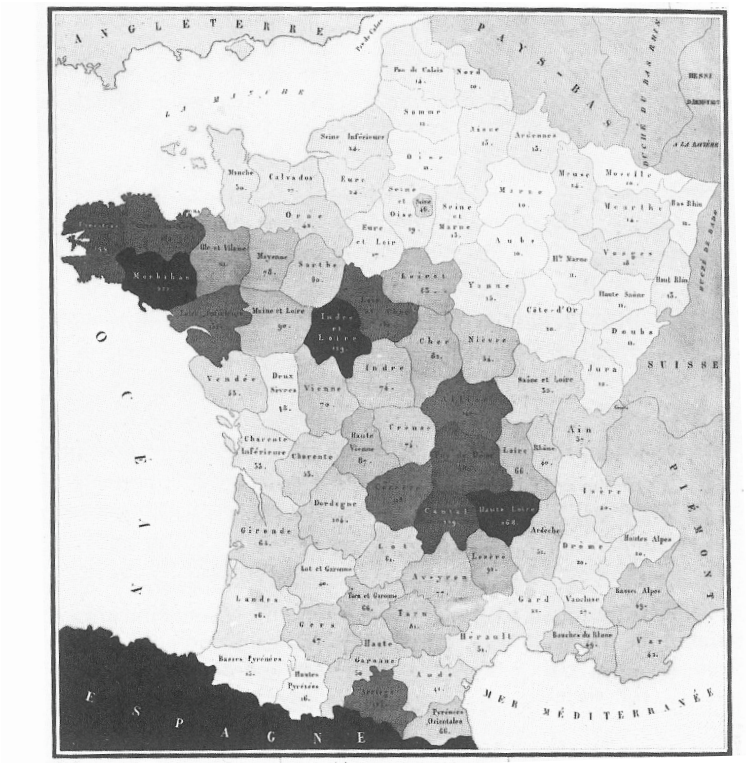


Figure 1 : Charles Dupin (1784-1873) : Carte figurative de l'instruction populaire de la France en 1827.

Cette carte présente déjà quelques particularités empiriques et techniques qui caractérisent également la production cartographique de la dialectométrie. Elle repose sur des données judicieusement collectées, visualise un sujet thématique bien défini (= le degré d'alphabétisation) et a été dressée moyennant la technique visuelle de la carte en plages (= carte choroplèthe). Son message cartographique repose sur une dizaine de teintes grises, réparties entre le noir complet et le blanc, dont le principe sémiotique est simple : plus c'est blanc, meilleure est l'alphabétisation des habitants des différents départements de la France.

Une considération sommaire du profil choroplèthe de la carte fait ressortir une ligne séparatrice située approximativement entre St-Malo et Genève. En 1827 et après, le message de cette carte a provoqué d'énormes conflits, voire des scandales, sur la scène intellectuelle, culturelle et politique de la France, et ceci à cause de l'inégalité évidente entre le Sud et le Nord de la Nation[9].

[9] Cf. à ce sujet la contribution de Roger Chartier (1992).

La « Statistique Départementale » reposait sur des piliers méthodiques très solides : il y avait déjà une théorie de départ – une sorte d'idée d'une « Belle au bois dormant », dont il fallait trouver la cachette ou les traces –, des techniques bien rodées pour la collecte des données et aussi une maîtrise cartographique toujours en essor pour mettre au point la meilleure des visualisations des recherches entreprises[10].

Évidemment, la galerie de telles cartes pourrait être facilement élargie. La production française respective en est très riche, pour ne pas dire exubérante[11].

À partir de 1860, il y avait même une cartographie statistique institutionnalisée, et pratiquée sous les auspices du Ministère des Travaux Publics, dont l'émanation la plus connue était l' « Album des statistiques graphiques[12] ». C'était une collection de cartes thématiques à couleurs, toutes en format in folio, qui avait le privilège d'être omniprésente dans tous les organismes et bureaux de l'État. En bref, la vie intellectuelle et politique de la France était imprégnée – au moment de la gestation de l'ALF – de cartes, d'idées cartographiques et de tous les prérequis méthodiques, théoriques et pratiques y ayant trait.

Et Jules Gilliéron : combien de tout ceci connaissait-il? Lui, qui pourtant était le fils d'un géologue? Evidemment, il est possible qu'il n'en sût rien, mais est-ce réellement vraisemblable?

Ma conviction est que l'entreprise de l'ALF représente, en dernière analyse, une des dernières émanations de la « Statistique départementale » telle qu'elle a été pratiquée tout au long du XIXe siècle – depuis la Révolution française et le Premier Empire – pour beaucoup d'autres variables, qu'elles soient d'origine humaine ou non.

Évidemment, depuis la période de l'ALF les méthodes de la géographie linguistique ont changé. Le grand dilemme devant lequel se trouvaient les auteurs d'atlas linguistiques du XXe siècle était la nature foncièrement bipolaire du travail d'atlas en tant que tel : ou bien collecter les données dialectales à l'aune de *l'espace* tout en mettant l'accent – dans le binôme *géographie linguistique* – sur le mot *géographie*, ou bien collecter les données à l'aune de la *faculté langagière* des locuteurs (dont on connaît le caractère hétéroclite) tout en mettant l'accent sur l'épithète *linguistique*.

Pour Gilliéron, la collecte de données exclusivement *basilectales* – issues d'un acte conscient de traduction de la part de témoins bilingues – était un prérequis de fer. Cette procédure lui semblait être la garante de la bonne *intercomparabilité* des données collectées. Cette idée – absolument normale au sein de toutes les sciences empiriques telles que la sociologie ou la psychologie – fut encore partagée et défendue par Albert Dauzat (1877-1955) au moment de la planification du « Nouvel Atlas linguistique de la France » (NALF)[13]. Or, nous savons bien que les premières réalisations

[10] Voir à ce sujet la grande analyse de Marie-Noëlle Bourguet de 1988.

[11] Pour une description très détaillée de ce contexte, je renvoie expressément à l'excellent livre du géographe français Gilles Palsky de 1996.

[12] Cf. Palsky (1996, 141 sq.).

[13] Pour le programme du NALF cf. Tuaillon (1983).

concrètes du NALF – toutes conçues et réalisées par d'éminents dialectologues de renom tels que Pierre Gardette (1906-1973), Jean Séguy (1914-1973) ou Pierre Nauton (1917-1970) – n'ont pas été faites sous le signe de l'intercomparabilité des données recueillies : ni entre elles, ni avec celles de l'ALF ou d'un atlas voisin, mais bien plutôt sous le signe de *l'authenticité*, tant des enquêtes – faites souvent en *patois* et partant sans le biais de la traduction bilingue consciente – que des réponses, recueillies non pas en forme standardisée, mais au cours d'une *conversation dirigée*. D'où l'inévitable augmentation, chaleureusement saluée par les dialectologues, de réponses doubles, triples et même au-delà, toutes hautement idiosyncrasiques, mais dépourvues de toute intercomparabilité.

Au début, la géographie linguistique romane n'a nullement été entravée par cette bipolarité. A preuve de cela, je cite la double dédicace, formulée par Walther von Wartburg (1888-1971) au début du premier volume du FEW, à l'intention de Wilhelm Meyer-Lübke (1861-1936), le traditionnaliste néogrammairien, et à Jules Gilliéron (1854-1926), l'innovateur empirique et théorique.

À la différence de Pierre Gardette, Jean Séguy, en tant qu'auteur de l' « Atlas linguistique de la Gascogne » (ALG), voulut longtemps respecter les principes prônés par Albert Dauzat et produire, de ce fait, « de l'intercomparable ». Il a fini par succomber d'abord à la tentation de *l'authenticisme* empirique, pour se convertir ensuite – surtout après l'expérience faite avec le quatrième volume de l'ALG – en *dialectométricien*.

Par le biais de l'analyse *quantitative* des données *qualitatives* débordantes contenues dans son ALG, il voulait en extraire une sorte de sève vitale cachée, comparable à la « Belle au bois dormant », dont on devine l'existence, mais ignore le domicile.

Aux dires de Xavier Ravier, son fidèle compagnon de travail, Jean Séguy voulait faire tenir (ou réduire) le gascon – entendu comme une masse de « 47 millions de chiures de mouche » – « dans une formule ou un schéma[14] ».

Ce qu'il est devenu de la *dialectométrie* depuis 1973, année du lancement du mot et aussi de la chose, j'en parlerai tout de suite. D'entrée de jeu, j'aimerais bien souligner encore un fait important. Les études de dialectométrie, dont il existe de nos jours beaucoup de variantes, ont démontré que sous les enchevêtrements souvent inextricables ou carrément chaotiques d'un nombre inouï d'aires linguistiques appartenant à toutes les catégories linguistiques imaginables, il se cache d'étonnantes régularités spatiales qui peuvent être considérées comme autant d'émanations de lois géolinguistiques très claires et simples qui, elles, constituent la contrepartie des « Lautgesetze » (lois linguistiques) d'ordre diachronique découvertes par les néogrammairiens de Leipzig dans le dernier quart du XIXe siècle.

[14] Cf. Ravier (1976, 390).

2. La dialectométrie : une méthode inductive et exploratoire

Au début des années 70 du XXe siècle, la dialectométrie (DM) est née sous l'impulsion de la curiosité de quelques géolinguistes[15] qui voulaient savoir si, derrière (ou : sous) l'apparent désordre des structures spatiales des cartes de n'importe quel atlas linguistique (AL), il existe un ordre enveloppant et, partant, ou bien « majeur » ou bien « caché ». Ce but ne pouvait être rejoint que par la synthèse, faite à la main[16] ou à l'aide de l'informatique alors en état de gestation[17], de beaucoup de cartes d'atlas. Par une vieille tradition de la philosophie occidentale, une telle procédure – qui part du plan du « particulier » pour arriver à celui du « général » – est appelée « induction ». Comme elle se lance aussi à la recherche de choses jusqu'alors inconnues, elle a par définition aussi une vocation « exploratoire ».

Après plus de quatre décennies d'intenses recherches-DM appliquées à des atlas linguistiques romans[18], allemands[19] et anglais[20], il est clair que l'ordre sous-jacent (c'est-à-dire « caché » ou « de profondeur ») assumé d'entrée de jeu, existe réellement et qu'il est l'émanation (ou la création) d'une activité particulière des humains. Au sein de l' « École dialectométrique de Salzbourg » (EDMS), il est question, dans ce contexte, depuis belle lurette de la « gestion basilectale de l'espace par l'homo loquens », considérée comme une des multiples facultés innées des humains et à laquelle est attribué le rang d'un « universal » du langage *à plein titre*.

C'est ainsi qu'en face d'un atlas linguistique (AL) à dialectométriser, nous nous sommes toujours posé les questions suivantes :

- Quelle est la variation géolinguistique *globale* répertoriée sur toutes les planches de l'AL en question, en l'occurrence de l'ALF ?
- Comment se reflète, sur toutes les planches de l'ALF, la *gestion basilectale de la Galloromania* de la part des Français, des Suisses romands et des Wallons? Pour préciser : nos recherches-DM relatives à l'ALF englobaient, certes, le réseau entier de l'ALF, mais ont dû se limiter, pour des raisons d'ordre économique, à l'analyse de quelque 40% des 1421 planches originales de la série A de l'ALF[21]. Ce pourcentage est pourtant suffisamment grand pour pouvoir en tirer des conclusions représentatives et, de ce fait, fiables.

---

15 Je renvoie, à ce sujet, à deux articles de Jean Séguy (1971 et 1973 : avec la mention du néologisme *dialectométrie*) ainsi qu'à mes propres contributions de 1971 (avec N. Winterleitner) et 1976.

16 C'était le cas pour Jean Séguy, décédé déjà en 1973.

17 J'avais le privilège de pouvoir me servir, dès le début, de l'informatique.

18 L'essentiel de nos recherches-DM romanes porte sur les grands atlas nationaux ALF, AIS (Italoromania), ALPI (Ibéroromania) et ALDC (domaine catalan).

19 Cf Goebl/Scherrer/Smečka (2013).

20 Cf. Goebl/Schiltz (1997).

21 Pour plus de détails sur la dialectométrisation des données de l'ALF, cf. Goebl (2002a et 2003).

3. Préparation cartographique du réseau d'enquête de l'ALF

Rappelons d'abord trois caractéristiques élémentaires de l'ALF :

- Le réseau de l'ALF comprend 638 localités[22] où Edmond Edmont[23] a mené 639 enquêtes.
- Jules Gilliéron a inauguré, par les modalités typographiques de la publication des données de l'ALF, la tradition – devenue plus tard caractéristique de la géographie linguistique *romane* en général – de présenter, sur chaque planche d'atlas, une synopse de toutes les transcriptions recueillies sur le terrain, relatives à un item particulier du questionnaire utilisé[24]. C'est le principe du « full text map » qui s'oppose à celui du « symbol map » utilisé par les *germanistes* et les *anglicistes*.
- Dès le début, la présentation des données transcrites originales sur la superficie d'une planche d'atlas a rendu nécessaire – pour les besoins de leur dépouillement rationnel – le recours à un instrument heuristique spécial, à savoir aux « cartes muettes ».

Ces faits expliquent pourquoi la « géographie linguistique » créée par Jules Gilliéron ne pouvait se faire sans un recours continu et massif à la pensée et aux méthodes d'une discipline située en dehors de la linguistique, à savoir de la *cartographie*. Le maniement et le remplissage des *cartes muettes* demandaient en outre, de la part des adeptes de la nouvelle méthode, le déploiement de *réflexions géo-classificatoires* visant à subdiviser les données originales d'une planche d'atlas en groupement spatiaux utiles suivant la logique d'un principe classificateur préalablement choisi. Les annales des études romanes fournissent une quantité énorme d'applications très réussies de cette démarche géo-classificatoire[25].

De même qu'aux temps de la publication de l'ALF quand l'imprimeur Protat Frères (de Mâcon) a mis en vente des cartes muettes du réseau de l'ALF pour permettre aux lecteurs de ce dernier d'y reporter le rendement de leurs réflexions géo-classificatoires, il est également de nos jours indispensable de créer, pour le réseau d'un AL en train d'être dialectométrisé, un support cartographique capable de recevoir les informations spatiales tant *qualitatives* que *quantitatives* utilisées au cours

---

22 Le nombre réel des *points d'enquête* de l'ALF est 638 et non pas 639 comme il est indiqué très souvent par erreur. L'ALF comporte cependant, sur chacune des 1 421 planches de la série A, les résultats transcrits de 639 *enquêtes* faites par Edmond Edmont. La solution de cette « énigme » est qu'Edmont avait réalisé deux enquêtes au point-ALF 284, Saint-Pol-sur-Ternoise, sa ville natale.

23 Rappelons qu'Edmond Edmont (1849-1926) était l'enquêteur de l'ALF.

24 Du temps de la publication des données de l'ALF, le principe combien heureux du « full text map » était bien loin d'être hors conteste. Un des adversaires les plus farouches de ce principe était Paul Meyer qui s'employait énergiquement pour une publication des données de l'ALF sous la forme de tableaux : cf. Storost (1994, 77). Quelle catastrophe pour la géographie linguistique si P. Meyer avait réussi à imposer ses visées saugrenues!!

25 Ceci concerne surtout les apports de l'onomasiologie romane qui, elle, se basait avant tout sur les atlas linguistiques.

de l'analyse-DM. Un tel support présuppose l'établissement d'un compartimentage polygonal du réseau en question, dit aussi « tessélation de Voronoï ».

La polygonisation de Voronoï se fait à travers les étapes suivantes[26] :

- triangulation du réseau des points d'enquête
- construction successive des médiatrices sur les côtés de triangle du réseau
- De la fusion de toutes les médiatrices ainsi construites naît le réseau polygonal dont nous avons besoin.

4. La taxation des données d'un atlas linguistique

Voir les cartes 1 et 2.

Au sein de l'EDMS, l'on entend par la « taxation » (des planches originales d'un AL) l'analyse d'un secteur déterminé des informations originales contenues dans une carte originale de l'AL examiné, et l'établissement successif d'une « image » des données originales respectives sous la forme d'une « carte de travail » (CT). Précisons que ce genre de travail classificatoire s'aligne parfaitement sur une tradition déjà ancienne en matière de linguistique romane qui avait été inaugurée par Jules Gilliéron lui-même[27].

Les CT issues d'une campagne de taxation disposent de structurations internes très variables qui, elles, peuvent osciller entre 2 et *n* « aires taxatoires ». L'élément linguistique de base qui régit chaque aire taxatoire, est appelée « taxat ». Au cours de la dialectométrisation de plus de 600 des 1 421 planches originales de l'ALF entreprise peu avant le tournant du millénaire à Salzbourg[28], nous en avons extrait 1 681 CT appartenant aux catégories de la *phonétique*, du *lexique* et de la *morpho-syntaxe*, dont la figure 2 montre la distribution quantitative.

---

26 Pour plus amples détails cf. Goebl (1981, 363 sq.) et Goebl (1984 I, 90 sq.).

27 Cf. – entre autres – Gilliéron (1918) : à la fin de cet ouvrage se trouve une typisation lexicale relative à la carte 1 (abeille) de l'ALF. Voir également Jaberg (1908), où l'on trouve toute une gamme de typisations coloriées de différentes planches de l'ALF.

28 Voir nos contributions de 2000, 2002a et 2003 qui en présentent les principes et quelques résultats.

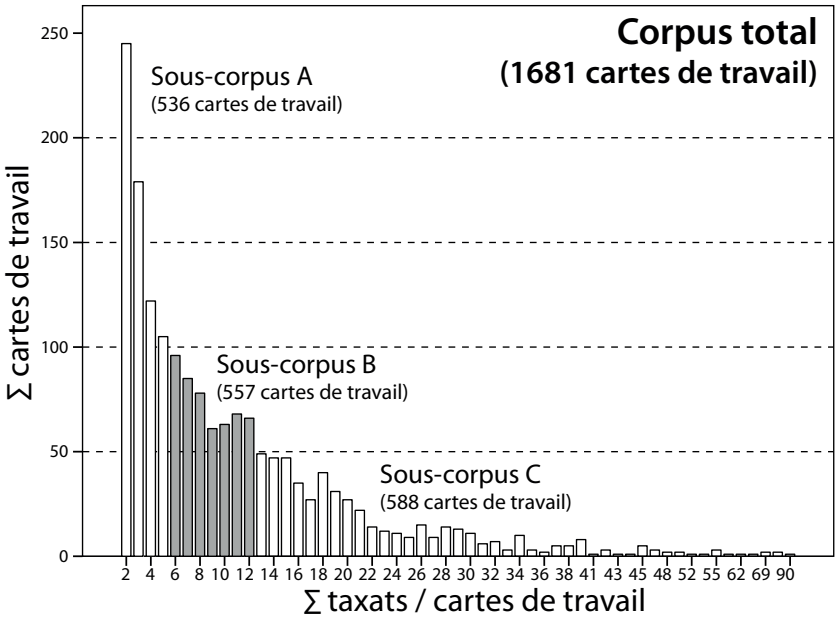


Figure 2 : Histogramme – pourvu de 55 colonnes – montrant, pour notre corpus-DM tiré de l'ALF, la relation entre le nombre de taxats par carte de travail (en abscisse) et le nombre absolu des cartes de travail (en ordonnée)

Les colonnes de l'histogramme en question représentent une courbe décroissante de gauche à droite[29]. Il s'agit de 55 colonnes de hauteurs inégales qui couvrent un espace « polynymique » allant de 2 taxats/CT jusqu'à 90 taxats/CT : voir l'abscisse sur la figure 2.

Les hauteurs de 55 colonnes signalent les nombres respectifs des CT analysées. Il y a donc 245 CT « *bi*-nymes » (= avec *deux* types linguistiques ou taxats), 179 CT « *tri*-nymes » (= avec *trois* taxats), 122 CT « *tétra*-nymes » (= avec quatre taxats) et seulement une CT de nature *90*-nyme. Elle se réfère d'ailleurs à la carte 545 de l'ALF, relative aux dénominations de la *fauvette*.

[29] Il s'agit d'une courbe exponentielle. Nous avons d'ailleurs découvert les mêmes régularités dans toutes nos analyses dialectométriques qui portent sur tous les grands atlas romans ainsi que sur ceux des études anglaises. Le caractère exponentiel de ces courbes a été utilisé, en 1985, par le linguiste quantitatif allemand Gabriel Altmann pour la définition de ce qu'il a appelé alors la « loi de Goebl ».

Le bilan taxatoire de notre analyse-DM de l'ALF se présente donc comme suit :

corpus	∑ cartes de travail (CT)	∑ aires taxatoires / taxats
total	1 681	18 047
phonétique	1 117	10 642
lexique	465	6 514
morpho-syntaxe	99	891

Dans l'ensemble, la richesse structurale des quelque 18 mille aires taxatoires analysées est très grande. N'oublions pas que la structure spatiale de chaque carte de travail (CT) peut varier dans quatre directions :

1. du nombre des taxats
2. de la taille des aires taxatoires
3. de la configuration des aires taxatoires et
4. de la localisation des aires taxatoires à l'intérieur du réseau d'atlas en question.

Il en résulte une variabilité vraiment déroutante. Une des conséquences de cette grande variabilité est le fait que les différentes aires taxatoires ne se recouvrent presque jamais. Il en résulte un effet d'*enchevêtrement* très complexe entre les aires taxatoires qui semble être typique de beaucoup de phénomènes spatialement répartis.

La carte 1 montre une des 1 117 CT *phonétiques* analysées. Avec ses 17 taxats *phonétiques* – qui symbolisent autant de succédanés du nexus vocalique latin -ātu (dans merc-ātu « marché ») – elle représente une CT déjà relativement complexe.

La carte 2, en revanche, appartient à notre corpus *lexical*, riche de seulement 465 CT : elle est franchement moins complexe[30].

Une des premières expériences que nous ayons faites au cours de nos travaux-DM concernait l'impact du nombre absolu des CT synthétisées par voie-DM. Nous avons constaté alors que les structures-DM de profondeur obtenues à l'aide de la synthèse de 1 681 CT (= corpus total) ressemblent de très près à celles obtenues avec 50%, 25% et 15% de cet effectif, à condition que l'extraction de l'urne pleine en ait été faite par voie aléatoire[31]. La même expérience a pu être réitérée avec d'autres atlas linguistiques romans et non romans.

De cette réalité, étonnante au premier coup d'œil, nous avons conclu que les structures-DM de profondeur mises au jour, sont imprégnées d'une grande *redondance*

[30] La distribution spatiale et l'origine des noms de jours de la semaine de la Galloromania ont été analysées à plusieurs reprises : cf. Gilliéron/Roques (1912, 85-108), Rohlfs (1949), Wartburg (1949) et Baehr (1952).

[31] Cf. Goebl (1984 I, 206-219).

interne qui, elle, serait comparable à celle dont disposent tous les systèmes ou signaux communicatifs, qu'ils soient de nature linguistique ou non.

Une autre réflexion concernait les conditions de la genèse de cette redondance dont l'origine était indubitablement humaine. Depuis un certain temps, nous sommes convaincus que toutes ces régularités quantitatives sont l'émanation directe d'un comportement particulier des dialectophones par rapport aux défis communicatifs de l'espace qu'ils habitent. Nous lui avons conféré l'appellation de « gestion (ou : aménagement) basilectale de l'espace par l'homo loquens »[32].

5. Brève présentation de la « chaîne dialectométrique » de l'EDMS

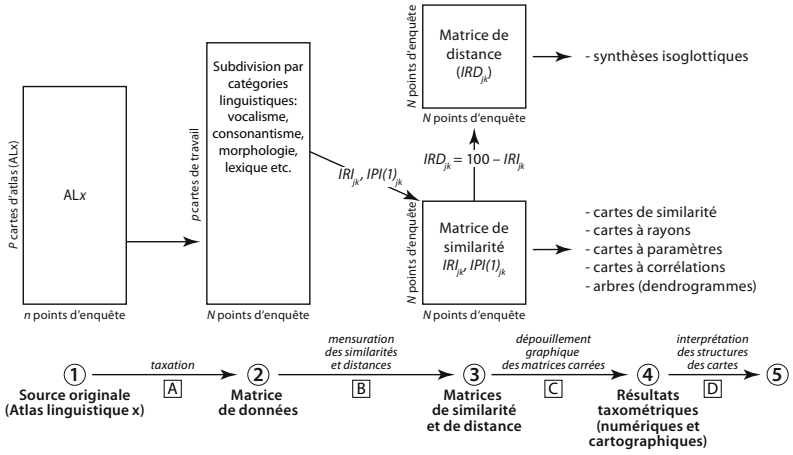


Figure 3: Diagramme de la chaîne méthodique utilisée par l' « Ecole Dialectométrique de Salzbourg »

Les méthodes de l'EDMS sont alignées en forme de « chaîne » : voir la figure 3. Le tableau suivant renseigne sur les fonctions des différents « chaînons » ou étapes de cet alignement[33].

---

32 Comparer à cela le titre de notre contribution de 2005 et la définition de ce concept dans Goebl (2002a, 5).

33 Pour une introduction détaillée aux buts et méthodes de la DM de Salzbourg nous renvoyons à Goebl (1984), vol. I-III (en allemand) et à nos contributions de 2000, 2002a, 2003 et 2012 (toutes en français). Voir en outre les adresses-web suivantes :

- bibliographie complète de nos travaux-DM :
  https ://www.sbg.ac.at/rom/people/prof/goebl/dm_publi.htm
- trois sites-DM de travail et de présentation :
  http ://www.dialectometry.com/
  http ://www.dialectometry.com/dmdocs/index.html
  http ://ald.sbg.ac.at/dm/

Étapes méthodiques	États des données en voie de traitement-DM	Commentaires
A	1-2	Choix d'un atlas linguistique [1] pour la dialectométrisation ; définition et application des principes taxatoires respectifs ; taxation [A] d'un nombre aussi grand que possible de cartes originales de l'atlas linguistique retenu ; saisie électronique successive des cartes de travail (CT) analysées et établissement de la matrice de données [2] (avec N points d'enquête et p cartes de travail) à l'aide du logiciel VDM[34]. Les données taxées et la matrice de données se situent au niveau de l'échelle métrologique nominale.
B	2-3	Choix et application d'une mesure *quantitative* pour déterminer les similarités (quantitatives) entre les N points d'enquête : ici : choix de l'IRI_{jk} (« Indice Relatif d'Identité »)[35] ; établissement consécutif de la matrice de similarité (carrée : N × N) [3, en bas]. Celle-ci contient la variation *qualitative* de la matrice de données après sa conversion dans une forme *quantitative*.
C	3-4	Transformation (partielle ou intégrale) du contenu *quantitatif* de la matrice de similarité en représentations et schémas visuels. Le choix des outils de visualisation dépend uniquement des besoins heuristiques du géolinguiste. En utilisant le logiciel VDM : choix de l'algorithme de visualisation approprié et du nombre des paliers chromatiques (intervalles).
D	4-5	Surtout par voie optique : interprétation et évaluation théoriques et pratiques [D] des structures spatiales des mises en cartes effectuées. Genèse de nouvelles représentations conceptuelles et théoriques [5] des données de départ [1].

34 VDM : „Visual DialectoMetry", logiciel dialectométrique créé par Edgar Haimerl (Seattle, EUA) entre 1997 et 1999, mis à jour continuellement depuis cette date.

35 L'indice de similarité standard de la DM de Salzbourg est l' « Indice Relatif d'Identité » (IRI_{jk}) auquel fait écho l' « Indice Relatif de Distance » (IRD_{jk}) selon la formule $IRI_{jk} + IRD_{jk} = 100$.

Précisons encore que dans cette démarche pluridisciplinaire, le savoir nécessaire pour l'établissement de la matrice de données et pour la définition de l'indice de similarité est à rechercher du côté de la « classification numérique » (dite aussi « automatique »)[36] alors que les méthodes pour la confection et la visualisation appropriée des cartes en couleurs sont fournies par la cartographie numérisée[37], appelée de nos jours aussi « visualistique » (en anglais : « visualistics »). Ces facilités n'existaient d'ailleurs pas encore en tant que telles au commencement de nos activités-DM.

Une des particularités de la procédure pratique de l'EDMS est de toujours privilégier la dimension heuristique traditionnelle de la géographie linguistique : à savoir le plan des cartes. Tous les résultats quantitatifs des calculs-DM effectués sont directement convertis en schémas visuels appropriés et offrant un accès aussi suggestif (et facile !!) que possible aux géolinguistes. Soulignons en outre que l'EDMS pratique également une sélection judicieuse des procédures statistiques utilisées tout en évitant l'application de celles dont la *complexité mathématique* n'est pas à la hauteur de leur *utilité géolinguistique*.

À partir de 1997, tous les calculs et toutes les visualisations-DM s'effectuent à l'aide d'un logiciel–DM appelé « Visual DialectoMetry » (VDM), qui avait été créé par notre ami Edgar Haimerl et qui, depuis cette date, a connu plusieurs remaniements et élargissements[38].

Dans la panoplie salzbourgeoise des indices de similarité, il y a, entre autres, aussi l'indice IPI_{jk} (« Indice Pondéré d'Identité ») qui dispose de certains avantages taxométriques et visualisateurs auxquels le dialectométricien peut avoir recours le cas échéant.

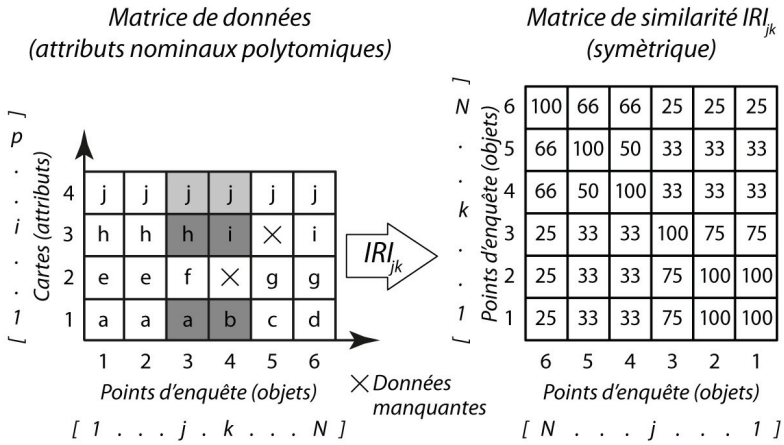
36 Cf., entre autres, les grands manuels taxométriques de Hans Hermann Bock (1974) ainsi que de Peter H. A. Sneath et Robert R. Sokal (1973).

37 Voir les manuels cartographiques de Roger Brunet (1987) et de Jean-Paul Bord (2012).

38 Le logiciel VDM est à la libre disposition de qui en fait la demande.

6. La mesure de similarité entre deux vecteurs de la matrice des données

Voir la figure 4.



$$IRI_{3,4} = 100 * \frac{\Sigma(COI)}{\Sigma(COI + COD)} = \frac{1}{1+2} = 33\%$$

Figure 4 : Schéma de la mensuration de la similarité linguistique entre les différents points d'enquête d'un atlas linguistique

$IRI_{3,4}$: Indice Relatif d'Identité entre les points d'enquête 3 et 4
COI : co-identité (entre les paires de taxat j et j)
COD : co-différences (entre les paires de taxat a et b ainsi que h et i)

L' « Indice Relatif d'Identité » (IRI_{jk}) mesure la proportion de taxats identiques entre deux vecteurs-points d'enquête (j et k) de la matrice de données. Théoriquement, les scores de l'IRI_{jk} peuvent varier entre 1 et 100 (%). Suivant la formule combinatoire N/2 (N - 1) et pour l'effectif N = 641, le bilan total des calculs à effectuer s'élève à 205 120 scores qui, par la suite, seront stockés dans la matrice de similarité carrée (aux dimensions N × N).

Signalons encore que nous avons ajouté, aux 638 points d'enquête *originaux* de l'ALF, trois points *artificiels* relatifs aux langues standard du *français*, de l'*italien* et du *catalan*[39] qui permettent d'étudier l'impact de ces trois registres linguistiques sur le tissu des dialectes de l'ALF. La taille de N s'élève donc à 641 unités.

La matrice de similarité, outre d'être carrée, est aussi parfaitement symétrique. La valeur-IRI de 33% calculée pour les localités factices 3 et 4 se retrouve donc des deux côtés de la diagonale de la matrice de similarité : voir à nouveau la figure 4.

[39] Cf. à ce propos Goebl 2000, 212.

Une fois la matrice de similarité établie, il faut que le dialectométricien-classificateur se mette à la dépouiller judicieusement : voir à ce propos l'étape C sur la figure 3. En plus, c'est le premier des nombreux pas *exploratoires* qui mènent des structures « de surface » aux structures de « profondeur ».

7. La mise en carte (visualisation) de scores quantitatifs : à l'exemple de plusieurs cartes de similarité

Voir les cartes 3-8.

Le défi heuristique à relever en la matière, est double[40] :

- du point de vue *statistique* :
  Il s'agit d'extraire un des N vecteurs de la matrice de similarité, de le dépouiller de la valeur relative à la diagonale qui est toujours égale de 1 (ou 100), et de visualiser les N-1 valeurs (ou scores) restantes.
- du point de vue *cartographique* :
  Il s'agit de *convertir* la *variabilité numérique* des N-1 scores de similarité extraits en une *variabilité visuelle analogue* pour fournir au géolinguiste-dialectométricien des informations aptes à être rapidement saisies par voie mentale et interprétées après du point de vue linguistique.

La « machine visualisatrice » employée à cet effet opère avec un certain nombre d'algorithmes de visualisation dont le plus utilisé à Salzbourg est l'algorithme MINMWMAX. Il répartit les N-1 valeurs à visualiser équitablement des deux côtés de la moyenne arithmétique (MA) de la distribution de similarité respective, tout en définissant des intervalles numériquement toujours égaux des deux côtés de la MA[41].

Il en résulte des profils iconiques remarquablement stables et bien ordonnés auxquels le logiciel VDM permet de conférer des silhouettes plus ou moins fines, allant de 2 à 18 paliers chromatiques. Les couleurs que nous utilisons obéissent à l'agencement chromatique du spectre solaire qui est d'ailleurs celui de l'arc-en-ciel.

En général, le bleu est associé aux valeurs numériques petites alors que le rouge renvoie aux scores maximaux. Chaque carte de similarité – comme d'ailleurs toutes nos visualisations – dispose, en outre, d'une légende numérique et d'un histogramme qui fournit des informations qui intéressent plutôt le *statisticien* que le *linguiste*.

---

40 Voir à ce propos Goebl 1981 et 1984 (vol. I, passim).

41 Voir à ce sujet surtout Goebl (1981, 1983 et 1984).

8. Structure de profondeur A : l'importance et le rôle heuristiques des cartes de similarité

Voir les cartes 3-8.

Les cartes 3-8 constituent six spécimens de cartes de similarité relatives à différents sous-corpus (équivalents à autant de matrices de données partielles) tirés de notre matrice de données-ALF intégrale.

Les cartes 3 et 4 montrent deux des 641 cartes de similarité calculées à partir de notre corpus intégral (contenant 1 681 CT). Elles disposent, comme chaque carte de similarité, d'un point de référence dont le polygone est toujours en blanc. Par la structuration – généralement très régulière – de leur profil choroplèthe, elles indiquent le décalage spatial des similarités linguistiques des N-1 points d'atlas restants par rapport aux points de référence préétablis : 394 (Auderville, Manche) et 778 (Agde, Hérault).

Une des grandes découvertes de la DM de Salzbourg est d'ailleurs l'étonnante régularité avec laquelle ce décalage s'inscrit dans l'espace. Et ceci malgré l'indéniable « chaotisme » qui se dégage de la considération particularisante de quelque 18 000 aires taxatoires réparties sur les 1 681 CT extraites de l'ALF[42].

En ligne générale et dans les conditions de visualisation déjà mentionnées, l'entourage immédiat du point de référence prend les couleurs *rouge* (= intervalle 6) et *orange* (= intervalle 5), qui constituent toujours une aire (quantitative) cohérente. Par contre, les couleurs *bleu foncé* (= intervalle 1) et *bleu clair* (= intervalle 2) caractérisent les « antipodes » de la carte de similarité. Si la visualisation a été effectuée à l'aide de six paliers chromatiques, les polygones en *jaune* (= intervalle 4) et *vert* (= intervalle 3) représentent des zones-*tampon* (ou de *transition*) entre – métaphoriquement parlant – les « meilleurs amis », d'un côté, et, de l'autre, les « pires ennemis » du basilecte du point de référence.

Du point de vue linguistique, toute carte de similarité donne une réponse quantitative et finement nuancée à la question de savoir quelle est la position de tel dialecte à l'intérieur d'un espace dialectal examiné. J'ajoute que de telles questions ont été posées très souvent (et depuis longtemps) au sein de la romanistique et d'autres philologies modernes.

Les cartes 3, 4 et 5 montrent des profils géotypologiques pour les domaines du normand, du languedocien et du berrichon. Les profils choroplèthes respectifs sont très différents, surtout en ce qui concerne les cartes 3 (de type d'oïl) et 4 (de type d'oc).

Évidemment, les trois profils de similarité que nous venons de voir ont démontré que le décalage de la similarité linguistique dans l'espace peut être très variable. Toujours est-il qu'il s'agit là d'un phénomène *humain* ou, pour être plus précis, d'une émanation directe de la « gestion basilectale de l'espace par l'homo loquens », évoquée au début de cet article.

---

[42] Cf. Goebl (2003, 65).

Or, il serait intéressant de comparer le décalage quantitatif produit dans l'espace par une activité d'origine *humaine* avec celui issu d'un modèle *géométrique* relatif à la même superficie. Un tel modèle consiste à définir les proximités aériennes réciproques entre les points de l'ALF à l'aide de la formule bien connue de Pythagore ($a^2 + b^2 = c^2$). La visualisation des 641 vecteurs de la matrice des proximités euclidiennes crée autant de profils choroplèthes absolument réguliers qui, eux, peuvent servir d'étalon pour une comparaison avec les calculs de similarité.

La carte 6 en constitue un spécimen relatif au point de référence 103 (Bué, Cher). En la comparant avec la carte 5, l'on comprend aisément que la gestion *basilectale* (= humaine) du réseau de l'ALF s'écarte considérablement de la gestion *géométrique* (= *euclidienne*) du même espace.

Or, la comparaison *visuelle* des deux profils choroplèthes multicolores peut être complétée par une évaluation *quantitative* des deux séries de 641 scores, relatives, d'un côté, aux *valeurs de similarité* (selon l'IRI) et, de l'autre, aux *scores de proximités* (selon la formule de Pythagore). Un tel calcul de corrélation peut être fait moyennant le coefficient de corrélation de [Auguste] Bravais et [Karl] Pearson r(BP). Pour les 2 × 641 scores qui sont à la base des cartes 5 et 6, il en résulte la valeur +0,697. Prise comme telle, cette valeur ne nous dit absolument rien. Ce n'est que dans le cadre de la DM corrélative que de telles valeurs peuvent revêtir une signification *linguistique* à proprement parler : voir le paragraphe 11.

Les cartes 7 et 8 reposent sur deux sous-corpus linguistiquement différents : alors que la carte 7 se réfère à notre sous-corpus *phonétique* (avec 1 117 CT), la carte 8 a été construite à l'aide de notre sous-corpus *lexical* (avec 465 CT).

L'allure de leurs profils choroplèthes est – malgré certaines similarités de base – quelque peu différente. On a l'impression que la stabilité interne du profil choroplèthe de la carte 7 (phonétique) dépasse celle du profil choroplèthe de la carte 8 (lexique).

9. Structure de profondeur B : l'analyse « interponctuelle » et la carte à cloisons

Voir les cartes 9 et 10.

L'analyse dite « interponctuelle » connaît deux versants : celui de la visualisation de *distances* et celui de la visualisation de *similarités* interponctuelles, c'est-à-dire relevées entre deux points d'enquête contigus. La première des deux visualisations, appelée « carte à cloisons », correspond exactement à la synthèse traditionnelle d'isoglosses destinée à dégager des bourrelets de frontière plus ou moins épais[43]. Il

[43] Voir les tableaux en appendice du travail séminal de Karl von Ettmayer (1924) où se trouve un grand nombre de schémas isoglottiques exécutés à la main et contenant des bourrelets de frontière de toute sorte. Voir aussi les cartes isoglottiques reproduites dans Rosenqvist 1919 et Jochnowitz 1973.

s'agit donc de la modélisation, faite avec les moyens de la DM, d'une technique déjà ancienne et, par conséquent, bien connue des géolinguistes.

La seconde visualisation, dite « carte à rayons », est, en revanche, de date récente. Du point de vue taxométrique, elle constitue le complément de la carte à cloisons, puisqu'elle visualise des *similarités* interponctuelles (s) dont le rapport avec les *distances* interponctuelles (d) est régi par la formule suivante : s + d = 100.

Le message linguistique des deux types de visualisation est également complémentaire. Alors que la carte à *cloisons* visualise, à travers l'épaisseur et le coloriage froid des côtés de *polygone*, le degré d' « aversion » qui existe entre deux locolectes contigus, la carte à *rayons* représente, par l'épaisseur et le coloriage chaud des côtés de *triangle*, des relations d' « amitié » à courte distance[44].

La structure iconique des cartes 9 et 10 ressort de l'interaction visuelle de 1 791 éléments graphiques (i. e. de côtés de polygone) et fournit une vision très nuancée des enchevêtrements interponctuels qui existent entre le Nord et le Sud de la France. Il n'offre cependant aucune trace d'une confrontation *linéiforme* – c'est-à-dire revêtant les apparences d'une « véritable frontière » – entre les domaines d'oïl et d'oc. Sur les deux cartes, l'œil perçoit, par contre, une vaste ceinture de passage, sillonnée d'effets de cloisonnements finement gradués.

Une fois de plus, la traversée du Nord au Sud semble être spatialement plus « ordonnée » sur la carte 9 (basée sur le sous-corpus *phonétique*) que sur la carte 10 (basée sur le sous-corpus *lexical*).

Encore une petite remarque d'ordre quantitatif : l'effectif total des scores de distance (selon l'IRD_{jk}) pris en compte sur les cartes à cloisons 9 et 10, comprend seulement 0,873% du total des scores disponibles dans les matrices de distance respectives[45]. La clarté très suggestive de ces deux cartes repose donc sur des effectifs très réduits de la matrice de distance sous-jacente. En termes de *taxométrie*, la *valeur classificatoire* des deux cartes est donc relativement *faible*.

Heureusement, il y a des méthodes taxométriques beaucoup plus « enveloppantes » et, de ce fait, plus sûres : voir les paragraphes suivants.

10. Structure de profondeur C : au-delà des cartes de similarité : les cartes à paramètre

Voir les cartes 11-14.

Par « paramètre » l'on entend, en matière de *statistique descriptive*, une des nombreuses valeurs « caractéristiques » qui dénotent une quelconque distribution de fréquence, tel que le *minimum*, le *maximum*, la *MA*, l'*écart-type*, la *symétrie*, etc. En

[44] Pour des spécimens de la carte à rayons, cf. Goebl (1983 et 2012, 137-138).

[45] Cette relation se calcule comme suit : 1 791 : 205 120 × 100.

considérant les histogrammes au bas des cartes 3 et 4, l'on constate aisément que la symétrie des distributions de similarité peut être très variable.

À y regarder de près, cette propriété *statistique* revêt aussi une grande importance *linguistique*. Sur la carte 4, l'histogramme de droite, relatif à un parler du languedocien maritime, est fortement biaisé vers la gauche. Ceci signifie que des 640 scores de similarité 432 (= 53 + 245 + 134) se trouvent au-dessous de la MA[46] et seulement 208 (= 100 + 76 + 32) au-dessus du même seuil[47].

Le bilan analogue pour l'histogramme de la carte 3 se présente comme suit :

- valeurs situées au-dessous de la MA : 272 (= 9 + 168 + 95)
- valeurs situées au-dessus de la MA : 368 (= 13 + 198 + 157)

Quand on admet qu'une carte de similarité reflète – au moins en théorie – la position communicative du dialecte d'un point d'atlas, les valeurs mentionnées ci-dessus suggèrent que la position communicative *générale*[48] du dialecte du point-ALF 778 (situé en Languedoc méridional) est moins bonne que celle du point-ALF 394 (situé dans le nord de la Manche) et que, par conséquent, la plus ou moins grande symétrie des histogrammes respectifs renvoie, en dernière analyse, à cette potentialité interactive.

Or, il s'est avéré que le meilleur indice pour saisir la *symétrie* d'une distribution de similarité et, avec elle, sa vertu *communicative*, est le « coefficient d'asymétrie de [Ronald A.] Fisher » (CAF). En calculant le CAF de chacune des N distributions de similarité et en visualisant par la suite les N scores ainsi dégagés, l'on obtient une carte choroplèthe finement articulée (voir les cartes 11-14) dont la valeur linguistique est considérable. En établissant la synthèse des CAF nous poussons – métaphoriquement parlant – encore plus avant notre « plongée exploratoire dans les profondeurs » des données de l'ALF.

Du point de vue statistique, les scores du CAF oscillent des deux côtés de la valeur 0 : ils prennent des valeurs *négatives* quand les scores de similarité (selon IRI_{jk}) s'accumulent entre le maximum et la MA, et des valeurs *positives* quand le cumul intervient entre le minimum et la MA.

Sur les cartes 11-14, les polygones en *bleu foncé* – équivalant à des scores négatifs du CAF – renvoient à des dialecticités *bien connectées* avec le reste du réseau, alors que les polygones *rouges* – correspondant toujours à des scores *positifs* du CAF – se réfèrent à des variétés *détachées*, *conservatrices* et *peu enclines* à s'ouvrir à des interactions et échanges linguistiques.

---

46 Il s'agit des valeurs appartenant aux intervalles 1-3 : sur la carte 4, elles sont toutes marquées par des couleurs « froides » (bleu foncé, bleu clair et vert).

47 Il s'agit des valeurs appartenant aux intervalles 4-6 : sur la carte 4, elles sont toutes marquées par des couleurs « chaudes » (jaune, orange et rouge).

48 Il s'agit de toujours mettre en rapport les différents locolectes de l'ALF avec l'ensemble de son réseau.

Depuis un certain temps, j'utilise le concept de *compromis linguistique* (dérivé du terme allemand *Sprachausgleich*) pour désigner le degré d'interaction qu'un locolecte entretient avec le reste de son réseau. Au vu des cartes 9 et 10, l'on peut donc dire que les locolectes marqués en bleu (foncé et clair) sont très interactifs, alors que le contraire est vrai pour les locolectes marqués en rouge.

L'agencement spatial des deux tonalités est d'ailleurs très éloquent. Alors que les polygones en bleu (foncé ou non) forment, d'un côté, un grand cercle fermé (autour du domaine d'oïl) et, de l'autre, un petit cercle à demi ouvert (relatif au francoprovençal) et évoquent par là l'idée d'une *irradiation spatiale* de grande envergure, les polygones rouges constituent des agglomérations *ponctuelles* suggérant l'idée d'immobilisme et de défense « tous azimuts ».

En effet, nous sommes en présence d'une analyse-DM très efficace dont l'utilité est avant tout d'ordre *diachronique*. C'est ainsi que le grand agencement circulaire bleu dans le nord évoque l'action irradiatrice continue des dialectes du type francilien, alors que la pince entrouverte bleue autour du domaine francoprovençal renvoie aux contacts et conflits linguistiques que ce géotype entretenait tant vers le nord que vers le sud.

Que le coloriage en bleu foncé – équivalent à des scores négatifs du CAF – soit en relation directe avec l'interaction linguistique tout court, est corroboré par le comportement du seul îlot linguistique compris dans le réseau de l'ALF, à savoir du P. 635 (Andraut, Gironde). L'on y trouve un dialecte oïlique de type saintongeais qui avait été transféré en territoire gascon au cours du XVe siècle. Comme il avait été exposé à toute sorte d'interactions au cours des derniers six siècles, il va de soi qu'il figure en bleu foncé.

Cette analyse-DM permet donc de repérer, au sein d'un réseau donné, des zones linguistiquement « mouvementées » et des régions relativement calmes ou même « inertes ».

Ajoutons encore que les quatre cartes en question reposent sur autant de sous-corpus différenciés du point de vue *quantitatif* et *qualitatif*.

- Pour les cartes 11 et 12 :

Les deux sous-corpus utilisés se distinguent par leur degré de *polynymie*. Alors que la carte 11 repose sur 536 CT dont les structures internes sont relativement simples, la carte 12 est basée sur 588 CT dont les structures sont très complexes. Pour mieux comprendre cet effet, il suffit de jeter un regard sur la figure 2. L'on y perçoit l'emplacement de trois sous-corpus (A, B et C) dont la structuration interne varie entre 2-5 taxats par CT (= sous-corpus *oligo*-nyme[49] A) et 13-90 taxats/CT (= sous-corpus *poikilo*-nyme[50] C).

Entre parenthèses : la carte 1, pourvue de 17 taxats et, partant, d'une structure plutôt complexe, appartient au sous-corpus C, alors que la carte 2, disposant de seulement trois taxats et, de ce fait, d'une structuration très simple, fait partie du sous-corpus A.

[49] Grec *olígos* „petit, peu".

[50] Grec *poikílos* „bigarré, polychrome".

Or, il s'est avéré, au cours de nos analyses-DM, que de la synthèse de beaucoup de CT *oligo*-nymes (= simples ou « claires ») ressortent pratiquement les mêmes structures de profondeur que l'on obtient moyennant la synthèse de beaucoup de CT *poikilo*-nymes (= complexes ou « chaotiques »). Cet effet quelque peu surprenant montre que les structures-DM de profondeur disposent d'une grande redondance interne, comparables en cela au comportement d'autres signaux anthropiques.

Il n'est donc pas étonnant que les cartes 11 et 12 montrent pratiquement la même structuration choroplèthe caractérisée par une circularité « oïlique » et une pince entrouverte « francoprovençale » en bleu foncé dans le nord, et par trois « balouards de résistance » en rouge situés dans le sud.

- Pour les cartes 13 et 14 :

Alors que le profil choroplèthe de la carte 13 – dressé à l'aide de 1 117 CT *phonétiques* – s'aligne grosso modo sur celui des cartes 11 et 12, le profil choroplèthe de la carte 14, reposant sur 465 CT *lexicales*, s'en écarte un peu. L'effet oïlique de circularité y est moins saillant : ce qui ressort avec beaucoup de clarté, c'est, d'un côté, une large bande d'*interaction* et de *compromis linguistique* qui se creuse, d'une façon très compacte, au beau milieu de la Galloromania, et, de l'autre, la réduction de la zone coloriée en rouge, orange et jaune, causée par le détachement du Limousin et de l'Auvergne du bloc occitano-catalan.

11. Structure de profondeur D : la dialectométrie corrélative

Voir les cartes 15-18.

La DM corrélative, qui est une des pièces fortes de l'arsenal méthodique de l'EDMS, existe depuis 2005[51] et a fourni depuis beaucoup de preuves de sa grande utilité heuristique. Le module de VDM relatif à la DM corrélative permet de calculer la corrélation entre deux matrices de similarité (etc.) à dimensions extérieures égales (= deux fois N × N), et de visualiser rapidement le rendement de ces calculs sous la forme d'une carte choroplèthe à couleurs avec N points (ou polygones).

La *nature phénoménologique* des deux variables à corréler peut varier à loisir : l'on peut corréler, entre elles, différentes *catégories linguistiques*, les mettre en relation, l'une après l'autre, avec l'*espace euclidien*, ou comparer, par voie de corrélation, les résultats numériques de différentes *mensurations de similarité*. Ce qui importe au cours de ces opérations statistico-visualisatrices, c'est de ne jamais perdre de vue leur finalité et signification *linguistiques* à proprement parler. Ajoutons encore que la profondeur des structures synthétiques rejointes par la DM corrélative, est très grande.

Nous commençons la description de l'efficacité de la DM corrélative par la présentation d'une corrélation cartographique entre les *similarités linguistiques* d'un côté et les *proximités euclidiennes* de l'autre : voir la carte 15. Le but en est d'étudier de plus

[51] Cf. Goebl (2005, passim).

près les multiples facettes du décalage spatial que subissent les scores de similarité linguistique des différentes cartes de similarité.

En reprenant la leçon donnée par les cartes 5 (= de *similarité*) et 6 (= de *proximité*) il est possible de calculer 641 scores de corrélation – toujours selon le coefficient de corrélation r(BP) – et de les visualiser par la suite suivant les normes habituelles de l'EDMS. Il en résulte la carte 15 qui, d'un côté, repose sur le corpus-ALF intégral (= 1 681 CT) et, de l'autre, sur les proximités euclidiennes calculées selon le théorème de Pythagore.

Le profil choroplèthe de la carte 15 est d'une clarté et régularité exemplaires. Pour bien pouvoir le comprendre en tant que *linguiste*, il faut avant tout connaître la signification *linguistique* de la variabilité *numérique* calculée, et du symbolisme chromatique y ayant trait.

Or, les teintes *chaudes* renvoient à une *harmonie intrinsèque* du déploiement des similarités linguistiques et des proximités euclidiennes dans l'espace, alors que les teintes *froides* signalent de fortes *frictions* et *divergences* entre ces deux « énergies » spatiales. Evidemment, de telles disharmonies sont dues à tout ce qui peut influer de l'extérieur sur l'évolution « naturelle » de la langue dans l'espace, soit en l'accélérant, soit en la retardant. Les facteurs en question sont bien entendu d'ordre *historique*, *politique* et *social*.

Dans cette perspective, la Galloromania se présente comme le théâtre d'une grande confrontation entre le Nord et le Sud où le premier a fini par prendre le dessus tout en déstabilisant graduellement le Sud, en passant par le Massif Central pour arriver aux Pyrénées. La ceinture médiane de la Galloromania, allant de la Gironde à l'ouest jusqu'à l'éventail du francoprovençal à l'est, constitue la zone des affrontements les plus violents.

Une fois de plus, la position de l'îlot linguistique du P.-ALF 635 (Andraut) – où l'on trouve le score-r(BP) minimal – nous fournit la preuve du bien fondé de notre supposition selon laquelle la divergence de fond entre les gestions *basilectale* et *euclidienne* de l'espace crée les scores de corrélation les plus bas et, partant, des coloriages en bleu foncé. Rappelons que la « Petite Gavacherie » dont Andraut (= P.-ALF 635) fait partie, avait été fondée au XVe siècle en territoire d'oc comme colonie linguistique d'oïl d'origine saintongeaise. Il va de soi que la fondation d'un îlot linguistique en terre linguistiquement étrangère constitue une violation éclatante de la relation « naturelle » entre les gestions *basilectale* et *euclidienne* de l'espace.

Évidemment, il est possible de répéter le même exploit-DM pour la *phonétique* et l'*espace* d'un côté (voir la carte 16), ainsi que pour le *lexique* et l'*espace* de l'autre (voir la carte 17). Les plus grandes divergences entre les profils choroplèthes des deux cartes à corrélations se situent en Provence, alors que pour le reste de la Galloromania, la distribution spatiale des corrélations entre les gestions linguistique et euclidienne de l'espace reste sensiblement la même.

Selon toute vraisemblance, l'allure éminemment régulière des cartes 15-17 est due, entre autres, à la grande ancienneté des divergences qui se sont manifestées, au cours des siècles, entre les gestions *linguistique* et *euclidienne* de la Galloromania.

La dernière des quatre cartes corrélatives que nous venons de présenter, se réfère à un problème éminemment *grammatical* : voir la carte 18. Il s'agit de savoir dans quelle mesure la gestion de l'espace moyennant deux *catégories linguistiques différentes* s'annonce comme convergente ou divergente.

Les mensurations et les visualisations y ayant trait fournissent un profil choroplèthe très clair : les gestions *lexicale* et *phonétique* divergent considérablement le long d'un couloir « delphino-forme » allant de la Gironde aux Alpes alors qu'elles convergent tant au nord qu'au sud de la Galloromania. À remarquer, une fois de plus, l'allure très régulière du profil choroplèthe dans son ensemble.

Ajoutons encore que le couloir *delphino*-forme mentionné ci-dessus a fait son apparition dans beaucoup d'autres analyses-DM relatives à l'ALF. Il est cependant bon de ne pas le confondre avec le tracé du « croissant » de mémoire ronjatienne. Ce dernier se situe plus au nord et dispose d'un contour beaucoup plus mince[52].

Précisons en outre que l'application de la DM corrélative à d'autres domaines linguistiques – romans ou autres – a fourni également des résultats très clairs et bien structurés : cf. à ce propos Goebl (2005, passim, et 2012, 155 sq.).

12. Structure de profondeur E : l'analyse dendrographique

Voir les cartes 19-20.

Le niveau E de l'analyse-DM des structures de profondeur concerne la DM dendrographique ou – dans la terminologie technique de la classification numérique (ou taxométrie) – la « classification ascendante hiérarchique » (CAH). La CAH est largement employée au sein de beaucoup de sciences naturelles et humaines. Elle sert au groupement algorithmique de N objets – définis, chacun, par p attributs – en structures hiérarchiques pourvues de ramifications binaires. Du point de vue statistique, il s'agit d'une analyse dite « multivariée ». Dans la chaîne-DM, la CAH ne représente cependant qu'une seule des nombreuses possibilités de dépouillement de la matrice de similarité (voir la figure 3). Vu la longue tradition de la *pensée arborescente* au sein de pratiquement toutes les branches de la linguistique, ce genre d'analyse offre un certain attrait « naturel » aussi pour les besoins de la géolinguistique.

Les algorithmes à utiliser (ou utilisables) dans le cadre de la CAH sont très variés. Ils opèrent, tout en exécutant des impératifs mathématiques préalablement définis, des fusions (ou agglomérations) binaires entre les N objets de la matrice de similarité en question. En général, leur action agglomératrice débute par les différences *mini-*

[52] Pour l'historique du *croissant*, voir le travail séminal de Guylaine Brun-Trigaud de 1990 : pour les contours géographiques traditionnels du « croissant » tel qu'il avait été défini en 1913 par Jules Ronjat (1864-1925), voir la carte à la page 274 (ibidem).

males qui existent entre les N^2 valeurs de la matrice de similarité pour en rejoindre les différences *maximales* : d'où le qualificatif « ascendant » de la CAH. Quant au choix de l'algorithme-CAH le plus approprié aux besoins de la DM, il n'y a pas de solution fixe ou « passe-partout ». C'est pourquoi le dialectométricien doit sélectionner, dans l'arsenal algorithmique de la CAH, la procédure dont les résultats correspondent le plus à ses propres besoins et visées.

Un des algorithmes-CAH qui répond le mieux à ces impératifs est celui qui a été proposé, en 1963, par le statisticien américain Joe Ward, Jr. Il fournit, en ligne générale, d'excellents résultats quand il est appliqué à des données de nature géolinguistique. Un des critères pour déterminer l'utilité *linguistique* d'un algorithme-CAH est la cohérence spatiale des différents clusters (= « chorèmes ») extraits de l'arbre calculé.

Évidemment, l'arbre en tant que configuration ramifiée ne saurait jamais suffire aux besoins d'un *géolinguiste* épris de schémas *cartographiques* et, par conséquent, *spatiaux*. Il faut donc projeter le rendement classificatoire de l'arbre dans l'espace. Ce transfert est nommé« spatialisation ». C'est ainsi que chacun des branchages de l'arbre (dénommés par nous „dendrèmes") retenus comme linguistiquement « significatifs », est transposé dans le secteur approprié du réseau en question qui, lui, est appelé à Salzbourg « chorème ». Inutile d'ajouter que le coloriage des dendrèmes correspond parfaitement à celui des chorèmes.

Le jeu alternatif des différents chorèmes disséminés dans l'espace établit donc une excellente classification des locolectes examinés, alors que le message hiérarchique concomitant de l'arbre permet d'y ajouter des nuances quantitatives – et aussi diachroniques – supplémentaires.

Ceci est dû au fait que tout arbre issu d'une CAH peut être utilisé également à une *reconstruction diachronique* telle qu'elle est conçue et pratiquée depuis belle lurette par la *lexicostatistique* et la *glottochronologie*[53]. La logique de cette perspective veut que la ramification croissante à partir de la racine de l'arbre corresponde au déploiement diachronique de la fragmentation linguistique du territoire en question.

Jetons à ce propos un coup d'œil sur la carte 19 : les deux embranchements de la première bifurcation correspondent, en effet, aux domaines d'oc (= dendrèmes/chorèmes 1-4) et d'oïl (= dendrèmes/chorèmes 5-8). La bifurcation suivante – toujours selon la logique de l'algorithme de Ward – engendre la séparation entre le francoprovençal (= dendrème/chorème 5) et le reste du domaine d'oïl (= dendrèmes/chorèmes 6-8). Evidemment, il est possible de continuer ce genre d'analyse jusqu'au niveau des feuilles. VDM en offre les possibilités techniques nécessaires.

Une autre particularité – d'ordre statistique – mérite également d'être mentionnée. C'est que les différents dendrèmes (et chorèmes) doivent leur genèse à la considération serrée de leur *variabilité quantitative interne*. C'est ainsi que naît une certaine hiérarchisation des dendrèmes/chorèmes en fonction de leur hétérogénéité quantita-

[53] Cf. à ce sujet Goebl (2003, 84-89).

tive interne (et aussi réciproque) : plus un dendrème est proche de la racine de l'arbre, plus grande en est l'hétérogénéité interne. Et vice versa.

Ceci signifie que, sur la carte 19, la variabilité numérique interne du dendrème/chorème relatif au domaine d'oïl (= dendrèmes/chorèmes 5-8) est moins grande que celle du dendrème/chorème relatif au domaine d'oc (= dendrèmes/chorèmes 1-4)[54].

Soulignons encore que les cartes 19 et 20 se distinguent par leurs corpus de base : la carte 19 repose sur notre sous-corpus *phonétique* (avec 1 117 CT) alors que la carte 20 émane d'une matrice de données *lexicale* (avec 465 CT). Il va de soi que les deux arborescences et leurs spatialisations ne sont pas complètement identiques. Ceci vaut surtout pour l'agencement (relationnel) des huit dendrèmes à l'*intérieur* des arbres correspondants.

Pour en faciliter la comparaison, nous avons conféré, sur les cartes 19 et 20, le même coloriage (et numérotation) aux chorèmes placés dans une situation géographique comparable, d'où il résulte cependant une numérotation non plus continue (c'est-à dire allant de 1 à 8) mais discontinue des différents dendrèmes de la carte 20. L'on peut en déduire que la première bifurcation aboutit, sur les deux cartes, aux mêmes résultats (= nette séparation entre Oc et Oïl), certes, mais aussi que la hiérarchisation des huit dendrèmes/chorèmes coloriés à l'intérieur de l'arbre de la carte 20 se distingue quelque peu de la situation analogue sur la carte 19. Toujours est-il que la cohérence spatiale des chorèmes telle qu'elle se manifeste près de la racine, persiste aussi au *milieu* de l'arbre et près des *feuilles*. Ceci prouve le haut degré de « classificabilité » de nos données.

13. Epilogue

Le fondement théorique de la DM de Salzbourg est l'hypothèse de l'existence de la gestion basilectale de l'espace par l'homo loquens dont les manifestations réelles – toutes saisissables par le biais d'un atlas linguistique de type ALF, AIS, ALPI ou ALDC – obéissent à certaines régularités voire lois générales qui ne peuvent être découvertes qu'à travers un certain nombre d'opérations *inductives* ou de *synthèse*. Il importe que les résultats de ces opérations de synthèse soient dûment visualisés afin qu'ils puissent déclencher les opérations mentales voulues.

Il est en outre très important de ne pas oublier qu'il y a, en dehors de la géolinguistique à proprement parler, de nombreuses sciences étroitement liées avec l'espace dont les méthodes géo-classificatoires sont très poches de celle de la DM, et avec lesquelles une coopération interdisciplinaire s'impose. Nous n'en citons ici que la *géographie humaine* et la *génétique des populations*[55]. Dans la rétrospective, il est bien

[54] La raison en est que la racine du méga-dendrème 5-8 est plus proche du tronc de l'arbre que celle du méga-dendrème 1-4.

[55] Citons, à titre d'exemple, la contribution interdisciplinaire de Scapoli / Goebl / Sobota / Mamolini / Barrai de 2005.

évident que de telles disponibilités – ou faudrait-il dire « velléités » – interdisciplinaires existaient déjà du temps de Coquebert de Montbret père et fils.

Il est cependant non moins évident que – du point de vue *empirique* et *méthodique* – un fil conducteur rouge s'étend de leurs activités à la DM actuelle.

Université de Salzbourg Hans GOEBL

Remerciements

- Préparation et confection de tous les clichés de cet article : Werner Goebl (Vienne)
- Contrôle linguistique de mon texte français : Stéphanie Lescure (Marbourg, Hesse)

Que mon fils Werner et Mme Lescure reçoivent ici l'expression de ma sincère et profonde gratitude.

Abréviations souvent utilisées

AL	atlas linguistique
CAF	coefficient d'asymétrie de [Ronald A.] Fisher
CAH	classification ascendante hiérarchique
CT	carte de travail
DM	dialectométrie, relative à la dialectométrie, dialectométrique
EDMS	Ecole dialectométrique de Salzbourg
IRD_{jk}	Indice Relatif de Distance (entre les points d'atlas j et k)
IRI_{jk}	Indice Relatif d'Identité (entre les points d'atlas j et k)
MA	moyenne arithmétique
MEDMW	algorithme d'intervallisation (centré sur la MA, avec la formation, des deux côtés de la MA, de *n* classes égales en nombre de points d'atlas)
MINMWMAX	algorithme d'intervallisation (centré sur la MA, avec la formation, des deux côtés de la MA, de *n* classes dont la largeur numérique est égale)
r(BP)	coefficient de corrélation de [Auguste] Bravais et [Karl] Pearson
VDM	Visual DialectoMetry (logiciel dialectométrique utilisé à Salzbourg ; librement disponible sur demande)

Références bibliographiques

AIS = Jaberg, Karl / Jud, Jakob, 1928-1940. *Sprach- und Sachatlas Italiens und der Südschweiz*, 8 vol., Zofingen, Ringier (réimpression : Nendeln, Kraus 1971).

ALDC = Veny, Joan / Pons i Griera, Lídia, 2001-2014. *Atles Lingüístic del Domini Català*, 7 vol., Barcelona, Institut d'Estudis Catalans.

ALF = Gilliéron, Jules / Edmont, Edmond, 1902-1910. *Atlas linguistique de la France*, 10 vol., Paris, Champion (réimpression : Bologne, Forni 1968).

ALPI = *Atlas lingüístico de la Península Ibérica*, 1962, vol. I : Fonética, Madrid, CSIC.

Altmann, Gabriel, 1985. « Die Entstehung diatopischer Varianten. Ein stochastisches Modell ». *Zeitschrift für Sprachwissenschaft* 4, 139-155.

Baehr, Rudolf, 1958. « Zu den romanischen Wochentagsnamen », in : Lausberg, Heinrich / Weinrich, Harald (ed.), *Romanica. Festschrift für Gerhard Rohlfs*, Halle, Niemeyer, 26-56.

Berschin, Helmut / Felixberger, Josef / Goebl, Hans, 2008[2]. *Französische Sprachgeschichte. Lateinische Basis, interne und externe Geschichte, sprachliche Gliederung Frankreichs. Mit einer Einführung in die historische Sprachwissenschaft*, Hildesheim, Olms.

Bock, Hans Hermann, 1974. *Automatische Klassifikation. Theoretische und praktische Methoden zur Gruppierung und Strukturierung von Daten (Cluster-Analyse)*, Göttingen, Vandenhoeck & Ruprecht.

Bord, Jean-Paul, 2012. *L'univers des cartes. La carte et le cartographe*, Paris, Belin.

Bourguet, Marie-Noëlle, 1988. *Déchiffrer la France. La statistique départementale à l'époque napoléonienne*, Paris, Editions des Archives Contemporaines.

Brunet, Roger, 1987. *La carte. Mode d'emploi, Paris*, Fayard.

Brun-Trigaud, Guylaine, 1990. *Le croissant : le concept et le mot. Contribution à l'histoire de la dialectologie française au XIXe siècle*, Lyon, Centre d'études linguistiques Jacques Goudet.

Brun-Trigaud, Guylaine / Le Berre, Yves / Le Dû, Jean, 2005. *Lectures de l'Atlas linguistique de la France de Gilliéron et Edmont. Du temps dans l'espace. Essai d'interprétation des cartes de l'Atlas linguistique de la France de Jules Gilliéron et Edmond Edmont augmenté de quelques cartes de l'Atlas linguistique de la Basse Bretagne de Pierre Le Roux*, Paris, CTHS.

Chartier, Roger, 1992. « La ligne Saint-Malo – Genève », in : Nora, Pierre (ed.), *Les lieux de mémoire, vol. 3 : Les France, tome 1 : Conflits et partages*, Paris, Gallimard, 738-775.

Ettmayer, Karl von, 1924. « Das Wesen der Dialektbildung, erläutert an den Dialekten Frankreichs », *Denkschriften* [= Mémoires] *der Akademie der Wissenschaften in Wien*, philosophisch-historische Klasse, 66/3, 1-56, 7 Tafeln [= tableaux].

FEW = Wartburg, Walther von (1922/1928 sq.). *Französisches etymologisches Wörterbuch*, Bonn, Bouvier.

Gilliéron, Jules, 1918. *Généalogie des mots qui désignent l'abeille d'après l'Atlas linguistique de la France,* Paris, Champion.

Gilliéron, Jules / Roques, Mario, 1912. *Etudes de géographie linguistique d'après l'Atlas linguistique de la France*, Paris, Champion.

Goebl, Hans, 1976. « La dialectométrie appliquée à L'ALF (Normandie) », in : Vàrvaro, Alberto (ed.), *Atti del XVI Congresso Internazionale di Linguistica e Filologia Romanza* (Naples 1974), Naples, Marchiaroli/Amsterdam, Benjamins, vol. 2, 165-195.

Goebl, Hans, 1981. « Eléments d'analyse dialectométrique (avec application à l'AIS) », *RLiR* 45, 349-420.

Goebl, Hans, 1983. « Parquet polygonal et treillis triangulaire. Les deux versants de la dialectométrie interponctuelle », *RLiR* 47, 353-412.

Goebl, Hans, 1984. *Dialektometrische Studien. Anhand italoromanischer, rätoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF*, Tübingen, Niemeyer, 3 vol.

Goebl, Hans, 2000. « La dialectométrisation de l'ALF : présentation des premiers résultats », *Linguistica* 40, 209-236.

Goebl, Hans, 2002a. « Analyse dialectométrique des structures de profondeur de l'ALF », *RLiR* 66, 5-63.

Goebl Hans, 2002b. « Sprachatlanten : woher? womit? wozu? Einige buntgemischte Reflexionen am Gartenzaun zwischen Romanistik und Germanistik », in : Anreiter, Peter / Ernst, Peter / Hausner, Isolde / Kalb, Helmut (ed.), *Namen, Sprachen und Kulturen. Imena, Jeziki in Kulture. Festschrift für Heinz Dieter Pohl zum 60. Geburtstag*, Vienne, Edition Präsens, 257-274.

Goebl, Hans, 2003. « Regards dialectométriques sur les données de l'Atlas linguistique de la France (ALF) : relations quantitatives et structures de profondeur », *Estudis Romànics* 25, 59-96.

Goebl, Hans, 2004. « VDM – Visual DialectoMetry. Vorstellung eines dialektometrischen Software-Pakets auf CD-ROM (mit Beispielen zu ALF und Dees 1980) », in : Dahmen, Wolfgang / Holtus, Günter / Kramer, Johannes / Metzeltin, Michael / Schweickard, Wolfgang / Winkelmann, Otto (ed.), *Romanistik und neue Medien*. Romanistisches Kolloquium XVI (Rauischholzhausen, 2000), Tübingen, Narr, 209-241.

Goebl, Hans, 2005. « La dialectométrie corrélative : un nouvel outil pour l'étude de l'aménagement dialectal de l'espace par l'homme », *RLiR* 69, 321-367.

Goebl, Hans, 2006. « Warum die Dialektometrie nur in einem roman(ist)ischen Forschungskontext entstehen konnte", in : Dahmen, Wolfgang / Holtus, Günter / Kramer, Johannes / Metzeltin, Michael / Schweickard, Wolfgang / Winkelmann, Otto (ed.), *Was kann eine vergleichende romanische Sprachwissenschaft heute (noch) leisten*. Romanistisches Kolloquium XX (Göttingen, 2004), Tübingen, Narr, 291-317.

Goebl, Hans, 2008. « La dialettometrizzazione integrale dell'AIS. Presentazione dei primi risultati », *RLiR* 72, 25-113.

Goebl, Hans, 2011. « Quo vadis, atlas linguistice? Einige wissenschaftshistorische und zeitgeistkritische Reflexionen zur atlasgestützten Geolinguistik », in : Schlaak, Claudia / Busse, Lena (ed.), *Sprachkontakte, Sprachvariation und Sprachwandel. Festschrift für Thomas Stehl zum 60. Geburtstag*, Tübingen, Narr, 5-27.

Goebl, Hans, 2012. « Introduction aux problèmes et méthodes de l' ‹Ecole dialectométrique de Salzbourg› (avec des exemples gallo-, italo- et ibéroromans) », in : Álvarez Pérez, Xosé Afonso / Carrilho, Ernestina / Magro, Catarina (ed.), *Proceedings of the International Symposium on Limits and Areas in Dialectology (LimiAr)*, Lisbonne, Centro de Linguística da Universidade de Lisboa, 117-166 (dans le web : <limiar.clul.ul.pt/proceedings/general_session/7_goebl.pdf>).

Goebl, Hans / Scherrer, Yves / Smečka, Pavel, 2013. « Kurzbericht über die Dialektometrisierung des Gesamtnetzes des ‹Sprachatlasses der deutschen Schweiz› (SDS) », in : Schneider-Wiejowski, Karina / Kellermeier-Rehbein, Birte / Haselhuber, Jakob (ed.), *Vielfalt, Variation und Stellung der deutschen Sprache* [Mélanges pour le 70e anniversaire de Ulrich Ammon], Berlin/Boston, Walter de Gruyter, 153-175.

Goebl, Hans / Schiltz, Guillaume, 1997. « A Dialectometrical Compilation of CLAE 1 and CLAE 2. Isoglosses and Dialect Integration », in : Viereck, Wolfgang / Ramisch, Heinrich (ed.), *The Computer Developed Linguistic Atlas of England*, vol. 2, Tübingen, Niemeyer, 13-21.

Goebl, Hans / Winterleitner, Norbert, 1971. « Projekt einer sprachstatistischen Auswertung von in Sprachatlanten gespeicherter linguistischer Information mit Hilfe elektronischer Rechenanlagen », *Linguistische Berichte* 14, 60-61.

Herrgen, Joachim, 2001. « Die Dialektologie des Deutschen », in : Auroux, Sylvain / Koerner, E. F. Konrad / Niederehe, Hans-Josef / Versteegh, Kees (ed.), *History of the Language Sciences / Geschichte der Sprachwissenschaften / Histoire des sciences du langage. An International Handbook on the Evolution of the Study of Language from the Beginning to the Present / Ein internationales Handbuch zur Entwicklung der Sprachforschung von den Anfängen bis zur Gegenwart / Manuel international sur l'évolution de l'étude du langage des origines à nos jours*, Berlin/New York, de Gruyter, vol. 2, 1513-1535.

Jaberg, Karl, 1908. *Sprachgeographie. Beitrag zum Verständnis des Atlas linguistique de la France* Aarau, Sauerländer (version espagnole : *Geografía lingüística. Ensayo de interpretación del « Atlas lingüístico de Francia »*. Traducción de Antonio Llorente y Manuel Alvar, Granada, Universidad de Granada, 1959).

Jochnowitz, George, 1973. *Dialect Boundaries and the Question of Franco-Provençal*, La Haye, Mouton.

Ködel, Sven, 2014. *Die Enquête Coquebert de Montbret (1806-1812). Die Sprachen und Dialekte Frankreichs und die Wahrnehmung der französischen Sprachlandschaft während des Ersten Kaiserreichs,* Bamberg.

Laboulais-Lesage, Isabelle, 1999. *Lectures et pratiques de l'espace. L'Itinéraire de Coquebert de Montbret, savant et grand commis d'Etat (1755-1831)*, Paris, Champion.

Ozouf-Marignier, Marie-Vic, 1989. *La formation des départements : la représentation du territoire français à la fin du 18ème siècle*, Paris, Editions de l'Ecole des Hautes Etudes en Sciences Sociales.

Palsky, Gilles, 1996. *Des chiffres et des cartes. Naissance et développement de la cartographie quantitative française au XIXe siècle*, Paris, CTHS.

Pop, Sever, 1950. *La dialectologie. Aperçu historique et méthodes d'enquêtes linguistiques*, Louvain, Chez l'auteur/Gembloux, Duculot.

Ravier, Xavier, 1976. « Jean Séguy et la traversée du langage gascon. Réflexions sur une topogenèse géolinguistique », *RLiR* 40, 389-402.

Rohlfs, Gerhard, 1949. « Les noms des jours de la semaine dans les langues romanes », in : *Miscelânea de filologia, literatura e história cultural, à memória de Francisco Adolfo Coelho (1847-1919)*, Lisbonne, Centro de Estudos Filológicos, 88-94.

Rosenqvist, Arvid, 1919. « Limites administratives et division dialectale de la France », *Neuphilologische Mitteilungen* 20, 87-119 (carte publiée à nouveau in : Berschin/Felixberger/Goebl 2008², 261).

Scapoli, Chiara / Goebl, Hans / Sobota, Slawomir / Mamolini, Elisabetta / Rodriguez-Larralde, Alvaro / Barrai, Italo, 2005. « Surnames and Dialects in France : Population structure and cultural evolution », *Journal of Theoretical Biology* 237, 75-86.

Séguy, Jean, 1971. « La relation entre la distance spatiale et la distance lexicale », *RLiR* 35, 335-357.

Séguy, Jean, 1973. « La dialectométrie dans l'Atlas linguistique de la Gascogne », *RLiR* 37, 1-24.

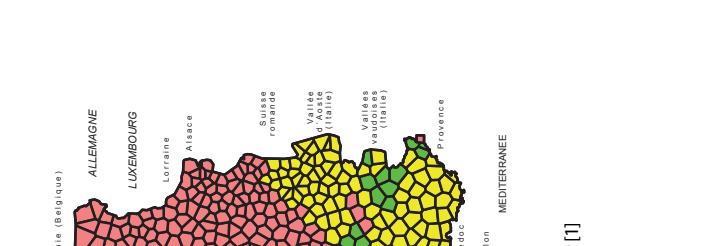
Sneath, Peter H. A. / Sokal, Robert R., 1973. *Numerical Taxonomy. The Principles and Practice of Numerical Classification*, San Francisco, W. H. Freeman.

Storost, Jürgen, 1994. «Noch einmal zur Diez-Stiftung», *RJb* 45, 74-84.

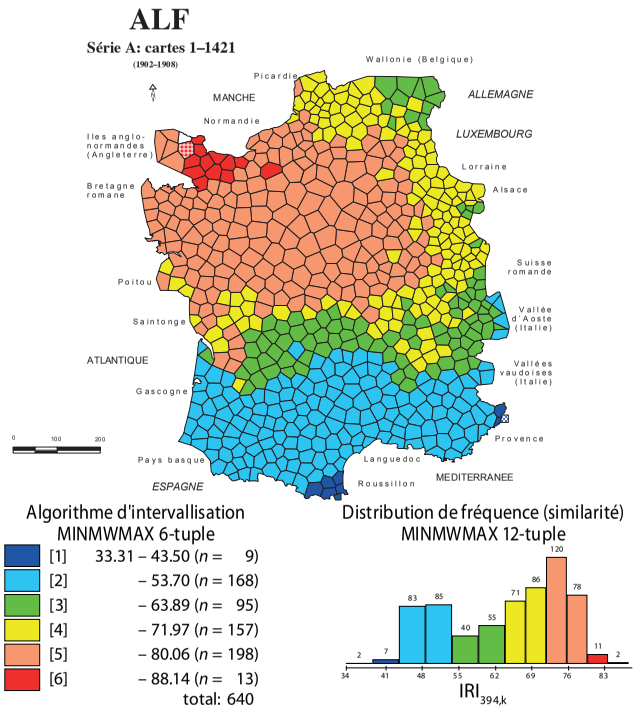
Tuaillon, Gaston, 1983. « Les atlas linguistiques régionaux de France », *BALI* III Serie, Dispensa n. 7, 68-89.

Wartburg, Walther von, 1949. «Los nombres de los días de la semana», *RFE* 33, 1-14.

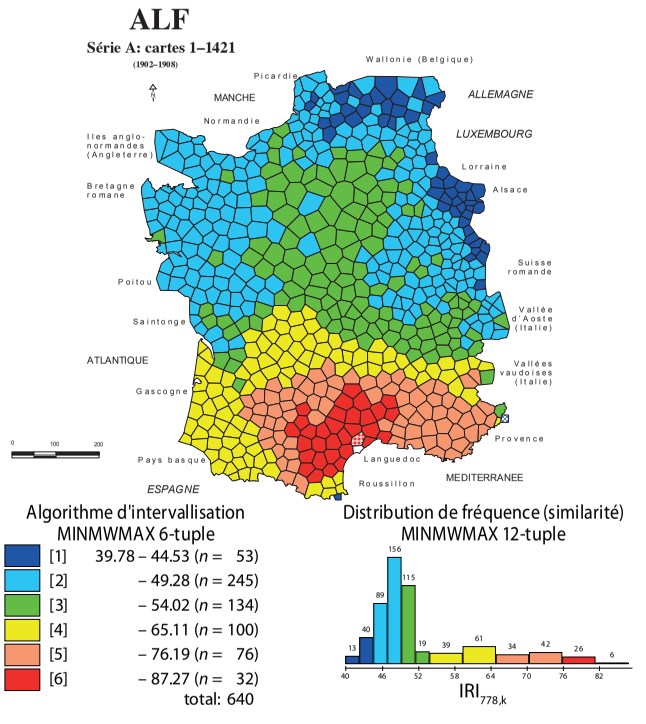
Carte 1: Exemple d'une carte de travail 17-nyme de nature *phonétique* (selon ALF 812 *marché*) : les succédanés du nexus latin -ÁTU



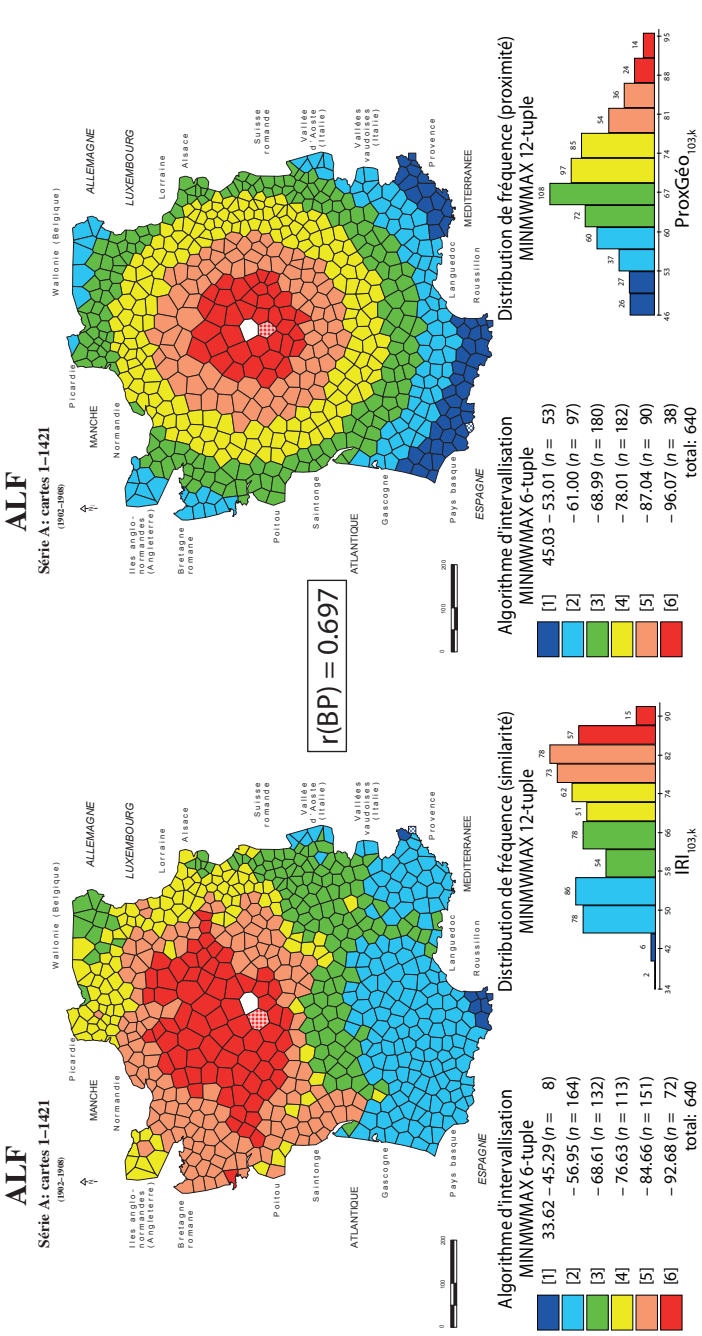
Carte 2: Exemple d'une carte de travail 3-nyme de nature *lexicale* (selon ALF 839 *mercredi*) : les réalisations lexicales du concept « mercredi »



Carte 3: Profil choroplèthe d'une carte de similarité relative au point-ALF 394 (Auderville, Manche)
Corpus: corpus intégral (1 681 cartes de travail)
Indice de similarité: IRI_{jk}
Algorithme d'intervallisation: MINMWMAX 6-tuple

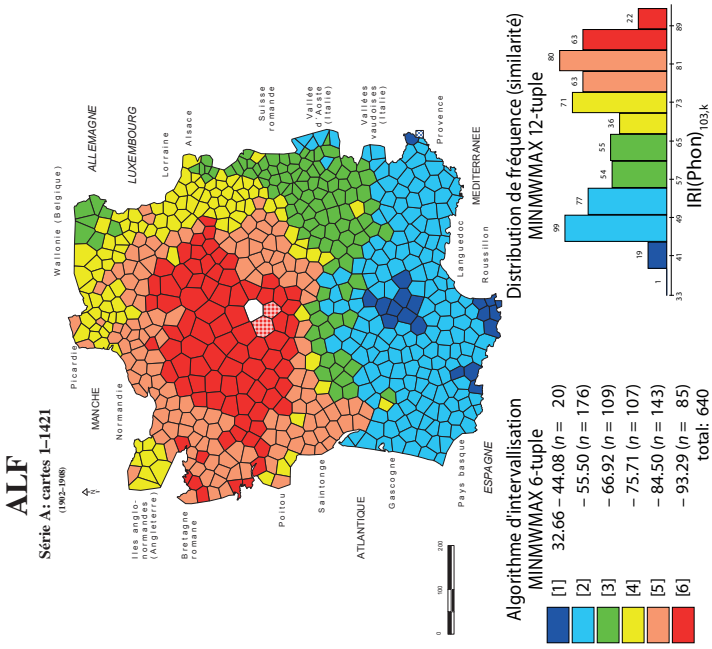


Carte 4: Profil choroplèthe d'une carte de similarité relative au point-ALF 778 (Agde, Hérault)
Corpus: corpus intégral (1 681 cartes de travail)
Indice de similarité: IRI_{jk}
Algorithme d'intervallisation: MINMWMAX 6-tuple

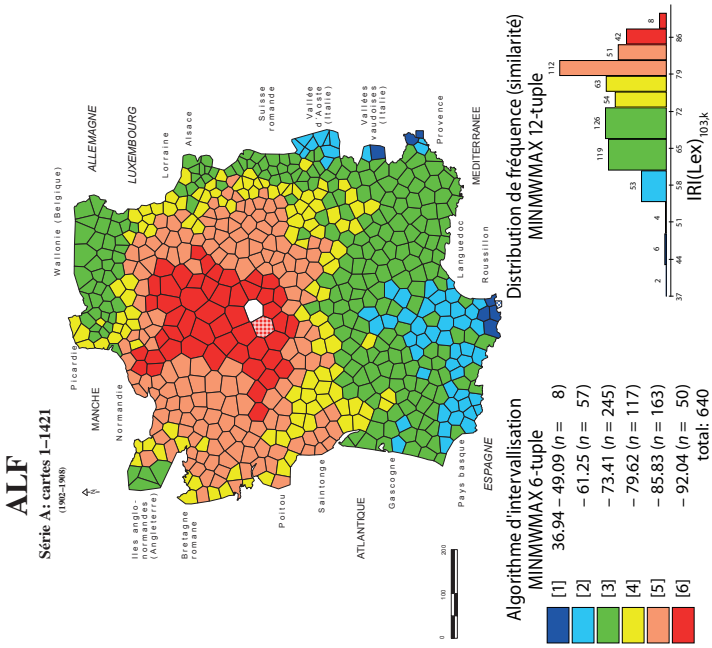


Carte 5: Profil choroplèthe d'une carte de similarité relative au point-ALF 103 (Bué, Sancerre, Cher)
Corpus: corpus intégral (1 681 cartes de travail)
Indice de similarité: IRI_{jk}
Algorithme d'intervallisation: MINMWMAX 6-tuple

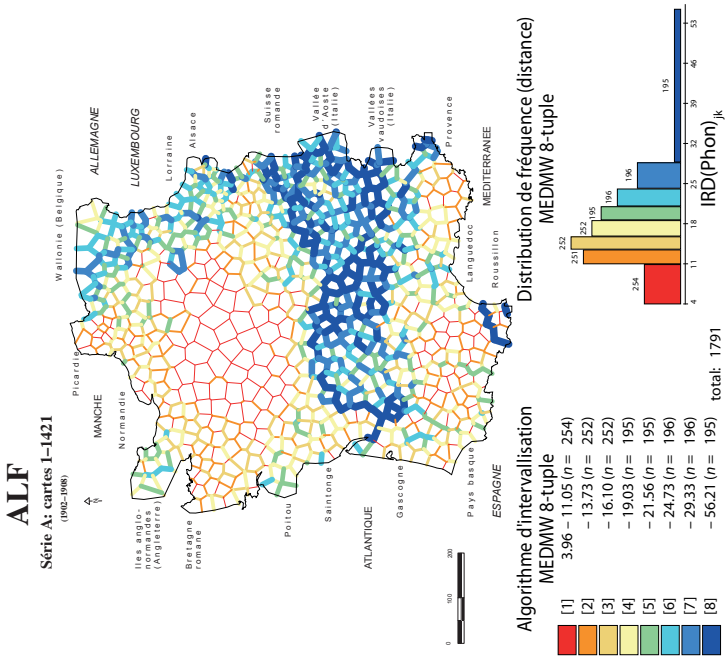
Carte 6: Profil choroplèthe d'une carte de proximité relative au point-ALF 103 (Bué, Sancerre, Cher)
Indice de proximité: théorème de Pythagore
Algorithme d'intervallisation: MINMWMAX 6-tuple



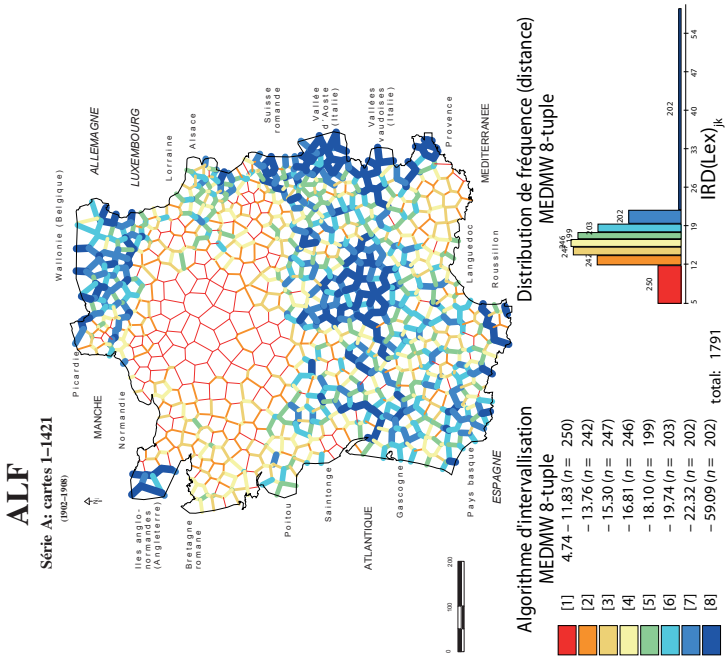
Carte 7: Profil choroplèthe d'une carte de similarité relative au point-ALF 103 (Bué, Sancerre, Cher)
Corpus: sous-corpus phonétique (1 117 cartes de travail)
Indice de similarité: IRI_{jk}
Algorithme d'intervallisation: MINMWMAX 6-tuple



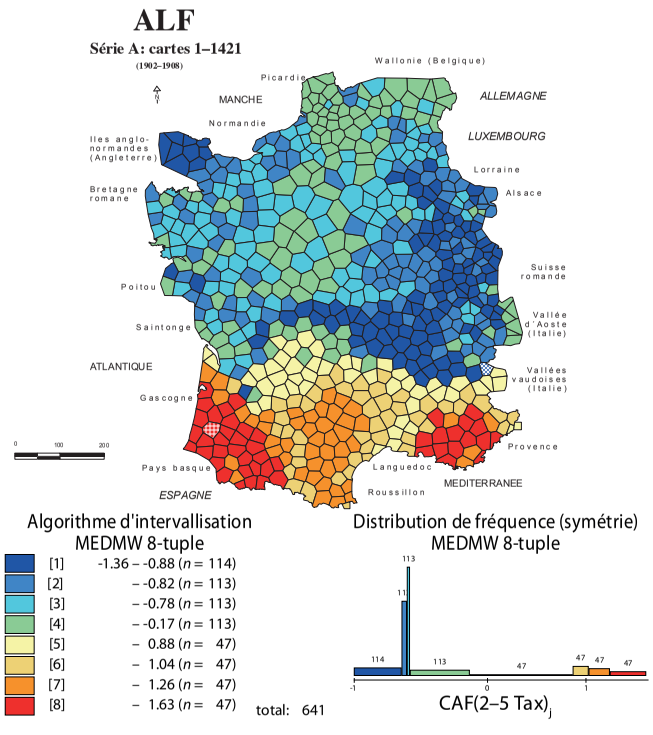
Carte 8: Profil choroplèthe d'une carte de similarité relative au point-ALF 103 (Bué, Sancerre, Cher)
Corpus: sous-corpus lexical (465 cartes de travail)
Indice de similarité: IRI_{jk}
Algorithme d'intervallisation: MINMWMAX 6-tuple



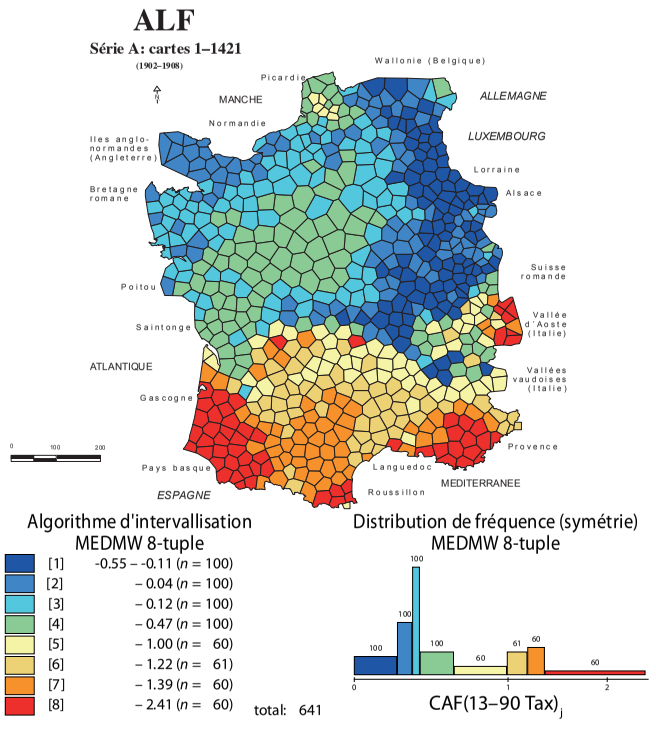
Carte 9: Analyse interponctuelle (carte à cloisons)
Corpus: sous-corpus phonétique (1 117 cartes de travail)
Indice de distance: IRD_{jk}
Algorithme d'intervallisation: MEDMW 8-tuple



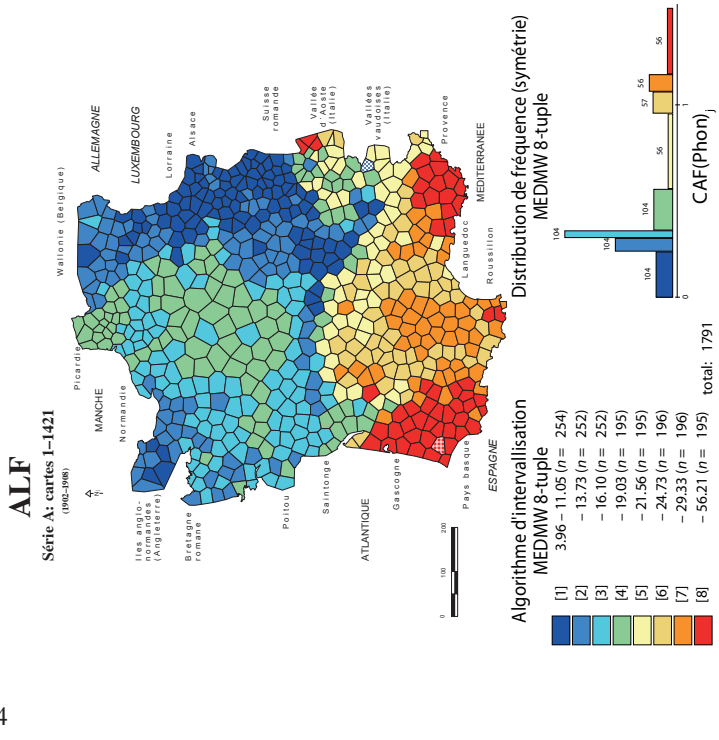
Carte 10: Analyse interponctuelle (carte à cloisons)
Corpus: sous-corpus lexical (465 cartes de travail)
Indice de distance: IRD_{jk}
Algorithme d'intervallisation: MEDMW 8-tuple



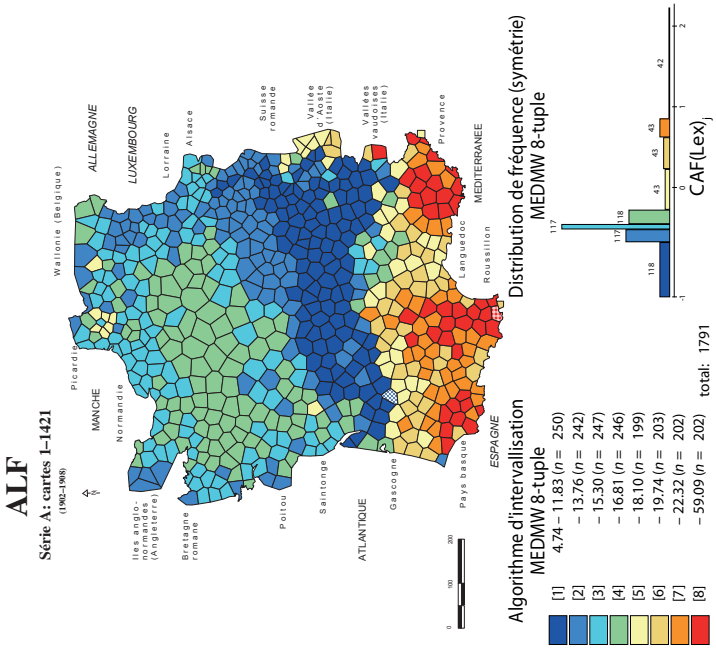
Carte 11: Synopse des scores du « coefficient d'asymétrie de Fisher » (CAF)
Corpus: sous-corpus A (entre 2 et 5 taxats par carte de travail): 536 cartes de travail
Indice de similarité: IRI_{jk}
Algorithme d'intervallisation: MEDMW 8-tuple



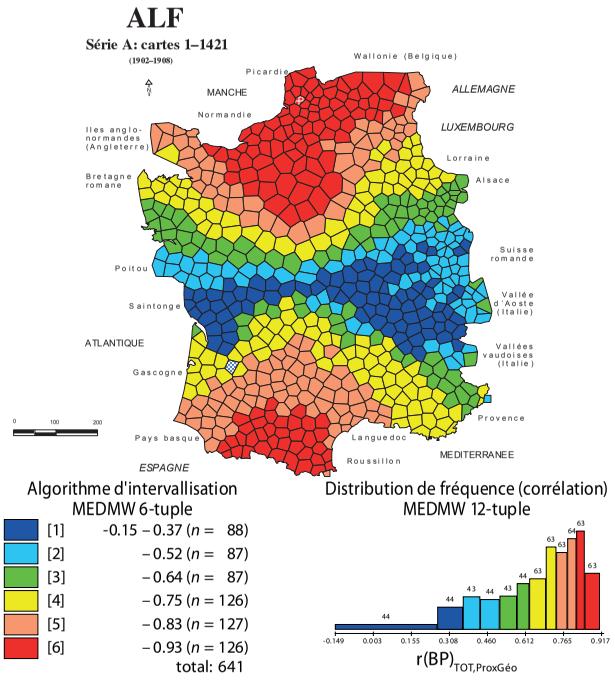
Carte 12: Synopse des scores du « coefficient d'asymétrie de Fisher » (CAF)
Corpus: sous-corpus C (entre 13 et 91 taxats par carte de travail): 588 cartes de travail
Indice de similarité: IRI_{jk}
Algorithme d'intervallisation: MEDMW 8-tuple



Carte 13: Synopse des scores du « coefficient d'asymétrie de Fisher » (CAF)
Corpus: sous-corpus phonétique (1 117 cartes de travail)
Indice de similarité: IRI_{jk}
Algorithme d'intervallisation: MEDMW 8-tuple

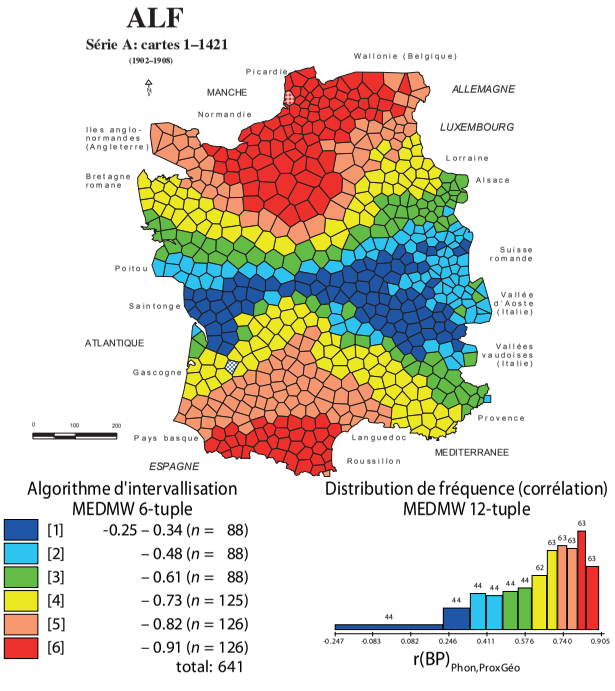


Carte 14: Synopse des scores du « coefficient d'asymétrie de Fisher » (CAF)
Corpus: sous-corpus lexical (465 cartes de travail)
Indice de similarité: IRI_{jk}
Algorithme d'intervallisation: MEDMW 8-tuple



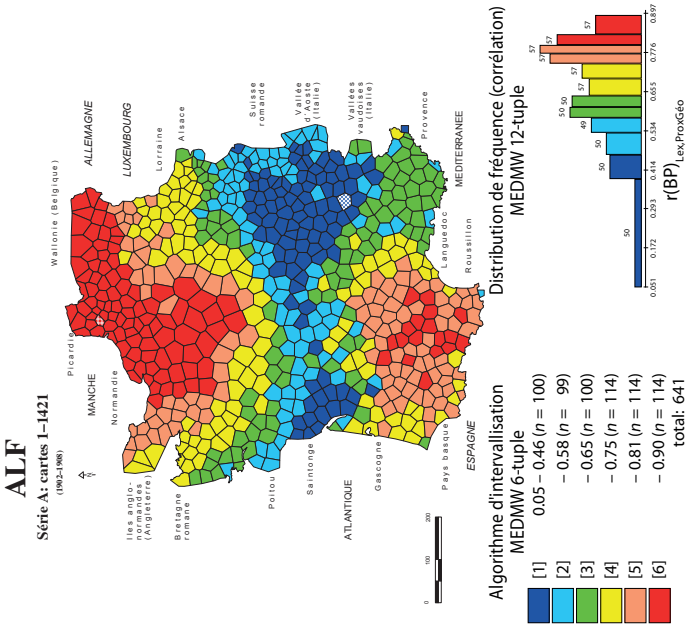
Carte 15: Profil choroplèthe d'une carte à corrélations
Matrices carrées corrélées:
- matrice de similarité (selon IRI_{jk})
  corpus intégral (1 681 cartes de travail)
- matrice de proximité (selon le théorème de Pythagore)

Algorithme d'intervallisation: MEDMW 6-tuple

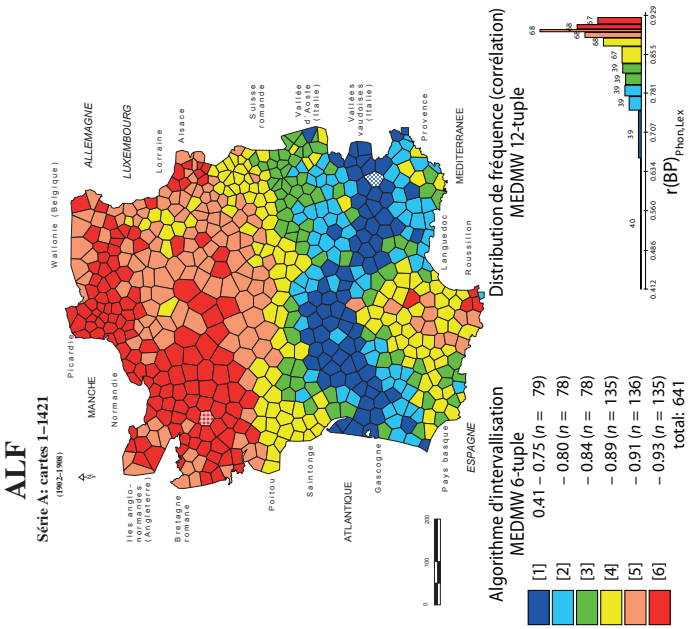


Carte 16: Profil choroplèthe d'une carte à corrélations
Matrices carrées corrélées:
- matrice de similarité (selon IRI_{jk})
  sous-corpus phonétique (1 117 cartes de travail)
- matrice de proximité (selon le théorème de Pythagore)

Algorithme d'intervallisation: MEDMW 6-tuple

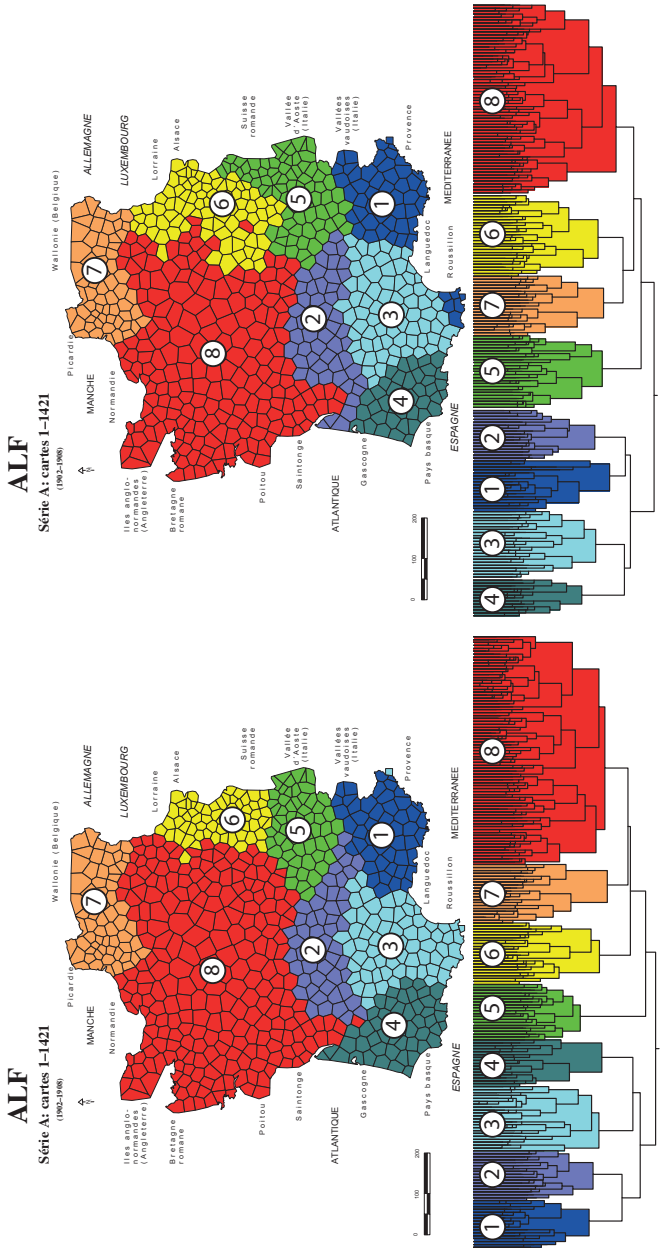


Carte 17: Profil choroplèthe d'une carte à corrélations
Matrices carrées corrélées:
- matrice de similarité (selon IRI_{jk})
  sous-corpus lexical (465 cartes de travail)
- matrice de proximité (selon le théorème de Pythagore)
  Algorithme d'intervallisation: MEDMW 6-tuple



Carte 18: Profil choroplèthe d'une carte à corrélations
Matrice de similarité (selon IRI_{jk}) corrélées:
- sous-corpus phonétique (1 117 cartes de travail)
- sous-corpus lexical (465 cartes de travail)

Algorithme d'intervallisation: MEDMW 6-tuple



Carte 19: Analyse dendrographique (selon l'algorithme de Joe Ward, Jr.)
- En bas: arbre (classification en 8 dendrèmes)
- En haut: spatialisation de l'arbre (classification en 8 chorèmes)

Indice de similarité: IR_{ijk}
Corpus: sous-corpus phonétique (1 117 cartes de travail)

Carte 20: Analyse dendrographique (selon l'algorithme de Joe Ward, Jr.)
- En bas: arbre (classification en 8 dendrèmes)
- En haut: spatialisation de l'arbre (classification en 8 chorèmes)

Indice de similarité: IR_{ijk}
Corpus: sous-corpus lexical (465 cartes de travail)

Philologie numérique et syntaxe historique

1. Principes généraux de la syntaxe historique et diachronique

Les nouveaux outils numériques, dont les linguistes peuvent bénéficier, mais aussi qu'ils peuvent faire évoluer en fonction de leurs besoins, ont-ils modifié nos comportements, et même plus fondamentalement notre démarche ? Ou bien ne s'agit-il que d'une couche de modernité qui, sans rien changer, a pour effet de faciliter un peu notre travail – « nouveaux outils, ancienne démarche » en quelque sorte ?

Je voudrais tenter de présenter ici de quelle façon, me semble-t-il, les possibilités ainsi ouvertes ont, en peu d'années, profondément affecté, modifié les pratiques, la démarche, et même la conception que le linguiste diachronicien se faisait de ses tâches et les théories que l'on pouvait formuler.

La linguistique *historique*, et la syntaxe *diachronique*[1] en particulier, ont pour but la recherche, non seulement des *régularités* (comme en synchronie), mais également et surtout des *irrégularités* de la langue d'une époque et d'un lieu (et en cela elle se distingue de la linguistique synchronique).

Les *régularités* concernent non seulement la grammaire en elle-même, mais également les variantes repérées et intégrées à la grammaire. Les *irrégularités* en revanche, ce sont *les variantes non attendues* ; il peut s'agir d'une *innovation ponctuelle* – un hapax, rencontré une seule fois et que l'on ne retrouvera plus, ou bien d'une *nouvelle forme* ou *construction* qui, de texte en texte, de témoin en témoin, perdure et s'installe – et dans ce cas on peut parler véritablement de *changement,* ou encore d'une *variation* dans l'emploi d'une forme ou construction – sa progression, sa régression -, ou enfin sa *disparition*.

Rappelons une autre spécificité de la linguistique diachronique ; elle travaille sur un objet linguistiquement singulier, la ou les langues de siècles passés, pour lesquels il n'existe plus de locuteur natif pour *valider*, ou invalider, l'analyse que peut en donner le linguiste. Dans ce domaine, l'introspection n'est d'aucun secours, non plus que la

---

[1] On distingue les deux noms (histoire, diachronie) et les adjectifs correspondants (historique, diachronique), par leur extension. Pour nous, la *linguistique historique* recouvre aussi bien l'histoire *interne* de la langue que son histoire *externe,* et l'approche diachronique ou même synchronique de la grammaire des époques anciennes. La *linguistique*, ou la *syntaxe*, *diachronique*, désigne de façon bien plus précise et restreinte uniquement l'étude de *l'évolution de la langue*, *ou de la syntaxe*.

connaissance que le spécialiste peut avoir acquise de sa fréquentation des textes de ces époques, ou des grammaires historiques qu'il a assimilées. Il faut rappeler d'emblée que *le linguiste diachronicien ne peut donc produire que des hypothèses, jamais de certitude.* Dès lors son seul recours est celui des textes et des documents *écrits*, jusqu'au moment où apparaît l'enregistrement qui conserve l'oral. Il s'ensuit que plus on a de témoins, plus les hypothèses sont solides, ce qui, à l'évidence, limite la portée et les acquis possibles du travail *'à la main'*. Pour le linguiste diachronicien, on le voit, d'emblée, le recours aux *corpus numérisés* est une façon de multiplier ses capacités d'exploration et donc de trouvailles, et de se donner ainsi de plus fortes possibilités pour la formulation d'hypothèses.

La *syntaxe historique* est une discipline jeune : ce n'est guère que depuis un siècle et demi qu'elle existe de façon autonome. Pendant quelques décennies elle s'est développée sur le terreau théorique de la linguistique synchronique, utilisant, en les gauchissant plus ou moins, les concepts et méthodes qui se développaient, avec quelle ampleur et quelle richesse, pour la linguistique synchronique (structuralisme, grammaire générative, sémantique fonctionnelle). Mais peu à peu ont vu le jour des théories spécifiquement dédiées à la diachronie, c'est-à-dire portant sur l'évolution des langues et le changement linguistique (grammaticalisation, réanalyse, analogie, contact entre langues).

Le *traitement numérique des textes* est, lui, bien plus jeune encore, il n'a guère que cinq décennies, ou plutôt, si l'on ne considère que le développement des outils nécessaires au linguiste, deux décennies. C'est dans les années 60 qu'a commencé à se développer par exemple le corpus de français historique *Frantext* ; mais les outils utiles ne sont venus qu'après.

Or ces deux domaines ont justement commencé à se développer – est-ce un hasard ? – en pleine période de réflexions nouvelles et extrêmement fécondes sur la production et la réception des textes (Zumthor, Jauss, Iser…).

La démarche des diachroniciens a dès lors pu bénéficier de ce triple 'progrès' – technique, théorique et interprétatif. Pour utiliser les multiples possibilités ainsi générées, tant dans le champ théorique que dans le champ des techniques, nous avons développé *une conception intégrée pour l'étude du 'changement linguistique' en syntaxe* : seul un travail d'équipe, long, régulier, solidaire, novateur a permis ces avancées (UMR 'ICAR'-Equipe ICAR-3 à Lyon, ANR 'SRCMF' Paris-Stuttgart, 'Corptef' Lyon-Paris).

2. Philologie numérique

2.1. Qu'appelle-t-on philologie numérique ('digital philology') ?

On range sous cette dénomination des activités relativement diverses, qui se déclinent sous plusieurs formes différentes, et en particulier pour ce qui nous intéresse sous deux formes complémentaires : la numérisation de textes existant déjà sous une forme

médiée différente (éditions imprimées), qui peuvent ainsi entrer dans la confection de corpus (nous y reviendrons) ; et d'autre part ce que l'on nomme 'édition numérique' ou 'édition électronique', qui désigne spécifiquement une édition conçue en fonction des outils désormais à notre disposition, et dont le support est d'emblée informatique.

Ces deux activités ont en commun de traiter des données numériques (qu'elles soient primaires, produites comme numériques, ou secondaires, c'est-à-dire numérisées) suivant une *méthode d'annotation numérique* qui permet de leur intégrer des informations nécessaires à une exploitation automatique, quantitative en particulier. Ces annotations concernent des informations ayant trait soit à *la mise en forme du texte*, soit à *son enrichissement linguistique*. Pour distinguer toutes les unités utiles à la mise en forme et à la structuration du texte (mots, phrases, vers, lignes, paragraphes, pages, folios, colonnes d'un manuscrit, frontières des passages en 'discours direct', etc.), on opère un balisage suivant les normes recommandées par la TEI (et éventuellement adaptées) dans le langage XML. Concernant l'enrichissement linguistique (pour les linguistes, les littéraires, les historiens aussi), on 'pose' des étiquettes morphologiques sur chaque forme (selon une liste de catégories explicite) ; et grâce à des développements récents, on est à même parfois d'opérer un 'parsage syntaxique' des textes, suivant des étiquettes fonctionnelles et syntaxiques définies. Enfin, il s'avère toujours utile de pratiquer une lemmatisation des formes du texte.

Par ailleurs, pour exploiter ces données numériques, il faut disposer d'un ou de plusieurs *outils de requêtes* permettant de travailler à partir de ces textes 'enrichis'.

Nous examinerons successivement les deux faces que nous avons évoquées, qui sont désormais à la base de nos recherches linguistiques : les corpus numérisés d'une part, et l'édition numérique de textes d'autre part.

2.2 La 'linguistique outillée' : constitution et exploitation de corpus

Un corpus est une entité 'raisonnée', dont la constitution et les possibilités sont adaptées au type de recherche visé. Dès les années 90, bien des publications ont traité de ces questions (voir par exemple Habert & al. 1997). Pour ce qui concerne la linguistique diachronique du français, plusieurs bases de données ont été constituées depuis quelques décennies, et FRANTEXT a été pionnier dans ce domaine. Généralement, ces bases de données sont constituées de textes déjà édités, que l'on numérise et balise. De ce point de vue, on choisit surtout des éditions fiables, peu interventionnistes, et même des éditions très fidèles, qu'on numérise en les balisant (mot, propositions, phrases, paragraphes, etc.), et éventuellement qu'on enrichit en y intégrant des étiquettes morphologiques ou syntaxiques. Une question de nature différente, mais non triviale car elle se pose encore régulièrement, concerne les droits juridiques de ces textes ; les éditions qui ne sont plus sous droits d'auteur ou d'éditeur échappent à cette difficulté, mais ce sont souvent des éditions anciennes, réalisées suivant des normes un peu rustiques; dans le cas contraire, il est nécessaire de conclure des accords avec les éditeurs.

Pour étudier l'histoire du français, le linguiste bénéficie de plusieurs recueils de textes ainsi numérisés : FRANTEXT, la BFM (Base de français médiéval de l'ENS de Lyon : http://*bfm.ens-lyon.fr ou http://txm.bfm-corpus.org/)*, le Nouveau corpus d'Amsterdam désormais hébergé à Stuttgart (http://www.uni-stuttgart.de/lingrom/stein/corpus), le corpus MCVF d'Ottawa (www.voies.uottawa.ca/corpus*), etc.*

2.3 Une édition fidèle à une « version usagée » d'un texte

L'autre voie pour constituer un corpus de français, en particulier pour la période la plus ancienne, est de réaliser directement des éditions numériques selon les normes les plus adéquates aux recherches en cours. Il faut alors choisir parmi des formes d'édition différentes: édition critique, avec un apparat critique éventuellement complexe ; ou édition d'un manuscrit choisi 'de base', dans la tradition 'bédiériste' ('continiste'? 'segriste'?). Cette deuxième option a l'avantage de fournir au linguiste un document dans l'état le plus proche du texte 'réel' tel qu'il a été copié et lu à l'époque de sa production. La transcription peut se faire – et ensuite s'afficher – par exemple sous trois formes, qui reflètent à des degrés divers la forme même du manuscrit :

- Édition 'courante' (ou 'normalisée'), où les abréviations sont résolues, les lettres ramistes insérées, les signes diacritiques (accents, point sur le *i,* cédille) employés selon les normes habituellement pratiquées pour l'édition de texte, les mots séparés comme dans la langue moderne.
- Édition 'diplomatique' n'introduisant aucune de ces modifications, mais n'imitant cependant pas les graphies spécifiques de l'époque (pour *s* et *r* en particulier).
- Édition 'facsimilaire', qui représente les variantes graphiques des lettres, et conserve les abréviations dans leur forme originelle.

Un exemple de ce type d'édition 'très fidèle', est celle que nous avons réalisée de la *Queste del saint Graal* (composée entre 1215 et 1225), à partir d'un manuscrit copié à la fin du XIIIe siècle sans doute, le Ms K dans la terminologie de Pauphilet 1921 (Cote : Ms P.A.77 de la BM de Lyon) (http://txm.bfm-corpus.org/txm: version librement téléchargeable selon la licence Creative Commons). Seule une édition de type 'bédiériste' et 'pragmatique' (pour reprendre le terme de F. Duval 2006, ce qui signifie : 'adéquat à la situation'), permet d'utiliser au maximum les possibilités du numérique, notre but étant de disposer d'un texte tel qu'il a été pratiqué, lu et peut-être recopié par les contemporains.

Parmi la cinquantaine de manuscrits qui nous sont parvenus de ce texte, dont une vingtaine copiés au XIIIe siècle, il n'était pas encore possible de réévaluer les stemmas et les regroupements proposés antérieurement (Pauphilet 1921 [1980], Bogdanow 2006). Aussi avons-nous choisi le même manuscrit de base que Pauphilet (1923), le Ms K. Cela avait l'avantage de pouvoir comparer les deux éditions : pour ce texted'environ 110 000 mots, il existe entre l'édition Pauphilet et la nôtre, faite près d'un siècle plus tard avec des moyens techniques et des principes un peu différents, plus de 20 000 différences. Elles concernent la ponctuation (près de 10 000 différences, ce qui n'est guère étonnant, car même les éditions 'diplomatiques' ne respectent pas toujours la

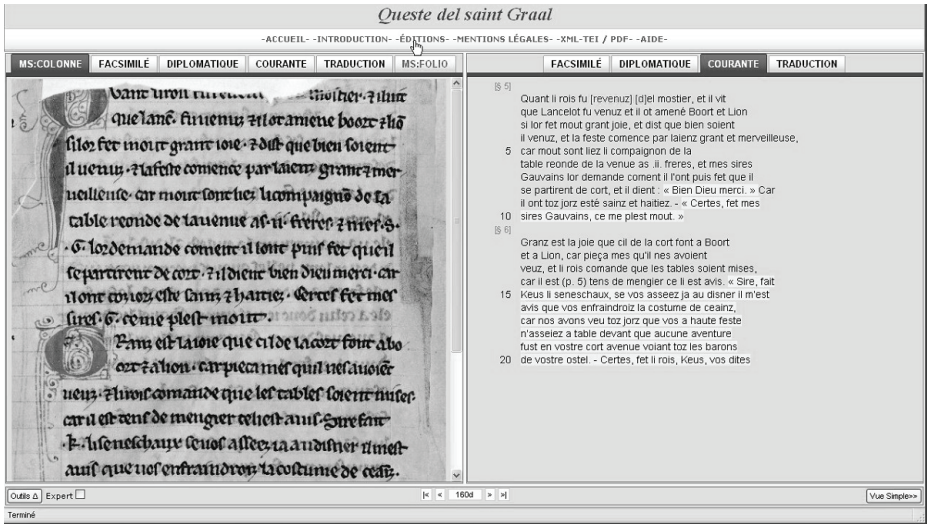
ponctuation du manuscrit, alors que souvent – c'est le cas du ms. K – elle est parfaitement acceptable pour nous), mais aussi des modifications opérées par Pauphilet dans la graphie des mots et leur segmentation, ou par la suppression ou l'ajout de mots ou expressions (d'après les mss de contrôle), par des changements dans l'ordre des mots (11 000 cas), et par un usage différent des majuscules (près de 3000 différences).

Ces trois versions 'courante', 'diplomatique' et 'facsimilaire', qui forment notre édition du même texte, peuvent être affichées, et lues en parallèle à l'écran grâce à une technique d'affichage multi-facettes en deux volets ; on peut jouer ainsi sur un co-affichage binaire entre 5 textes : les trois versions éditées, la photo du folio manuscrit correspondant, et également la traduction du passage édité.

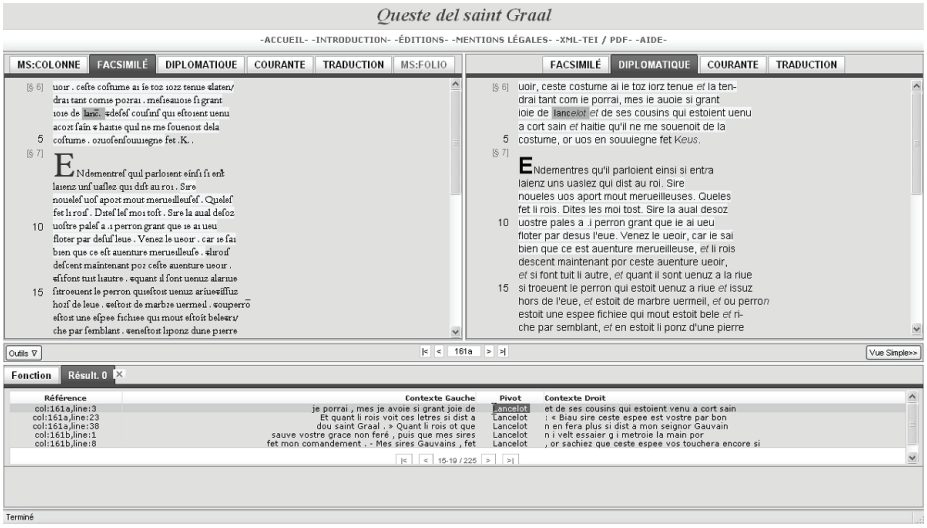
Il faut souligner qu'un tel résultat n'est possible que grâce à un travail d'équipe très poussé et multi-spécialisé.

Voici ce que l'utilisateur peut voir sur son écran en affichant par ex. une colonne du manuscrit K et son édition courante :

Edition numérique de la Queste del saint Graal (ms K de Lyon) (C. Marchello-Nizia & A. Lavrentev) (ENS Lyon) online http://txm.bfm-corpus.org/txm (téléchargeable)



Ou bien, comme ci-dessous, deux versions différentes, facsimilaire et diplomatique, avec au bas la réalisation d'une requête sur les occurrences de 'Lancelot' – qui apparaît surligné de rouge dans le texte :



2.4 L'outil de requêtes et de calculs TXM

A partir de tous les textes dont nous disposons, soit numérisés secondairement, soit éditions numériques, nous utilisons le nouveau logiciel TXM (Serge Heiden, Lyon, 2011) pour effectuer des requêtes et des calculs. Cela nécessite d'avoir préalablement constitué un corpus balisé et étiqueté suivant un corps de normes flexibles et adaptables (XML-TEI), et de savoir utiliser le langage de requêtes adéquat.

Nous donnerons quelques exemples pour illustrer le fait que ce logiciel TXM permet certes les requêtes les plus simples, concernant par exemple un mot unique, ou encore la cooccurrence de deux ou plusieurs mots, mais aussi des requêtes bien plus complexes, combinant tout à la fois mots et catégories grammaticales.

Voici quelques exemples de requêtes (CQP) effectuées avec le logiciel TXM :

- Pour obtenir toutes les occurrences d'un mot précis (dans un texte précis, ou dans tout le corpus): soit en entrant simplement le mot dans la fenêtre dédiée, par ex. *chevalier,* ou en utilisant la formule canonique [word=»chevalier»] ; et si on veut également les formes avec -*s* final, la formule est : [word=»chevalier.*»]
- Ce mot en 'discours direct' : [word=»chevalier» & q=»[123]»] (où q='quotation')
- Si l'on recherche les occurrences de l'adverbe de phrase *si,* on demandera ses emplois devant un verbe conjugué : [word=»si» & q=»[123]»] {0, 2} [pos= «VERcjg»]
- A partir non plus de mots, mais de catégories (pos='part of speech'), on peut demander par exemple les adjectifs qualificatifs postposés à un nom, en 'discours direct', entre le Xe et le XIIIe siècle, ce qui s'exprime ainsi : [pos=»NOMcom»] [pos=»ADJqua» & _.text_siecle=»10|11|12» & q=»[123]»]

- Si l'on recherche les occurrences du pronom sujet indéfini ON entre le IXe et le XVIe siècle la formule sera : [word!=»m'|s'|il|je|j'|elle»%c] [word=»on|l'on|om|lom|un| l'un|um|l'um|en|l'en|nen|em|l'em|hom
|l'hom|hon|l'hon|l'an|an» %c & ((pos!=»PRE|DET.* « & fropos!=»PRE|DET.*» & _.text_siecle =»10|11|12|13|14|15») (frpos!=»PRP|NUM|DET:ART» & _.text_siecle=»16»))]

En combinant de façon encore plus complexe mots, ponctuations et catégories, et en jouant sur leurs exclusions, nous pouvons même aboutir à peu de choses près à la liste des propositions dans la *Chanson de Roland* (éd. Moignet) :

> [word!=»AOI»] []* ([word=»,»][word=»\p{Lu}.*» & word!=»AOI»] | [word=»tant»%c] [word=»cum|cume»] |[word=»,»] [word=»il|jo|si|sim|sel|sem|dist»]|[(word=»[.!?:;-]») | (word=»ki.*|qu.*|dunt|dun|sel|sem|sim»%c & pos=»PROrel|CONsub.*|DETrel|DETint») | (word=»Si») | (word=»sin»%c) | (word=»Ne|N'|unches»)]) within s

En un clic ou presque, en tout cas très rapidement, on obtient un tableau Excel de 4878 lignes, soit autant de propositions potentielles ; après corrections (près d'une dizaine d'heures tout de même), on parvient à un résultat correct de 4974 propositions pour l'ensemble de la *Chanson de Roland* – soit moins de 2% de bruit, ce qui est peu.

2.5 En quoi ces avancées techniques renouvellent-elles l'approche des faits syntaxiques ?

La rapidité, la quantité d'informations traitées (plusieurs millions de mots actuellement), le nombre des facteurs interagissant combinés, la complexité des possibilités de requêtes, atteignent désormais des niveaux jusque là inconnus, et qui étaient bien évidemment inatteignables par des recherches 'à la main'. Et pour l'étude de la syntaxe, qui est un champ complexe où priment les combinaisons de termes et même de construction, il est fondamental de disposer d'un outil qui puisse combiner et hiérarchiser plusieurs facteurs, et qui puisse effectuer non seulement des requêtes sur des termes ou des catégories, mais également des calculs complexes à travers des tris croisés de plusieurs facteurs. Dès lors qu'on dispose de ces possibilités, la pratique se modifie, et les hypothèses que l'on peut formuler en sont modifiées dans leur nature - une très grande complexité devenant possible, ainsi que leurs résultats, puisqu'on parvient à une granularité très grande. Grâce à la rapidité des calculs, on teste aussitôt une hypothèse (par exemple l'existence d'une construction), et les régularités ou les impossibilités apparaissent assez vite, puisqu'est offerte la possibilité de jouer sur les contextes, en modifiant l'un après l'autre chacun des facteurs de la requête.

Cette facilité pour fouiller de très gros corpus permet le repérage quasi immédiat des *innovations* (préludant à un possible changement), mais aussi des *raréfactions* (obsolescences), ou des *absences* (disparitions). Le diachronicien suit ainsi le *rythme* d'un changement, et dans tous ses contextes ; or on sait qu'une innovation apparaît toujours dans un contexte précis, et gagne ensuite successivement d'autres contextes.

On gagne également ainsi la possibilité de *comparer* la grammaire de textes, d'auteurs, d'époques, de dialectes ou sociolectes, ou de genres différents – à la seule condition que ces paramètres aient été renseignés. Ces possibilités ont permis, par exemple, de 'pister' l'apparition, puis le développement sémantique du pronom indéfini *on,* ou bien l'histoire croisée, liée, des adverbes quantifieurs / intensifieurs *moult / très / beaucoup,* à travers les stades d'apparition, de changements de catégorie, d'extension ou de disparition ; de telles explorations exigent le traitement de millions de mots sur quelques siècles.

Nous illustrerons ces développements en syntaxe par quelques exemples pris dans trois domaines de recherche en cours.

3. Syntaxe historique et utilisation des corpus

Nous évoquerons trois domaines dans lesquels les apports de la 'philologie numérique' se sont révélés déterminants : la *ponctuation*, *l'ordre des mots*, le *discours direct* (ou '*oral représenté*'). Pour chacun des cas, nous ferons préalablement un bref 'état des lieux'.

3.1 Ponctuation et syntaxe : deux cas liés au discours direct, le comma et l'incise

Pour un état de langue ancien, antérieur à l'existence de l'enregistrement, et en l'absence de locuteur natif, il paraît impossible de formuler des hypothèses sur l'intonation des locuteurs. Et des travaux, en nombre réduit cependant (Llamas-Pombo, Lavrentiev), ont montré qu'une étude systématique des éléments segmentants que sont les signes de ponctuation aboutit à des enseignements parfois très riches au plan syntaxique.

La détection de certains phénomènes, rendue possible par l'existence d'éditions numérisées respectueuses de la ponctuation scribale, a permis, dans les deux cas que l'on va voir, de formuler des hypothèses neuves concernant la *syntaxe* et *l'intonation*. Ainsi, dans certains manuscrits français médiévaux, l'emploi du *comma* semble porteur d'une indication à l'adresse du lecteur concernant le type d'intonation à adopter. Et, inversement, l'absence de ponctuation dans des contextes où la langue moderne ponctue systématiquement, invite peut-être à réinterpréter ces éléments. Mais sans l'accès à une transcription numérique très fidèle d'une part, et à un traitement automatique de toutes les unités et de leurs contextes d'autre part, ce type d'analyse n'est guère possible.

Le *comma* est assez peu utilisé au Moyen-Âge – mais peut-être passe-t-il souvent inaperçu ? Et quand il est repéré, il est interprété comme une ponctuation forte, et transcrit soit par « ; », soit par un point d'exclamation « ! » (dont il est l'ancêtre semble-t-il).

Notre édition du ms. K de la *Queste del saint Graal,* très respectueuse de la ponctuation du copiste, permet de repérer trois signes : l'un très courant, suivi ou non de majuscule, le *punctum,* les deux autres bien plus rares : le *punctum interrogatiuum,* et le *comma* (*punctum eleuatum,* que nous transcrivons imitativement par **.'**).

Sur les 60 folios occupés par le texte de la *Queste,* on a relevé 195 occurrences du *comma* : quelle en est la valeur ? Il se rencontre principalement dans trois contextes assez différents :

a) 67 commas suivent immédiatement *Ha / ha* ou *Hé / hé* ; ils sont en discours direct et leur valeur est clairement exclamative ;

b) 18 sont employés entre deux membres de phrase corrélés, tels *aussi com…aussi,* ou entre deux propositions parallèles : les exemples sont extraits des versions 'courante' ou 'diplomatique' de l'édition :

«… car ausi com vos l'avez par achaison de fame .' ausi le perdroiz vos par damoisele.» (*Queste* éd. cour. § 64, fol 171c : 'De même que vous l'avez obtenu grâce à une femme, de même le perdrez-vous à cause d'une demoiselle')

«… car li filz n'amoit le pere .' ne li peres l'enfant.» (*Queste* éd. cour. § 50, fol 168d)

c) 82 sont employés entre une subordonnée initiale en *Quant, Se, Por ce que, qui que*, et la principale :

« Et se uos uolez sauoir donc ces granz auentures sont par tantes foiz auenues .' alons a lui moi *et* uos » (*Queste* éd. dipl. § 42, fol 167a : 'Et si vous voulez savoir d'où ces grandes aventures sont venues tant de fois, allons trouver cet homme, vous et moi.')

« Mais quant tu ne cuideras pas que tu en puisses eschaper .' lors descoeure la croiz *et* di… » (*Queste* éd. dipl. § 45, fol 167c : 'Mais quand tu ne croiras pas pouvoir en réchapper, alors découvre la croix et prononce ces paroles :..')

«… por ce que li lyons estoit meins mesfesanz que li serpenz .' corui je sus au serpent et l'ocis. » (*Queste* éd. courr. § 132, fol 183a)

Nous constatons que 75% des emplois du comma dans ce manuscrit se trouvent dans un épisode au discours direct. Cela pourrait être interprété comme une indication en vue d'une lecture à haute voix d'un passage qui se donne justement pour de l'oral, la prise de parole d'un personnage. Le comma indique qu'il existe bien à cet endroit une pause, mais sur laquelle il ne faut pas laisser retomber la voix, car un lien syntaxique et sémantique fort lie les deux portions de l'énoncé (corrélations et parallélismes : *de même que…, de même ; parce que…, alors ; quand…, alors)*..

Quant à l'incise, tout éditeur d'un texte médiéval sait qu'il 'faut' la ponctuer en l'entourant de virgules, de même qu'il est habituel d'introduire des guillemets pour indiquer le changement de locuteur. C'est ce que nous avons fait dans la version 'courante' de notre édition de la *Queste.* Mais l'usage du copiste du ms. K est tout autre : l'incise *fet X* n'est presque jamais balisée par un signe de ponctuation :

et li rendent son salu, biau seignor fait il quele auenture uos a ca amenez. Sire fet M*e*s S*ires* G*auuains* la grant fain *et* le grant desirrier que nos auions de parler a uos (*Queste* éd. dipl. fol 197a, § 192)

Quelle explication fournir à cette étonnante absence de démarquage ? Nous proposerons de voir dans l'incise, justement, une simple marque de changement de niveau énonciatif : l'auteur n'est plus l'instance de discours, c'est un personnage qui l'est

temporairement devenu, et c'est ce seul fait que marque l'incise. Deux phénomènes nous conduisent à cette interprétation. D'une part le fait que, dans ce texte, l'incise a toujours cette même forme canonique (relativement nouvelle alors) : verbe *faire* employé au présent, toujours avec un sujet exprimé et postposé – ce qui est doublement marqué : *fet il, fet li rois* : l'expression est tout à fait formulaire. Et d'autre part, on constate que l'incise ne cooccurre pas avec une annonce de discours direct (du type : *et dist au roi*) ; les deux marques sont exclusives l'une de l'autre (à quelques rares exceptions près). Il s'agit simplement de marquer le décalage énonciatif, et l'une des formules suffit, toutes deux étant parfaitement figées dans ce rôle. Dès lors, point besoin de ponctuer, *fet il* a presque une valeur adverbiale.

3.2 De l'ordre des mots

C'est sans doute, depuis un siècle et demi, le domaine de la syntaxe qui a suscité les plus nombreuses études et où les acquis sont les plus conséquents. Dans un premier temps, l'analyse a d'abord été centrée sur le pôle verbal et ses arguments (S, V, O : cf J.-J. Le Coultre, 1875, et L. Foulet, 1967 [1919 / 1930]). Puis l'essor de la grammaire générative a placé l'accent sur la relation S-V, avec la question du sujet non exprimé (ou 'sujet-nul' dans cette problématique) et de la position du verbe (V2), et depuis les premières explorations générativistes de la diachronie dans les années 85, l'attention a été totalement occupée par la périphérie gauche du verbe et sa cartographie, pour laquelle des affinements sont en cours.

La possibilité du traitement rapide de gros corpus a renouvelé et infiniment élargi le corps des énoncés étudiés, et a permis à d'autres orientations de se dessiner. Entre autres points, le nombre des facteurs pris en compte s'est multiplié, et en conséquence la chronologie des changements et leur définition se sont affinées. Désormais, on est à même d'analyser l'ensemble des éléments de la proposition, non seulement S et V, mais aussi l'objet et l'attribut, et tous les éléments, qu'ils soient ou non dans la valence verbale. Les avancées dans l'étude de la typologie des langues du monde ont permis de reformuler la question de la position et de l'optionalité du sujet et de l'objet dans l'histoire des langues et du français en particulier.

Un grand nombre d'acquis peut servir désormais de socle à des études nouvelles. On sait en effet que la syntaxe des constituants propositionnels du français à époque très ancienne, du IXe au XIIIe siècle, se caractérise par huit traits :

1) Le verbe est majoritairement en seconde position dans tous les types de proposition, mais aucune des autres positions n'est exclue (V1, V3, V4), la position finale méritant une étude en soi, car jusqu'au XIIIe siècle, le verbe final est fréquent en subordonnée.
2) La non-expression du sujet domine, plus ou moins largement jusqu'au XIIIe s., sauf dans certaines propositions subordonnées et à certaines personnes de l'interlocution où son expression domine (Vance 1997, Prévost 2001, Detges 2003).
3) Lorsqu'il est exprimé, le sujet est majoritairement antéposé au verbe, mais sa postposition n'est pas rare.

4) Dès le XIIe s., l'expression du sujet (nominal et pronominal) et son antéposition au verbe croissent parallèlement.
5) En revanche, dès l'origine du français, l'expression de l'objet est obligatoire – sa non-expression étant d'une part très limitée et d'autre part jamais obligatoire (Schoesler 2001, Marchello-Nizia 2009, Rouquier / Marchello-Nizia 2012, 2013).
6) Pour le sujet et le verbe, leur syntaxe est dissymétrique entre les déclaratives et les subordonnées.
7) Les non-régies (indépendantes et principales) déclaratives sont le type de proposition le plus fréquent dans tous les textes.
8) Parmi les non-régies, les types les moins fréquents (interrogatives, injonctives, incises) ont d'emblée ou acquièrent dès l'époque ancienne une syntaxe spécifique.

Concernant l'objet direct, on sait que dès l'origine (IXe s.) l'objet pronominal est cliticisé au verbe qu'il précède immédiatement, sa postposition existe cependant mais est fort rare. Son expression était obligatoire, sa non-expression restant très limitée, ne concernant que l'objet pronom personnel de 3e personne - et même alors l'objet peut être exprimé. Quant à l'objet nominal, il peut s'antéposer ou se postposer au verbe, mais il se postpose majoritairement.

Mais récemment, cette fonction 'objet' qui était restée assez peu explorée en syntaxe diachronique, a bénéficié de nombreux acquis nouveaux, tant pour sa chronologie, que pour l'hypothèse de mono-actancialité. On sait désormais que l'objet nominal (nom, pronoms non personnels) est majoritairement antéposé au verbe avant le milieu du XIe s. (Rouquier / Marchello-Nizia 2012, 2013). Mais il se postpose majoritairement, en toute proposition, dès le milieu du XIe siècle (*Vie de saint Alexis*). Et dès le début du XIIIe siècle, la postposition de l'objet nominal est devenue la règle en prose (97% des objets nominaux sont postposés au verbe dans la *Queste*). Il est par ailleurs devenu contigu au verbe (*Paieront une contravention* ceux qui..), alors que le sujet continue de pouvoir en être séparé.

Enfin, grâce au traitement automatique d'un grand nombre de textes, l'hypothèse d'une 'monoactancialité' du français a pu se formuler. En effet, avant le XIIIe s., le sujet est absent dans plus de 50% des propositions verbales ; mais il apparaît qu'il est encore plus fréquemment absent dans les propositions à verbes transitifs (Rouquier / Marchello-Nizia 2012, 2013). En effet, l'objet étant obligatoirement exprimé (Schoesler 2001), et le sujet étant optionnel, il en résulte que si dans une proposition à verbe transitif il n'y a qu'un seul argument, c'est nécessairement l'objet. Dès lors, peut-on dire que l'ancien français était une langue monoactancielle orientée vers l'objet ? Peut-être, puisque complémentairement on constate que le français moderne présente une situation inverse, l'expression du sujet y étant obligatoire, mais non celle de l'objet (Lambrecht / Lemoine 1996, Larjvaara 2000). Resterait à expliquer le passage d'un type de mono-actancialité à un autre.

3.3 L' « oral représenté » et son importance pour décrire le changement linguistique

L'insertion d'un discours dans un récit, et spécialement dans un *récit écrit*, est un phénomène courant, bien repéré et maintes fois étudié depuis plus d'un siècle (Marchello-Nizia 2012) : il s'agit d'un épisode de 'parole', donné comme oral (fictionnel ou ayant réellement eu lieu) et prononcé par un personnage. Les études sur le 'styte direct', puis sur le 'discours direct', ont montré que ce changement énonciatif fait l'objet d'un marquage externe par des formules consacrées, avant l'invention d'un signe diacritique spécial, les guillemets, aux 16e-17e siècles. Mais anciennement, seules des formules annoncent d'une part l'arrivée d'un passage en 'discours direct', d'autre part sa clôture avec un retour au récit (cet élément-là n'est généralement pas étudié par les linguistes), et parfois le changement de locuteur est rappelé dans le discours même par une incise qui y est insérée suivant une syntaxe très spécifique. Ces trois formules, nous les nommons *l'annonce*, *le rappel*, et *l'incise*. Un tel marquage externe du discours direct, c'est-à-dire du discours rapporté se donnant comme l'exacte reproduction de paroles ayant été prononcées oralement, semble présent dans les langues du monde de toutes époques. On désigne souvent désormais par l'expression '*oral représenté*' les épisodes de discours direct spécifiques de l'écrit, et spécialement le discours direct inséré dans du récit et annoncé comme du discours rapporté.

Plusieurs études se sont donné pour but de décrire comment l'auteur différencie les deux plans énonciatifs, celui du 'récit' et celui de l''oral' (fictionnel ou non). Des particularités linguistiques de l' 'oral représenté' en ancien et moyen français ont déjà été repérées : parmi les premiers, P. Guiraud (1967) avait mis en évidence un statut différent et de la déclinaison, et des démonstratifs, dans les passages en discours direct. Par la suite, les études de B. Cerquiglini (1981), P. Zumthor (2000 [1972] , 1983), et plus tard et de façon plus approfondie D. Lagorgette (1998, 2006) et S. Marnette (2005, 2006), ont contribué à caractériser de façon de plus en plus précise cette grammaire de l'*oral représenté* : l'interjection, l'apostrophe dans ses diverses modalités, les réponses *oie, oïl, naje, nennil*, les exclamations comme *Ha, Hé,* la formule de serment et de véridiction *si m'aïst Diex* (Marchello-Nizia 1985), ne se trouvent guère que dans ce type d'épisodes.

Mais outre les morphèmes et expressions qui sont réservés à la représentation de l'oral, d'autres phénomènes grammaticaux distinguent les deux modes du récit et du discours direct, non par une exclusivité d'emploi, mais simplement par une fréquence plus ou moins forte dans chacun des deux types discursifs. Grâce à l'étiquetage morphologique de gros corpus, des requêtes automatiques contrastives menées sur le 'récit' d'une part et sur l' 'oral représenté' d'autre part ont conduit à mettre en évidence l'existence de deux grammaires partiellement différentes, chacune ayant ses « constructions préférées », pour reprendre l'expression de Glikman / Mazziotta (Guillot & al. à par., Glikman / Mazziotta à par., Donaldson à par.) ; un travail en cours sur l'emploi des temps verbaux a permis, pour l'instant à partir d'un corpus réduit, de formuler l'hypothèse que ce n'est pas le présent par exemple qui est le temps

« préféré » du discours direct, mais le futur. Déjà, Marnette (2006) avait mis en évidence six traits distinguant le 'discours direct' du récit. Plus récemment, sur un corpus bien plus vaste de 61 textes étiquetés et traités automatiquement (par 'partitions' sous TXT et analyses contrastives), Guillot / et al. ont mis en évidence 22 traits par lesquels on peut caractériser le discours direct (DD) et, ce qui est nouveau, 19 particularités du 'non-DD', ainsi qu'une série de traits non sensibles à cette distinction.

Il est certain que cette pratique d'une comparaison systématique est riche de découvertes. Ainsi par exemple, les deux paradigmes de démonstratifs CIL et CIST se rencontrent tous deux en récit et en discours direct (question qui avait troublé Guiraud…), alors qu'en très ancien français les formes en *ist* (qui disparaissent après le XIIe siècle) ne se trouvent qu'en discours direct. Mais la fréquence de CIL et de CIST dans chacune des deux modalités d'écrit est inversée : CIL est le démonstratif de loin le plus fréquent à l'écrit, CIST y étant peu fréquemment employé, essentiellement avec une valeur mémorielle et dans un contexte précis, celui du retour au récit après un discours direct (Guillot 2004) ; et inversement, à l'oral représenté c'est CIST qui domine, et dans des contextes qui ont permis de circonscrire plus exactement qu'on ne l'avait fait auparavant la valeur exacte, pragmatique, de ce paradigme : il désigne les éléments que le locuteur définit comme appartenant à sa sphère de locuteur (Marchello-Nizia 2004), tandis que CIL s'emploie, à l'inverse, pour caractériser un élément de la situation comme n'appartenant pas à la sphère du locuteur (Marchello-Nizia 2006). Cela, seule l'analyse automatique d'un très grand nombre de textes variés a permis de le décrire et surtout de l'interpréter.

Nous l'avons vu, le discours direct inséré en récit et balisé par les éléments que sont l'annonce, ou l'incise, et le rappel, peut lui-même, en tant que phénomène, faire l'objet d'une étude diachronique. Dans les plus anciens textes français, on rencontre surtout des monologues (*Vie de saint Léger*), et les dialogues restent relativement peu nombreux (*Sermon sur Jonas, Passion de Clermont, Vie de saint Alexis, Chanson de Roland).* Mais dès les premiers romans du milieu du XIIe siècle, ils deviennent courants et même fréquents. Et on rencontre même dès lors un discours direct enchâssé dans un autre (*Chanson de Roland* 377-391), et un siècle plus tard, dans la prose dela *Queste del saint Graal,* on atteint trois niveaux d'enchâssement énonciatif (éd. en ligne Marchello-Nizia / Lavrentiev, fol 167b, 168a, 190b).

Une ultime question se pose, cruciale pour la linguistique diachronique – et cette question a été bien des fois débattue aux débuts de la linguistique historique. Dès lors qu'on admet communément que c'est dans la pratique de la 'parole' (Saussure) et plus spécialement dans les réalisations orales d'une langue qu'apparaissent les innovations qui, par la suite, peuvent se pérenniser en 'changements', est-il légitime de considérer ce que nous avons nommé l' 'oral représenté' comme un lieu où les changements seront mieux détectables qu'ailleurs ? Cette représentation hypothétique des pratiques orales n'est-elle pas un leurre pour le diachronicien ? Nous avons tenté (Marchello-Nizia à par.) de reprendre cette question et d'apporter quelques arguments empiriques en faveur d'une exploration ciblée des représentations de l'oral, en

l'absence de tout accès direct à celui-ci. Dans cette même perspective, Detges (2003) avait émis l'hypothèse qu'un changement syntaxique capital en français, le développement de l'expression du sujet, et d'abord du sujet pronominal, puis sa généralisation, aurait pris son essor en discours direct, dans l'emploi systématiquement 'autothématisé' du pronom sujet *je*. Nous avons montré que plusieurs changements dont on a pu récemment repérer les premières attestations montrent que celles-ci se situent très majoritairement dans des épisodes de discours direct, qu'il s'agisse d'interactions verbales en récit, ou bien d'interventions directes de l'auteur dans son texte. On évoquera simplement le cas de deux morphèmes quantifieurs / intensifieurs du français, *très* et *beaucoup,* qui ne sont pas hérités du latin, mais se sont développés longtemps après, et ont fini par remplacer le morphème *moult* (qui subsiste dans la majorité des langues romanes). Les plus anciens emplois de *biau cop* ou *grant cop* à valeur possiblement quantifiante (et non plus simplement métaphorique) sont attestés dès le XIIIe siècle et le début du XIVe siècle, et sont liés à une énonciation en première personne :

> Nos engins getoient au leur et les leurs aus nostres, mes onques n'oÿ dire que les nostres feissent *biau cop*.
>
> (Jean de Joinville, *Vie de saint Louis,* éd. Monfrin, § 193 : 'Nos engins tiraient contre les leurs et les leurs contre les nôtres, mais *je n'ai jamais entendu dire* que les nôtres aient réalisé *un beau coup'*/ 'aient abouti à *beaucoup de résultats'*)

> Miex le devons faire *biau cop*
> De nos tamps que ne faisoit Job
> Qui fu devant le tamps de grasce.
> (*Livre de Job*, ms G, composé à la fin du XIIIe siècle, v. 1207-9 : '*Nous devons* nous conduire beaucoup mieux à présent que ne le faisait Job, qui vivait avant l'époque chrétienne')

Quant à l'autre marqueur, *très,* dont les premiers emplois comme intensifieur d'adjectif se rencontrent dans des textes du second tiers du XIIe siècle, on constate que dès cette époque un emploi lui est en quelque sorte réservé puisqu'on n'y rencontre jamais *moult* en cette position : c'est en début d'épisodes d' 'oral représenté', en apostrophe, comme intensif d'un adjectif 'affectif' épithète d'un nom de personne :

> « biax tres dous pere, et vos, bele tres douce mere… »
>
> (Floire et Blancheflor, court récit en vers écrit vers 1160, v. 3213)

Ce type d'apostrophe est fréquent ensuite, chez Chrétien de Troyes en particulier :

> si li dist : " Ma tres chiere dame,
> vos qui…"
> (Chrétien de Troyes, *Yvain*, éd. Roques, v. 2551 : 'Et il lui dit : « Ma très chère dame, vous qui… »)

Comme parfois dans le cas d'innovations, la forme introduite est destinée à occuper un emploi nouveau : avec le développement de la représentation des interactions orales insérées dans les récits, l'apostrophe, qui ouvre en général ces épisodes, s'est

peu à peu constituée en formule canonique, et *très* y a trouvé sa niche de développement, avant de se généraliser aux autres contextes.

L'exploitation ciblée des outils et techniques numériques dans le traitement des textes a ouvert la voie vers de nouvelles frontières, comme les exemples ci-dessus évoqués l'ont montré.

Elle permet une exploration approfondie des changements linguistiques, et rend possible la formulation d'hypothèses jusqu'alors hors d'atteinte concernant l'évolution des langues, et en particulier celle-ci, très importante pour la linguistique diachronique : après les discussions sur l'oral lieu privilégié de l'innovation, ne doit-on pas considérer l' 'oral représenté' comme un type discursif témoin privilégié du changement linguistique ?

UMR 5191 ICAR Christiane MARCHELLO-NIZIA

Bibliographie

Adams, Marianne P., 1987. *Old French, Null Subjects and Verb Second Phenomena*, Ph. D. : U. C. L.A.

Bogdanow, Fanny (ed.), 2006. *La Quête du Saint-Graal, roman en prose du XIIIe siècle. Texte établi et présenté* par F. B., traduction par Anne Berrie. Paris, Librairie générale française (Livre de poche. Lettres gothiques).

Cerquiglini, Bernard, 1981. *La parole médiévale : discours, syntaxe, texte*, Paris, Editions de Minuit.

Detges, Ulrich, 2003. «Du sujet parlant au sujet grammatical. L'obligatorisation des pronoms sujets en ancien français dans une perspective pragmatique», *Verbum*, 25, 307-333.

Detges, Ulrich, 2008. «How useful is case morphology? The loss of the Old French two-case system within a theory of Preferred Argument Structure», in: Barddal, J. / Chelliah, S. (ed.), *The Role of Semantic, Pragmatic, and Discourse Factors in the Development of Case, Studies in Language Companion Series*, Amsterdam, Benjamins, 93-120.

Donaldson, Bryan, 2014. «Socio-Stylistic Reflexes of Syntactic Change in Old French», Journal of French Language Studies 24, 319-345.

Duval, Frédéric, 2006. «La philologie française, pragmatique avant tout ?», in : Duval, F. (éd.), *Pratiques philologiques en Europe*, Paris, Ecole des Chartes, 115-150.

Ecole nationale des Chartes, Coordination : Vielliard, Françoise / Guyotjeannin, Olivier / Bourgain, Pascale, 2001-2005. *Conseils pour l'édition des textes médiévaux* : Fascicules I, II et III, Paris, Ecole nationale des Chartes.

Foulet, Lucien, 1967 [1930]. *Petite syntaxe de l'ancien français*, Paris, Champion.

Glikman, Julie / Mazziotta, Nicolas (2013). «Représentation de l'oral et structures syntaxiques dans la prose de la Queste del saint Graal (1225-1230)», in : Lagorgette, Dominique / Pierre *Larrivée* (ed.), 43-64.

Guillot, Céline / Lavrentiev, Alexei / Pincemin, Bénédicte / Heiden, Serge, 2013. «Le discours direct au Moyen Age : vers une définition et une méthodologie d'analyse» , in : Lagorgette, Dominique / Pierre *Larrivée* (ed.), 17-41.

Guiraud, Pierre, 1967. «L'assiette du nom dans la Chanson de Roland : II : le démonstratif», Romania 88, 59-83.

Habert, Benoît / Nazarenko, Adeline / Salem, André, 1997. *Les linguistiques de corpus*, Paris, A. Colin.

Heiden, Serge / Decorde, Matthieu / Bertrand, Lauranne, 2013. *Manuel de TXM. Version 0.7,* Lyon, ENS de Lyon. <http://txm.sourceforge.net/doc/manual/manual.xhtml>.

Iser, Wolfgang, 1985 [1972]. *L'Acte de lecture : théorie de l'effet esthétique* (trad. de *Der Akt des Lesens*), Bruxelles, Mardaga.

Jauss, Hans-Robert, 1978. *Pour une esthétique de la réception*, Paris, Gallimard.

Lagorgette, Dominique, 1998. *Désignatifs et termes d'adresse dans quelques textes en moyen français*, Université-Paris-10, thèse de doctorat.

Lagorgette, Dominique, 2006. «Etude des fonctions pragmatiques des termes d'adresse en ancien et moyen français», in : Taavitsainen, I. / Härmä J. / Korhonen, J. (ed.) *Dimensions du dialogisme*, Mémoires de la Société Néophilologique de Helsinki, LXVI, 315-338.

Lagorgette, Dominique / Larrivée, Pierre (ed.), 2013. *Représentations du sens linguistique 5*, Chambéry, Ed. Université de Savoie.

Lambrecht, Knud / Lemoine, K. 1996. «Vers une grammaire des compléments d'objet zéro en français parlé», *TCLC* 9, 270-310.

Larjavaara, Meri, 2000. *Présence ou absence de l'objet : limites du possible en français contemporain*, Helsinki, Academia Scientiarum Fennica.

Lavrentiev, Alexei, 2008. *Manuel d'encodage XML-TEI étendu des transcriptions de manuscrits dans le projet BFM-Manuscrits. Version 2.1*, Lyon, ENS de Lyon. <http://ccfm.ens-lyon.fr/IMG/pdf/BFM-Mss_Encodage-XML.pdf>.

Lavrentiev, Alexei, 2009. *Tendances de la ponctuation dans les manuscrits et incunables français en prose, du XIIIe au XVe siècle*. Thèse de doctorat en sciences du langage, Lyon, École normale supérieure Lettres et sciences humaines. <http://tel.archives-ouvertes.fr/tel-00494914/>.

Lavrentiev, Alexei / Marchello-Nizia, Christiane, 2011. «The Development of the Demonstrative Determiner CE in the Manuscripts of Chrétien de Troyes' *Lancelot* or *Le Chevalier de la Charrette*», in : Greco, G. L. / Thorington, E. M. (ed.), *Dame Philology's Charrette, Approaching Medieval Textuality through Chrétien's* Lancelot. *Essays in Memory of Karl D. Uitti,* Tempe (Arizona), ACMRS, 127-140.

Le Coultre, Jean-Jules, 1875. *De l'ordre des mots dans Crestien de Troyes,* Thèse, Dresde.

Llamas Pombo, Elena, 2007. «Réflexions méthodologiques pour l'étude de la ponctuation médiévale», in : Lavrentiev, A. (ed.), *Systèmes graphiques de manuscrits médiévaux et incunables français : ponctuation, segmentation, graphies. Actes de la journée d'étude de Lyon, ENS LSH, 6 juin 2005*, Chambéry, Université de Savoie, 11-48.

Marchello-Nizia, Christiane, 2004. «Deixis and subjectivity: the semantics of demonstratives in Old French (9th-12th century)», *Journal of Pragmatics* 37/1, 43-68.

Marchello-Nizia, Christiane, 2006. «Du subjectif au spatial : l'évolution des formes et du sens des démonstratifs en français», *Langue française* 152, 114-126.

Marchello-Nizia, Christiane, 2007. «Le comma dans un manuscrit du 13e s. : grammaticalisation d'un marqueur de corrélation, ou marquage d'une intonation montante ?», in : Bertrand, O. / Prévost, S. / Charolles, M. / François, J. / Schnedecker, C. (ed.), *Discours, diachronie, stylistique du français, Etudes en hommage à Bernard Combettes,* Berne, Peter Lang, 293-305.

Marchello-Nizia, Christiane, 2009. «L'ordre des constituants phrastiques en français : une évolution conduite par l'objet ?» in : Havu, E. / Helkkula, M. / Tuomarla, U. (ed.), *Du côté des langues romanes, Mélanges en l'honneur de Juhani Härmä,* Helsinki, Mémoires de la Société Néophilologique de Helsinki LXXVII, 33-54.

Marchello-Nizia, Christiane, 2012. «L' «oral représenté» en français médiéval, un accès construit à une face cachée des langues mortes», in : Guillot, C. / et al. (ed.), *Le Changement en français. Etudes de linguistique diachronique*, Berne, Peter Lang, 247-264.

Marchello-Nizia, Christiane, 2014. «L'importance spécifique de l' "oral représenté" pour la linguistique diachronique» , in : Ayres-Bennett, Wendy / Rainsford, Thomas M. (ed.), *L'histoire du français. État des lieux et perspectives*, Paris, Garnier, 161-174..

Marchello-Nizia, Christiane / Lavrentiev, Alexei, 2013. *Queste del saint Graal. Édition numérique interactive du manuscrit de Lyon (Bibliothèque municipale, P.A. 77)*, Lyon, ENS de Lyon. <http://txm.bfm-corpus.org/txm>.

Marnette, Sophie, 2005. *Speech and Thought Representation in French*, Amsterdam-Philadelphia, John Benjamins Publishing Company.

Marnette, Sophie, 2006. «La signalisation du discours rapporté en français médiéval», *Langue française* 149, 131-147.

Moignet, Gérard (ed.), 1969. *La Chanson de Roland,* Texte original et traduction par G. M. Paris, Bordas.

Pauphilet, Alfred, 1921 [1980 Reprints]. *La tradition manuscrite et l'établissement du texte de la Queste del Saint Graal, attribuée à Gautier Map*, Paris, Champion.

Pauphilet, Alfred (ed.), 1923. *La Queste del Saint Graal, roman du XIIIe siècle*, Paris, Champion (Classiques français du Moyen Âge).

Prévost, Sophie, 2001. *La postposition du sujet en français aux XVe et XVIe siècles,* Paris, Editions du CNRS.

Rouquier, Magali / Marchello-Nizia, Christiane, 2012. «De (S)OV à SVO en français : où et quand ? L'ordre des constituants propositionnels dans la *Passion de Clermont* et la *Vie de saint Alexis»*, in : Dufresne, M. (ed.) *Constructions en changement. Hommage à Paul Hirschbühler*, Presses de l'Université de Laval, 111-155.

Rouquier, Magali / Marchello-Nizia, Christiane, 2013. «La position de l'objet direct nominal et l'ordre des mots dans la *Passion de Clermont,* la *Vie de saint Léger* et la *Vie de saint Alexis*», *ACILPR* XXVI, 333-343.

Schoesler, Lene, 2001. «The coding of the subject-object distinction from Latin to Modern French», in: Faarlund, J. T. (ed), *Grammatical Relations in Change*, Amsterdam-Philadelphia, John Benjamins, 273-302.

Vance, Barbara, 1997. *Syntactic Change in Medieval French: Verb-Second and Null Subjects*, Dordrecht, Kluwer Academic Publishers.

Zumthor, Paul, 1983. *Introduction à la poésie orale,* Paris, Le Seuil.

Zumthor, Paul, 2000 [1972]. *Essai de poétique médiévale,* Paris, Le Seuil (Points).

Pour une réconciliation entre théorie et pratique : le cas de la sémantique historique[1]

Je l'ai dit déjà en d'autres occasions, et, comme c'est un sujet qui me tient à cœur, j'y insiste, et je répète volontiers qu'il y a toujours un grand décalage entre la théorie linguistique et la mise en pratique de cette théorie, à savoir dans notre cas précis entre l'analyse du changement lexical comme il est expliqué dans la théorie de la sémantique historique et l'application de cette analyse dans la lexicographie historique. Tandis que le théoricien a la liberté de s'élever dans les hauteurs de son esprit, de choisir le matériel de démonstration qui lui convient et de laisser de côté ce qui le dérange ou ce qui le retiendrait trop longtemps, le lexicographe, lui, se voit dans l'obligation d'analyser le matériel dans sa totalité pour essayer d'en donner une image qui s'approche le plus possible de la réalité.

Ce qui est toujours captivant et ce qui fait toujours plaisir et ce que nous ferons aujourd'hui ensemble, c'est-à-dire moi, je le ferai pour vous, c'est une confrontation des acquis de la théorie sémantique avec les données des dictionnaires et d'en comparer les résultats qui, et je peux le dire de prime abord, ne sont pas toujours favorables pour les lexicographes. Je dois ajouter tout de suite que je vais me restreindre dans ce qui suit au seul champ du français médiéval, mais un coup d'œil dans les dictionnaires historiques d'autres langues romanes montre une situation tout-à-fait comparable.

Comme objet de mon analyse j'ai choisi le mot *jeu*, un mot d'un sémantisme assez complexe dont s'est servi déjà Ludwig Wittgenstein pour développer ses idées sur les ressemblances familiales des mots. Il sera donc question de montrer à l'exemple de ce mot les différents procédés qui ont produit de nouvelles acceptions. Si Wittgenstein a dit : « Nous voyons un réseau complexe de similitudes qui s'entrecroisent et s'enveloppent les unes les autres »[2], je pense qu'au-delà de ces similitudes, on peut être aujourd'hui bien plus précis dans la description des changements qui ont lieu dans le cadre du champ sémantique d'un mot. Pour faire cette description, nous disposons du côté théorique, depuis plusieurs années déjà, par exemple du travail magistral du regretté Andreas Blank sur les principes du changement lexical, un livre qui doit ou devrait obligatoirement intéresser chaque lexicographe. Si je vois bien, il n'y a aucun travail qui ait dépassé cette analyse de Blank.

---

1 Dans cette version écrite, j'ai maintenu délibérément le caractère oral de ma conférence.

2 « Wir sehen ein kompliziertes Netz von Ähnlichkeiten, die einander übergreifen und kreuzen. » Wittgenstein 2001, § 66.

Du côté pratique je prends comme points de repère les trois dictionnaires qui s'occupent actuellement du français médiéval, à savoir l'*Anglo-Norman Dictionary* (AND)[3], le *Dictionnaire du Moyen Français* (DMF) ainsi que le *Dictionnaire étymologique de l'ancien français* (DEAF). Je laisse de côté délibérément d'autres dictionnaires qu'on pourrait inclure dans cette analyse, comme le dictionnaire de Godefroy, celui de Tobler et Lommatzsch et aussi le FEW, le *Französisches Etymologisches Wörterbuch* de von Wartburg, puisque ce sont des dictionnaires dont le futur se trouve déjà dans le passé, si je puis ainsi dire, et puisque pour ces dictionnaires, l'analyse et la description sémantique visaient des buts différents. Le *Trésor de la langue française* (TLF) sera cité deux fois.

La façon dont je vais procéder dans la suite, est assez simple. A l'exemple du mot *jeu* je vais analyser plusieurs phrases de l'ancien et du moyen français suivant les critères proposés par Blank et par d'autres, pour voir ensuite ce qu'en ont fait les dictionnaires qui nous intéressent ici. Mais voyons d'abord la structure sémantique générale résultant de l'analyse du mot *jeu* dans les dictionnaires en question qui sont - pour vous les 'ramentevoir', si vous me permettez ce joli vieux mot - l'*Anglo-Norman Dictionary*, le *Dictionnaire du Moyen Français* et le *Dictionnaire étymologique de l'ancien français*. Dans l'AND nous avons une structure quadripartite qui consiste en 1° les jeux de hasard, le jeu d'échecs et la stratégie pour jouer ce jeu, 2° jeu d'enfant et chose insignifiante, 3° amusement, divertissement organisé ou non, présentation de tours d'adresse (par des jongleurs, bouffons, etc.), jeu tout court, cortège, chasse, marqué comme terme cynégétique, au figuré jeu et activité, et finalement 4° enjouement sans souci, gaîté, et, par euphémisme, rapport sexuel[4]. Cette structure est suivie d'un grand paragraphe qui contient 18 collocations du type *en ju*, *par ju*, mais aussi *ju d'amor*, *ju de enfance*, *ju d'Olimpiades* etc. C'est donc un peu une sorte de fourre-tout d'un caractère assez hétérogène. On aurait peut-être mieux fait de ranger ces collocations sous les sens numérotés respectifs.

Le DMF donne d'abord sous I. un sens général "activité à laquelle on s'adonne librement", qui comprend un premier groupe A. "activité sans règle établie et sans enjeu" avec des acceptions plus précises qui sont 1. "activité qui n'a d'autre but que le plaisir qu'elle procure ; divertissement, amusement", 2. au figuré "ce qui présente un ou plusieurs caractères du divertissement" et 3. par métonymie "joie, plaisir". Vient ensuite un groupe B. "activité supposant des règles définissant un succès et un échec" avec l'acception 1. "activité réglée, comprenant des règles, et où l'on perd ou gagne", ensuite une catégorie 2. par métonymie, où sont rangés "partie", "lieu où l'on joue", "ce qui sert à jouer", "nombre de points obtenus par un joueur au cours d'une partie de dés" et "somme d'argent engagée dans une partie d'un jeu où l'on risque de

[3] Les sigles employés ici sont ceux du DEAF, v. sous http://www.deaf-page.de/bibl_neu.htm.

[4] 1 game (of chance) ; (chess) game ; (chess) strategy, method of playing a game 2 (children's) game ; (fig.) trifle, unimportant matter 3 diversion, (organised) entertainment ; performance of tricks (by jugglers, jesters, etc.) ; play, pageant ; (ven.) sport of hunting ; (fig.) game, activity 4 light-hearted cheerfulness, merriment ; (euph.) sexual intercourse.

l'argent". Nous avons après, sur le même niveau que l'"activité supposant des règles définissant un succès et un échec" un groupe C. – par métonymie "façon de jouer" et, au figuré, "façon d'agir". Viennent sous II. - "Spectacle" les groupes A. - "Activité attrayante, numéro présenté par une personne", B. - "Représentation théâtrale d'une pièce en vers (dramatique ou comique)", C. - "Joutes sportives livrées devant des spectateurs" et D. - comme terme historique de l'histoire grecque "compétition" et de l'histoire romaine "spectacles et réjouissances collectives organisées périodiquement". Nous avons dans ce dictionnaire une grande bipartition avec une subdivision en sept groupes avec des précisions et cas spéciaux en 125 paragraphes au total.

Dans le DEAF, il y a d'abord un sens général 1° "divertissement d'un caractère ludique", ensuite 2° par métonymie "attitude d'esprit, passagère ou durable, caractérisée par la joie et par le fait de prendre à la légère tout ce qui arrive, manifestation de cette attitude", 3° par métaphore *jeu / jeu d'amour* etc. "échanges de caresses amoureuses pouvant inclure l'acte sexuel", 4° "activité ludique organisée par un système de règles définissant un succès et un échec, un gain et une perte", 5° "activité ludique célébrée lors d'une fête ou d'un tournoi et réalisée avec ou sans armes (joute, danse, etc.)", 6° "composition dramatique destinée à être représentée, représentation de cette composition", 7° par métaphore "action (de Fortune ou d'Amor) qui se déroule sans que l'on puisse en comprendre les différents pas", 8° "possibilité où l'on peut ou doit choisir". Nous avons ici, sur un même niveau, huit sens, avec une subdivision en 89 paragraphes.

Ce premier coup d'œil montre déjà une approche assez divergente des dictionnaires pour structurer le sémantisme complexe de notre mot. Il me semble pourtant que l'on peut établir un sens de base, soit dans le style du DMF "activité à laquelle on s'adonne librement" avec la précision "activité qui n'a d'autre but que le plaisir qu'elle procure ; divertissement, amusement", soit dans le style du DEAF "divertissement d'un caractère ludique". Dans la structure de l'AND cela correspond probablement au sens quatre, en excluant les activités sexuelles. Ce sens général vaut pour un contexte comme le suivant du *Roman de Rou* – dans la scène en question, le dîner est terminé, et maintenant, le roi – *En nobles dames e en beles E en corteises dameiseles Torna son deduit e s'entente, C'est un gieu qui mult atalente* (ca.1170, RouH III 10262 "il dirigea sa distraction et son attention vers des dames nobles et belles et vers des demoiselles courtoises, c'est un divertissement qui cause beaucoup de plaisir"). Et il vaut pour le contexte suivant, tiré de l'*Espinette amoureuse* de Jean Froissart: *Souvent pensoie, sus et jus. Et alefois a aucuns jus, As quels s'esbatent jone gent, Jeuoie de coer liet et gent* (ca.1369, FroissEspF 1141 "Souvent je pensais à ceci et à cela, et parfois je m'adonnais aussi, d'un cœur léger et joyeux, à des divertissements auxquels se livrent les jeunes gens").

Je passe à un contexte qui se trouve dans le fabliau *Le clerc qui fu repus derriere l'escrin* et que voici: *En Haynau ot une bourgoise En une ville, assez courtoise, Plaine de jeu et de soulas, K'Amours le tenoit en ses las* (13^{e}s., NoomenFabl n°119,9 "Dans le Hainaut, il y avait dans une ville une bourgeoise assez courtoise qui était portée sur

le divertissement et le plaisir, puisque Cupidon la tenait dans ses lacs"). Les éditeurs définissent cette attestation par "frivolité". A mon avis, *jeu* désigne ici une attitude d'esprit qu'il faut voir dans une certaine relation avec le concept de 'divertissement d'un caractère ludique' que l'on peut considérer aussi comme l'archisème du mot *jeu*. La relation qui unit les deux acceptions de "divertissement" et de "attitude d'esprit" est basée, au sens le plus large, sur une métonymie. Pour être plus précis et pour se servir de la terminologie proposée par les théoriciens, on pourrait parler d'une relation de contiguïté, ce qui veut dire qu'il y a une certaine proximité ou un certain rapport entre les deux acceptions auxquels est due la possibilité de les associer. Il s'agit donc de la métonymie 'ACTION – ATTITUDE D'ESPRIT QUI EST À LA BASE DE CETTT ACTION'. Et j'en déduis la définition: par métonymie "attitude d'esprit, passagère ou durable, caractérisée par la joie et par le fait de prendre à la légère tout ce qui arrive, manifestation de cette attitude", qui est la deuxième dans la structure du DEAF, où la métonymie est explicitée dans une note. Dans les deux autres dictionnaires il n'y a rien de comparable.

Néanmoins, on y trouve des collocations du type *en jeu*, *par jeu* au sens de "par plaisanterie" que j'ai rangées dans l'article du DEAF sous cette acception que vous avez toujours devant les yeux. Un exemple typique se trouve dans l'*Estoire des Engleis* de Geoffroi Gaimar: *Li reis par giu e par gabel Levat le pan de sun mantel* (ca.1139, GaimarB 3793 "Par plaisanterie et pour s'amuser, le roi leva le bord de son manteau (il est question du manteau de la reine)". Cette attestation est citée par l'AND et définie "in a light-hearted manner, playfuly" et elle se trouve dans le paragraphe de collocations mentionné entre *en ju* et *ju circeis*, terme historique au sens de "(Roman) circus". Le DMF enregistre également *par jeu* au sens de "par plaisanterie" avec des contextes comparables, et range la collocation sous I.A.2. au fig. "Ce qui présente un ou plusieurs caractères du divertissement", ce qui est assez proche du DEAF. Dans ce cas-là, on peut dire que les trois dictionnaires ont une vue unanime des faits sémantiques.

Je passe tout de suite aux attestations dans lesquelles *jeu* tout seul ou employé dans des syntagmes se trouve dans des contextes où sont échangées des caresses amoureuses et où il est question de rapports sexuels. Les acceptions qui en résultent sont dues à des procédés métaphoriques: l'action d'échanger des caresses et l'acte sexuel sont comparés dans une métaphore conceptuelle à un divertissement d'un caractère ludique: 'L'ACTE SEXUEL EST UN DIVERTISSEMENT D'UN CARACTÈRE LUDIQUE'. C'est une métaphore qui se comprend aisément, mais qui ne s'impose pas nécessairement.

Je vous rappelle que pour désigner l'acte sexuel nous avons des métaphores où celui-ci est comparé à une joute ('L'ACTE SEXUEL EST COMPARABLE À UNE JOUTE': *behorder*), à un acte de travail ('L'ACTE SEXUEL EST UN ACTE DE TRAVAIL': *faire sa besogne*), à un acte de violence ('L'ACTE SEXUEL EST UN ACTE DE VIOLENCE': *fergier* "mettre des entraves (à qn)"), etc., et même à l'action centrale de la vie ('L'ACTE SEXUEL EST L'ACTION CENTRALE DE LA VIE': *le faire*). Ma définition est donc: 3° par métaphore *jeu* / *jeu d'amor* etc. "échange de caresses amoureuses pouvant inclure l'acte sexuel" et ceci

est dit dans le dictionnaire sans expliciter la métaphore. Voilà un contexte parlant. Dans le *Roman d'Aiol et de Mirabel* on lit: *Car vous tornés vers moi, jovente bele! Se vous volés baisier n'autre ju faire, J'ai tres bien en talent que je vous serve* (2^e^m. 12^e^s., Aiol[1]F 2174 "Tournez-vous donc vers moi, belle fille ! Si vous voulez un baiser ou une autre caresse, je veux volontiers vous servir"). C'était un exemple avec *jeu* tout court, dans le suivant, la métaphore est élargie d'un aspect tactile. Dans le *Lai de l'ombre* on lit: *De tel geu com l'en fet des mains Estoit ele dame et il mestre* (ca.1221, OmbreB[2] 948 "Les caresses que l'on se fait avec les mains, ils savaient très bien les échanger tous les deux") avec une variante intéressante: *De tel geu com l'en fet des mains* (var. ms. ca.1285 *De tel geu con on fet de rains*) *Estoit ele dame et il mestre.* Je traduis la variante littéralement: "Les caresses que l'on se fait avec les reins, ils savaient très bien les échanger tous les deux". Ce *jeu qu'on fait des reins* est apparemment l'acte sexuel – le DMF donne un exemple pour *jeu des reins* d'environ 1500 – et, dans notre lai, le copiste de ce manuscrit a intentionnellement changé le mot et avec ceci aussi la chose. Avec *jeu des reins* nous entrons dans le domaine de l'euphémisme, c'est-à-dire d'une expression que l'on emploie pour atténuer son propos. Nous avons ici le cas spécial d'une métonymie, un emploi euphémique partant d'une métonymie de proximité, puisque la partie des reins n'est pas trop éloignée des parties génitales: 'L'ACTE SEXUEL ES UN DIVERTISSEMENT D'UN CARACTÈRE LUDIQUE OU LES PARTIES GÉNITALES, DÉSIGNÉES PAR EUPHÉMISME PAR LA PARTIE DES REINS, JOUENT LE RÔLE PRINCIPAL'. On peut se poser la question de savoir si *jeu* et les syntagmes respectifs qui désignent l'échange de caresses et l'acte sexuel ne se trouvent pas toujours dans des emplois euphémiques, et je reconnais que la classification 'métaphore' ou 'métonymie' dans ce dernier cas, ou 'euphémisme' a parfois quelque chose de subjectif, et qu'elle est parfois une question de goût. Comment cela? Voyons le prochain exemple, tiré de *Jehan et Blonde* de Philippe de Beaumanoir: *Aprés itex mox s'entrebaisent, De tous les jus d'amours s'aaisent Fors d'un que loiatés despit* (prob. ca.1235, BeaumJBlL 1520 "Après ces mots, ils échangent des baisers et s'adonnent à toute sorte de caresses amoureuses sauf à celle que l'honnêteté méprise"). Ici la métaphore inclut le rôle important attribué à l'amour qui – dans le cadre d'une intersubjectivité reconnue – est certainement plus proche de l'échange de caresses que les reins. C'est pourquoi *jeu d'amour* reste pour moi, cette fois-ci, dans le domaine de la métaphore. Un contexte comparable se trouve dans les chansons politiques anglo-normandes: *Des chanoynes seculers, Qe dames servent volenters […] chescun frere a sa sorour Deit fere le giw d'amour* (ca.1300, AspinChansPol XII 126 "Des chanoines séculiers, chaque frère doit faire l'amour à sa sœur"). L'AND cite ce contexte et le range sous la définition "love play (often used as a euphemism for sexual intercourse)" et considère *jeu d'amour* comme euphémisme. Le DMF donne sous la définition "amour physique" toute une série de syntagmes comme *le jeu gaillard*, *le jeu de nature*, ou *le jeu des bas instruments* sans parler de métaphore ou d'euphémisme. Dans les exemples cités dans le DMF il est certainement question de l'acte sexuel comme dans d'autres syntagmes du type *faire le jug uimain* (3^e^q. 12^e^s., GirRossDécH 8684) où l'euphémisme part de la métaphore 'L'ACTE SEXUEL EST UN DIVERTISSEMENT D'UN CARACTÈRE LUDIQUE ET QUE EST INHÉRENT

À LA NATURE HUMAINE'. L'exemple suivant vient encore de *Aiol et Mirabel*: *Cele nuit voirement a joie s'esbanissent, S'il font ju de cortine, ne vous mervelliés mie, Dusc'al demain al jor que l'aube est esclairie* (2emm.12es, Aiol2F 10963 "Cette nuit, ils ont en effet beaucoup de plaisir, ils font l'amour, ne vous en étonnez pas, jusqu'à l'aube du jour"). Et là, nous avons un emploi euphémique évident basé sur la métaphore 'L'ACTE SEXUEL EST UN DIVERTISSEMENT D'UN CARACTÈRE LUDIQUE QUI A LIEU À HUIS CLOS'. Un exemple encore, tiré d'une pastourelle d'Ernous li Viele: *Par force sos moi la mis, Demanois le ju françois Li fis a mon talant* (13es., RomPast III 6,41 "Je l'ai mise avec violence au-dessous de moi, et tout de suite, j'ai couché avec elle à mon gré"), et nous avons ici encore un euphémisme partant de la belle métaphore 'L'ACTE SEXUEL EST UN DIVERTISSEMENT D'UN CARACTÈRE LUDIQUE QUI ES PRATIQUÉ DE PRÉFÉRENCE OU LE MIEUX PAR LES FRANÇAIS'. En tout cas, dans la désignation des échanges de caresses il y a un certain flou, et il reste toujours des cas douteux comme dans l'exemple suivant tiré de la *Clef d'amour*: *les dames se seulent trere Es liez lor basses et estendre Pour les geus savorous attendre* (prob. 1280, ClefD 1888 "les dames ont l'habitude de se rendre et de s'étendre sur le lit de leurs servantes pour attendre les jeux savoureux"). Ici, il n'est pas évident de savoir ce que sont les jeux savoureux, c'est peut-être un cas pour les ressemblances familiales de Wittgenstein. Je me demande pourtant si la définition dépend ou doit dépendre de l'imagination du lexicographe ou s'il faudrait établir une catégorie - je ne voudrais pas parler de définition - pour de tels cas qui sont pratiquement impossibles à discerner. Il va de soi que les dictionnaires n'ont pas une telle catégorie et que le théoricien évite élégamment des exemples qui ne seraient pas univoques.

Avec ceci, je passe au sens numéro 4° du DEAF qui comprend tout ce que l'on appelle aujourd'hui les jeux de société: DEAF 4° "activité ludique organisée par un système de règles définissant un succès et un échec, un gain et une perte". Je vous rappelle que c'est le sens numéro 1 dans l'AND: 1° game (of chance); (chess) game; (chess) strategy, method of playing a game, où il faudrait pourvoir l'acception de "stratégie, méthode" d'un marquage ; j'y reviendrai. Et c'est le sens qui ouvre, dans le DMF, le groupe "B. - Activité supposant des règles définissant un succès et un échec" , avec 1. "Activité réglée, comprenant des règles, et où l'on perd ou gagne" et 2. P. méton. Nous avons là une restriction du sens général qui se comprend aisément, mais cette restriction n'est mentionnée dans aucun des dictionnaires. C'est une acception peu spectaculaire qui en résulte, mais nous y verrons plusieurs emplois métonymiques. Voyons d'abord des contextes qui témoignent de l'estime différente que l'on portait à cette sorte de jeux. Dans le premier, tiré du *Poème moralisé*, les jeux – on peut le deviner – sont l'œuvre du diable: *Es eschac et es taubles at trop de vaniteit, Mult [par] fut liés li diable cant cel geu ot troveit ; Par cest geu sunt mains home confondus et dampneis* (ca.1200, PoèmeMorB 3226/27 "Dans le jeu d'échecs et dans les jeux de table il y a trop de vanité. Le diable était bien content quand il avait inventé les jeux. Et par cette sorte de jeu beaucoup de gens ont été corrompus et honnis"). Dans le suivant, tiré d'Urbain le Courtois et bien qu'il s'agisse ici également d'une sorte d'enseignement moral, la connaissance des jeux fait plutôt preuve de l'homme du monde: *A eschés jue et a tables Et a touz autres juez covenablez. Qui a plusurs jues*

se set deduire Ceo est grant honur et nurture (ms. 14^{e}s., ParsonsCourt E 230/231 “Il joue aux échecs et aux jeux de table et à toute autre sorte de jeux convenants. Quand on sait s’amuser avec plusieurs jeux, c’est un bon signe d’honneur et d’éducation”).

Parmi les scènes où l’on joue aux dés, il y en a une, dans le *Jeu de saint Nicolas* de Jehan Bodel, où *jeu* se trouve dans un emploi métonymique: *Biaus dous amis, coi que tu aies, Tu n’en goutas et si le paies : J’ai quaernes, le plus mal jeu !* (ca. 1195, BodelNicH 309 “Cher ami, quoi que tu aies, tu n’en profites pas et tu paieras. J’ai un double de quatre, le coup le plus mauvais”). *Quaerne* est un terme de jeu qui désigne le coup où chacun des deux dés montre quatre points, donc un très bon coup, est *le plus mal jeu* est ici une ironie, avec laquelle le joueur, Auberon, se moque de son copain. Le changement sémantique qui s’est produit ici, ou que l’auteur a introduit intentionnellement, se place dans le cadre d’une forme particulière de la métonymie, à savoir de la synecdoque : la désignation du jeu est prise pour le coup, donc la partie pour une partie du tout, le résultat est donc : *jeu* par synecdoque “coup de dés”. Cet exemple est pris du DEAF, et dans le DMF nous trouvons un contexte comparable, où, dans le grand paragraphe intitulé ‘par métonymie’, *jeu* est défini ‘par métonymie’ par “nombre de points obtenus par un joueur au cours d’une partie de dés” avec le contexte suivant, tiré de la Passion d’Auvergne : *Or sa, Malque, recontons donc. Deux six et quatre seze font ; C’est mon jeu. Mes quines et six est le tien, que ne vault rien pis Ne mieulx ; par ainsi il est per* (1477, PassAuvR 204 “Alors, Malque, comptons encore une fois. Deux fois six et quatre font seize. C’est mon coup. Mais un double de cinq et six, c’est le tien, qui n’est ni plus mauvais ni meilleur ; c’est donc match nul”). On dirait que dans ce cas, la tendance de la description des deux dictionnaires est assez similaire, mais les résultats sont quand-même assez divergents.

Le même procédé d’une synecdoque a eu lieu aussi en parlant du jeu d’échecs, où *jeu* peut prendre, par synecdoque, le sens de “suite stratégique de coups d’échecs”. Cet emploi de *jeu* est largement attesté à l’époque ancienne, et il est aussi enregistré dans l’AND, seulement, nous l’avons vu, sans marquage stylistique. Un exemple tiré d’un traité d’échecs : *En ceste giu n’ad poynt de matesoun* (13e s., EchecsRoyH 27 “Avec cette stratégie de jouer il n’y aura pas de mat”). Là encore, nous avons la désignation du tout pour une partie du tout.

Toujours en parlant du jeu d’échecs, on trouve un autre emploi métonymique de *jeu* qui, aujourd’hui, est courant, mais qui est bien attesté déjà en ancien français. Voyons un exemple, tiré du Roman de Horn : *A itaunt sunt asis, lur gius ont ordonez* (ca. 1170, HornP 2758 “Alors ils se sont assis et ont ordonné leurs pièces d’échecs”). Le sens de *jeu* est ici “ensemble des pièces nécessaires à un ou à plusieurs joueurs pour réaliser un jeu déterminé”. Si j’ai parlé d’un emploi métonymique, celui-ci est encore à préciser. Le procédé de la désignation du tout pour les pièces matérielles qui permettent la réalisation du tout est certes à la limite de la synecdoque. Mais il me semble plus prudent de parler ici d’une relation de contiguïté, où cette dernière est basée sur les rapports étroits entre le jeu et les pièces.

Une synecdoque des plus claires se trouve dans l’acception suivante : *jeu* par synec-

doque "partie de jeu" (DEAF), où la désignation du tout pour une partie du tout est évidente. Cette acception est très bien attestée, par exemple dans le Roman de Horn : *s'asiet a l'eschekier. Quatre geus ad jowé* (ca. 1170, HornP 2728 "il s'assoit à l'échiquier et a joué quatre parties"), cité par le DEAF et par l'AND sous : (chess) "game" sans plus, ou dans un document de 1404 : *...debat se meut entre eulx sur un de leurs jeux, telement que le dit Gilet dist au dit Chauvet que il avoit perdu un jeu, et icellui Chauvet lui respondi que non avoit* (Doc. Poitou G, t. 7,50 "une querelle s'établit sur une de leurs parties du jeu, de sorte que le dit Gilet dit au dit Chauvet qu'il avait perdu une partie, et ce Chauvet lui répondit qu'il n'avoit pas perdu"), cité par le DMF sous : 'par métonymie' "partie'. Ici, les trois dictionnaires sont en principe d'un même avis, mais les résultats se présentent sous trois formes différentes.

Jehan Bodel emploie les termes *petit jeu* "partie avec un petit enjeu" (BodelNicH 291) et *bon jeu* "partie avec un grand enjeu" (BodelNicH 1055), mais on peut se demander si nous avons ici affaire à des syntagmes lexicalisés ou si *jeu* avec cette acception peut se combiner aussi avec d'autres adjectifs, sans que ce soit à retenir dans un dictionnaire. Ceci vaut aussi pour une tournure comme *avoir le plus beau d'un jeu* qui se lit dans le Roman du Comte d'Artois dans la phrase suivante : *...ja sçavez vous que depuis que cy vindrent voz ennemis que riens ou tout peu y ont fait de leur prouffit et comme l'en dit communement quant on a le plus beau d'un jeu on ne le doit point laissier* (mil. 15e s., ComteArtS 38 "vous savez déjà que, depuis que vos ennemis sont venus ici, ils n'ont rien ou très peu fait à leur profit et, comme on dit communément, on ne doit pas arrêter son jeu quand cela marche très bien") et qui est donnée dans le DMF sous la rubrique 'par métonymie', mais sans définition. S'il y a ici quelque chose de remarquable, ceci réside dans la tournure plus générale *avoir le plus beau de qch.* qui n'a rien à voir avec *jeu*.

Dans un proverbe, nous avons encore un emploi métonymique, où la désignation de l'activité est employée pour dénommer le système des règles de cette activité : *Qui en jeu entre, jeu consente* (13e s., ProvM 1914 "Qui commence un jeu, en accepte les règles"). C'est le deuxième *jeu* qui désigne le système des règles dont j'ai parlé. Avec une petite modification, ce proverbe se trouve aussi dans la chanson de geste de *Garin le Loherain* : *Qui en gieu entre, bien doit gieu consentir* (4e q. 12e s., GarLorrV 8961 "Qui commence un jeu, doit en accepter les règles"). Il semble que cette acception "système de règles qui définissent un jeu" n'est pas attestée ailleurs et ce ne sont que le *Trésor de la langue française* et le DEAF qui l'ont enregistrée.

Nous restons dans le sens de "activité ludique" et passons tout de même quasiment au contraire. Ce qu'il y a d'attrayant dans *jeu* et ce pourquoi j'ai choisi ce mot comme exemple de mon propos, c'est que c'est un des mots rares qui ont vraiment un spectre sémantique impressionnant et où l'on peut observer des changements sémantiques de toute sorte. D'abord un contexte. Dans le *Roman du Comte d'Artois* on lit : *Grant bien firent a la contesse d'Artois lez doulcez parollez de son seigneur oÿr ainsy amiablement dire et au vray entendre ellez servoient bien a son jeu, sy luy respondy, comme celle qui par grant sens et subtillité conduisoit sa poursieute...* (mil. 15e s.,

ComteArtS 138 “Cela fit du bien à la comtesse d’Artois d’entendre les aimables paroles de son mari, et, à vrai dire, elles rendaient de bons services à son propos, et elle lui répondit comme une femme qui, avec beaucoup de sagesse et de subtilité, poursuivait son but ”). Cet exemple est pris du DMF qui le range dans un paragraphe : ‘P. anal. [Désigne une situation ou une action mettant en jeu des intérêts divergents]’ et définit “entreprise visant à servir les intérêts propres de quelqu’un”. Et cette définition correspond à peu près à celle du DEAF donnée pour des contextes comparables : par métaphore “action sérieuse avec laquelle on vise à servir ses propres intérêts”. Ce que les exemples cités dans les dictionnaires ont en commun, c’est que les situations décrites n’ont rien d’une activité ludique et que les acteurs poursuivent un certain but. L’acception que nous avons ici est basée sur une métaphore conceptuelle que l’on peut décrire ainsi : ‘LES ACTIONS SÉRIEUSES DE LA VIE SONT COMPARABLES À DES ACTIVITÉS LUDIQUES OÙ L’ON ESSAIE DE GAGNER’. Nous sommes ici près de l’antiphrase où un mot est employé dans un sens contraire au sens originaire. Un autre exemple pour cet emploi se trouve dans un poème de Huon de Saint-Quentin : *il lor fist le gieu Kaïn Qui son frere ocist en la fin* (1221, HuonQuJérS 220 “il procéda envers eux comme Caïn qui, finalement, tua son frère”).

Dans les textes de notre littérature médiévale, nous avons toute une série de locutions qui sont à rattacher au sens de “action qui vise à servir les intérêts propres de quelqu’un”. Au moins, le DEAF et le DMF en donnent une trentaine du type *avoir le jeu* „avoir du succès dans ce que l’on est en train de faire” (DEAF), *estre un fort jeu* “constituer une rude tâche” (DMF), *faire le jeu a* (aucun) “jouer un mauvais tour à (qn)” (DEAF), etc. L’AND donne seulement *hanter sun ju* “to carry on one’s game”, définition qui, en tout cas, n’est pas univoque. Comme la plupart de ces locutions ne sont documentées que par une seule attestation, il faut toujours se demander, si elles sont vraiment lexicalisées. Regardons quelques exemples de plus près. Dans le DMF, il y a un contexte tiré de *Jehan de Saintré* d’Antoine de la Sale : *Estant Saintré en la grace du roy, de la royne, des Seigneurs et de Madame, et de tous autres, pour abregier, le plus amé, le plus honnoré escuier de France a cause de sa grant doulceur et humilité, et aussi de sa largesse, qui aide bien au jeu* (av. 1456, JSaintréM 187 “Comme Saintré jouissait de la grâce du roi et de la reine, des seigneurs et de Madame et de tous les autres, bref, comme il était l’écuyer le plus estimé et le plus honoré de la France, grâce à sa clémence et son humilité, et aussi à sa largesse, ce qui est d’un grand secours dans une telle situation”). Le DMF en déduit la locution *aider au jeu* “faciliter les choses”.

Dans le fabliau du Prestre comporté, la femme prépare la visite de son amant qui est le prêtre, et on lit : *li bains ert ja sour le fu Et li capon mis en l’espoi. Li priestres aresta molt poi, Ki n’a soing de son ju deffaire* (1e m. 13e s., PrestreCompN 53 “le bain était déjà sur le feu et le chapon était déjà à la broche. Le prêtre ne tarda pas beaucoup qui n’avait pas l’intention de rater cette occasion”) et le DEAF en tire la locution *desfaire son jeu* “renoncer à avoir du succès dans ce que l’on est en train de faire”. Comme dernier exemple – et comme je l’ai dit, il y en a bien d’autres – j’ai choisi *gaaignier le jeu par aler* „trouver son salut dans la fuite” qui est attesté dans le fabliau de

Haimet et de Barat : (les deux croient qu'ils ont vu leur père pendu, et ils sont frappés d'effroi) *Le gieu gaaignent par aller : Endui sont en fuie touchié* (ca. 1195, HaimBarNo 412 "Ils trouvent leur salut dans la fuite: tous les deux se mettent à s'enfuir"). Pour cette locution nous avons encore une deuxième attestation qui se trouve dans le fabliau *Du chevalier qui fist parler les cons* : *Et le prestres, sanz mot soner, Gaaigne le gieu par aller, Qui s'enfuit par une charriere : Pour cent mars ne tornast arriere* (13e s., NoomenFabl n° 15,302 "Et le prêtre, sans dire aucun mot, trouve son salut dans la fuite, car il s'enfuit par une voie charretière: même pour cent marcs il n'aurait pas fait demi-tour"). Cette attestation se trouve dans le DEAF déjà sous GAAIGNIER dans la locution *gaaignier le jeu* qui y est définie "l'emporter au jeu", sans commentaire, ce qui ne va certainement pas, puisqu'ici il n'est pas question d'un jeu.

Revenons au jeu en tant qu'activité ludique. Il est certainement indiqué de séparer des jeux que j'ai appelés 'jeux de société' les activités qui se réalisent ou plutôt qui se célèbrent à l'occasion d'un tournoi ou d'une grande fête publique. Voyons ce que disent les textes et ce qu'en font les dictionnaires. D'abord un contexte de l'*Estoire des Engleis* de Gaimar : *un naim aveit Ki […] saveit saillir e tumber E autres gius plusur juer* (ca. 1139, GaimarB 3988 "il y avait un nain qui savait sauter et danser et faire plusieurs autres numéros"). Il est rangé par l'AND sous "performance of tricks (by jugglers, jesters etc.)". Le DMF cite du *Registre criminel du Châtelet de Paris* le passage suivant : *…elle s'en ala en la compaign[i]e d'icelle Jehennette veoir les jeux que les menesterelz fesoyent ès hales* (1389-1392, Reg. crim. Chât., I, 265 "elle s'en alla en la compagnie de cette Jehanette pour voir les représentations que les ménestrels faisaient dans les halles"), et définit sous II. 'Spectacle' A. "activité attrayante, numéro présenté par une personne". Pour l'ancien français, je cite un contexte du roman de Brut : *Un jur firent Troïen feste A la manière de lur geste ; Caroles faiseient e geus Pur la joie des novels lieus U il esteient assené* (1155, BrutA 1079 "Un jour les Troyens faisaient une fête à la manière de leurs ancêtres. Il y avait des caroles et des présentations de toute sorte parce qu'ils se réjouissaient des lieux nouveaux où ils s'étaient rencontrés"). Comme il ne résulte pas de ce dernier contexte de quelle sorte de représentations il s'agit, je me suis contenté de définir pour le DEAF "activité ludique célébrée lors d'une fête ou d'un tournoi et réalisé avec ou sans armes (joute, danse, etc.)". C'est assez général et peut paraître un peu décevant, mais chaque précision risque de fausser le sens. Un exemple seulement. Dans le *Roman d'Eneas* se trouve le contexte suivant : *Eneas fist une grant feste Et geus a la tombe son père : L'annivarsaire idonkes ere* (ca. 1160, EneasS[1] 2159 "Énéas organisa une grande fête avec des représentations de toute sorte près de la tombe de son père : c'était son anniversaire"). Le TLF définit dans sa notice historique pour cette attestation "compétitions sportives". A mon avis, rien n'indique cette restriction à un caractère sportif des jeux en question et on ne peut et on ne doit pas préciser davantage.

Le DMF donne sous la rubrique 'spectacle' un sens séparé : C. "joutes sportives livrées devant des spectateurs" avec une seule attestation prise de la *Prise d'Alexandrie* de Guillaume de Machaut : *Et puis moult bien [le roi] les festia, Et fist jouster en*

leur presence Ses chevaliers maint cop de lance. Li Sarrazin se mervilloient Coment il ne s'entretuoient; Car il sont dou gieu desapris, Pour ce qu'il ne l'ont pas apris (ca. 1372, GuillMachPriseM 4178 "Ensuite le roi leur fit un grande fête et fit jouter ses chevaliers en leur présence. Les Sarrasins s'étonnèrent qu'ils ne se tuaient pas les uns les autres, parce qu'ils ignoraient cette sorte de jeu qu'ils n'avaient pas appris"). Ici encore, je me demande si la définition ne prend pas trop en considération le contexte. Autrement, il faudrait la pourvoir d'un marquage comme restriction de sens.

Restons encore dans le spectre des jeux représentés et regardons de plus près quelques attestations chez Robert de Clari dans son récit de la quatrième croisade. Dans le contexte, il est question des jeux qui se pratiquaient dans l'hippodrome de Constantinople à l'époque : *Se i avoit deus jus ensanle, quant on juoit, si se wagioient ensanle li empereres et l'empeerris que li uns des jus giueroit miex de l'autre, et tout chil ensement qui les gius esgardoient* (1216, RobClariL XC 11/13 "Quand il y avait deux équipes au jeu, l'empereur et l'impératrice faisaient un pari que l'une des équipes jouerait mieux que l'autre, et ainsi le faisaient tous les autres qui regardaient le spectacle"). Ce dernier *gius* désigne les activités elles-mêmes, mais dans les deux autres attestations nous avons évidemment un emploi métonymique pour les équipes qui jouent. C'est encore par une relation de contiguïté que la désignation de l'activité est prise pour désigner les acteurs. Elle se compare à celle que nous avons vue tout à l'heure, où *jeu* était employé pour désigner l'ensemble des pièces nécessaire pour un jeu. La définition qui en résulte est donc par métonymie "équipe qui participe à un *jeu*".

Dans le même paragraphe chez Robert de Clari nous lisons : *il i avoit une plache, qui pres estoit du palais de Bouke de Lion, que on apeloit les Jus l'empereeur.* […] *entor chele plache si i avoit bien trente degrés ou quarante, la ou li Griu montoient pour eswarder les jus* (1216, RobClariL XC 4 "il y avait une place qui était près du palais de Bouche de Lion et que l'on appelait 'les jeux de l'empereur'. Autour de cette place il y avait trente ou quarante marches que les Grecs montaient pour regarder les spectacles"). Une fois de plus, nous avons ici un emploi métonymique avec une relation de contiguïté : la désignation de l'activité est prise pour désigner le lieu où cette activité est réalisée. Comme *les Jus l'empereeur* est employé ici en tant que nom propre, il n'y aura pas de définition entre guillemets dans le dictionnaire : par métonymie *les Jeux l'emperëor* nom donné à l'hippodrome de Constantinople.

Passons à la prochaine acception, et regardons d'abord un contexte, pris de la légende de l'empereur Constant : *li plus poissant del mont Qui Furent sour se roee monté Ont esté de li desmonté Et gieté en la boe jus. Mout par est desgissés ses jus* (2e m. 13e s., EmpConstOctC 26 "Les plus puissants du monde qui étaient montés sur sa roue, c'est-à-dire la roue de Fortune, étaient précipités par elle et poussés en bas dans la boue. Son procédé est très bizarre"). Dans ce contexte, nous avons affaire au procédé de Fortune qui distribue, suivant l'ordre de la providence divine, bonheur et malheur. Mais cet ordre est inintelligible pour les pauvres mortels. Il en est de même avec le procédé de Cupidon, dieu de l'amour, qui lance ses dards quand il veut et où il veut, sans que l'homme ait la possibilité de comprendre pourquoi ou pourquoi pas.

Regardons un exemple pour ce jeu de Cupidon, pris du roman des Saxons. Après une querelle pleine de malentendus, Sebile dit à Baudoin : *Tel sont li geu d'Amors ; qi an connoit la voie ? An rampones poignaz son san amploie. Dure chose est d'Amors qant l'uns l'autre guerroie* (fin 12e s., SaisnLB 3846 "Ainsi est le procédé de Cupidon ; qui en connaît la voie ? En mauvaises paroles Cupidon fait usage de son bon sens. C'est difficile avec Cupidon quand l'un et l'autre se font la guerre"). Dans les deux attestations nous avons un emploi métaphorique, emploi qui est basé sur la métaphore : 'LES PROCÉDÉS DE FORTUNE ET DE CUPIDON SONT UN JEU OÙ L'ON NE PEUT PAS SAVOIR QUI VA GAGNER ET QUI VA PERDRE'. Et par conséquent, la définition du DEAF se lit comme suit : par métaphore "action (de Fortune ou de Cupidon) qui se déroule sans que l'on puisse en comprendre les différents pas". Cette définition est à mon avis aussi valable pour l'attestation suivante du *Livre du voir dit* de Guillaume de Machaut : *Je* [Fortune] *afflue et me depart sans bonne, Telz est mes jeus ou je me donne* (1364, GuillMachVoirI 8214 [c'est Fortune qui parle] "J'afflue et je me retire sans cesse, telle est la façon dont je procède"). Mais le DMF, qui donne cette citation, la range, avec d'autres citations comparables, sous la définition générale "activité qui n'a d'autre but que le plaisir qu'elle procure ; divertissement, amusement" avec l'ajout : [en parlant de Fortune]. L'AND ne parle ni de Fortune ni de Cupidon.

Mais ce dictionnaire est le seul à enregistrer *jeu* comme terme cynégétique : (ven.) "sport of hunting" avec deux exemples dont un que voici, tiré des lettres et pétitions anglo-normandes : *j'ay entendu q'il vous a plu de prendre mez chiens ovec les autres quant en alastez pur tuer le cerf et huissiez un tresbeal jeu* (1396, LettrOxfL 272,17 "j'ai entendu qu'il vous a plu de prendre mes chiens avec les autres quand vous êtes partis pour tuer le cerf et que vous avez eu une très belle chasse"). Et nous avons ici encore un emploi métaphorique où la métaphore peut être décrite ainsi : 'LA CHASSE EST UNE ACTIVITÉ QUI, MALGRÉ SON CARACTÈRE LÉTAL POUR LE GIBIER, A QUELQUE CHOSE DE LUDIQUE', au moins pour les chasseurs. Je vous rappelle que la chasse était, déjà au moyen-âge, un événement social de premier ordre.

Je passe une dernière fois au DMF pour expliciter une dernière métonymie. La phrase qui me sert d'exemple est tirée d'un document du *Trésor des chartes du comté de Rethel* : *...pour paver la nuefve cuisine, la cuisine delés la grande salle, la petite cuisine dessous la garderobe de monseigneur, le gieux de paume* (doc. 1409/1410, ChRethelS 2,607 "pour paver la nouvelle cuisine, la cuisine à côté de la grande salle, la petite cuisine au-dessous la garderobe de monsieur, la salle du jeu de paume"). Le DMF range l'attestation à juste titre sous : 'p. méton.' "lieu où l'on joue" et donne encore, pour préciser : *Jeu de paume* "espace délimité et aménagé pour la pratique du jeu de paume". On pourrait ajouter qu'il s'agit d'une synecdoque où la désignation de l'activité est employée pour désigner le lieu où cette activité se réalise.

Il y a encore bien d'autres sens qu'il faudrait énumérer ici, comme tout ce qui concerne la pièce théâtrale et sa représentation – je ne nomme que le *Jeu de saint Nicolas* que nous avons vu tout à l'heure. Il manque le sens de "ce qui sert à jouer", donc l'emploi métonymique pour le jeu de cartes, attesté depuis le 15e siècle seu-

lement. Il manque le sens de "manière de jouer d'un instrument de musique", et il manque évidemment *le jeu parti*, ce "genre lyrique en forme de débat entre deux trouvères sur un sujet de casuistique amoureuse et à la fin duquel un juge se prononce en faveur d'une des positions soutenues ; pièce de ce genre". Cette acception est due à une restriction de sens partant de la locution *partir un jeu* "prendre une décision" qui, quant à elle, se rattache au sens métaphorique "situation dans laquelle on peut ou doit choisir" basé sur la métaphore : 'LA VIE ES COMPARABLE À DES SITUATIONS OMME DANS UN JEU, DANS LESQUELLES IL FAUT PRENDRE UNE DÉCISION POUR LE PROCHAIN COUP'.

Je m'arrête ici pour ne pas vous priver de votre dé-jeu-ner qui, bien sûr, n'a rien à voir avec les jeux dont je vous ai parlés, mais qui peut pourtant, au moins parfois, avoir quelque chose de divertissant. Et j'en viens à ma conclusion.

1° Indépendamment du nombre absolu des attestations – et je n'ai cité que quelques-unes des milliers qui existent pour le mot *jeu* – les anciens textes nous fournissent des énoncés au niveau de la *parole* – *parole* au sens saussurien, bien entendu – desquels il faut filtrer les acceptions appartenant à la *langue*. Le lexicographe ne doit pas, sur la base d'un seul contexte, établir un sens nouveau et, par conséquent, une définition nouvelle, s'il n'a pas de bons arguments. Je vous rappelle seulement les cas de "frivolité" et de "compétitions sportives" mentionnés tout à l'heure.

2° Dès que l'on constate une nouvelle acception, on doit encore se méfier : est-elle vraiment raisonnable dans le champ sémantique en question ? Et si le champ sémantique d'un mot – comme dans le cas de *jeu* – est assez vaste et une nouvelle acception paraît probable, il faut toujours se demander quel est le rapport entre cette nouvelle acception et une acception déjà existante. Tout changement sémantique doit pouvoir s'expliquer par rapport à une acception préexistante en accord avec les possibilités du changement sémantique.

3° La sémantique théorique met à notre disposition l'explication des différents types de changement de sens. A part la métaphore et les différentes formes de la métonymie, nous avons l'intensification et la restriction de sens, l'antiphrase, l'analogie, etc. Le lexicographe doit essayer de comprendre à quel type de changement il a affaire. Si j'ai dit au début de mes propos qu'il y a toujours un décalage entre la théorie et sa mise en pratique, nous avons pu et dû constater qu'il y a parfois un décalage considérable dans la classification et dans la description des matériaux dans les dictionnaires mêmes. Parfois, on ne peut pas s'empêcher d'avoir l'impression que c'est plutôt le hasard qui y règne. Une analyse plus poussée de la part des lexicographes serait donc la bienvenue.

4° Je rêve d'un Dictionnaire du Français médiéval qui soit en mesure de réunir les matériaux rassemblés dans les trois dictionnaires AND, DEAF et DMF et qui soit en mesure de leur donner une structure sémantique unifiée[5]. Je rêve d'une structure sémantique unifiée qui soit basée sur les acquis de la théorie sémantique et qui n'hésite pas à montrer plus explicitement et avec plus de conséquence le fonctionnement

[5] Il va de soi que l'apport de chaque dictionnaire doit être et rester reconnaissable.

du changement lexical pendant les cinq siècles qui forment le cadre chronologique de ces dictionnaires. Peut-être un jour, un tel travail d'unification et de restructuration sera réalisable au sein d'un grand projet lexicographique européen, mais, comme je l'ai dit, c'est un rêve.

Académie des Sciences de Heidelberg Thomas STÄDTLER

Bibliographie

Blank, Andreas, 1997. *Prinzipien des lexikalischen Bedeutungswandels am Beispiel der romanischen Sprachen*, Tübingen, Niemeyer.

Wittgenstein, Ludwig, 2001. *Philosophische Untersuchungen*. Kritisch-genetische Edition. Hgg. von Joachim Schulte. Frankfurt/M, Suhrkamp.

Romains et Germains entre Moselle et Rhin avant l'an mille

C'est avec grand plaisir que j'ai accepté de parler de ce sujet dans une conférence plénière à Nancy cette année. Trois raisons m'y ont amené : d'abord, en 1974 au congrès de notre Société à Naples, j'ai pu parler du superstrat germanique dans les langues romanes. Seconde raison : à l'université de Sarrebruck, depuis plus de 20 ans, nous avons organisé pendant le semestre d'hiver un séminaire interdisciplinaire entre historiens, archéologues, géologues, germanistes et romanistes traitant le sujet : les problèmes de la colonisation le long de la frontière linguistique entre Germains et Romains, principalement en Lorraine. Enfin la troisième raison : à chaque congrès de notre Société, il y a un sujet local : cette fois-ci nous siégeons à Nancy en Lorraine. L'occasion m'est donc donnée de parler de la constitution linguistique de cette région. Il se trouve que j'ai publié aussi, en 1973, dans notre *Revue de Linguistique Romane*, une contribution : « La répartition géographique des éléments franciques en gallo-roman ». Aujourd'hui, 40 ans plus tard, je voudrais vous présenter un tour d'horizon qui montre les progrès dans ce domaine de recherche.

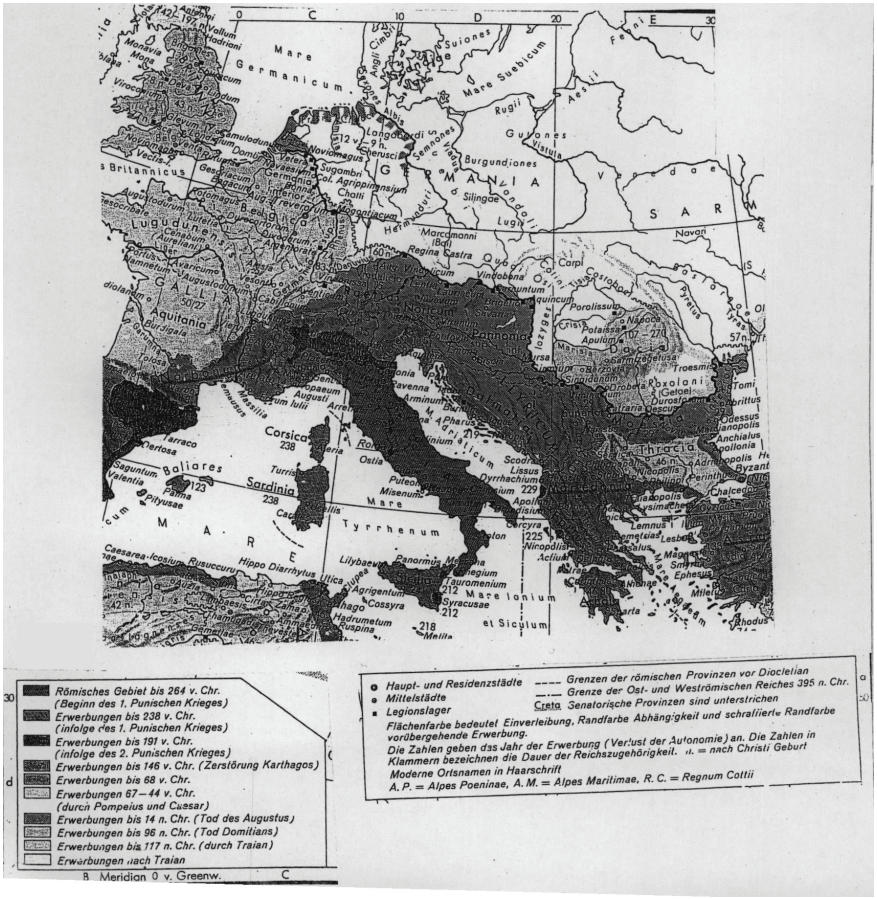
Il y a 40 ans, je pouvais me fonder sur une nouvelle base de travaux linguistiques : d'une part la publication des volumes 16 (1959), 17 (1966) et 15 (1969) du FEW concernant les éléments germaniques et d'autre part, en 1970, la refonte du premier volume de la *Romania Germanica* de Gamillscheg intitulé *Les Francs.* Aujourd'hui, nous disposons d'excellentes publications récentes qui traitent de l'histoire des Francs, des Alamans, ainsi que des synthèses archéologiques et des études linguistiques de mes amis germanistes Wolfgang Haubrichs et Wolfgang Kleiber.

Considérons d'abord l'extension de la région traitée : la zone entre la Moselle et le Rhin ; à l'époque romaine, une partie de la *Belgica* I et II et de la *Germania* avec la ville de Trèves, capitale de l'Empire romain vers l'an 300 sous Constantin le Grand, et avec Metz, autre centre important au bord de la Moselle, plus tard capitale de l'Austrasie, sous les Mérovingiens, et avec la ville romaine de Mayence, point de défense au bord du Rhin.

Jusqu'à la première partie du 5e s., le Rhin constitua la frontière entre la *Germania* et la *Romania* et, au moment de la plus grande expansion, se trouvait à l'est de ce fleuve le limes, qui incluait aussi les *agri decumates* (Cf. carte 1).

L'époque qui nous intéresse est celle de la décomposition de l'Empire romain, de l'époque franque, mérovingienne et carolingienne jusqu'à l'arrivée des Capétiens au 10e siècle. C'est aussi l'époque du passage du latin tardif à l'ancien français et, du côté-

germanique, de la constitution de l'ancien haut allemand avec les premiers documents linguistiques et littéraires.



Carte 1 : Extension de l'empire romain

Notre point de départ : l'Empire romain au 3e siècle au moment des premières invasions des tribus germaniques, c'est-à-dire de 257 à 261, avec la conséquence qu'entre 259 et 260 les Romains abandonnèrent le *limes germanicus* avec les *agri decumates*. Cette région abandonnée fut occupée ensuite par les Alamans. Les nouvelles frontières de l'Empire furent donc le Rhin et le Danube. Quant à notre région entre Moselle et Rhin, elle appartint aux Francs qui devinrent, en tant que *foederati*, les nouveaux voisins des Romains dans la *Belgica secunda*.

Les invasions des Francs au 4e siècle coïncidèrent avec les crises internes de l'Empire romain : l'usurpation de Maxentius (352) et la guerre civile entre Maximus et Theodosius I (388). Eugen Ewig dans son livre *Les Mérovingiens et le règne des*

Francs écrit : au 4e siècle, « une symbiose s'installait à la cour de Trèves entre l'Empire et les Francs. À partir de la fin du 3e siècle, des prisonniers barbares furent obligés au service militaire des Romains sous les noms de *laeti* » et en 358 l'empereur Julien les rendit sédentaires (*cf.* Ewig 1988, 11). Ces Francs furent intégrés dans l'armée romaine comme *dediticii* et beaucoup parmi eux servaient dans l'*exercitus Gallicanus* dont quatre généraux furent des Francs : Merobaudus (372-83), Richomer (382-94), Bauto (383-388) et Arbogast (388-394).

La situation politique en Gaule changea quand l'empereur Honorius (395-423) déplaça la résidence impériale de Trèves à Arles puis à Milan et enfin à Ravenne. Après la réorganisation de la défense le long du Rhin (413-415), les tribus germaniques sous le commandement du *magister equitum per Gallias* participèrent elles-mêmes à la défense de la frontière rhénane en qualité de *foederati* sans pouvoir empêcher le pillage de Trèves en 413 par des Francs. En 428 Aetius 'le dernier Romain' réussit à rétablir l'ordre, mais céda en 436 des zones de colonisation aux Francs sur les rives occidentales du Rhin.

Dans la bataille décisive contre Attila en 451, les Francs luttèrent aux côtés d'Aetius. Après sa mort, les Francs rhénans avancèrent dans la province de Mayence (*Germania* I) et s'emparèrent temporairement de la métropole mosellane de Trèves.

En 486, après la fin de l'Empire romain occidental, les Francs rhénans occupèrent de nouveau Trèves et la province rhénane qui alors fut administrée par le *comes* Arbogast, descendant du général romain homonyme, lui aussi officier romain. L'extension franque de cette époque se reflète dans l'expression *Francia rhinesis* du cosmographe de Ravenne (8e siècle) qui indique comme délimitation méridionale Mayence et Toul.

En 482, quand le Franc salien Clovis prit la succession de son père Childerich, Syagrius, qui sera vaincu par Clovis en 487, régnait en France. Sa défaite fut la victoire du *rex Francorum* sur le *rex Romanorum*, et donc la victoire des Francs sur les Gallo-romains, et ce qui assura à Clovis l'hégémonie sur tous les Francs. Il devenait ainsi le fondateur de la dynastie des Mérovingiens (482-751).

En 496 et 497, il y eut une guerre entre les Francs rhénans et les Alamans parce que ces derniers, après avoir traversé le Rhin, s'infiltraient dans la *Francia rinense*. Il est probable que Clovis accorda son aide aux Francs rhénans et remporta la victoire décisive près de Tolbiac. Grégoire de Tours nous relate que Clovis au point culminant de la bataille implora le Christ, le dieu de sa femme, la Bourguignonne Chrodechild, en promettant sa conversion au christianisme s'il sortait vainqueur. C'est ainsi que la religion chrétienne des Francs fut le catholicisme et non l'arianisme des Goths.

Pour pouvoir évaluer les contacts des Germains avec les Romains entre Moselle et Rhin du 4e au 6e siècle, époque décisive pour la romanisation et la germanisation de cette région, examinons d'abord la situation régionale vue sous la perspective des quatre villes importantes situées aux bords de la Moselle et du Rhin : c'est-à-dire de Trèves (*Augusta Treverorum*), Metz (*Divodurum*), Mayence (*Mogontiacum*) et Worms (*Borbetomagus*). L'essor de la capitale, Trèves, commença à la fin du 3e siècle

et surtout sous Constantin le Grand (313-335) et s'acheva avec le déplacement de la préfecture des Gaules à Arles entre 395 et 407. Des invasions germaniques au milieu du 4e siècle interrompirent le premier apogée. Pour le 5e siècle Salvien parle de quatre conquêtes des Francs qui finirent au commencement du 6e siècle par l'incorporation de Trèves dans le royaume des Mérovingiens.

À partir du 5e siècle on assiste à un déclin de cette ville, marqué par une dépopulation sensible : les murailles romaines de Trèves entourent 285 hectares par opposition p. ex. à celles de Metz avec ses 70 hectares. Trèves à l'époque impériale comptait peut-être 30 000 habitants qui se réduisirent à quelques milliers au 5e siècle. Metz est nommée dans la liste des villes détruites en 451 par les troupes d'Attila.

Mogontiacum (Mayence) était le nom d'un camp pour deux légions romaines, une forteresse qui au milieu du 3e siècle fut entourée d'une muraille. Au milieu du 4e siècle, Mayence fut temporairement conquise par les Germains. Après la victoire de Julien en 357 sur les Alamans près de Strasbourg, le pouvoir romain fut rétabli. Au 4e siècle, la ville était le centre de la flotte romaine sur le Rhin. Mais une attaque des Alamans en 368, des Burgondes en 411 et le martyre de l'évêque Aureus lors de l'invasion des Huns (436 ou 450) annoncèrent le déclin de cette ville romaine.

Wormatia était un *castellum romanum* du premier siècle (Grünewald 2008, 87) qui exista jusqu'à la fin du 5e siècle. Il semble qu'en 415 dans la région de Worms les Burgondes étaient établis en fonction de *foederati* sans laisser toutefois de traces archéologiques si l'on en croit l'état de la recherche actuelle. Cependant la source du *Nibelungenlied* en moyen haut allemand et des témoins onomastiques sont des indices de leur présence dans la région de Worms. En 435/36 Aetius, avec l'aide des Huns, détruisit leur royaume et transféra les Burgondes survivants en *Sapaudia* (Savoie) dans les environs de Genève. L'organisation romaine semble être terminée en 455 après l'inspection d'Avitus, plus tard empereur romain. Vers 500 des Germains, soit des Alamans ou des Francs, s'établirent dans cette région.

En conclusion : après une floraison à l'époque romaine, ces villes romaines survécurent aux 4e et 5e siècle, après les grandes invasions, et se reconstituèrent à l'époque mérovingienne, grâce aussi à l'organisation ecclésiastique chrétienne et à leurs églises.

Pour témoigner de la continuité de la population, l'archéologie et la numismatique ont une importance primordiale et des publications archéologiques des dernières années sont d'une grande utilité ; je pense aux Mélanges Fingerlin de cette année (Bierbrauer 2012), à ceux de Frauke Stein l'année passée (Stein 2011) et aux magnifiques catalogues des expositions concernant les Romains, les Francs et les Alamans.

1. La *Romania germanica* : la rencontre des Francs et des Romains : conséquences linguistiques

Passons à la partie proprement linguistique de la rencontre des Francs et des Romains entre Moselle et Rhin. Aujourd'hui, la limite linguistique entre allemand

et français traverse cette région et d'après les données historiques, on se demande pourquoi, après le règne des Francs, des Mérovingiens et des Carolingiens, après la domination politique des peuples germaniques, la langue française est toujours la langue dominante dans cette région. Les données linguistiques reflètent sans doute des données ethniques et des civilisations différentes. Si aujourd'hui la frontière linguistique est linéaire, il en allait autrement à l'époque de son établissement. Pour la partie occidentale, il faut parler d'une *Germania submersa* et pour la partie orientale d'une *Romania submersa*.

1.1 *Romania submersa*

Comme j'ai essayé de le montrer dans ma partie introductive, on trouvait des Germains dans la partie à l'ouest du Rhin, dans la *Germania* I et II, comme ethnie non romaine, à partir du 3e/4e siècle. Les premiers arrivés étaient les *laeti ;* Zosime parle des λετοί, ἔθνος Γαλατικόν.

Dans LexMA 5,1612 Wirth explique le terme comme une « dénomination pour des membres d'un colonat militaire dans l'antiquité tardive, constitué d'éléments germaniques en Gaule dont l'existence est prouvée depuis. Leur emploi militaire est important pour la politique démographique ». Nous ne savons pas si, dans notre région, il existait des lètes ou non. Pour le 4e siècle, il faut cependant supposer qu'une couche sociale élevée, appartenant à la noblesse sénatoriale ou militaire, est bilingue ; on peut au moins communiquer soit dans une langue germanique soit en latin tardif régional. Le bilinguisme continue à l'époque des Mérovingiens où certaines évolutions phonétiques se rapprochent déjà du protoroman et de l'ancien haut allemand.

Au 7e siècle, il y a dans notre région des zones où le protoroman domine ou bien au contraire où c'est l'ancien haut allemand qui est à la première place. L'évolution de *t-* initial > *ts-*, p. ex. *Taberna* > *Zabern* dans des toponymes est instructive pour la germanisation. C'est une évolution phonétique qui s'est achevée avant l'an 600. Si *t-* s'est conservé dans des toponymes, on peut conclure que vers 600 cette évolution germanique ne s'est pas encore réalisée dans cette zone : *Zabern* (Alsace) et *Zeutern* par opposition avec *Taben*, *Tholey*, *Theley* qui conservent le *t-* latin.On peut donc dire que l'évolution de *t-* > *ts-* des Francs était déjà achevée avant la germanisation de ces îlots romans. Cette germanisation se passait après l'an 600. Il faut cependant séparer l'assibilation romane et *t-* > *ts-* de la mutation germanique : *ti* > *ts*, p.ex. **Martiacum* > *Merzig* qui constitue l'assibilation romane et *Taberna* > *Zabern* pour la mutation germanique.

On peut fixer un peu plus tard le changement de *-k-* > *-ch-* comme dans l'ancien francique *maken* > *machen* (cfr. angl. *to make*) dans le cours du 7e s. et vers 700. Etant donné que des centaines de toponymes en *-acu* > *-ach* connaissent ce changement, ce suffixe toponomastique constitue un indice chronologique important pour le 8e siècle.

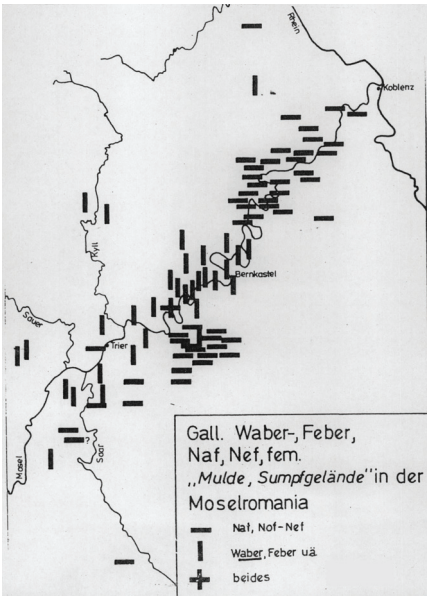
Si dans certaines régions de la Moselle, de la Sarre et en Lorraine il y a des îlots linguistiques, on peut dire que dans ces régions les Romans ont survécu plus longtemps.

La constitution de la frontière linguistique de notre région dépend de plusieurs facteurs : de l'espace géographique, des terrains non favorables à la culture, de l'essartement au Moyen Âge et de l'organisation ecclésiastique et administrative sous les Mérovingiens et les Carolingiens.

1.2 Germanisation de la *Romania* mosellane

A l'origine, il y avait des zones bilingues distribuées comme une peau de léopard et qui se développaient en îlots linguistiques. Des exemples pour une germanisation échelonnée entre le 7e et le 10e siècle existent dans la région de Tholey en Sarre, dans la zone rhénane de Mayen (Eifel) et près de l'embouchure de la Moselle et de celle du Rhin. La plus grande de ces zones est celle que constitue la *Romania* mosellane autour de Trèves. Ces îlots de romanité prolongée furent absorbés à l'époque carolingienne, à l'exception de la *Romania* mosellane qui se maintint peut-être jusqu'au 10e siècle. Il y a vingt ans, je supposais pour la *Romania* mosellane une continuité jusqu'au 12e siècle parce que le toponyme *Chasnot* montrait la diphtongaison *castanetum* > *castaneit* > *castanoit* > *castanot*, une évolution phonétique appuyée par une forme wallonne du 12e siècle.

Vu que Martina Pitz avait trouvé en 2008 une attestation *Fontanodo* évidemment antérieure dans un document original de Paris pour les années 825-829, ma première attestation ne peut plus être maintenue. La dernière évolution phonétique attestée dans la *Romania* mosellane est donc celle de la palatalisation de [a] en syllabe ouverte : *pratu* > **prait* > *preit* > *pre* qui normalement est datée du 9e siècle ; cf. carte 2 ::



Carte 2

J'accepte donc la conclusion de Martina Pitz avait formulée en 2008 (Pitz 2008, 450) (je traduis) :

> Ainsi, il est évident que le 9e siècle constitue une période décisive. Avec la formation du royaume de Lorraine après l'an 843 et le partage de la Lotharingie, résultant du traité de Meersen, la séparation des langues, en marche depuis le 8e siècle, atteignit un nouveau dynamisme qui accéléra la formation de cette zone de cristallisation des deux grands groupes de langues galloroman et germanique et fit fondre en une seule les zones d'interférences situées loin de la frontière linguistique. C'est au plus tard à ce moment-là que le roman mosellan comme moyen de communication se rétrécit et que la fragmentation en divers îlots à l'intérieur du francique se réalisa. C'est alors que le roman mosellan se développa comme 'Haussprache', puisqu'il ne pouvait plus absorber les évolutions récentes du galloroman.

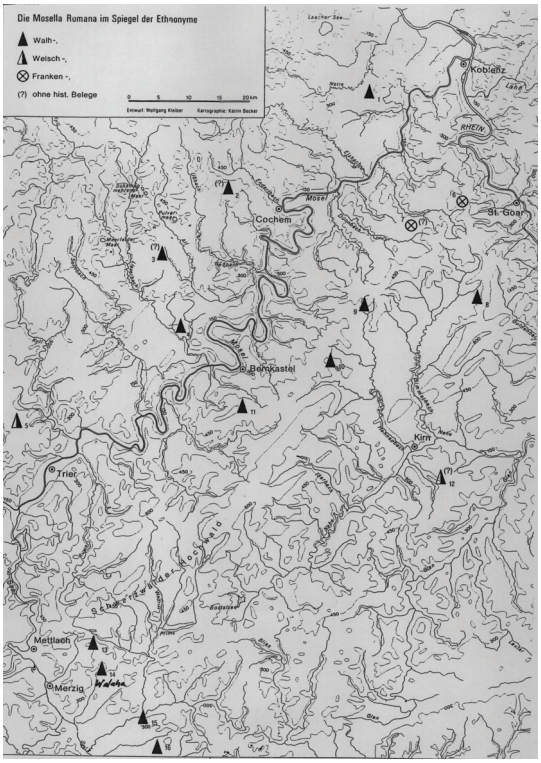
Un autre indice chronologique non négligeable pour l'époque carolingienne est constitué par la métaphonie [a] > [e] suivi d'un *i* ou *y* en ancien haut allemand, par exemple pour la dénomination d'un ruisseau : *Seffersbach* (près de Merzig/Saar) : *Sefferne* (1217) < **Savirna*. La germanisation de cette zone romane au 8e siècle était en train de se réaliser par opposition au noyau roman de la vallée de la Moselle. Ce nom de ruisseau contient également un autre phénomène phonétique intéressant. Jusqu'au 8e siècle *[v]* est adapté en *[w]*, p. ex. lat. *vinum* > all. *Wein*. À partir de cette date, nous notons une substitution de *w* > *f*. Un autre nom de ruisseau *Vokara > *Wochera* (1084) montre que la germanisation a probablement eu lieu aux 7e/8e sièclex au moment où le changement de *k* > *ch* était encore en vigueur. Les changements phonétiques en ancien haut allemand permettent de situer approximativement le moment de la germanisation.

Autre exemple : la forme *Merzig* au bord de la Sarre, qui est attestée comme *Merceche* en 1059. La forme romane, attestée en latin médiéval *Marciacum* (949) révèle le nom latin du propriétaire foncier *Martius* + *-iacu* (suffixe qui indique la propriété du fondateur).

Si vous regardez la carte 3 (*cf. infra*), vous trouvez près de Merzig le toponyme *Walaha*, aujourd'hui *Wahlen*, au 11e siècle, *Wala/Valmünster* < *Walamonasterii* au 10e siècle. Sa base est constituée d'un ethnonyme **walh-oz* qui correspond à la dénomination *Volcae*, pour une tribu celtique (attestée chez Jules César). Puis, par un élargissement sémantique, le mot a désigné les Celtes en général et finalement on appela 'Walen' les romans, colons en Gaule. Dans une glose contemporaine, nous trouvons *Romanus* pour *Walh*. Christa Jochum-Godglück dans son article tout récent « Walchensiedlungsnamen und ihre historische Aussagekraft » :

> Evidemment, l'endroit préféré de ces toponymes en *Wahlen* c'est la proximité de routes romaines, de centres d'appui ou de trafic tardo-romaines ou bien des propriétés royales franciques ou leurs propriétés fiscales » (Jochum-Godglück 2012, 206).

Un toponyme en *Wala* se trouve normalement à la périphérie d'un îlot roman, c'est la dénomination pour les Romans donnée par des non-romans ; les toponymes en *Walen* signalent donc des positions de recul de zones romanes[1] ; cf. la carte 3:



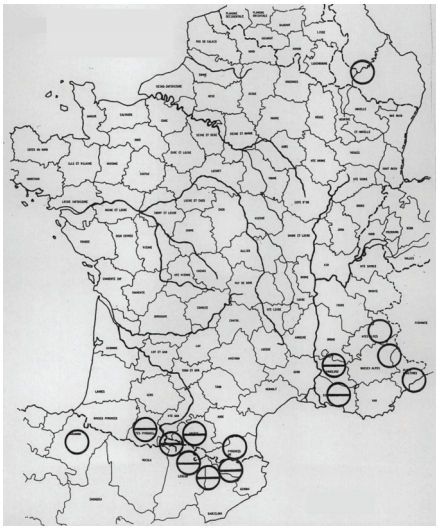
Carte 3

Pour caractériser les îlots linguistiques, il y a normalement aussi des éléments lexicaux reliques, des restes pétrifiés qui font présupposer une extension géolinguistique autrefois plus étendue. Prenons donc un exemple tiré d'une conférence de Johannes Hubschmid (Hubschmid 1983).

Dans cet article – inspiré par un travail de Jakob Jud (Jud 1917) – Hubschmid a montré comment des mots d'origine latine se sont conservés, dans des zones de langues non-romanes, dans une zone frontière germanique. Je me limite à l'exemple de *horreum* conservé comme toponyme figé (*Oeren*, cloître de Trêves), mais qui, dans d'autres zones de la Romania, existe aussi comme appellatif.

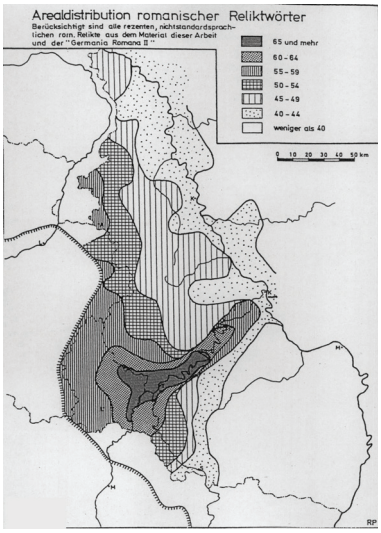
[1] Cp. Bach (1981, 182): « In der Regel sind es wohl Kelto-Romanen, die in den von germanischen Stämmen den Römern entrissenen Gebieten zurückgeblieben waren ».

Il s'agit de vestiges lexicaux qui souvent sont pétrifiés dans des toponymes et qui, comme appellatifs, sont remplacés par d'autres mots p.ex. *horreum* par *spicarium* et plus tard par *grenier* ; cf. carte 4 :



Carte 4 : *horreum* (Hubschmid 1983, 120)

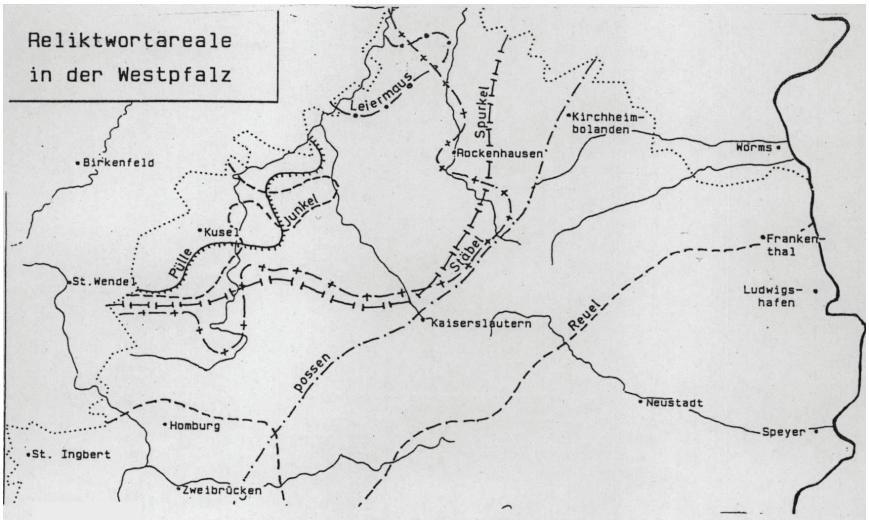
En plus, un îlot linguistique est caractérisé par une quantité d'emprunts qui continue dans la langue de superstrat. Pour la Romania mosellane, Wolfgang Kleiber et son disciple Rudolf Post ont examiné ces emprunts ; *cf.* carte 5 :



Carte 5 : distribution aréale des mots romans reliques

Sur la carte 5, vous voyez à l'évidence qu'il y a un noyau dans la vallée inférieure de la Moselle avec comme centre Trèves, où la densité dépasse 65 et davantage. C'est la région viticole où nous trouvons p.ex. *feiel* (< VIALE) 'petit chemin dans la vigne', *gimme* (< GEMMA) 'bouton (de la vigne)', *käbe* (< CAPUT) 'vigne', c'est-à-dire des expressions techniques de la viticulture. Un autre vestige modèle est ⌜*gotte*⌝ (< GUTTA) 'ruisselet' qui se trouve également dans l'îlot linguistique de la Forêt noire et dans les Vosges.

Rudolf Post (Post 1989) a examiné des emprunts frontaliers en exploitant le *Pfälzisches Wörterbuch*. Normalement, il s'agit d'emprunts au galloroman comme STABULUM > *Schafstäbel* 'gîte des moutons', JUNCUS > *Junk*, *Jonk* 'jonc', GLIS > *Leiermaus* 'souris', etc. La carte annexe (carte 6) indique la distribution aréale au nord du Palatinat :

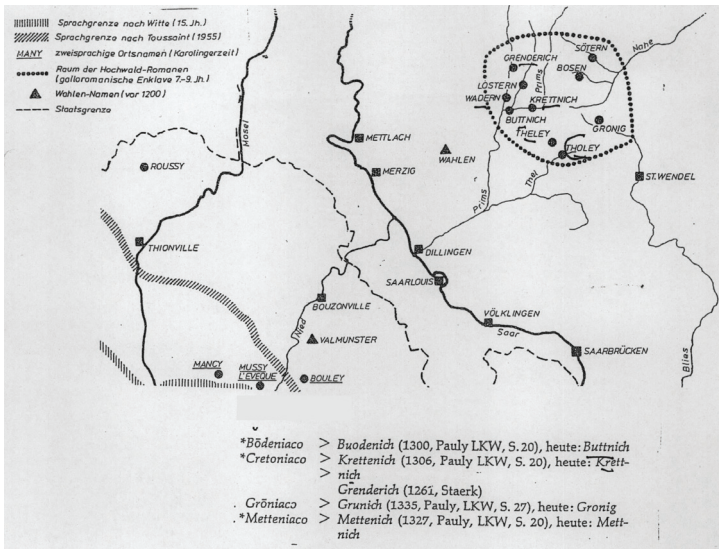


Carte 6 : distribution aréale au nord du Palatinat

Comme exemple d'un deuxième îlot linguistique de la *Romania submersa* de notre région, j'ai choisi celui du Hochwald, la région de Tholey et Theley, îlot beaucoup plus réduit que l'îlot mosellan et absorbé par le dialecte allemand déjà à l'époque carolingienne (carte 7).

Des toponymes typiques de cette région de Theley, Tholey avec conservation de *t-*, et *Grenderich, Krettnich, Buttnich* avec le changement de *-k* > *-ch* ; *Wadern* avec conservation de *w-* :

**Bōdeniaco* > *Buodenich* (1300 [Pauly 1965, 20]), aujourd'hui : *Buttnich*
**Cretonacio* > *Krettenich* (1306 [Pauly 1965, 20]), aujourd'hui : *Krettnich*
Grenderich (1262 [Staerk 1976])
Grōnacio > *Grunich* (1335 [Pauly 1965, 27]), aujourd'hui : *Gronig*
**Metteniaco* > *Mettenich* (1327 [Pauly 1965, 20]), aujourd'hui : *Mettnich*



Carte 7

En 1985, j'avais écrit (je traduis le texte allemand [Pfister 1985, 290]) :

> Étant donné que dans la région du Hochwald (Tholey) où l'habitat germanique n'a pas laissé de traces et où l'on ne connaît pas de sépulture des époques mérovingienne et carolingienne, la continuité onomastique est garantie par l'hydronomie préromane, et que la région devait donc avoir été peuplée de manière continue, il faut admettre la seule explication plausible : dans cette région forestière isolée avec peu de terrains fertiles et cultivés, ce ne sont pas des Germains qui vivaient, mais des une population romane résiduelle.

En plus nous avons *Wadrill*, attesté au 10e siècle sous la forme *Waderella* avec accentuation finale (*Wadríll*), indice d'une prononciation romane par opposition à la prononciation germanique *Wádern*. La forme *Wadrill* correspond à *Kastellaún* et *Riól* dans la Romania mosellane.

L'abbaye de Tholey possède un document extraordinaire qui nous informe sur la fondation de l'abbaye, une charte de l'année 634 (copie du 10e siècle) qui contient le testament du doyen verdunois, le noble *Adalgisil Grimo*, qui constitue la fondation ecclésiastique d'un grand propriétaire terrien mérovingien dont la sphère d'interêt allait de la région mosellane de Trèves, avec Tholey, jusqu'à Tours au bord de la Loire, et de Verdun en descendant la Meuse jusqu'à Liège.

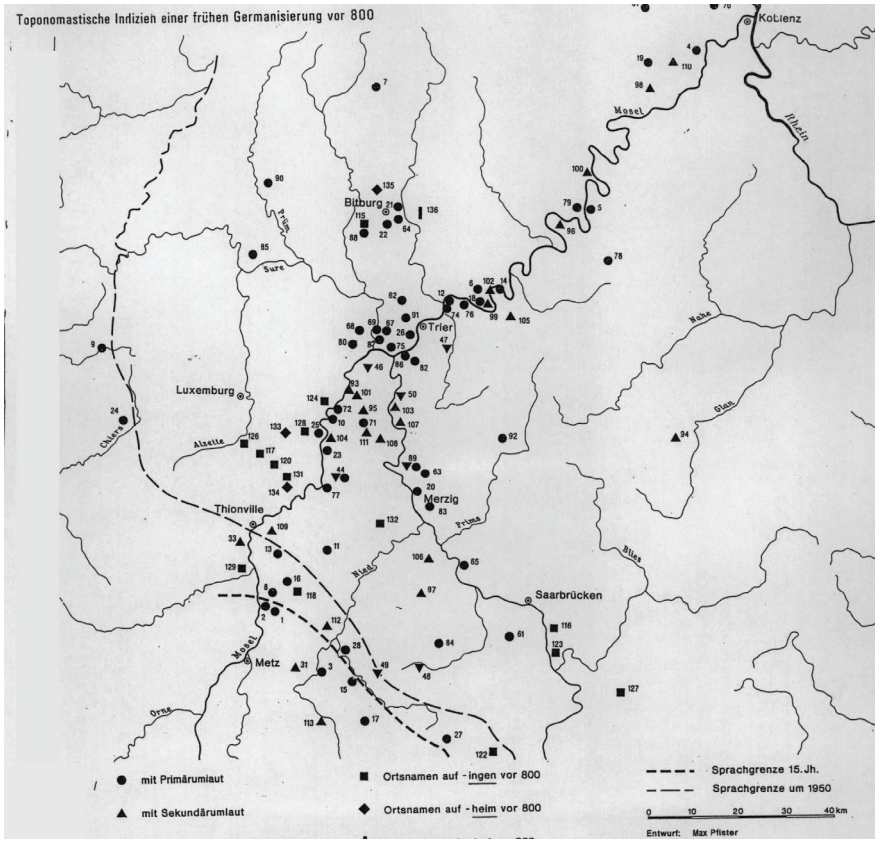
Tholey était le poste le plus important à l'est de l`évêché de Verdun et montre déjà pour le premier tiers du 7e siècle les relations ecclésiastiques entre Verdun et Tholey. Le nom même de *Tholey* se trouve trois fois dans le testament de Grimo :

> l. 31 *locum verum cognominante Domo et castrum* <u>*Theulegio*</u> *sectum in Vosago*
> l. 51 *in predicto loco Doma aut* <u>*Toleio*</u>

Toujours en corrélation avec *Domu*/*Doma* qui est peut-être la correspondance préromane de *Theulegio*/*Taulegius*/*Toleio*. Probablement l'origine est-elle *TEGULEIUM, dérivé de TEGULA. La base supposée **teulei* signifierait donc 'édifice avec un toit de tuiles'.

1.3 *Germania submersa*

Comme il existe à l'est de la frontière linguistique des îlots linguistiques, il existe également à l'ouest de cette ligne une colonisation franque particulièrement intense (carte 8). Il y a p. ex. des toponymes en *-ingen* avant l'an 800, des colonies franques probablement mérovingiennes. De vrais îlots linguistiques germaniques se trouvent pourtant plus à l'ouest dans la région de la Meuse (près de Verdun et de Saint-Mihiel) et n'entrent plus dans notre zone examinée entre la Moselle et le Rhin; *cf.* la carte :



Carte 8

Quant aux exemples lorrains en *-ingen* avec un nom germanique dans la première partie, citons

Bassing (Moselle, canton Dieuze) : *Badugis-ingas* > *Badgisingas* (764)
Gisselfingen/*Gélucourt* (Moselle, canton Dieuze) : **Gisalwulf-ingas* > *marca Gisalfinco* (787)

et deux autres exemples lorrains où le toponyme roman conserve *a* devant *i*, tandis que la forme franque montre le changement de *a* devant *i* > *e* ('Primärumlaut') :

Antilly/all. *Enterchen* (Moselle, canton Vigy) : **Antiliacum* > *Antillei* (1241, original)
Many/all. *Merchen* (Moselle, canton Faulquemont) : **Mariniacum* > *in villa Mernicha* (981, copie) ; *Marnei* (1162, original).

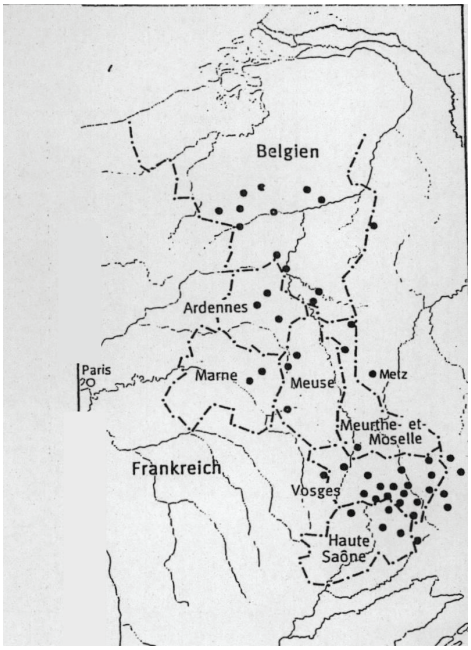
Étant donné les formes *Enterchen* et *Mernicha* avec le changement de *a* devant *i* > *e* il faut supposer que pour ces endroits la germanisation est achevée à l'époque carolingienne.

Quant à la durée de l'îlot linguistique germanique dans les Argonnes près de Verdun, mon ami Wolfgang Haubrichs (1992, 666) écrit :

> L'existence de ces îlots semble être en rapport avec l'activité de certaines formes de colonisations au 7e siècle, de la royauté, des maires du palais et de grandes familles de la noblesse. Leur extension temporaire ne doit pas être surestimée.

Dans les années 30 du siècle passé quelques chercheurs – p. ex. l'historien Petri (1937) pensaient que la colonisation francique s'étendait jusqu'à la Loire, thèse ajourd'hui abandonnée. Gamillscheg dans sa *Romania Germanica* avait cependant reconnu déjà à la même époque que la pertinence de la toponymie pour résoudre le problème de la colonisation franque est limitée et il avait mis avec raison l'accent sur le lexique germanique, sur les emprunts au francique du lexique français. Ce sont aussi Jakob Jud et Walther von Wartburg qui parlaient d'une pénétration profonde du lexique francique en français. Il faut cependant reconnaître que ces emprunts ont une valeur inégale : un mot comme fr. *maréchal* provient d'une couche francique de l'administration mérovingienne avec une vaste extension ; l'élément *-macher* (< MACERIU 'mur', p. ex. dans des toponymes [*Auersmacher*] peut provenir d'un emprunt régional également bien répandu comme appellatif. Pertinentes cependant sont des expressions rurales comme lorrain *ran de porc* 'toit à porcs' (FEW 16,237a) < a.francique *hrannô* (attesté comme *chranne* dans la Lex Salica). Ce mot est attesté en Wallonie belge, dans les Ardennes, en Champagne orientale, en Lorraine et dans les Vosges comme le montre la carte 9 (*cf. infra*).

L'exemple ancien francique *hrannô* est également intéressant du point de vue phonétique. La consonne francique *h* dans la combinaison *hl* et *hr* dans les noms propres comme *Hlud-wig*, *Hrod-bert* présente des difficultés de prononciation pour un Roman. La romanisation d'un tel nom se réalise par une substitution : *hl-* et *hr-* deviennent *chl-* et *chr-*, p.ex. *Chlotha-charius* et *Chr[o]de-gar* dans une charte originale de St. Denis (584/629 [Haubrichs 2009, 85]). Dans l'ancien haut allemand du 8e siècle [h] devant [r/l] s'affaiblit, de sorte qu'une couche d'emprunt tardif se présente sans *h*, c'est-à-dire *hranne* > afr. *ran*.



Carte 9 : extension de l'ancien francique *hrann* "toit à porcs" dans les dialectes orientaux. Par M. Pfister ; réalisation graphique I. Bell

De telles substitutions de sons exigent des contacts bilingues d'une longue durée. Ces contacts bilingues sont aussi assurés par des couples onomastiques. Nous avons déjà rencontré *Gisselfingen*/*Gélucourt* :; *Antilly*/*Enterchen* et *Many*/*Merchen*. Ces doublets toponomastiques, étudiés dans la thèse exemplaire de Maria Besse (Besse 1997) sont d'un grand intérêt parce qu'elles contiennent souvent des indices chronologiques. En Sarre p. ex. près de Besseringen il y a un toponyme *ad Buntendele* (milieu 10e s., copie 11e s.), attesté comme *Ponte* en 1249, *Punten* (en 1360). Il s'agit probablement d'un lat. PONTO, -ONIS 'bac' (Besse 1997, 80). Étant donné que *p* initial germanisé dans cette région donnerait *pf* avec la mutation consonantique (*palatium* > *Pfalz*) et que le *-t-* de *Ponte* est conservé, on peut dire que la germanisation a dû se produire après l'an 600.

1.4 Amalgame de civilisations (acculturation ; hybridismes)

Après avoir parlé de la *Romania submersa* et de la *Germania submersa*, de la rencontre des ethnies romanes et germaniques, des doublets onomastiques, il est évident que dans une grande partie de la Lorraine il y avait un bilinguisme au moins à l'époque mérovingienne et en partie encore sous les carolingiens.

Martina Pitz écrit dans les Mélanges Stein (Pitz/Stein 2000, 365) :

> La connaissance de la langue française apparaissant comme une composante de la culture et de l'éducation de la noblesse germanophone de la région dans sa recherche d'un idéal imprégné des valeurs culturelles du monde roman […], on peut penser que les gouvernements de l'espace frontalier étaient souvent bilingues […], ou tout au moins disposaient d'une administration qui pouvait communiquer dans leur langue maternelle avec les sujets de l'autre langue […].

Dans les cimetières du 4e au 9e s. de notre région nous pouvons normalement saisir les différences entre les rites mortuaires germaniques et ceux des Romains. Pour un Franc l'existence après la mort continue. C'est pourquoi il doit être enterré avec ses armes et ses bijoux, qui indiquent son statut social. Son cadavre reposait normalement dans un sarcophage orienté de l'ouest vers l'est. Les Romains par contre étaient incinérés ou, s'ils étaient chrétiens, portaient le linceul mortuaire. Dans l'article instructif de l'archéologue Frauke Stein (2011d), l'auteur distingue le rite mortuaire A, celui des Francs, et le rite B, gallo-romain. Pour le rite A, les sépultures pour hommes contiennent les armes *spatha*, saxe, lance et bouclier qui correspondent à l'époque et à l'état social et des parties du costume festif ; les sépultures pour femmes les fibules, fermeture de la ceinture et ses pendants, jarretière, accessoires de chaussures.

Le rite mortuaire B gallo-romain est plus simple, en grande partie sans accessoires, pour les hommes rarement une arme, une francisque ou un saxe, une boucle ou garniture de ceinture, pour les femmes galloromanes normalement sans fibules et comme bijoux des perles et des anneaux.

Pour l'assimilation du rite mortuaire A (francique) et B (gallo-romain) on peut noter un décalage chronologique qui va de l'ouest vers l'est : par exemple entre Argonnes et Meuse, l'assimilation se réalise au commencement du 7e siècle et à l'est de la Sarre après la moitié du 7e siècle. Il est clair que l'archéologie a des problèmes à partir du 8e siècle quand les cimetières par rangées n'existent plus et que les accessoires font défaut. En plus, l'identité de la population galloromane est menacée du moment que la partie orientale de la Romania est isolée et est en train de perdre ses bases culturelles. On peut dire qu'à partir de la seconde moitié du 9e siècle, la limite linéaire de la frontière linguistique est plus ou moins fixe et l'acculturation est terminée : à l'est se constitue l'allemand, à l'ouest le français.

Après l'archéologie, passons à un autre domaine de recherche où l'on peut observer la rencontre du monde germain et du monde galloroman : c'est l'onomastique. Je m'appuie sur un travail de Wolfgang Haubrichs (Haubrichs 2004a). En comparant les noms de personnes galloromans du 5e au 8e siècle à ceux de la tradition antique, Haubrichs a montré une augmentation des noms thériophores (*Aper, Ursus, Leo, Lupus*) qui rappellent la force des bêtes sauvages de combat, puis des noms de bon augure (*Felix, Vitalis, Bona,* etc.), les noms chrétiens et théophores comme *Redemtus, Dominicus, Deodatus*, *Cristina, Bonifatius* avec le rappel des noms de Saints (*Sebastianus, Laurentius, Martinus, Remigius*).

Quant aux noms germaniques, on peut dire que les noms de personnes les plus anciens sont normalement des noms parlants :

Segestes 'victorieux' avec son frère *Segi-mer* 'célèbre par sa victoire' et son fils *Segi-mund* 'main victorieuse', *Theuda-rik* 'souverain du peuple' ou bien **Hluda-wih* 'combattant célèbre' > *Chlodwig.*

Ou bien dans le nom de Chlothar, fils de Clovis, nous voyons la substitution d'un groupe de consonnes difficile à prononcer **Hluda-hari* > *Chlotharius.*

La société bilingue se manifeste dans des formations hybrides ; *cf.* pour la région entre Moselle et Rhin :

Blan-mundi < BLANDUS 'agréable' et germ. **mundaz* 'protecteur', moine du cloître Murbach sous l'évêque Baldobert (751-765). Ce couvent abrite des moines en partie galloromans. *Dona-bertus* < *dona* 'seigneur' *-bertus* < germ. *berhtaz* 'brillant, célèbre', moine du cloître Gorze sous l'abbé *Chrodegang* (765ca.)

Fide-ardus < **Fide-hard* (< germ. **harduz* 'fort'), abbé à Tholey (8e s.)

Justu-bertus < *Justus-*berhtaz*, moine à Marmoutier (Alsace), probablement 8e s.

Urse-ram < **Ursa-hrabna* < germ. **hrabna* 'corbeau', moine au cloître de Hornbach (Palatinat) sous l'abbé *Wirundus* (8e s.).

Ces noms lorrains hybrides sont des phénomènes provenant d'une région de contact, bilingue au 8e siècle.

Après l'onomastique, examinons quelques cas d'hybrides en toponymie. Martina Pitz a publié dans notre *Revue de Linguistique Romane* un excellent article (Pitz 2002b). Le type *Avri-court* est un amalgame : *avri* < germ. *Albfrid* combiné avec le suffixe roman *-court* < COHORTEM. Pitz écrit avec raison :

Quoi qu'il en soit il ne fait plus l'ombre d'un doute que la genèse *linguistique* des noms en *-court* est en relation directe avec cette évolution historique et culturelle que la recherche a désignée comme la *symbiose romano-franque* [note : *cf.* Prinz 1994].

Martina Pitz a répéré toutes les mentions de noms en *-court* dans les chartes originales antérieures à l'année 800 pour le nord de la France. Par exemple :

6	CLA 15/595	751 original	*Ebroaldocurte*
14	CLA 16/623	777 original	*Sicramno curte*

Quant à *Ebroaldocurte/Sicramnocurte*, M. Pitz note avec raison qu'on ne doit pas interpréter cet *-o* comme une mauvaise latinisation mais comme un authentique réflexe du cas régime roman.

Un autre exemple :

Aboncourt (comm. Metzervisse) : 1147 cop. *Epindorf* (AD Mos H 1714 141v°) ; 1176 or. *Ebbendorph* (AD Mos H 1742-1b) ; 1277 or. *Aboncourt* (BNF Coll. Lorr 976-28 ; 1310 or. *Aboncourt* (BNF Coll. Lorr 976-57 ; 1348 or. *Aboncourt* (AD Mos 7 F 468 ; 1480 or. *Ebendorff* (AD MM B 1937) ; 1636 or. *Endorff alias Aboncourt* (AD MM B 392-4bis) ;

< **Abbône curte*, NP germ. *Abbo* (**ab-*) pour les formes romanes en *-court* ;

< **Ebben dorf*, NP **Abio* > *Ebbo*, pour les formes en *-dorf.*

Avec l'explication de l'auteur :

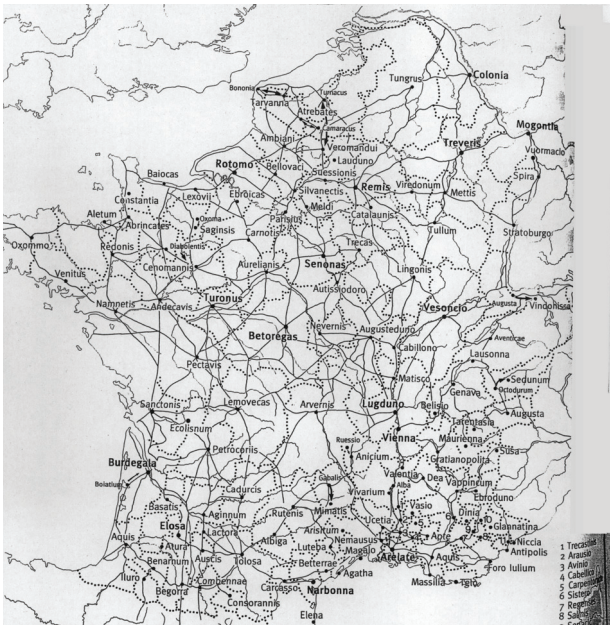
> Le couple Aboncourt-Endorf (n° 2) a fait désespérer des générations de toponymistes qui s'expliquaient mal la présence d'un *-e-* dans la forme allemande, donc d'un *-a-* suivi de *-i-* ou yod provoquant un *Umlaut*, alors que la forme romane n'en contenait pas. Les formes allemandes ne peuvent effectivement s'expliquer en conformité avec les lois phonétiques du vieux haut allemand qu'en posant **Ebben dorf*, ce qui suppose un éponyme **Abjo*, devenu *Ebbo* par suite du *Umlaut* et de la gémination de la consonne provoqués par l'influence d'un yod.

Faire ces recherches pour chacune des 88 attestations du type *Avricourt* comme Martina Pitz l'a fait, c'est un travail pénible mais indispensable.

2. Christianisation et organisation ecclésiastique

Passons à un autre facteur important pour la formation de la nouvelle identité culturelle comme résultat de la rencontre entre Galloromania et Germania sous les règnes mérovingien et carolingien : la christianisation et l'organisation ecclésiastique entre le Rhin et la Moselle.

Commençons par la carte 10 (par Jochen Martin, dans Jedin / Latourette / Martin 1990, carte 22) :



Carte 10 : l'Église occidentale jusqu'aux alentours de 600

À l'époque de Clovis, notre région était divisée entre les archevêchés *Treveris* et *Magontia* et les évêchés *Mettis*, *Tullum*, *Stratsburgo*, *Spira* e *Vuormacio*. Le com-

mencement de l'ère chrétienne en Gaule est marqué par l'année 313, date de l'édit de tolérance de Constantin. Les signatures du protocole du synode de Cologne en 346 nous montrent qu'il y avait à cette date-là des évêques à Cologne, Mayence, Worms, Spire, Strasbourg et Bâle.

À la fin du 4e siècle, sous l'empereur Theodose I, empereur de 379 à 395, le christianisme devint l'unique religion d'État. Au siècle suivant cependant l'Église chrétienne subit une crise existentielle.

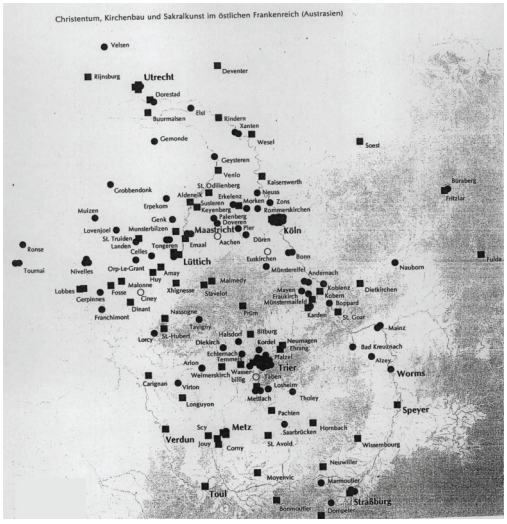
Les invasions des tribus germaniques et des Huns firent s'écrouler l'organisation ecclésiastique qui, après la conversion de Clovis, était à reconstituer. Aux 6e et 7e siècle la christianisation des Francs avançait lentement à partir de centres comme Metz, Mayence, Worms, Toul et Strasbourg.

Pour prouver la continuité du christianisme après la chute de l'Empire romain et au commencement du règne mérovingien, nous disposons aussi des inscriptions funéraires comprenant des symboles chrétiens (croix, colombes, etc.). Dans l'article de Boppert/Haubrichs ces témoignages importants proviennent de la région de Trèves, de Cologne, de Mayence et sont à dater surtout entre le commencement du 5e siècle et le 6e siècle.

La christianisation s'observe également dans les rites mortuaires : pendant la seconde moitié du 7e siècle, la population franque accepte de plus en plus les visions de l'autre monde chrétien et abandonne la coutume de doter les hommes ensevelis d'armes diverses en proportion de leur richesse et d'enterrer les femmes avec leurs parures. Finalement, vers la fin du 7e s., on renonce aux sépultures avec accessoires et l'endroit d'inhumation n'est plus hors de l'habitat mais au cimetière de l'église paroissiale. C'est ainsi que la source archéologique la plus importante du début du Moyen Âge se tarit. Aux 6e et 7e siècles, sous les rois mérovingiens, l'église impériale commence à se développer. Le pouvoir des évêques s'accroît. Au 6e siècle, les évêques descendaient de l'aristocratie, des sénateurs romains, et devenaient ainsi un lien entre les représentants nouveaux du pouvoir politique franc et la population galloromane. D'après le droit ecclésiastique, l'évêque était à la fois le possesseur de la fortune ecclésiastique de son diocèse et le seigneur des communautés ecclésiastiques et des monastères de sa ville épiscopale. Les églises à l'intérieur d'une ville devenaient de nouveaux centres.

Si nous regardons la carte des édifices ecclésiastiques à l'époque des Mérovingiens, nous voyons l'importance de l'ancienne limite de l'Empire romain, qui se prolonge encore dans les siècles suivant la chute de l'empire : si nous nous limitons à notre zone entre le Rhin et la Moselle la majorité des églises se trouve à Trèves, à Metz, à Strasbourg.

Sous les Mérovingiens, Metz acquit une importance particulière comme capitale de l'Austrasie; *cf.* la carte 11 :



Carte 11

L'évangélisation de cette ville paraît remonter au dernier quart du 3e s. parce que le cinquième évêque est attesté en 346. Dans la première moitié du 8e s., la ville comptait trent-neuf églises au moins parmi lesquelles – écrit Pierre-Edouard Wagner [Merzig-Wadern 1992, 217] – « certaines fondations pouvaient remonter aux premiers temps de l'église messine (fin IIIe siècle – début IVe siècle) ». L'église des Saints-Apôtres p. ex. s'élevait sur la nécropole antique. Après 641, on y avait transféré le corps de Saint Arnoul, maire du palais d'Austrasie, évêque de Metz et ancêtre de la dynastie carolingienne. À partir de 584, les rois mérovingiens résident fréquemment à Metz jusqu'en 629, date du transfert de la capitale à Paris par Dagobert.

Au 8e s., deux principales familles de la noblesse s'emparèrent du pouvoir, en Austrasie d'abord, puis dans l'ensemble du royaume franc : celle d'Arnoul, maire du palais et évêque de Metz, puis la noble famille de Pépin, son successeur à la mairie du palais. C'est pourquoi Metz fut l'objet d'une attention particulière de la part des Pippinides.

La Meuse et la Moselle constituaient au début du Moyen Âge les axes privilégiés du commerce de la Méditerranée à la mer du Nord et un passage de la *vita s. Martini* signale l'activité des bateliers sur la Moselle assurant le transport du sel de Metz à Trèves. Dans la vallée de la Seille Vic et Marsal, centres de l'exploitation du sel, sont attestés dès le 6e s. Dans la politique du 9e s., Metz jouait un rôle important. Par ex., c'est dans l'abbatiale Saint-Arnoul à Metz que l'empereur fut inhumé en 840. Dans l'organisation ecclésiastique, il y avait une concurrence entre Metz et Trèves qui finit par la subordination de Metz, de Toul et de Verdun à l'archevêché de Trèves.

En 655, l'abbaye de Wissembourg en Alsace fut fondée d'après la règle de Benoît et Colomban. L'abbaye de Wissembourg, particulièrement privilégiée, était un noyau d'expansion dans la seconde moitié du 7e s. favorisé par les activités des Etichones,

les puissants ducs d'Alsace. L'irlandais Colomban, venu en Gaule en 592, fonda au sud des Vosges plusieurs monastères dont le plus important fut celui de Luxeuil. Sous Dagobert Ier, roi de 623 à 639, il y eut une étroite collaboration avec l'organisation monastique des irlandais et de leurs disciples, ce qui profita considérablement au développement de la mission catholique.

L'abbaye de Lorsch, fondée en 762 et léguée en 772 à Charlemagne qui l'éleva au rang d'abbaye royale, centre riche et bien doté de l'embouchure du Rhin jusqu'au lac de Constance, devint un autre centre important.

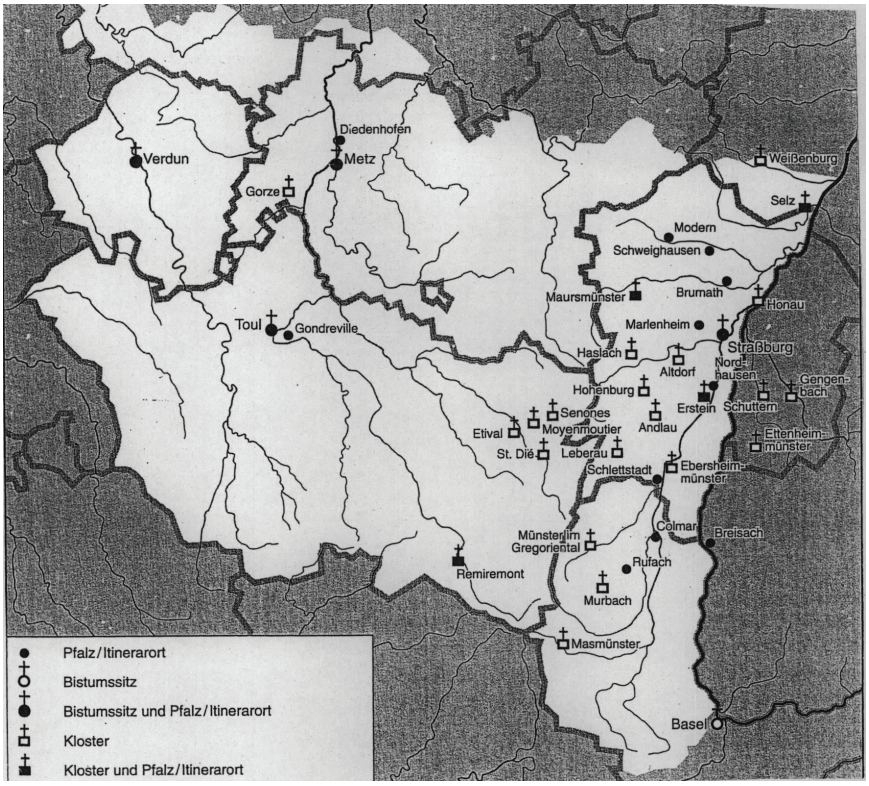
L'importance de la partie méridionale de l'espace que nous examinons, à savoir l'Alsace et, en particulier, Strasbourg se voit aussi dans les serments de Strasbourg (842), premier document en ancien français et un des premiers en ancien haut-allemand, tous les deux édités et commentés de manière exemplaire par Holtus et Gärtner (1995).

Carte 12

Le traité de Meersen en 870 entre Charles et Louis montre l'intérêt de ces deux seigneurs pour la région des Vosges : l'évêché et le comté de Toul et l'abbaye de Senones furent attribués à Charles, Remiremont, Moyenmoutier, Saint-Dié et l'évêché de Strasbourg à Louis l'Allemand.

La carte de la Lorraine de 888 à 923 (carte 12) montre que notre zone ne comprend que la partie méridionale de la Lorraine avec inclusion de l'Alsace en considérant la frontière sous Charles le Simple. Le « royaume soi-disant intermédiaire » n'a existé que du traité de Verdun (843) jusqu'à la mort du roi Zwentibold (900).

L'importance de l'Alsace sous les Carolingiens et au 10e s. se voit sur la carte 13 (évêchés, monastères et sièges royaux en Alsace et dans ses environs) :



Carte 13

À noter les nombreuses fondations d'abbayes organisées par Trèves et Metz dans les Vosges : Remiremont, Saint-Dié, Etival, Senones, Moyenmoutier et dans la partie orientale des Vosges, des fondations également initiées par Metz : Marmoutier et Munster dans le Val Saint-Grégoire. Il est évident aussi que la route de la Moselle de

Gondreville à Remiremont constituait un axe important du trafic et les fondations des monastères servaient leur sécurité.

De plus, vers l'an 700, Eticho, l'ancêtre des Etichones, fut le fondateur de nombreuses abbayes en Alsace (Hohenburg sur le Mont Sainte Odile, Ebersmunster, Honau et Murbach). Cette dernière, Murbach, fut particulièrement favorisée par Charlemagne. Louis le Pieux séjourna cinq fois à Remiremont et l'Alsace devint, à cette époque, le centre du royaume.

3. Passons aux conclusions :

« Germains et Romains entre Moselle et Rhin » constitue un très vaste sujet qui doit tenir compte de rencontres particulièrement mouvementées entre le 3e et le 10e siècle. Le même phénomène de déclin de la *Romania* à l'ouest du Rhin et de la formation d'une civilisation germanique dans cette zone se retrouve également dans les Pays Bas, dans une partie de la Belgique, au Luxembourg, et dans la Suisse alémanique. Cette rencontre de deux civilisations différentes est un long processus d'amalgame et d'acculturation dans des sphères bien différentes : dans les rites mortuaires, dans les langues galloromanes et germaniques avec leurs toponymes, leurs anthroponymes et leurs emprunts linguistiques, dans des documents juridiques, administratifs et littéraires. C'est pourquoi il faut une vue d'ensemble qui exige l'effort commun d'une recherche interdisciplinaire qui, aujourd'hui, devrait être possible dans une Europe unie.

Président d'honneur de la Société de Linguistique Romane
Professeur émérite à l'Université de la Sarre Max PFISTER

Références bibliographiques

Ament, Hermann, 1996. «Die Franken in den Römerstädten der Rheinzone», in : *Die Franken – Wegbereiter Europas / Les Francs – Pionniers de l'Europe*, Mainz, Philipp von Zabern, 129-137.

Bach, Adolf, 19812. *Deutsche Namenkunde II : Die deutschen Ortsnamen*, Heidelberg, Winter.

Besse, Maria, 1997. *Namenpaare an der Sprachgrenze. Eine lautchronologische Untersuchung zu zweisprachigen Ortsnamen im Norden und Süden der deutsch-französischen Sprachgrenze*, Tübingen, Niemeyer.

Bierbrauer, Volker, 1996. «Romanen im fränkischen Siedelgebiet», in : *Die Franken – Wegbereiter Europas / Les Francs – Pionniers de l'Europe*, Mainz, Philipp von Zabern, 110-120.

Bierbrauer, Volker, 2012. «Christliche Jenseitsvorstellungen und romanische Beigabensitten vom 5. bis zum 6./7. Jahrhundert», *Grosso Modo – Quellen und Funde aus Spätantike und Mittelalter. Festschrift für Gerhard Fingerlin zum 75. Geburtstag*, Weinstadt, Greiner, 39-50.

Boppert, Walburg / Haubrichs, Wolfgang, 1998/1999. «Frühgeschichtlicher Grabstein des Aigettheus aus Worms», *Mainzer Archäologische Zeitschrift* 5/6, 229-240.

Buchmüller-Pfaff, Monika, 1990. *Siedlungsnamen zwischen Spätantike und frühem Mittelalter. Die -(i)acum-Namen der römischen Provinz Belgica Prima*, Tübingen, Niemeyer.

Buchmüller-Pfaff, Monika, 1991. «Namen im Grenzland. Methoden, Aspekte und Zielsetzung in der Erforschung der lothringisch-saarländischen Toponomastik», *Francia* 18, 165-194.

Engels, Heinz. 1961. *Die Ortsnamen an Mosel, Sauer und Saar und ihre Bedeutung für eine Besiedlungsgeschichte*, Trier (Schriftenreihe zur Trierischen Landesgeschichte und Volkskunde 7).

Ewig, Eugen, 1954. *Trier im Merowingerreich. Civitas, Stadt, Bistum*, Trier, Paulinus-Verlag.

Ewig, Eugen, 1980. «Frühes Mittelalter», in : Petri, Franz / Droege, Georg (ed.), *Rheinische Geschichte in drei Bänden*, vol. 1, 2, Düsseldorf, Schwann.

Ewig, Eugen, 1988. *Die Merowinger und das Frankenreich*, Stuttgart, W. Kohlhammer.

Fingerlin, Gerhard, 1974. «Zur alamannischen Siedlungsgeschichte des 3.-7. Jahrhunderts», in : Hübener, Wolfgang (ed.), *Die Alemannen in der Frühzeit*, Bühl (Baden), Konkordia, 45-88.

Gärtner, Kurt / Holtus, Günter, 1995. «Die erste deutsch-französische ‹Parallelurkunde›. Zur Überlieferung und Sprache der Straßburger Eide», in : Gärtner, Kurt / Holtus, Günter (ed.), *Beiträge zum Sprachkontakt und zu den Urkundensprachen zwischen Maas und Rhein*, Trier, Verlag Trierer Historische Forschungen, 97-127.

Gamillscheg, Ernst, 19702. *Romania Germanica. Sprach- und Siedlungsgeschichte der Germanen auf dem Boden des alten Römerreiches*, Berlin / New York, de Gruyter.

Geuenich, Dieter, 1997. *Geschichte der Alemannen*, Stuttgart, Kohlhammer.

Geuenich, Dieter / Haubrichs, Wolfgang / Jarnut, Jörg, 1997. *Nomen et gens. Zur historischen Aussagekraft frühmittelalterlicher Personennamen*, Berlin/New York, de Gruyter.

Grünewald, Mathilde, 1996. «Worms zwischen Burgunden und Saliern», in : *Die Franken – Wegbereiter Europas / Les Francs – Pionniers de l'Europe*, Mainz, Philipp von Zabern, 160-162.

Grünewald, Mathilde, 2012. «Die vermeintliche Völkerlawine der Neujahrsnacht 406/407», in : *Grosso Modo – Quellen und Funde aus Spätantike und Mittelalter. Festschrift für Gerhard Fingerlin zum 75. Geburtstag*, Weinstadt, Greiner, 1-6.

Haubrichs, Wolfgang, 1983 : «Siedlungsnamen und frühe Raumorganisation im oberen Saargau», in : Haubrichs, Wolfgang / Ramge, Hans, *Zwischen den Sprachen. Siedlungs- und Flurnamen in germanisch-romanischen Grenzgebieten*, *Beiträge des Saarbrücker Kolloquiums vom 9.-11. Oktober 1980*, Saarbrücken, SDV, 221-287.

Haubrichs, Wolfgang, 1992. «Germania Submersa. Zu Fragen der Quantität und Dauer germanischer Siedlungsinseln im romanischen Lothringen und in Südbelgien», in : Burger, Harald et al. (ed.), *Verborum Amor. Studien zur Geschichte und Kunst der deutschen Sprache. Festschrift Stefan Sonderegger*, Berlin/New York, de Gruyter, 633-666.

Haubrichs, Wolfgang, 1997. «Galloromanische Kontinuität zwischen unterer Saar und Mosel. Problematik und Chancen einer Auswertung der toponymischen Zeugnisse aus germanistischer Perspektive», in : Holtus, Günter / Kramer, Johannes / Schweickard, Wolfgang (ed.), *Italica et Romanica. Festschrift für Max Pfister*, Tübingen, Niemeyer, 3, 211-237.

Haubrichs, Wolfgang, 1998a. «Fränkische Lehnwörter, Ortsnamen und Personennamen im Nordwesten der Gallia. Die 'Germania submersa' als Quelle der Sprach- und Siedlungsgeschichte», in : Geuenich, Dieter (ed.), *Die Franken und die Alemannen bis zur «Schlacht bei Zülpich» (496/497)*, Berlin/New York, de Gruyter, 102-129.

Haubrichs, Wolfgang, 1998b. «Romanen an Rhein und Mosel. Onomastische Reflexionen», in : Ernst, Peter / Patocka, Franz (ed.), *Deutsche Sprache in Raum und Zeit. Festschrift für Peter Wiesinger*, Wien, Edition Praesens, 379-413.

Haubrichs, Wolfgang, 2004a. «Romano-germanische Hybridnamen des frühen Mittelalters nördlich der Alpen», in : Hägermann, Dieter (ed.), *Akkulturation. Probleme einer germanisch-romanischen Kultursynthese in Spätantike und frühem Mittelalter*, Berlin/New York, de Gruyter, 179-203.

Haubrichs, Wolfgang, 2004b. «Identität und Name. Akkulturationsvorgänge in Namen und die Traditionsgesellschaften des frühen Mittelalters», in : Pohl, Walter (ed.), *Die Suche nach den Ursprüngen. Von der Bedeutung des frühen Mittelalters*, Wien, Verlag der Österreichischen Akademie der Wissenschaften, 85-105.

Haubrichs, Wolfgang, 2008. «Hybridität und Integration. Vom Siegeszug und Untergang des germanischen Personennamensystems in der Romania», in : Dahmen, Wolfgang (ed.), *Zur Bedeutung der Namenkunde für die Romanistik, Romanistisches Kolloquium XXII*, Tübingen, Narr, 87-140.

Haubrichs, Wolfgang, 2009. «Sprachliche Integration, Sprachinseln und Sprachgrenzbildung im Bereich der östlichen Gallia», in : Kolzer, Theo / Schiefer, Rudolf (ed.), *Von der Spätantike zum frühen Mittelalter : Kontinuitäten und Brüche, Konzeptionen und Befunde*, Ostfildern, Thorbeke, 61-99.

Haubrichs, Wolfgang / Pfister, Max, 1983. «Tholey – Name und Geschichte», *Tholeyer Brief* 11, 13-18.

Haubrichs, Wolfgang / Pfister, Max, 2001. «Die Prümer Romania», in : Bentzinger, Rudolf / Nübling, Damaris / Steffens, Rudolf (ed.), *Sprachgeschichte – Dialektologie – Onomastik – Volkskunde, Beiträge zum Kolloquium am 3./4. Dezember 1999 an der Johannes Gutenberg-Universität (Wolfgang Kleiber zum 70. Geburtstag)*, Stuttgart, Steiner, 169-195.

Hubschmid, Johannes, 1981. «Etymologische und sprachgeographische Bemerkungen zu romanischen Ortsnamen aus dem Moselland», *Zeitschrift für romanische Philologie* 97, 499-503.

Hubschmid, Johannes, 1983. «Ortsnamenforschung im germanisch-romanischen Grenzgebiet und romanische Sprachgeographie», in : Haubrichs, Wolfgang / Ramge, Hans, *Zwischen den Sprachen. Siedlungs- und Flurnamen in germanisch-romanischen Grenzgebieten, Beiträge des Saarbrücker Kolloquiums vom 9.-11. Oktober 1980*, Saarbrücken, SDV, 89-120.

Irsigler, Franz / Löffler, Günter (ed.), 1982–. *Geschichtlicher Atlas der Rheinlande*, Im Auftrag der Gesellschaft für Rheinische Geschichtskunde in Verbindung mit dem Landschaftsverband Rheinland hg. von F. I. und G. L., Köln, Rheinland (Publikationen der Gesellschaft für Rheinische Geschichtskunde, XII. Abteilung 1a. N.F.).

Jedin, Hubert / Latourette, Kenneth Scott / Martin, Jochen, 1990. *Atlas d'histoire de l'Eglise. Les églises chrétiennes hier et aujourd'hui*, Turnhout, Brepols.

Jochum-Godglück, Christa, 2012. «Walchensiedlungsnamen und ihre historische Aussagekraft», in : Fehr, Hubert / Heitmeier, Irmtraut (ed.), *Die Anfänge Bayerns. Von Rätien und Noricum zur frühmittelalterlichen Baiovaria*, St. Ottilien, EOS-Verlag, 197-217.

Jud, Jakob, 1917. «Probleme der altromanischen Wortgeographie», *Zeitschrift für romanische Philologie* 38, 1-75.

Jungandreas, Wolfgang : «Deutsche und Romanen an Mosel und Saar», in : *Fragen der Landes- und Volksforschung in den Gebieten um Mosel und Saar. Protokoll der Tagung der Arbeitsgemeinschaft für westdeutsche Landes- und Volksforschung (Bonn) in Saarburg vom 28.9.-1.10.1953*, Bonn, 13-17.

Jungandreas, Wolfgang, 1955. «Ein romanischer Dialekt an der Mosel zwischen Eifel und Hunsrück um 1200», *Zeitschrift für romanische Philologie* 71, 414-421.

Jungandreas, Wolfgang, 1962. *Historisches Lexikon der Siedlungs- und Flurnamen des Mosellandes*, Trier, Lintz (Schriftenreihe zur Trierischen Landesgeschichte und Volkskunde 8).

Jungandreas, Wolfgang, 1971. «Die Moselromanen. Die romanische Moselenklave während der fränkischen Landnahme», *Zeitschrift für romanische Philologie* 87, 32-73.

Jungandreas, Wolfgang, 1979. *Zur Geschichte des Moselromanischen. Studien zur Lautchronologie und zur Winzerlexik*, Wiesbaden, Steiner (Mainzer Studien zur Sprach- und Volksforschung 3).

Kleiber, Wolfgang, 1973. «Zwischen Antike und Mittelalter. Das Kontinuitätsproblem in Südwestdeutschland im Lichte der Sprachgeschichtsforschung», *Frühmittelalterliche Studien* 7, 27-52.

Kleiber, Wolfgang, 1974. «Die romanische Sprachinsel an der Mosel im Spiegel der Reliktwörter», *Kurtrierisches Jahrbuch* 14, 16-32.

Kleiber, Wolfgang, 1979. «Waber/Feber – Naf/Nef. Zwei moselländische Flurnamen gallischer Herkunft.», in : *Mosel, Eifel, Hunsrück. Der Landkreis Cochem-Zell. Landschaft – Kultur – Geschichte - Wirtschaft. Zum 10jährigen Bestehen des Kreises Cochem*, Landkreis Cochem-Zell.

Kleiber, Wolfgang, 1983. «Das moselromanische Substrat im Lichte der Toponymie und Dialektologie.», in : Haubrichs, Wolfgang / Ramge, Hans, *Zwischen den Sprachen. Siedlungs- und Flurnamen in germanisch-romanischen Grenzgebieten, Beiträge des Saarbrücker Kolloquiums vom 9.-11. Oktober 1980*, Saarbrücken, SDV, 153-192.

Kleiber, Wolfgang, 1985. «Probleme romanisch-germanischen Sprachkontakts an der Mosel vornehmlich im Bereich der Prosodie von Toponymen.», in : Schützeichel, Rudolf (ed.), *Giessener Flurnamen-Kolloquium 1984*, Heidelberg, Winter, 582-545.

Kleiber, Wolfgang / Pfister, Max, 1992. *Aspekte und Probleme der römisch-germanischen Kontinuität. Sprachkontinuität an Mosel, Mittel- und Oberrhein sowie im Schwarzwald*, Stuttgart, Steiner.

Müller, Wulf, 1983. «Une ancienne zone de contact : Le Val d'Orbey (Haute Alsace)», in : Haubrichs, Wolfgang / Ramge, Hans, *Zwischen den Sprachen. Siedlungs- und Flurnamen in germanisch-romanischen Grenzgebieten, Beiträge des Saarbrücker Kolloquiums vom 9.-11. Oktober 1980*, Saarbrücken, SDV, 313-341.

Müller, Wulf, 1986. «Zu den romanischen Namen des Mosellandes», *Rheinische Vierteljahrsblätter* 50, 1-15.

Parisse, Michel, 1995. «La Lotharingie : Naissance d'un espace politique, Lotharinghia», in : Herrmann, Hans-Walter / Schneider, Reinhard, *Lotharingia, Eine europäische Kernlandschaft um da Jahr 1000 / Une région au centre de l'Europe autour de l'an Mille*, Saarbrücken, SDV, 31-48.

Pauly, Ferdinand, 1965. *Siedlung und Pfarrorganisation im alten Erzbistum Trier*, Bd. 4, *Das Landkapitel Wadrill*, Bonn, L. Röhrscheid.

Petri, Franz, 1937. *Germanisches Volkserbe in Wallonien und Nordfrankreich : die fränkische Landnahme in Frankreich und die Bildung der westlichen Sprachgrenze*, Bonn, L. Röhrscheid.

Petri, Franz, 1977. *Die fränkische Landnahme und die Entstehung der germanisch-romanischen Sprachgrenze in der interdisziplinären Diskussion*, Darmstadt, Wissenschaftliche Buchgesellschaft.

Pfister, Max, 1972. «Die sprachlichen Berührungen zwischen Franken und Galloromanen. Forschungsbericht anhand von Ernst Gamillscheg, Romania Germanica. Sprach- und Siedlungsgeschichte der Germanen auf dem Boden des alten Römerreiches», *Zeitschrift für romanische Philologie* 88, 175-193.

Pfister, Max, 1973. «La répartition géographique des éléments franciques en gallo-roman», *Revue de linguistique romane* 37, 126-149.

Pfister, Max, 1978. «Le superstrat germanique dans les langues romanes.», in : *Atti del XIV Congresso Internazionale di Linguistica e Filologia Romanza (Napoli 15-20 aprile 1974)*, Napoli/Amsterdam, Macchiaroli/Benjamins, I, 49-97.

Pfister, Max, 1983. «Galloromanische Relikte in der Toponomastik Ostlothringens und des Saarlandes.», in : Haubrichs, Wolfgang / Ramge, Hans, *Zwischen den Sprachen. Siedlungs- und Flurnamen in germanisch-romanischen Grenzgebieten, Beiträge des Saarbrücker Kolloquiums vom 9.-11. Oktober 1980*, Saarbrücken, SDV, 121-152.

Pfister, Max, 1985. «Tholey und der saarländisch-lothringische Raum zwischen Romania und Germania. Sprachwissenschaftliche und ortsnamenkundliche Überlegungen», in : *Tholey [634 - 1984], Wissenschaftliche Vorträge gehalten aus Anlaß des 1350-jährigen Jubiläums von Gemeinde und Abtei Tholey*, St. Ottilien, EOS-Verlag, 282-298.

Pfister, Max, 1990. «Die Moselromania und die romanischen Reliktzonen im Hochwald-Mittelrheingebiet und im Schwarzwald», in : Kramer, Johannes / Winkelmann, Otto (ed.), *Das Galloromanische in Deutschland*, Wilhelmsfeld (Pro Lingua 8), 11-32.

Pfister, Max, 1995. «Die sprachliche Situation zwischen Maas und Rhein im Frühmittelalter» in : Gärtner, Kurt / Holtus, Günter (ed.), *Beiträge zum Sprachkontakt und zu den Urkundensprachen zwischen Maas und Rhein*, Trier, Verlag Trierer Historische Forschungen, 61-96.

Pfister, Max / Haubrichs, Wolfgang, 1998. «Romania – Germania : Die Bedeutung von Ortsnamen für die Sprachgeschichte im Grenzgebiet zweier Sprachen», in : Debus, Friedhelm (ed.), *Romania – Germania : Die Bedeutung von Ortsnamen für die Sprachgeschichte im Grenzgebiet zweier Sprachen* (Jahrespreise 1996 und 1997 der Henning-Kaufmann-Stiftung zur Förderung der Deutschen Namenforschung auf sprachgeschichtlicher Grundlage), Heidelberg, Winter, 35-61.

Pitz, Martina, 1999. «Toponymie zwischen den Sprachen. Ortsnamen als Instrument landes- und siedlungsgeschichtlicher Forschung im lothringischen Sprachgrenzraum», in : Brogiato, Heinz Peter (ed.), *Geographische Namen in ihrer Bedeutung für die landeskundliche Forschung und Darstellung*, Trier, Dokumentationszentrum für Deutsche Landeskunde, Universität Trier, 67-95.

Pitz, Martina, 2001a. «La genèse de la frontière des langues en Lorraine. Éléments pour un argumentaire philologique et toponymique», in : Demarolle, Jeanne-Marie (ed.), *Frontières en Europe occidentale et médiane de l'Antiquité à l'An 2000*, Metz, Centre de recherche histoire et civilisation de l'Université de Metz, 73-107.

Pitz, Martina, 2001b. «Genuine Übersetzungspaare primärer Siedlungsnamen an der lothringischen Sprachgrenze : Überlegungen zu ihrem sprach- und siedlungsgeschichtlichen Aussagewert», *Onoma* 36, 255- 288.

Pitz, Martina, 2002a. «In pago Albensel/Albechowa. Onomasiologische Reflexe frühmittelalterlicher Mischsiedlung von Franken und Romanen im heute romanophonen Lothringen», in : Ernst, Peter (ed.), *Ortsnamen und Siedlungsgeschichte, Akten des Symposiums in Wien vom 28.-30.9.2000*, Heidelberg, Winter, 95-106.

Pitz, Martina, 2002b. «Nouvelles données pour l'anthroponymie de la Galloromania : les toponymes mérovingiens du type Avricourt», *Revue de linguistique romane* 66, 421-449.

Pitz, Martina, 2003. «Zewen und Ziefen, Tawern und Tafers : Zur Integration von romanischem [v] als [w] oder [f] in vorgermanischen Ortsnamen der linksrheinischen Germanisierungszonen des römische Reiches», in : Anreiter, Peter / Plangg, Guntram A. (ed.), *Namen in Grenzregionen. Tagungsband des internationalen onomastischen Symposiums in Klingenthal / Elsass (7.-11.5.2001)*, Wien, Edition Praesens, 101-138.

Pitz, Martina, 2008. «Zu Genese und Lebensdauer der romanischen Sprachinseln an Mosel und Mittelrhein», in : Greule, Albrecht (ed.), *Studien zu Literatur, Sprache und Geschichte in Europa. Wolfgang Haubrichs zum 65. Geburtstag gewidmet*, St. Ingbert, Röhrig, 439-450.

Pitz, Martina / Stein, Frauke, 2000. «Genèse linguistique d'une région frontalière. Les environs de Forbach et Sarreguemines», *Les Cahiers Lorrains* 19/3, 365-412.

Post, Rudolf, 1982. *Romanische Entlehnungen in den westmitteldeutschen Mundarten. Diatopische, diachronische und diastratische Untersuchungen zur sprachlichen Interferenz am Beispiel des landwirtschaftlichen Sachwortschatzes*, Wiesbaden, Steiner.

Post, Rudolf, 1985. *Lehn- und Reliktwörter im Rheinland. Karte X, 1.1-X, 1.4 zum Geschichtlichen Atlas der Rheinlande*, 2. Lieferung, Köln, Rheinland-Verlag.

Post, Rudolf, 1989. «Galloromanische Reliktwortareale und Grenzlautentlehnungen im Pfälzischen», in : Greule, Albrecht (ed.), *Sprache, Literatur, Kultur. Studien zu ihrer Geschichte im deutschen Süden und Westen. Wolfgang Kleiber zu seinem 60. Geburtstag gewidmet*, Stuttgart, Steiner, 161-174.

Prinz, Friedrich, 1994. «Formen, Phasen, und Regionen des Übergangs von der Spätantique zum Frühmittelalter: Reliktkultur – neue Ethnika – interkulturelle Synthese im Frankenreich», in: Franz Staab (ed.), *Zur Kontinuität zwischen Antike und Mittelalter am Oberrhein*, Sigmaringen, Thorbecke, 171-192.

Staerk, Dieter, 1976. *Die Wüstungen des Saarlandes. Beiträge zur Siedlungsgeschichte des Saarraumes vom Frühmittelalter bis zur französischen Revolution*, Saarbrücken, Minerva-Verlag Thinnes und Nolte.

Stein, Frauke, 1974 «Franken und Romanen in Lothringen», in : Kossack, G. / Ulbert, G. (ed.), *Studien zur vor- und frühgeschichtlichen Archäologie. Festschrift für Joachim Werner zum 65. Geburtstag*, Teil II : *Frühmittelalter*, München, Beck, 579-589.

Stein, Frauke, 2011a. «Frühmittelalterliche Bevölkerungsverhältnisse im Saar-Mosel-Raum. Voraussetzungen der Ausbildung der deutsch-französischen Sprachgrenze», in : Frauke Stein, *Franken und Romanen*. Aufsätze aus 25 Jahren Forschung, anlässlich ihres 75. Geburtstages ausgewählt und herausgegeben von Rolf Hachmann und Volker Bierbrauer, Bonn, 179-200.

Stein, Frauke, 2011b. «Kulturelle Ausgleichsprozesse zwischen Franken und Romanen im 7. Jahrhundert. Eine archäologische Untersuchung zu den Verhaltensweisen der Bestattungsgemeinschaft von †Rency/Renzig bei Audun-le-Tiche in Lothringen» in : Frauke Stein, *Franken und Romanen*. Aufsätze aus 25 Jahren Forschung, anlässlich ihres 75. Geburtstages ausgewählt und herausgegeben von Rolf Hachmann und Volker Bierbrauer, Bonn, 325-350.

Stein, Frauke, 2011c. «L'adoption d'anthroponymes germaniques par les populations autochtones de la Galloromania à la lumière des données archéologiques. L'exemple de la Lorraine romane», in : Frauke Stein, *Franken und Romanen*. Aufsätze aus 25 Jahren Forschung, anlässlich ihres 75. Geburtstages ausgewählt und herausgegeben von Rolf Hachmann und Volker Bierbrauer, Bonn, 395-420.

Stein, Frauke, 2011d. «Ergebnisse zur Interferenz zwischen Franken und Romanen im frühen Mittelalter anhand des Gräberfeldes bei Audun-le-Tiche (F, Dép. Moselle)», in : Frauke Stein, *Franken und Romanen*. Aufsätze aus 25 Jahren Forschung, anlässlich ihres 75. Geburtstages ausgewählt und herausgegeben von Rolf Hachmann und Volker Bierbrauer, Bonn, 367-393.

Wieczorek, Alfried, 1996. «Die Ausbreitung der fränkischen Herrschaft in den Rheinlanden vor und seit Chlodwig I.», in : *Die Franken – Wegbereiter Europas / Les Francs – Pionniers de l'Europe*, Mainz, Philipp von Zabern, 241-260.

Table ronde « Quels corpus et quels outils d'exploitation de corpus pour les études de linguistique et philologie romanes : l'unité de la romanistique »

1. Pierre Kunstman : introduction

Quand les organisateurs de notre congrès m'ont demandé, l'an dernier, d'animer la présente table ronde sur les questions « Quels corpus et quels outils d'exploitation de corpus », il s'agissait bien entendu dans leur esprit de corpus et d'outils électroniques. Non que les corpus statiques imprimés soient aujourd'hui sans valeur : ils sont encore utilisés, quoique rarement. J'avoue d'ailleurs avoir commis moi-même un travail de ce genre. Petite anecdote : il y a plus d'une trentaine d'années, ma *Concordance analytique de la Mort le Roi Artu* fut publiée aux Éditions de l'Université d'Ottawa en deux gros volumes, soit un total de pas moins de 2 000 pages (Kunstmann 1982). Le regretté Kurt Baldinger, l'un des phares de la linguistique romane, dans un compte rendu qu'il fit paraître de mes deux volumes dans la revue de notre société, parla de *Heidenarbeit*, travail de bénédictin. L'expression ne manqua pas de me surprendre, car j'avais alors pris résolument le parti du numérique et je procédai avec une vitesse et une précision que notre confrère n'avait sans doute pas pu imaginer. Nous assistions alors aux nouvelles *Noces de Philologie et de Mercure*, à l'« alliance heureuse du génie informatique et de l'érudition textuelle » (Kunstmann 2012, 51). Les nouvelles technologies ont apporté au linguiste et au philologue des aides considérables et des perspectives nouvelles, qui peuvent dérouter ou donner le vertige. Si elles ne sauraient remplacer le jugement critique de l'expert, celui-ci, de son côté, ne saurait non plus se passer d'elles à chaque étape de sa recherche.

Aussi avons-nous choisi d'aborder cinq grandes questions qui nous semblent cruciales pour nos travaux, nos disciplines et nos champs d'investigation. Sans vouloir empiéter sur les contributions de nos intervenants, j'exprimerai brièvement et en termes généraux, pour chacun des thèmes retenus, ses caractères principaux et ses enjeux. Question préliminaire toutefois : qu'est-ce qu'un corpus ? Pour la plupart des spécialistes, ce n'est pas seulement un ensemble de mots ou de textes, mais un regroupement structuré suivant des critères linguistiques. Le corpus, à proprement parler, se distingue donc de la simple base textuelle ainsi que de l'hypertexte, dont le système le plus vaste est la toile d'araignée mondiale, le WWW. Pour ma part, j'adopterai la définition proposée par F. Rastier (2004, 32) : « un corpus est un regroupement struc-

turé de textes intégraux, documentés, éventuellement enrichis par des étiquetages, et rassemblés : (i) de manière théorique réflexive en tenant compte des discours et des genres, et (ii) de manière pratique en vue d'une gamme d'applications ». Définition qui me permet, par une transition naturelle, d'introduire les deux premiers thèmes de la table ronde.

1.1 L'annotation de corpus

L'annotation de corpus : dans les textes ou autour des textes. Le balisage interne constitue un apport hautement appréciable ; mais il n'est pas indispensable : un corpus brut est parfois préférable à un corpus mal codifié. La première étape de la codification est la lemmatisation, opération jadis très longue quand elle était effectuée manuellement, mais beaucoup plus rapide de nos jours avec l'aide de l'ordinateur. Les résultats sont presque parfaits pour les textes contemporains, mais ils sont aussi, dans certains cas, remarquables pour les textes anciens. Je tiens à mentionner l'outil conçu et développé par G. Souvay, au laboratoire qui nous accueille ici à Nancy : le LGeRM (Lemmes, Graphies et Règles Morphologiques), qui permet de « faciliter la consultation d'un dictionnaire, l'interrogation et la lemmatisation de textes médiévaux et trouve des applications dans l'édition électronique de manuscrits et la construction automatique de glossaires » (Souvay / Pierrel 2009, 149). La deuxième étape du balisage interne est l'étiquetage morphologique et parfois aussi syntaxique : le premier est relativement facile à effectuer, le second constitue une opération plus complexe. Enfin le balisage externe est absolument indispensable : protocoles de sélection et d'établissement de documents, constitution d'en-têtes permettant une identification précise des documents et de leurs auteurs. Qu'on me permette de citer, à titre d'exemple, le travail d'une collègue de mon université, F. Martineau : son corpus MCVF (Modéliser le changement : les voies du français) comprend deux millions et demi de mots couvrant les périodes de l'ancien français au français classique, annoté suivant le protocole TEI et étiqueté pour la morphologie et la syntaxe, avec en-tête permettant une identification des textes et des auteurs ; le corpus est accessible en ligne sur le site du projet.

1.2 La comparaison entre corpus

Il importe de distinguer si le travail comparatif s'effectue dans le cadre d'une même langue ou entre deux ou plusieurs langues romanes. Dans le premier cas, on s'intéressera surtout au phénomène de la variation : les quatre types traditionnels de Coseriu (diachronique, diatopique, diastratique, diaphasique), auxquels s'ajoutent la différence entre écrit et oral, celle entre les différents canaux de transmission de la parole (les nouvelles technologies brouillant souvent la limite entre oral et écrit) et même la distinction entre les sexes (diagénique). En ce qui a trait à la comparaison entre deux ou plusieurs langues romanes, il serait bon de s'interroger sur l'utilité respective des corpus parallèles (le même texte dans plusieurs langues) et des corpus comparables (différents textes de plusieurs langues, qu'on peut rapprocher suivant plusieurs paramètres).

1.3 L'édition de textes et de documents sonores

L'édition de textes et de documents sonores pose des problèmes considérables. Après la publication récente des meilleurs *tweets* de B. Pivot (*Les tweets sont des chats*, Pivot 2013), on ne saurait négliger le ' gazouillis ', ce type d'énoncé si répandu maintenant de Bucarest à Valparaiso. Comment aborder aussi la production textuelle sur le Web ?

1.4 La pérennisation des documents

La perpetuazione dei documenti è un problema cruciale ed anche angoscioso per chi non voglia lavorare con materiali effimeri: come conservare, diffondere, transcodificare ciò che ha richiesto delle settimane, dei mesi, ed anche spesso degli anni di lavoro?

1.5 Les études linguistiques portant sur les langues romanes

Por fin, ya que los organizadores quisieron que subrayáramos la unidad de la romanística, conviene examinar la importancia de los corpus multilingües para los estudios contrastivos en la área de la Romania (Europa, África y Américas) e interrogarse sobre el papel de nuestra Sociedad para una mejor coordinación de los estudios romances.

2. L'annotation de corpus

L'annotation de corpus : dans les textes (lemmatisation, étiquetage morphosyntaxique, gloses diverses) et autour des textes (protocoles de sélection et d'établissement des documents, constitution d'en-têtes permettant une identification précise des documents et de leurs auteurs). Vers une standardisation.

2.1 Point de vue d'Andrea Bozzi

Il tema dell'annotazione dei testi, in realtà, permette di affrontare un tema ancora più vasto, ovvero il rapporto fra la Linguistica Computazionale, che da anni è impegnata nello studio di procedure di analisi e annotazione, e gli studi di carattere filologico-computazionale, nella prospettiva di mettere a disposizione dei ricercatori un'infrastruttura adeguata alle proprie attività. L'elemento centrale riguarda l'adozione di standard per la codifica e la rappresentazione dei dati. Si tratta di un tema molto sentito da tempo e discusso anche a livello Europeo: lo *Standing Committee for the Humanities* della *European Science Foundation* (ESF), per fare solo un esempio, ha prodotto un documento interessante due anni fa, utile e ben documentato (si veda European Science Foundation 2011). Il fatto di avere a disposizione sistemi di codifica standard, come la TEI (*Text Encoding Initiative*, oggi disponibile anche nel formato xml)[1], aumenta la possibilità di interscambio dei dati realizzati in centri diversi

[1] Il manuale completo è disponibile all'indirizzo <www.tei-c.org/Guidelines>.

dal momento che l'analisi linguistica e filologica, ivi comprese le informazioni strutturate degli apparati critici e della bibliografia, segue criteri di marcatura espliciti e condivisi. È tuttavia necessario sottolineare due aspetti della questione, in particolare per quanto riguarda la TEI:

(1) La molteplicità delle attività e delle finalità di studio e di ricerca sui testi rende molto improbabile ottenere una standardizzazione esaustiva e da tutti condivisa. È illusorio pensare che gli interventi, nel campo delle varie discipline filologiche, siano suscettibili di una strutturazione preventiva costringendo, di fatto, i singoli ricercatori a seguire comportamenti e norme completamente definite a priori. Questo atteggiamento, purtroppo, è stato ed è ancora vivo nella visione universalistica dei cultori della TEI e, a mio giudizio, esso non favorisce la conversione al digitale di coloro i quali ancora indugiano a servirsi di mezzi di indagine tradizionali anche su documenti già disponibili in formato numerico grazie all'aggiornamento tecnologico che si è verificato nelle biblioteche e negli archivi di tutto il mondo.

(2) Gli attuali sistemi di marcatura (come, appunto, la TEI che è diffusa a livello planetario) sono poco orientati all'utente finale inesperto e necessitano di una serie di interventi per migliorare il modo con cui essi siano impiegati senza l'ausilio di personale esperto. È impensabile che chi si trovi ad affrontare problemi linguistici o editoriali talvolta molto complessi distolga il proprio impegno per accedere a liste di marcatori fra i quali scegliere quelli prescritti per ogni specifica situazione. Pur ribadendo che la macchina potrà fornire risultati importanti solo se disporrà di parti codificate bene ed in forma omogenea, è tuttavia indispensabile che si metta lo studioso nelle condizioni migliori e meno defatiganti per intervenire in questa delicata fase di preparazione dei dati. Egli deve disporre di un menù a forma di catalogo per selezionare, con la sola pressione del tasto del mouse, la voce da utilizzare e deve essere compito del sistema inserire il codice di marcatura ad essa corrispondente nella posizione indicata della versione digitale del testo.

Una ulteriore interessante prospettiva che si offre ai ricercatori è costituita, come si diceva, dalle infrastrutture di ricerca dove, tuttavia, deve ancora essere maggiormente rappresentata la comunità di studiosi che operano nel campo della filologia e della critica del testo. Ad oggi, infatti, né CLARIN né DARIAH[2], le due infrastrutture più sviluppate nel campo delle Scienze Umane, sono in grado di fornire risposte sufficienti ai bisogni della filologia del testo, nonostante l'enorme sviluppo della tecnologia digitale imponga urgentemente di trovare soluzioni adeguate in proposito.

2.2 Point de vue de Giovanni Parodi

En el marco de esta mesa redonda, resulta muy oportuno reflexionar acerca de los requisitos que se debe contar para la constitución de un corpus. Así, un referente importante es la propuesta de EAGLES (1996), en la que se propone recomendaciones para que un corpus pueda considerarse como tal: (1) el corpus debe ser lo más

[2] Si vedano i siti ufficiali dei due progetti <www.clarin.eu> e <www.dariah.eu>.

extenso posible de acuerdo con las tecnologías disponibles en cada época; (2) debe incluir ejemplos de amplia gama de materiales en función de ser lo más representativo posible; (3) debe existir una clasificación intermedia en los géneros entre el corpus en total y las muestras individuales; (4) las muestras deben de ser tamaños similares; (5) el corpus, como un todo, debe tener una procedencia clara.

En esta misma línea, algunos autores sugieren cuatro ventajas para adoptar una aproximación basada en corpus: (1) adecuada representación del discurso en su forma de ocurrencia natural en muestras amplias y representativas a partir de textos originales; (2) procesamiento lingüístico (semi-)automático de los textos mediante el uso de computadores (Biber / Reppen / Clark / Walter 2001); ello permite análisis más amplios y profundos de los textos mediante conjuntos de rasgos lingüísticos caracterizadores; (3) mayor confiabilidad y certeza en los análisis cuantitativos de los rasgos lingüísticos en grandes muestras de textos; (4) posibilidad de resultados acumulativos y replicables: posteriores investigaciones pueden utilizar los mismos corpus u otros pueden ser analizados con las mismas herramientas computacionales.

Ahora bien, ya sea si el corpus debe ser necesariamente de tipo digital o si aun es factible pensar en un conjunto de textos en papel, otros rasgos resultan más relevantes. También se hace evidente que el asunto de la extensión cobra importancia. Seguramente se dirá que ello depende en gran medida de los objetivos de la investigación. Sin embargo, sin importar que rasgos se desea privilegiar, si se busca un proceso de investigación sinérgico con resultados de índole acumulativa y posibilidad de replicación, resulta indudable que se debe adherir a la mayoría de las indicaciones propuestas.

Desde este marco se pueden identificar en mi opinión, al menos, ocho características relevantes, llegado el momento de construir y comprender los alcances de un corpus. Ellas se listan a continuación sin mediar ningún sesgo jerárquico. Al mismo tiempo, cabe destacar que este conjunto no está cerrado: (1) extensión, (2) formato, (3) representatividad, (4) diversificación, (5) marcado o etiquetado morfosintáctico, (6) procedencia, (7) tamaño de las muestras, y (8) clasificación y adscripciones de tipo disciplinar, genérico, temático, etc. (Parodi 2010a).

Estas características se constituyen en principios a tener en cuenta, dependiendo de los objetivos de cada investigador y de las posibilidades tecnológicas al alcance. No obstante lo anterior, el concepto de corpus muestra una característica que se hace relevante y pone en evidencia ciertas complejidades: aquella denominada representatividad. Es bien sabido que incluso los grandes corpus no logran dar cuenta de la lengua como un todo ni tampoco se pretende que así sea. La lengua en su dinamismo y heterogeneidad es mucho más rica de lo que se puede imaginar y no logra ser captada en un solo corpus, por gigantesco que sea su tamaño. Como se sabe, un corpus puede ofrecer información detallada acerca de una lengua particular, pero es imposible recolectar un corpus que abarque toda una lengua. Si ese fuera el caso, sería necesario recolectar todos los usos de dicha lengua. De este modo, se debe siempre tener presente que un corpus es sólo una colección finita de, en el caso de la lengua, un universo infinito.

Ahora bien, en el foco del tema central, desde el punto de vista de la anotación, sin duda, el contar con etiquetadores morfosintácticos automáticos o semi-automáticos se vuelve un asunto fundamental. Un tema central en este punto lo constituye el tipo de gramática que subyace a esos marcajes morfosintácticos y los sustentos tanto teóricos como empíricos que dan origen a las categorías y nomenclaturas empleadas. Como se sabe, no existe un criterio conceptual unívoco de gramática y esta posible diversidad da origen a modelos diversos, hecho que afecta los tipos de *taggers* y *parsers* existentes.

Paralelamente, existe un tipo estadístico computacional que se ha hecho muy popular en los últimos años. Este es conocido por su sigla en inglés: LSA (*Latent Semantic Analysis*). Este sistema de análisis semántico, que no presta atención a la morfología ni a la sintaxis, opera sobre grandes cantidades de textos planos no marcados y fue desarrollado inicialmente para el idioma inglés (Deerwester / Dumais / Furnas / Landauer & Harshman 1990; Dumais 1994). En la actualidad también se cuenta con desarrollos para el español (Venegas 2007; Gutiérrez 2005).

Como se aprecia, contar con marcadores morfosintácticos presenta una serie de interesantes desafíos, los cuales han ido encontrando lentamente soluciones. También es posible trabajar con corpus no-marcados y se consiguen interesantes análisis. El problema de qué gramática emplear para que soporte los etiquetadores permanece como un factor que debe superarse para que la estandarización de principios pueda lograrse. Del mismo, las características de qué es un corpus y cómo conformarlo también debería buscar principios de estandarización, con el fin de que los datos recabados sean efectivamente comparables.

2.3 Point de vue de Jean-Marie Pierrel

Le coût de définition et de production de ressources linguistiques de qualité (corpus, corpus annotés, dictionnaires et lexiques) est important et c'est un gâchis de vouloir, pour chaque projet de recherche, redéfinir l'ensemble des ressources dont on a besoin. À titre d'exemple, l'établissement d'une base de données textuelle telle Frantext (cf. Bernard / Lecomte / Dendien / Pierrel 2002) s'est chiffré en dizaines de personnes-an, alors même que les annotations sur cette base demeurent minimales. Il convient donc de prendre conscience que sans une véritable mutualisation de telles ressources, chaque équipe de recherche ou chaque chercheur se verrait dans l'obligation de tout réinventer, alors même que nul ne peut être spécialiste de tout. Ceci est particulièrement vrai dans un domaine aussi vaste que la linguistique ou la philologie des langues romanes qui nécessite d'aborder des aspects aussi divers que le lexique, la syntaxe, la sémantique, la pragmatique.

Une telle mutualisation de ressources implique non seulement le développement et l'usage de normes ou de standards pour les ressources linguistiques et textuelles, mais aussi une prise de conscience de cette nécessité d'une meilleure mutualisation des ressources (corpus, dictionnaires et lexiques).

2.3.1 Concernant ces aspects « normes et standards ». – Aujourd'hui, dans le cadre des documents textuels numériques, XML s'impose, tant pour les métadonnées que pour les données. Mais au-delà d'une simple utilisation d'un langage commun, il convient aussi de partager, voire normaliser, les schémas de balisage (ou DTD). C'est en effet au niveau de ces schémas qu'on est en mesure de définir proprement les balises utilisées, leurs attributs et leurs valeurs. Sans recommandations précises sur des standards à utiliser, chaque équipe risque de définir ses propres balises et schémas, et la mutualisation ou la réexploitation de ce balisage par une autre équipe risque donc d'être difficile, voire impossible. Deux recommandations s'imposent donc : (1) expliciter le mieux possible et de façon très précise son schéma de balisage et (2) si possible le définir en accord avec les recommandations internationales.

Un travail important est mené au niveau international dans le cadre du comité technique TC 37 de l'ISO (le sous-comité dédié aux ressources linguistiques et à leur normalisation, SC4 : http://www.tc37sc4.org) ou dans le cadre de la Text Encoding Initiative (TEI : www.tei-c.org), consortium international qui définit des recommandations de codage de ressources textuelles. Notre laboratoire ATILF participe à ces deux initiatives. Mais ne nous leurrons pas, nous ne sommes pas encore à un stade où ces recommandations répondent entièrement aux besoins très diversifiés des chercheurs. Très souvent, nous sommes confrontés à la nécessité de définir des extensions à ces schémas ou recommandations. Il est alors nécessaire de coopérer au sein de ces consortiums internationaux pour tout à la fois faire évoluer ces recommandations et faire prendre en compte les spécificités de nos besoins.

2.3.2 Conditions pour un meilleur partage des ressources (corpus, lexiques et dictionnaires). – Une des premières conditions à un vrai partage et à une mutualisation de ressources (corpus annotés, lexiques, dictionnaires) réside aussi dans un changement de mentalité. Trop souvent encore, nous rencontrons des chercheurs faisant preuve d'une frilosité quasi maladive pour partager leurs ressources linguistiques et corpus. Les justifications qu'ils nous donnent sont très diverses, mais toutes aussi peu convaincantes. On peut citer entre autres :

- La crainte d'être dépossédés de leurs productions. Beaucoup de chercheurs restent 'assis' sur leurs productions, refusant de diffuser leurs résultats ou ne les diffusant que partiellement, rendant ainsi leur réexploitation par d'autres quasi impossible.
- L'espoir, souvent vain, d'une valorisation contractuelle ou industrielle de leurs résultats. C'est en particulier le cas pour de vastes ressources lexicographiques ou lexicales et cela conduit à un gâchis énorme : souvent plusieurs équipes sont amenées à refaire un travail déjà fait par d'autres. On pourrait, à ce propos, se poser la question de savoir s'il ne conviendrait pas pour les langues d'opter pour une philosophie comparable à celle retenue en génomique, qui conduisit à rendre accessible à tous le décodage du génome humain.
- La volonté d'une publication papier avant toute ouverture de ces ressources, en particulier pour les travaux d'édition de textes anciens. Si ce souhait est à première vue le plus défendable, il convient de noter que les clauses contractuelles d'édition chez des éditeurs privés empêchent souvent ensuite une valorisation et un partage informatique de ces ressources via le Web.

Une recommandation s'impose donc : définir des métadonnées précisant les divers contributeurs à ces ressources et respecter une éthique scientifique en ne s'appropriant en aucune manière les travaux d'autrui. Cela est possible à condition de veiller à (1) enrichir sans jamais les dénaturer les métadonnées d'une ressource que l'on peut être amené à enrichir et (2) citer systématiquement les ressources utilisées dans nos travaux de recherche. Pour ce faire, une identification précise de ces ressources au travers d'un identifiant persistant de données ou PID (cf. http://www.pidconsortium.eu) est alors nécessaire. Sous ces conditions, on peut espérer à terme créer une adhésion véritable aux efforts de mutualisation de ressources linguistiques dans des entrepôts fiables de ces données.

2.4 Point de vue d'Achim Stein

La standardisation de la représentation textuelle proprement dite me semble être chose faite : les standards XML et TEI sont largement répandus, et la plupart des outils de traitement et d'exploitation récents utilisent ces formats. Les questions que j'aimerais soulever sont en partie techniques, mais surtout linguistiques, puisque je suis utilisateur, non pas développeur dans ce domaine.

Sur le plan technique, il est important de souligner que c'est justement le progrès dans le domaine de la standardisation qui permet d'ajouter à un même texte des annotations multiples et multicouches : transcription de l'oral (segmentale et suprasegmentale), standardisation d'un dialecte, lemmatisation, étiquetage morphologique, annotation syntaxique, balisage des entités nommées, etc. Certes, ces possibilités soulèvent également des questions : techniques, comme celle des conflits de segmentation ou de hiérarchie entre les couches d'annotation ; linguistiques, comme celle de l'interdépendance ou de l'indépendance des couches d'annotation. Mais d'un point de vue linguistique, la conséquence la plus importante est sûrement la mise en place d'une pluralité, voir d'une 'convivialité' : l'état de l'art permet à plusieurs modèles, et même à plusieurs interprétations linguistiques, de coexister au sein d'une même ressource.

Sur le plan linguistique, j'aimerais évoquer la relation entre le standard d'annotation et le modèle linguistique. Tout d'abord, *système d'annotation* n'équivaut pas à *modèle linguistique*, mais les affinités entre standards d'annotation et modèles linguistiques sont parfois fortes, par exemple dans l'annotation phonétique entre la transcription de l'intonation et le modèle TOBI, ou bien dans l'annotation syntaxique entre la grammaire à constituants et le modèle créé par l'Université de Pennsylvanie, ou encore entre les corpus arborés dépendanciels et le modèle CoNLL. Ces modèles sont largement répandus, et ils ont généré un certain nombre d'outils de requête et de traitement, sans pour autant recourir au format XML. Dans certains cas, il n'est même pas possible de conserver les informations éventuellement encodées en XML ou TEI dans le texte brut.

Je constate donc que, dans le domaine des corpus, on voit plus ou moins se reproduire ce qui s'est passé dans le domaine du développement des logiciels d'application : nous sommes témoins de la naissance de certains standards ' de fait ' (plutôt que ' de droit '),

établis par des projets de recherche de plus ou moins grande envergure et maintenus par les outils qui, souvent, ont été développés parallèlement aux ressources textuelles.

3. La comparaison entre corpus

Dans le cadre d'une même langue selon diverses considérations (époques, régions, styles, situations, etc. ; oral/écrit), pour mieux cerner la variation. Entre deux ou plusieurs langues romanes : utilité respective des corpus parallèles et des corpus comparables.

3.1 Point de vue de Giovanni Parodi

Con el fin de permitir una adecuada comparación entre corpus, se hace importante que los corpus y sus textos constitutivos se encuentren debidamente identificados y catalogados. Esto es, un corpus debe mostrar más de alguna clasificación de la colección que recoge, ya sea de índole temática, de registro, de género o de disciplina. Tanto para fines de investigación como de almacenamiento, todo corpus debería estar descrito y sus rasgos deberían estar disponibles y accesibles.

En este contexto, la constitución de un corpus debería, preferentemente, contar con la posibilidad de disponer de otros tipos de corpus de naturaleza diversa en alguna dimensión. Ello permite la comparación y, de este modo, el contraste hace emerger características distintivas y prototípicas que, de otro modo, sería imposible llegar a descubrir. En este sentido, la recolección de un solo y muy focalizado corpus, por amplio que sea, no brindará una gran riqueza en su descripción, salvo que ya se cuente con otros corpus disponibles previamente y, así, la comparación emerja con mayor facilidad. Esta perspectiva de estudio ha dado origen a lo que se ha llamado el enfoque multidimensional.

El Análisis Multirasgos (AMR) y Análisis Multidimensiones (AMD) fue creado originalmente como un método analítico para el estudio detallado de las variaciones entre registros. Ello permite describir y comparar múltiples textos de diversos corpus en estudio. Este enfoque metodológico fue inicialmente desarrollado, según Biber (1988 ; 1994), para (1) determinar los patrones lingüísticos sobresalientes y en co-ocurrencia en una lengua, desde una perspectiva empírica cuantitativa, y (2) comparar registros orales y escritos en un espacio lingüístico definido por aquellos patrones en co-ocurrencia.

Un aspecto clave para la descripción profunda de los textos de un corpus es el enfoque comparativo, es decir, construyendo una base de datos de corpus variados y a partir de registros diversificados. Diversos investigadores destacan que la investigación más reciente revela la importancia de emplear la comparación de corpus con registros de naturaleza diversificada para develar diferencias significativas sin importar necesariamente el tamaño de las muestras. Biber (1988) ha sido uno de los pioneros en mostrar la relevancia de la comparación de registros, ya que la descripción

de un determinado corpus se enriquece y se hacen evidentes sus rasgos prototípicos por medio de la comparación y el contraste. La mera descripción de un objeto aislado es menos informativa y no logra revelar ciertas características intrínsecas del mismo. Muchas de ellas emergen y adquieren relevancia al ser puestas en contexto, por ejemplo, con otros registros.

Desde este marco, buscando la comparación sistemática, Biber (1988) y Biber / Conrad / Reppen (1998) han dado cuenta de interesantes variaciones sistemáticas de orden gramatical y léxico en diversos registros del inglés oral y escrito. Por su parte, Parodi (2007) también ha dado cuenta, entre otras, de variaciones entre oralidad y escritura para el español. Dos hallazgos entre muchos de los reportados parecen relevantes: por un lado, los rasgos lingüísticos individuales presentan una ocurrencia diversa en variados registros; por otro, los mismos o similares rasgos lingüísticos pueden tener funciones diferentes al entrar en combinaciones con otros rasgos y, al aparecer, en textos pertenecientes a registros diversos. En este sentido, una de las fortalezas de este enfoque metodológico se funda en un principio lingüístico comunicativo que resulta extremadamente sensato: la variación entre registros no se explica únicamente por un solo parámetro o dimensión, lo que equivale a sostener que existen múltiples distinciones situacionales entre registros. Dicho de otro modo, no es posible pensar que un rasgo lingüístico o, incluso, unos pocos de ellos puedan explicar exclusivamente una determinada variación entre registros (por ejemplo: oral/escrito, formal/informal). Las investigaciones en que se ha aplicado el análisis multivariado han revelado que diferentes dimensiones se construyen a partir de conjuntos diferentes de rasgos lingüísticos co-ocurrentes, reflejando así diversas interpretaciones funcionales subyacentes (por ejemplo: objetividad, abstracción de información, modalización). Del mismo modo, las tradicionales distinciones de índole más dicotómica (interactivo/no-interactivo) se ven desafiadas por los estudios con análisis multidimensional, ya que se ha llegado a demostrar que existe un *continuum* de variación lingüística a lo largo de los registros. Por supuesto, esto último es concordante con las investigaciones que adscriben a la idea de categorías de limites difusos (*fuzzy categories*) y que hoy en día tienen gran aceptación entre la comunidad científica.

3.2 Point de vue de Jean-Marie Pierrel

Les besoins de codage standardisé de corpus abordés déjà précédemment (cf. ci-dessus 2.) sont importants dans ce cadre. Que cela soit pour des corpus comparables (en termes de genres, époques, langues, etc.) ou pour des corpus parallèles multilingues, dont on imagine volontiers qu'ils ne peuvent être élaborés par un chercheur seul ou même par une seule équipe, les conditions de mutualisation des ressources prennent ici une importance toute particulière. Trois points méritent à ce niveau toute notre attention : (1) les formats, (2) les référentiels de vocabulaires utilisés et (3) la mise en place d'une interconnexion simple entre nos entrepôts de données.

3.2.1. Concernant les formats. – Nous rappellerons ici simplement la nécessité qu'il y a d'utiliser des formats standard (XML pour le langage de base, TEI pour la défi-

nition des schémas d'annotation, UTF8 pour le codage des caractères par exemple). À cela s'ajoute bien sûr la nécessité de disposer de métadonnées précises permettant de caractériser les corpus mis à disposition de la communauté. Au-delà de simples métadonnées de type bibliographique, il importe de préciser dans ces métadonnées :

- la langue (ou les langues dans le cas de corpus parallèles) ;
- la date ou l'époque d'écriture ou de production de la ressource primaire et pas uniquement l'époque d'édition du document (à titre d'exemple, l'édition critique de la première traduction en français par Raoul de Presles de La Cité de Dieu de saint Augustin – 5 volumes contenant les 22 livres de l'œuvre – réalisée par l'équipe du projet ERC animé par notre collègue Olivier Bertrand doit clairement indiquer la date originale de cette traduction, 1371–1375, et non pas uniquement sa date de publication, qui sera dans ce cas 2013–, cf. Bertrand et al. 2013) ;
- le genre, etc.

Mais cela ne suffit pas. Encore faut-il pouvoir unifier ou déclarer des vocabulaires de données issus si possible de référentiels communs.

3.2.2 Concernant les référentiels de vocabulaires utilisés. – Lorsqu'on définit des métadonnées (genres, langues, etc.) ou des annotations sur des données primaires (annotations morphosyntaxique, parties de discours, etc.) et afin de faciliter leur réutilisation par d'autres, il apparaît indispensable d'utiliser un vocabulaire contrôlé et de le référencer, voire mieux, d'utiliser un vocabulaire contrôlé existant. Là encore, diverses initiatives internationales vont en ce sens en proposant des registres de catégories de données. ISOcat (cf. <www.isocat.org>) en est un bon exemple. ISOcat est un registre central de catégories de données pour tous les concepts pertinents en linguistique et dans le domaine des ressources linguistiques, y compris les catégories de métadonnées. L'interface Web ISOcat (cf. <http://www.isocat.org>) permet de rechercher des catégories de données qui vous intéressent, de sélectionner celles que vous souhaitez utiliser, et, dans le cas où aucune catégorie de données ne correspond à vos besoins, de définir et proposer votre catégorie de données dans ISOcat.

Là encore, il convient de s'organiser pour que la communauté des linguistes et philologues romanistes soit plus présente dans les groupes de travail à l'initiative de ces standardisations. La meilleure standardisation dans nos domaines sera en effet celle que nous définirons ensemble !

3.3 Point de vue d'Achim Stein

Malgré les efforts de standardisation sur le plan formel (XML, TEI, etc.), la comparaison de corpus se heurtera toujours, et nécessairement, aux différents types d'informations linguistiques présentes dans les corpus. Ces différences sont, à mon avis, non seulement inévitables, mais aussi nécessaires si les corpus sont censés refléter l'état de l'art et le progrès linguistique. Elles sont causées d'une part par les catégories spécifiques qui sont nécessaires pour annoter certaines variétés (diachroniques, diastratiques, diamésiques, etc.) et de l'autre par les différents modèles linguistiques qui sont mis en œuvre dans l'analyse des corpus.

La deuxième cause me semble incontournable. La plupart des corpus sont créés pour faire avancer ou pour faciliter la recherche linguistique dans un certain domaine. L'objectif primaire n'étant alors pas la couche 'texte', mais la couche 'annotation', il est normal que cette dernière reflète les objectifs et par conséquent les modèles linguistiques du projet. La déclaration des catégories sur ISOcat (mentionnée par J.-M. Pierrel) est une mesure importante et nécessaire, mais qui devra sans doute être étendue, puisque les couches d'annotation plus 'élevées' (syntaxe, sémantique) reposent sur des interactions complexes et ne peuvent être que difficilement ramenées à la définition de catégories isolées.

Tant que les théories linguistiques évolueront (ce que j'espère vivement), nous ne pourrons de toute façon pas éviter le problème de l'hétérogénéité des corpus. Mais nous pourrions l'amoindrir, par exemple en publiant systématiquement non seulement une documentation cohérente du modèle, mais aussi des corpus vérifiés (*gold standard*), facilitant la réutilisation d'un modèle ou la création d'outils de conversion.

3.4 Point de vue d'Andrea Bozzi

Il problema di una rappresentazione standardizzata dei fenomeni linguistici è di vitale importanza anche per la creazione di corpora paralleli multilingui. Per esempio, la codifica dei valori grammaticali (noti col nome di POS, *part of speech*) attribuiti da un sistema manuale, automatico o semiautomatico (ad esempio, un analizzatore morfologico, un sistema di lemmatizzazione, un 'parser' sintattico, un 'disambiguatore' semantico per i casi di omografia fra forme identiche, ma appartenenti a lemmi differenti, ecc.) può essere ormai eseguita con sistemi di marcatura condivisi. Mi sembra che i tempi siano ormai maturi, visti i risultati sempre più positivi ottenuti nell'ambito di progetti internazionali nel campo della linguistica computazionale, per diffondere maggiormente le informazioni sulle tecniche, i metodi e le procedure che hanno consentito progressi così consistenti affinché siano utilizzati anche dai cultori delle discipline filologiche i quali, in linea di massima, ne hanno informazioni solo di carattere generale. Mi permetto di ricordare che alcuni programmi realizzati, appunto, nell'ambito della linguistica computazionale sono stati originati dalla necessità di costituire corpora multilingui paralleli: i primi risalgono addirittura al periodo in cui la Commissione Europea finanziò ricerche nell'ambito del progetto EUROTRA (fine anni '70) per la realizzazione di programmi di traduzione automatica fra lingue comunitarie (in un primo tempo sette e successivamente nove).

Questa iniziativa ultradecennale, anche se non fu in grado di produrre molti dei risultati attesi, ebbe tuttavia il merito di dare avvio a corpora paralleli la cui annotazione multilivello (soprattutto morfologica, sintattica e lessicale) era propedeutica alla preparazione dei programmi di traduzione vera e propria. L'evoluzione tecnologica, sia per quanto riguarda il software che l'hardware, ha reso i sistemi di allineamento fra testi paralleli in varie lingue molto più efficienti e, soprattutto, veloci: i metodi statistici e stocastici possono venire applicati su quantità di dati incomparabilmente superiori a quelli sui quali si operava negli anni '80, con la conseguenza di

poter disporre di risultati quantitativamente e qualitativamente determinanti per lo studio della variazione fra una lingua ed un'altra. Ne hanno tratto vantaggio, soprattutto, le valutazioni di carattere morfosintattico e semantico. Ritengo che le strategie e le procedure oggi disponibili possano venire sperimentalmente applicate anche a corpora di lingue romanze di epoca medievale per avere a disposizione dati di comparazione concreti e in numero sufficiente a confermare, o eventualmente rivedere, la spiegazione di fenomeni evolutivi rispetto al latino, soprattutto quando si tratti di divergenze attribuibili a variazioni geolinguistiche o a condizionamenti socioculturali. Tali corpora e le annotazioni eseguite con l'aiuto di sistemi computerizzati consentono, inoltre, di indagare meglio anche le varietà linguistiche nell'ambito di un medesimo idioma per comprendere il rapporto sincronico e diacronico che lega un sistema linguistico alle proprie variazioni (come, per esempio, le variazioni dialettali o quelle intervenute, come si è detto, per cause socio-culturali).

4. L'édition de textes et de documents sonores

L'édition de textes (des codex médiévaux au corpus universel de la toile d'araignée mondiale WWW) et de documents sonores (de l'enquête sur le terrain traditionnelle au foisonnement de la production audiovisuelle) : que retenir et comment traiter ? L'établissement de lexiques de ces documents.

4.1 Point de vue de Jean-Marie Pierrel

L'apport de la linguistique de corpus à la compréhension des phénomènes langagiers est aujourd'hui devenu fondamental. Le nombre d'énoncés qu'entend et produit une personne durant sa vie est très grand. Grâce à l'augmentation de la variété et de la taille des corpus, il est aujourd'hui devenu possible de démontrer les faits langagiers à l'aide d'exemples attestés en grand nombre et de tester les propositions de la linguistique. Pour cela, un grand nombre de corpus contrôlés, bien décrits et variés, est nécessaire.

Non spécialiste de l'édition ni de textes écrits ni de documents sonores, je voudrais ici à Nancy rappeler les nombreuses contributions importantes à la lexicographie française qui ont vu le jour grâce à une exploitation raisonnée de corpus informatisés à Nancy, au sein du Centre de Recherche pour un Trésor de la Langue Française (CRTLF), puis de l'Institut National de la Langue Française (INaLF) et enfin de notre laboratoire, l'ATILF, qui se veut être aujourd'hui le digne successeur de ces deux structures.

Initiées au départ autour du projet de *Trésor de la Langue française* (TLF), premier dictionnaire se fondant sur une méthodologie systématique d'analyse des usages effectifs des mots de notre langue à travers l'exploitation d'une vaste base de données textuelles dont la saisie a débuté dès les années 60, ces études nancéiennes en lexicographie française se sont poursuivies, au-delà de la rédaction du TLF, suivant deux orientations complémentaires : la lexicographie historique, fortement appuyée sur la notion de corpus, et la valorisation informatique des ressources produites. Ces

orientations ont provoqué sur le plan des études lexicales une véritable révolution qui fit de l'informatique un outil indispensable pour (1) étudier le lexique et ses propriétés à travers l'exploitation intelligente de textes et de documents, (2) structurer et normaliser les connaissances lexicales et lexicographiques et (3) valoriser, partager et mutualiser les résultats de la recherche sur le lexique de notre langue, trop souvent encore dispersés.

Rappelons à titre d'exemple quelques-unes de ces contributions : le *Trésor de la Langue Française Informatisé* ou TLFi (cf. Dendien / Pierrel 2003), qui correspond à la valorisation informatique, sous forme de CDRom et sur le Web, du dictionnaire de référence qu'est le TLF. Au-delà d'une simple informatisation de la version papier du TLF, ce projet a permis de poser les bases d'une exploration nouvelle de données lexicographiques autorisant des parcours vraiment novateurs de ces matériaux et de donner une seconde vie à ce grand dictionnaire qui aujourd'hui, à n'en point douter, est le dictionnaire institutionnel français le plus consulté sur le Web. Le *Dictionnaire du Moyen Français* ou DMF (cf. DFMF 2012 ; Martin / Gerner / Souvay 2010 ; Städtler 2010), qui articule fortement valorisation informatique et lexicographie historique. Ayant comme objectif de combler une lacune existante entre le Tobler-Lommatzsch en amont et le Huguet en aval, ce dictionnaire se propose de décrire le lexique français de la période 1330–1500 et s'appuie sur un concept de lexicographie évolutive (Martin 2008) que l'informatique autorise aujourd'hui. D'autres développements récents en lexicographie historique concernent le programme TLF-Étym (cf. Steinfeld / Andronache 2011), qui se propose de réviser progressivement les notices intitulées 'Etymologie et histoire' du TLF ; la Base des mots fantômes (cf. Steinfeld 2007– ; 2013), ressource de métalexicographie critique dont l'ambition est de recenser ces pseudo-lexèmes disposant à tort d'un statut lexicographique ('ces mots qui n'existent pas'), les sens fantômes et les lemmatisations erronées qui se trouvent dans les dictionnaires historiques et étymologiques français de référence, mais dont on ne trouve aucune attestation véritable dans les grands fonds de textes historiques numérisés. Enfin le Portail lexical du Centre National de Ressources Textuelles et Lexicales (CNRTL ; cf. www.cnrtl.fr/portail), créé par le CNRS au sein de notre laboratoire. Ce portail lexical a pour vocation de valoriser et de partager, en priorité avec la communauté scientifique, un ensemble de données issues des travaux de recherche sur le lexique français. Projet évolutif, cette base de connaissances lexicales propose, à partir d'une unité lexicale, d'intégrer un maximum de connaissances disponibles.

Comme nous l'avons montré dans Pierrel / Buchi (2009), la mise à disposition très large des ressources linguistiques du DMF a permis des avancées scientifiques dans des domaines non entrevus lors de la conception de cette ressource, par exemple la pragmatique et la morphologie constructionnelle. C'est pourquoi à la question souvent posée de savoir quelles ressources retenir pour une mise à disposition de l'ensemble de la communauté scientifique, ma réponse est simple : toute ressource définie et construite dans le respect d'une bonne démarche linguistique et philologique justifie une telle diffusion.

4.2 Point de vue d'Achim Stein

En abordant les questions d'édition, nous nous plaçons dans une logique plus large, non exclusivement réservée aux linguistes. Car souvent, les corpus proviennent d'autres domaines (littérature, histoire, sociologie, médias, etc.) ou bien ils sont d'origine linguistique, mais intéressent ces autres domaines. Ici, nous retrouvons donc le sujet de l'annotation multicouche, et des principes favorisant une annotation 'débarquée', techniquement indépendante de la ressource (*stand-off annotation*).

Mais la conséquence la plus importante de cette ouverture vers d'autres domaines est, à mon avis, une contrainte qui s'impose à nous, linguistes ou créateurs de corpus. Cette contrainte dit que tout utilisateur d'un corpus, linguiste ou non, doit être capable d'évaluer le texte et son annotation de manière réfléchie et critique. Et les conséquences concernent bien entendu non seulement l'édition du texte, mais aussi son traitement ultérieur, c'est-à-dire les différents types d'annotation qui s'y ajoutent ainsi que les ressources externes (lexicales ou autres) qui se greffent sur ce texte.

Deux exemples :

- Représenter la fiabilité de l'annotation, dans l'annotation même ou par les outils qui servent à exploiter ce type d'annotation. Concrètement, un utilisateur qui obtient le résultat d'une requête qui repose sur certaines catégories, morphologiques ou syntaxiques, devrait être conscient du taux d'erreur de cette requête. Dans le meilleur des cas, l'outil d'exploitation est informé sur les taux d'erreur des outils d'annotation, parce que ces derniers, justement, les ont ajoutés à la ressource, parallèlement à la déclaration des catégories, dans l'en-tête du document, par exemple.
- Représenter les ambiguïtés d'analyse dans l'annotation. Le problème ressemble au précédent : les outils d'annotation peuvent réduire le nombre des décisions prises lors de l'analyse en annotant les alternatives. Ceci augmente le 'bruit' dans le résultat d'une requête et pose des problèmes de représentation aux standards d'annotation.

Bref, tout comme l'éditeur d'une édition critique documente les décisions qu'il prend et les problèmes qu'il rencontre, l'annotateur d'un corpus, humain ou machine, doit, lui aussi, documenter ses décisions et les intégrer, si possible, dans la ressource.

4.3 Point de vue d'Andrea Bozzi

Lo sviluppo della tecnologia digitale e del Web impone di porre l'attenzione su due aspetti:

(1) Esiste una differenza fra edizione elettronica mediante strumenti ipertestuali/multimediali e edizione assistita da calcolatore mediante applicazioni Web anche di tipo collaborativo.

(2) Esistono ormai molti modi diversi di produrre il lessico dei documenti digitali (dall'ʻindex lemmatum' all'indice ontologico-concettuale sulla base delle tecniche del 'semantic web') e delle modalità con cui si strutturano le descrizioni delle singole voci (dalla descrizione non strutturata alla descrizione secondo i metodi del lessico generativo computazionale)[3].

[3] Si veda Pustejovsky (1995); una interessante applicazione di questo metodo al lessico di Ferdinand de Saussure si legge in Ruimy / Piccini / Giovannetti (2012, 1043-1056).

Per quanto riguarda il primo punto, la marcatura da parte dell'editore critico di informazioni di tipo diverso (il testo, l'immagine digitale della fonte, i dati di apparato, la bibliografia, gli eventuali passi paralleli, le note critiche, ecc.) mediante l'utilizzo di 'link' ipertestuali produce edizioni elettroniche (sul Web o su supporti ottici) ove il ruolo del sistema informatico è piuttosto banale e limitato. Esso, infatti, non fa altro che valutare le marcature e offrire all'utilizzatore una lettura dei dati secondo la sequenza con la quale egli intende 'navigarli' (dal testo all'immagine, oppure dal testo all'apparato delle varianti, oppure ancora dalle varianti alla bibliografia, ecc.). L'edizione assistita da calcolatore, invece, è definibile come filologia computazionale in senso stretto poiché, su richiesta dello studioso, la macchina interviene attivamente nel processo critico-editoriale mettendo in moto componenti software tanto più numerosi quanto più completa è l'applicazione che li comprende. Essi, cooperando come moduli interrelati, producono risultati mirati e quantitativamente rilevanti grazie ai quali si mette in condizioni lo studioso di effettuare interventi critici più sicuri e documentati. Tale approccio facilita la registrazione di apparati, la costituzione del testo e, grazie a tecniche di 'clustering' e di 'information visualisation', anche la produzione di ipotesi stemmatiche[4].

Per quanto riguarda il lessico, anche in questo caso i sistemi di elaborazione elettronica dei dati testuali hanno da decenni sviluppato programmi molto efficienti di indicizzazione per la produzione di liste alfabetiche delle forme. Essi, a loro volta, sono stati completati da moduli di analisi morfologica in grado di attribuire automaticamente le forme flesse ai lemmi, spesso con informazioni sufficienti per la disambiguazione delle forme omografe derivate da lemmi etimologicamente diversi. Si sono in tal modo resi disponibili sia 'indices verborum' di autori o di interi 'corpora', ma anche 'indices lemmatum' utilissimi alla preparazione di studi e ricerche lessicali. Un aspetto correlato con quest'ultimo consiste nella descrizione dei valori semantici che un lemma possiede nei testi del corpus o, in ultima analisi, nel dizionario generale della lingua di riferimento. Nel corso delle ricerche in linguistica computazionale si sono susseguiti vari criteri adatti a tale scopo, ma quello che sembra offrire ottime capacità descrittive, indipendentemente dalla lingua al quale esso si applica, si basa sulla teoria del lessico generativo. Essa appare molto più versatile del noto WordNet (ItalWordNet per l'italiano e EuroWordNet per le relazioni fra elementi semantici fra più lingue europee, mediante il parallelo fra gruppi di sinonimi)[5]. La connessione fra il lessico semanticamente organizzato e l'archivio dei testi, poi, consente di effettuare ricerche molto dettagliate e mirate al ritrovamento solo di quei passi che documentano un valore semantico specifico, escludendone altri.

---

4 Un esperimento in tal senso è descritto in Corradini (2005, 355-368). Più di recente ha sviluppato tecniche di stemmatica computazionale Camps (2013a e 2013b).

5 Si vedano il sito ufficiale di WordNet (<www.illc.uva.nl/EuroWordNet>) e Rodríguez *et al.* (1998, 117-152).

4.4 Point de vue de Giovanni Parodi

¿Qué mantener en el registro de un análisis de corpus? y ¿cómo procesar automática o manualmente textos de naturaleza diversa? Ambas preguntas conllevan desafíos y proyecciones para el tratamiento de corpus e imponen una agenda programática a la que ciertamente se debe atender.

A la primera pregunta ya se abordaba en los puntos precedentes, pues un asunto relevante es de qué tipo de marcaje se puede disponer para un estudio de corpus. Clásicamente los etiquetadores morfosintácticos han estado en el foco y su adecuado desempeño sigue siendo materia de investigación y desarrollos. Alcanzar un 100% de marcaje fiable es un desafío. Otros tipos de marcas más textuales y más discursivas también estan siendo estudiadas y constituyen núcleo de atención. Por otro lado, también considerábamos que información extratextual de tipo contextual es relevante y debe ser codificada: tal como procedencia, fecha de recolección, lugar, etc. También información de orden temático, tipo de género discursivo, registro, etc.

En cuanto al procesamiento computacional de textos de variada naturaleza se ha abierto un campo de proporciones tremendo. Por una parte, el análisis de textos orales impone retos muy motivadores: dar cuenta de los rasgos de la oralidad y su contexto de producción (rasgos de índole pragmática) es un terreno en franco desarrollo. Por otra parte, el denominado re-descubrimiento de lo obvio, es decir, la característica multimodal de todo texto y no solo su constitución a partir de palabras (sistema verbal), sino de múltiples sistemas semióticos (gráfico, matemático, tipográfico, entre otros), ha creado un ámbito de investigación muy promisorio (Parodi 2010b). Incluso ya no solo se abordan los denominados *textos estáticos* (en formato papel o tipo PDF en computador), sino que los *textos dinámicos* (cine, video, hipertextos, textos web, etc.) ofrecen un área de estudio innovadora.

5. La pérennisation des documents

La pérennisation des documents : comment conserver, mais aussi diffuser le plus librement possible un maximum de documents soigneusement traités ; rôle des diverses institutions pour l'archivage des corpus, leur transcodage régulier, leur mise à la disposition du public au fil des ans.

5.1 Point de vue d'Achim Stein

Je ne répéterai pas l'importance des grandes initiatives européennes, DARIAH, CLARIN etc., qu'A. Bozzi et J.-M. Pierrel ont évoquée. Mais il serait important de signaler plus clairement les liens entre la Société de Linguistique Romane et ces initiatives, en tant qu'institution ou à travers les projets de recherche menés par ses membres. Une fois de plus, je vais distinguer entre la pérennisation des document statiques d'une part, qui concerne la philologie et dont mes collègues reparleront, et les outils et les méthodes, dynamiques et davantage liés à la recherche linguistique, de

l'autre. Concernant ce deuxième point, je rappelle que le projet européen CLARIN crée des plates-formes qui permettent de traiter – baliser et annoter – les corpus en ligne, sur des serveurs décentralisés. Les outils de traitement sont donc disponibles sur les serveurs, qui permettent aux utilisateurs de créer des chaînes de traitement.

L'avantage de ce concept est tout d'abord que les outils sont ainsi accessibles à tout le monde, et qu'ils ne nécessitent aucune installation technique. Le deuxième effet, plus important à long terme, est le fait que la longévité des outils et des méthodes est ainsi garantie. À mon avis, c'est cette facilité d'accès qui va produire un effet de standardisation, dans la mesure où elle permettra au plus grand nombre de produire des textes pourvus d'une annotation de qualité.

5.2 Point de vue d'Andrea Bozzi

Nel corso degli ultimi anni si è in più occasioni lamentata l'incertezza della permanenza dei dati sui supporti fisici adottati e si sono levate critiche sulla relativa tecnologia (prima magnetica, come ricorderanno tutti coloro che hanno lavorato con i nastri sui quali erano registrate le informazioni, poi digitale). Secondo la mia personale esperienza queste osservazioni non sono sempre giustificate poiché lavori che io stesso ho eseguito alla fine degli anni '70, opportunamente riversati sui nuovi mezzi di archiviazione mano a mano che essi erano immessi sul mercato e venivano acquisiti nei grandi centri di elaborazione dati, sono ancora integri e su essi possono essere applicati i nuovi e più potenti sistemi di trattamento e di analisi.

Un altro aspetto, piuttosto, deve essere preso in considerazione e riguarda la riconversione degli elementi di marcatura affinché da quelli originari, spesso sviluppati e utilizzati da un singolo, si possa ottenere una ri-codificazione dei testi secondo metodi standardizzati o, almeno, compatibili con gli standard condivisi a livello internazionale. Questo aspetto ci riconduce a quanto ho detto relativamente al primo tema trattato e al quale io attribuisco un grande valore. La condivisione di sistemi di codifica, soprattutto di fronte all'affermazione indiscussa del Web come infrastruttura comunicativa mondiale, offre l'opportunità alle scienze umane e alle comunità di ricercatori che in esse operano di mettere a frutto conoscenze, dati, programmi, risultati troppo spesso isolati gli uni rispetto agli altri. Le scienze umanistiche digitali potranno avere molti più elementi di aggregazione rispetto al passato e assumeranno un modo di agire simile a quello che caratterizza altri domini scientifici (in particolare, la fisica e la genetica). Affinché questa mia osservazione non rimanga vaga, ripeto in questa occasione un esempio che ho altre volte citato. La scelta di una lezione per stabilire un testo occitano medievale di argomento medico farmaceutico potrebbe avvalersi del raffronto con vasti archivi di lingue romanze (testi e dizionari) consultabili sul Web per verificare attestazioni o interrogare eventuali studi e saggi monografici. Se queste operazioni di raffronto sono sempre state fatte da studiosi scrupolosi ed attenti sugli strumenti cartacei disponibili, oggi l'indagine può (o potrebbe) essere condotta in maniera molto più agevole e rapida solo se i dati sono resi accessibili in linea e se essi hanno una esplicita e condivisa rappresentazione nella rete. Per questa ragione

sento il dovere di complimentarmi vivamente con il lavoro svolto a Nancy dall'ATILF che ospita questo convegno per il grande lavoro svolto proprio in questa direzione e plaudo a tutte le iniziative di valore scientifico che seguono questo esempio virtuoso, grazie agli sforzi di ricercatori di tutto il mondo. Per questa serie di motivi l''open data' e, per quanto riguarda i sistemi e le applicazioni computazionali, la loro disponibilità 'open source' senza pagamento di diritti di uso o riproduzione, è la reale prospettiva del futuro, soprattutto per i lavori che sono stati condotti da studiosi sulla base di finanziamenti pubblici. Le istituzioni pubbliche e le società scientifiche hanno, pertanto, il compito di:

- garantire i 'credits' e l''authorship' della documentazione (edizioni, sistemi, lessici, ecc.) immessa in rete su portali istituzionali e a questa documentazione deve essere attribuita adeguata validità come prodotto della ricerca valutabile;
- garantire la transcodifica dei valori propri di un qualunque sistema proprietario verso un sistema standard fra i quali quello che si è affermato, nel settore degli studi del testo, è la TEI;
- garantire una maggiore attenzione rispetto a quanto non è avvenuto fino ad oggi verso le infrastrutture di ricerca invitando a tenere in considerazione, nelle fasi di progettazione, anche le esigenze dei filologi oltre a quelle, oggi prioritarie, dei linguisti che operano su lingue vive moderne, degli storici, dei sociologi, ecc.

5.3 Point de vue de Giovanni Parodi

Para la adecuada permanencia y disponibilidad de los corpus se hace fundamental contar con interfaces computacionales en línea, construidas por medio de sitios web. Si los corpus son accesibles libremente en línea y se les puede consultar en formato 'marcado', se asegura su perpetuidad y se garantiza que el trabajo realizado sea productivo. Lamentablemente aún existen corpus que tienen carácter privado o cuyo acceso no es gratuito. O, incluso, que se recolectan pero nunca logran estar en un sistema abierto que permita el libre acceso y consulta.

Es evidente que existen costos económicos asociados a esta disponibilidad en línea. Es de esperar que instituciones universitarias y de investigación logren apoyar estos proyectos y se vaya paulatinamente haciendo más factible el acceso gratuito en línea.

En el caso de la lengua española, diversas universidades e instituciones gubernamentales de apoyo a la investigación han respaldado la creación de sitios web que alojan y permiten el acceso a corpus digitales. También se han creado interfaces de consulta que permiten a los investigadores obtener datos estadísticos de los textos analizados así como descripciones de estos corpus. La labor de la Real Academia Española (RAE) ha sido significativa en este sentido en los últimos años. La RAE mantiene acceso en línea y de modo gratuito de importantes corpus (<www.rae.es>). Lo mismo se puede decir del IULA de la Universidad Pompeu Fabra y otras tantas universidades españolas.

En el caso de Chile, una importante tarea en la investigación de corpus ha venido realizando la Escuela Lingüística de Valparaíso (ELV: <www.elv.cl>) de la Pontificia Universidad Católica de Valparaíso con la creación de El Grial (<www.elgrial.cl>). El Grial

es tanto una herramienta de etiqueta morfosintáctico para textos en español como también una plataforma en web de consulta de corpus textuales. En esta plataforma se obtiene información detallada de cada uno de los corpus disponibles gratuitamente en línea. Una de sus muchas fortalezas es el acceso a corpus diversificados, es decir, de disciplinas diversas, y separados, entre otros, por géneros del discurso y temáticas.

5.4 Point de vue de Jean-Marie Pierrel

L'importance d'un plus grand partage de nos ressources linguistiques, en particulier de nos corpus, au sein de la communauté de linguistique et philologie romanes, plaide pour la mise en place d'entrepôts de données fiables. En France, nous venons de bénéficier d'un soutien conséquent (2,6 millions d'euros HT) dans le cadre du Programme d'Investissement d'Avenir (PIA) de l'État français pour construire un équipement d'excellence ORTOLANG (Open Resources and Tools for LANGuage, <www.ortolang.fr>) dont nous assurons la responsabilité à Nancy.

Une telle infrastructure a pour but de proposer un réservoir fiable de données (corpus, lexiques, dictionnaires, etc.) et d'outils sur la langue et son traitement clairement disponibles et documentés qui valorise le français et les langues de France à travers un partage des connaissances sur ces langues accumulées par les laboratoires publics. Un tel équipement a aussi pour objectif de généraliser et d'assurer la pérennisation des efforts entrepris par les Centres de Ressources Numériques linguistiques mis en place par le CNRS (cf. Pierrel / Petitjean 2007) : CNRTL (Centre de Ressources Textuelles et Lexicales, <www.cnrtl.fr>) et SLDR (Speech and Language Data Repository, <http://sldr.org>, anciennement CRDO-Aix). Enfin, il servira de plate-forme technique sur la langue française, écrite et orale, support des actions de coordination menées par la grande Infrastructure de recherche française Huma-Num (<http ://www.huma-num.fr>), fusion d'ADONIS et de CORPUS, dans laquelle nos laboratoires et centres de ressources sont fortement impliqués. ORTOLANG se propose donc de mettre en place un processus permettant à une donnée linguistique, une fois créée, d'être cataloguée, éventuellement améliorée (voire corrigée), puis diffusée et enfin archivée. Le modèle d'ORTOLANG reprend les entités de base du modèle OAIS, en précisant le cycle de correction / enrichissement des données, rendu possible par l'archivage intermédiaire. Les fonctions envisagées pour cet équipement sont donc :

- Permettre, au travers d'une véritable mutualisation, à la recherche sur l'analyse, la modélisation et le traitement automatique de notre langue de se hisser au meilleur niveau international.
- Aider à la création de données : faciliter l'accès à des instruments permettant l'acquisition ou la création de données (par exemple, numérisation, chambre sourde, caméra rapide, mouvements oculaires, électro-encéphalographie, articulographe, etc.).
- Enrichir les données : plusieurs outils permettant d'enrichir automatiquement les données brutes (étiquetage morphosyntaxique, analyses prosodiques, syntaxiques, etc.) seront proposés sur la plate-forme ORTOLANG.
- Accompagner les auteurs sur les standards, les normes et les recommandations internationales actuelles : XML, TEI, LMF, MAF et SYNAF ; enrichissement des ressources et des outils.

- Identifier les données : catalogage des ressources et outils existants à travers un ensemble de métadonnées normalisées ; contrôle et validation des ressources et des outils.
- Assurer la pérennité des données au travers d'une double fonction de stockage sécurisé et d'archivage pérenne : stockage, maintenance et curation des ressources et des outils ; archivage pérenne, à travers la solution mise en place par la TGIR Huma-Num en lien avec le CINES.
- Assurer la diffusion : aide et accompagnement des utilisateurs et mise en place des procédures permettant à des utilisateurs de la plate-forme d'exploiter les ressources et outils mutualisés sans avoir à se soucier de leur localisation et implantation géographiques.

Au niveau européen, au cours des dernières années, diverses initiatives ont vu le jour pour permettre une concertation et une mise en réseau d'entrepôts fiables de données. On peut signaler entre autres les deux infrastructures de recherche que sont DARIAH (Digital Research Infrastructure for the Arts and the Humanities, <www.dariah.eu>) et CLARIN (Common Language Resources and Technology Infrastructure, <www.clarin.eu>). Il paraît indispensable que nos communautés s'impliquent plus dans ces initiatives.

Il serait en particulier souhaitable de pouvoir disposer d'un véritable réseau d'entrepôt de données de ce type regroupant les diverses langues romanes. L'initiative lancée à l'occasion de ce congrès par notre collègue Antonio Pioletti, président de la Société italienne de philologie romane, de lancer un projet de Conseil International d'Études Romanes, qui réunirait les principales associations scientifiques actives dans ce domaine, peut être l'occasion de progresser vers cet objectif. Nous plaidons pour que, dans le programme d'un tel Conseil International d'Études Romanes, émerge, entre autres, un projet de mise en place d'un réseau numérique regroupant nos ressources et connecté aux réseaux européens tels que celui que propose CLARIN (Wittenburg et al. 2010).

Pour notre part, en nous appuyant sur l'Equipex ORTOLANG que nous animons depuis Nancy, nous proposons, si la communauté juge intéressante une telle proposition, de participer activement à une telle initiative.

6. Les études linguistiques portant sur les langues romanes

Les études linguistiques portant sur les langues romanes : importance des corpus multilingues pour les études contrastives dans le domaine roman, en particulier pour des branches de notre discipline encore peu développées. Rôle de notre Société pour une meilleure coordination des études romanes.

6.1 Point de vue d'Andrea Bozzi

A costo di correre il rischio di ripetere valutazioni che ho già espresso nel corso del dibattito sugli altri temi della tavola rotonda, credo con convinzione che gli studi sulle lingue romanze possono trarre un grande vantaggio se i loro cultori siano disposti a prestare molta attenzione anche a quanto avviene almeno in altri due settori del 'Computing in the Humanities': la linguistica computazionale e la filologia computazionale.

La nostra Società presta certamente attenzione all'innovazione tecnologica e nei convegni internazionali ha da tempo inserito sessioni che coprono i nuovi strumenti, le nuove tecnologie, le basi di dati a supporto della ricerca linguistica e filologica, i programmi di elaborazione che sono usati con successo in ambito accademico. Tuttavia si dovrebbe forse fare qualche cosa di più per invitare i responsabili dell'educazione universitaria almeno dei paesi della comunità a favorire l'inserimento, nel curriculum di studi di linguistica e filologia romanza, di corsi a maggiore carattere tecnologico ove la linguistica e la filologia si possano arricchire dei contributi messi ormai a disposizione dalla scienza dell'informazione, dagli studi statistici, dalle metodologie di annotazione e codifica, dall'elaborazione dei dati. Con questo non intendo affatto sostenere che lo statuto delle nostre discipline debba cambiare o essere totalmente modificato, ma un aggiornamento si impone per non vederle agonizzare in un mondo in profonda trasformazione e con studenti che vi si dedicano in numero sempre minore.

Quanto ho detto nella discussione circa il terzo tema, per esempio, consente di riproporre metodi di analisi comparativa fra testi e lessici consultabili in linea nell'ambito di un corso universitario che si avvalga di accesso a strumenti computazionali disponibili e approvati scientificamente da parte di un collegio di docenti. Si tratterebbe, quindi, di offrire a docenti e discenti applicazioni grazie alle quali sia possibile fare annotazioni, scrivere piccoli saggi, eseguire analisi linguistiche, trovare citazioni o elementi di intertestualità. La formazione avverrebbe su due binari: quello classico della lezione frontale fra professore e allievi e quello 'strumentale' fra professore, allievi e sistemi informatici pronti ad offrire dati su cui ragionare e registrare i risultati delle osservazioni fatte. L'aspetto partecipativo, che è un elemento fondamentale nel Web, avrebbe in tal caso elementi decisamente positivi, anche perché gli allievi potrebbero cooperare alla valutazione dei fenomeni incontrati in un testo anche al di fuori dell'aula, in un ambiente virtuale.

La tecnologia è pronta per tutto questo scenario, ma mancano, probabilmente, alcune condizioni favorevoli la maggior parte delle quali non dipendono direttamente dalle nostre intenzioni: la Société de Linguistique Romane ha però tutto il prestigio internazionale necessario per aiutare, se lo vorrà, a seguire questa direzione, l'unica, a mio avviso, che è in grado di garantire un futuro ai suoi antichi e gloriosi settori di studio. Sono al corrente di quello che avviene in molti dei paesi europei a proposito degli studi umanistici, ed è di recente pubblicazione su un quotidiano italiano di larghissima diffusione un articolo che lamenta la crisi di iscrizioni al liceo classico, unica scuola secondaria superiore in Italia che, nell'offerta didattica, abbia lo studio della lingua greca e latina.

Questa tendenza è iniziata alcuni anni fa e non sono pochi i casi nei quali, in alcune città italiane, prestigiosi e storici licei siano stati accorpati ad altre scuole superiori per carenza di iscritti. Come dicevo, le cause sono molteplici, ma un possibile rimedio è l'introduzione di forme di aggiornamento che prevedano l'utilizzo sensato di innovazioni tecnologiche, ancora troppo limitate se non addirittura escluse o adottate sporadicamente e senza un piano formativo che si conformi anche alla presenza

di esse in forma organica. Io credo che sia, pertanto, necessario intervenire su due fronti: uno nell'ambito della formazione universitaria ed uno ad un livello più basso. L'innovazione tecnologica e l'adozione di nuovi parametri educativi sono una garanzia di sopravvivenza dal momento che avranno un ruolo determinante nella formazione di nuove generazioni di filologi, dotati di un bagaglio di conoscenze, anche tecniche, che permetteranno alle nostre discipline di venire curate e sviluppate nell'ambito di una cultura che ormai è divenuta digitale.

6.2 Point de vue de Giovanni Parodi

Los estudios de corpus han experimentado un auge tremendo en los últimos años y aunque, como dicen algunos, los corpus siempre han sido parte de la investigación lingüística, nunca como hasta ahora se ha logrado tanta sistematicidad metodológica en torno a objetivos comunes. La expansión de la denominada lingüística de corpus en una renovada mirada ha dado pie a magníficas discusiones teóricas y ha abierto escenarios para fortalecer el estudio de la lengua en uso y de la variación en todo sentido; así, los denominados estudios variacionistas que ahora incluyen tanto la variación disciplinar y la de géneros del discurso ha traído atención a aspectos de las lenguas que hasta poco no eran de interés de los investigadores.

La disponibilidad de soporte electrónico para los corpus y el albergarlos en sitios web de acceso abierto y gratuito constituyen oportunidades formidables para el estudio comparativo de las lenguas romances. Ello puede contribuir de modo certero a llevar adelante estudios empíricos basados en corpus que arrojen resultados robustos como nunca antes; también se lograría superar trabajos a partir de textos ejemplares y con escasa proyección, dado lo limitado de algunos corpus. Al mismo tiempo, la existencia de asociaciones como la que acoge esta mesa de discusión con interlocutores de diferentes lenguas y orígenes es una prueba de que se cuenta con interés de colaborar y aportar al desarrollo de la investigación en este campo disciplinar. Así, tanto la lingüística de corpus como los estudios basados en corpus, en un sentido amplio, muestran proyecciones vigorosas.

6.3 Point de vue de Jean-Marie Pierrel

Depuis une dizaine d'années, le paysage de la recherche en linguistique a largement évolué grâce à l'apparition d'importants corpus linguistiques aisément disponibles sur Internet. Si l'existence d'une linguistique de corpus n'est pas nouvelle, cette évolution de l'accès aux données dynamise de manière très importante le domaine, permet de démontrer l'importance, du point de vue fondamental, de la notion de variation, et autorise de grandes avancées dans la modélisation des théories exemplaristes ou dites basées sur l'usage.

Si, avant les années 2000, le paradigme générativiste dominait et conduisait à voir les théories et les modèles linguistiques comme fondamentalement sous-déterminés par les données factuelles, ce n'est plus le cas aujourd'hui. Ce sont d'abord les tra-

vaux psycholinguistiques d'observation longitudinale, et spécialement ceux menés sur les acquisitions précoces qui ont ébranlé le paradigme cognitiviste chomskyen en documentant une hétérogénéité et une variabilité intrinsèque très importantes et peu compatibles avec l'innéisme de la grammaire universelle. Ces travaux ont récemment rencontré les problématiques de la linguistique variationniste conduites indépendamment depuis plusieurs décennies. La confrontation avec les analyses du changement linguistique en temps réel a par ailleurs souligné l'importance des dynamiques qui structurent, forment et déforment les systèmes linguistiques dans le temps. Enfin, le développement des travaux contrastifs et typologiques a conduit à relativiser la portée des grandes hypothèses universalistes au profit d'une description plus fine et plus précise des données observées. Dans chacun des domaines et des sous-domaines de la linguistique, la notion d'usages ou de pratiques attestés a ainsi été remise au premier plan, induisant un rapport nouveau aux modélisations explicatives et aux formalisations.

Ces théories sont basées sur la notion de constructions, qui sont des associations entre forme et fonction. Les constructions peuvent être extrêmement variées, allant de formes figées (un mot, une holophrase, une expression idiomatique) à des structures plus générales (par exemple la structure transitive sujet-verbe-objet), en passant par de nombreux intermédiaires plus ou moins généralisés (par exemple la construction ' c'est X ', où X peut prendre n'importe quelle forme ; ou la construction ' X aime V_{inf} ', où X et V_{inf} sont mutuellement contraints). Les constructions peuvent se combiner pour produire des formes langagières de tout niveau de complexité. De telles théories permettent de modéliser la variété à tous les niveaux, de l'interlocuteur à l'intralocuteur. Elles font évoluer le système de catégorisation mis en place sur les exemplaires connus en élargissant sa base empirique, en modifiant le poids fréquentiel d'une série d'exemplaires, en favorisant la formation d'une construction plus générale que celles qui étaient disponibles sous la forme d'exemplaires auparavant.

L'apport de la linguistique de corpus à la compréhension des phénomènes langagiers est donc devenu fondamental. Le nombre d'énoncés qu'entend et produit une personne durant sa vie est très grand. Grâce à l'augmentation de la variété et de la taille des corpus, il est aujourd'hui devenu possible de démontrer les faits langagiers à l'aide d'exemples attestés en grand nombre et de tester les propositions de la linguistique et de la psycholinguistique. Pour cela, un grand nombre de corpus contrôlés, bien décrits et variés, est nécessaire.

De mon point de vue, la Société de Linguistique Romane a un rôle crucial à jouer pour accompagner la communauté des chercheurs dans cette évolution fondamentale de nos approches au travers d'initiatives à lancer :

- Pour mieux prendre en compte les traitements informatiques et la linguistique de corpus dans nos formations en linguistique et philologie romanes. En ce sens il convient sans doute de réfléchir pour mieux attirer et accueillir des spécialistes du traitement automatique des langues au sein des futures éditions du CILPR.
- Pour inciter la communauté de linguistique et philologie romanes à un véritable partage de ressources et d'outils en ce domaine sous le double aspect de l'évolution des mentalités, mais

aussi de l'accompagnement à la mise en place de plates-formes technologiques de gestion et de mutualisation de telles ressources au sein des diverses communautés linguistiques des langues romanes.

– Pour valoriser cette activité de production des documents et ressources numériques dans l'évaluation des chercheurs et accompagner ou mettre en place des structures servant à la fois de validation et de diffusion de ces productions. C'est en partie du moins le rôle que le CNRS a confié aux Centres Nationaux de Ressources, dont le CNRTL, il y a quelques années, et qui en France va être repris par l'équipement d'excellence ORTOLANG que nous animons à Nancy.

6.4 Point de vue d'Achim Stein

Nous vivons une époque de transition où les résultats de la recherche philologique moderne sont publiés sur support numérique. Nous allons vivre une époque où même l'objet de la philologie, le texte primaire, ne sera plus publié ou ne sera plus accessible sur support papier.

Dans les deux domaines qui nous intéressent, la linguistique et la philologie, il faut relever deux défis :

(1) La linguistique théorique doit se justifier face à un traitement purement automatique des langues qui ne repose plus sur les symboles, ni sur les structures, mais uniquement sur le traitement probabiliste d'un grand nombre de données.

(2) La philologie historique doit veiller à ce que ses procédures méthodiques s'adaptent, aussi bien que les textes eux-mêmes, aux nouveaux supports. Je me bornerai à proposer que le CILPR représente de manière plus importante le domaine de la linguistique computationnelle des langues romanes. La section 'Travaux en cours' attire telle ou telle contribution issue de ce domaine, mais elle ne suffit pas à rentrer dans le détail des questions méthodologiques dont nous avons souligné l'importance ici.

Certes, mon propos n'est pas de créer un forum pour les linguistes computationnels spécialisés, puisque le CILPR ne saurait de toute façon pas concurrencer les grandes conférences nationales et internationales dans ce domaine. Mais ce qui me semble important, ce serait la création d'un niveau d'interface entre ces spécialistes et les philologues romanistes.

Concrètement, les travaux computationnels devraient être intégrés d'une manière plus systématique dans les sections thématiques – tout d'abord, parce qu'il est parfois difficile de trancher entre 'linguistique' et 'computationnel', mais surtout parce qu'ilest important d'assurer un maximum de communication entre les deux communautés à l'intérieur des domaines thématiques.

7. Pierre Kunstmann : bilan et conclusion

On a insisté, dans cette table ronde, sur la représentativité des corpus. La standardisation est en grande partie accomplie, même si on peut penser qu'elle ne sera jamais exhaustive. Il convient d'opter pour une mutualisation franche et massive, en respectant toutefois l'éthique scientifique et la simple déontologie.

On a constaté l'utilité de l'analyse à traits ou dimensions multiples, la nécessité

d'un vocabulaire contrôlé pour définir les métadonnées, vocabulaire qu'il convient d'enrichir au fur et à mesure de l'évolution des théories et modèles linguistiques. Il importe de diffuser l'information sur les progrès de la linguistique computationnelle auprès des utilisateurs naturels, linguistes et philologues.

Pour les textes statiques ou dynamiques, il faut préciser clairement comment on procède, quelles sont les marges d'erreur des outils de traitement automatique et quel est le pourcentage de bruit auquel on peut raisonnablement s'attendre.

En ce qui concerne l'archivage et la diffusion des données, il semble qu'on avance sur un bon chemin, autant avec le projet ORTOLANG ou le CNRTL à Nancy qu'avec les initiatives de la Real Academia Española ou la plate-forme EL GRIAL pour l'espagnol, par exemple.

Mais tout cela serait à étendre à l'ensemble des langues romanes. Il serait bon que dans les programmes de linguistique de nos universités, une large part soit faite au traitement automatique des langues et que nos futurs congrès, à l'image de celui-ci, s'ouvrent aussi largement à la linguistique computationnelle des langues romanes.

Université d'Ottawa — Pierre KUNSTMANN (président)
Istituto di Linguistica Computazionale 'Antonio Zampolli', Pise — Andrea BOZZI
Université Pontificale Catholique de Valparaiso — Giovanni PARODI
ATILF (CNRS/Université de Lorraine) — Jean-Marie PIERREL
Université de Stuttgart — Achim STEIN

8. Références bibliographiques

Bernard, Pascale / Lecomte, Josette / Dendien, Jacques / Pierrel, Jean-Marie, 2002. «Computerized linguistic resources of the research laboratory ATILF for lexical and textual analysis : Frantext, TLFi, and the software Stella», in: Suarez Araujo, Carmen Paz / González Rodríguez, Manuel (ed.), *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002)*, Las Palmas de Gran Canaria, European Language Resources Association, vol. 3, 1090-1098.

Bertrand, Olivier et al., 2013. La Cité de Dieu *de saint Augustin traduite par Raoul de Presles (1371–1375), édition du ms BNF, fr. 22912*, vol. 1/1, Paris, Champion.

Biber, Douglas, 1988. *Variation across speech and writing*, Cambridge, Cambridge University Press.

Biber, Douglas, 1994. «Using register-diversified corpora for general language studies», in: Armstrong, Susan (ed.), *Using large corpora*, Cambridge, MIT Press, 180-201.

Biber, Douglas / Conrad, Susan / Reppen, Randi, 1998. *Corpus linguistics. Investigating language structure and use*, Cambridge, Cambridge University Press.

Biber, Douglas / Reppen, Randi / Clark, Victoria / Walter, Jenia, 2001. «Representing spoken language in university settings: The design and construction of the spoken component of the T2K-SWAL Corpus», in: Simpson, Rita / Swales, John (ed.), *Corpus Linguistics in North America*, Ann Arbor, University of Michigan Press, 48-57.

Camps, Jean-Baptiste, 2013a. «Detecting Contaminations in Textual Traditions. Computer Assisted and Traditional Methods», communication présentée au International Medieval

Congress (Leeds, 3 juillet 2013), <www.academia.edu/3825633/Detecting_Contaminations_in_Textual_Traditions_Computer_Assisted_and_Traditional_Methods>.

Camps, Jean-Baptiste, 2013b. « Sélection des lieux variants et construction d'un stemma : nouvelles expérimentations », communication présentée au XXVIIe Congrès international de linguistique et de philologie romanes (Nancy, 15-20 juillet 2013).

CNRTL = 2005–. *Centre National de Ressources Textuelles et Lexicales*, Nancy, ATILF, <http://www.cnrtl.fr/>.

Corradini, Maria Sofia, 2005. « Formalisation des variantes à des fins computationelles : vérification de l'hypothèse expérimentale sur un texte occitan », in : Buckley, Ann / Billy, Dominique (ed.), *Études de langue et de littérature médiévales offertes à Peter T. Ricketts*, Turnhout, Brepols, 355-368.

Deerwester, Scott / Dumais, Susan / Furnas, George / Landauer, Thomas / Harshman, Richard, 1990. « Indexing by latent semantic analysis », *Journal of the American Society for Information Science* 41/6, 391-407, <http://www.stanford.edu/class/linguist289/lsi.pdf>.

Dendien, Jacques / Pierrel, Jean-Marie, 2003. « Le Trésor de la Langue Française informatisé : un exemple d'informatisation d'un dictionnaire de langue de référence », *Traitement Automatique des Langues* 44/2, 11-37.

DMF2012 = Martin, Robert / Bazin-Tacchella, Sylvie (dir.), 2012. *Dictionnaire du Moyen Français (DMF2012)*, Nancy, ATILF, <http://www.atilf.fr/dmf>.

Dumais, Susan, 1994. « Latent semantic indexing (LSI) and TREC-2 », <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.54.5141&rep=rep1&type=pdf>.

EAGLES, 1996. *Synopsis and comparison of morphosyntactic phenomena encoded in lexicons and corpora. A common proposal and applications to European languages*, Pise, ILC-CNR.

European Science Foundation, 2011. *Science Policy Briefing 42 : Research Infrastructures in the Digital Humanities*, Strasbourg, European Science Foundation, <http://www.esf.org/fileadmin/Public_documents/Publications/spb42_RI_DigitalHumanities.pdf>.

Frantext = ATILF, 2002–. *Outil de consultation de ressources informatisées sur la langue française Frantext*, Nancy, ATILF, <http://www.frantext.fr>.

Gutiérrez, Rosa María, 2005. « Análisis Semántico Latente : ¿ Teoría psicológica del significado ? », *Revista Signos. Estudios de Lingüística* 38/59, 303-323.

Kunstmann, Pierre, 1982. *Concordance analytique de La Mort le roi Artu*, en collaboration avec Martin Dubé, Ottawa, Éditions de l'Université d'Ottawa, 2 vol.

Kunstmann, Pierre, 2012. « L'électronique à l'aide de l'éditeur : miracle ou mirage ? Bilan de quatorze années de travaux au LFA », *Perspectives Médiévales* 34, 1-51, <http://peme.revues.org/2245>.

Martin, Robert, 2008. « Perspectives de la lexicographie informatisée », in : Durand, Jacques / Habert, Benoît / Laks, Bernard (ed.), *Congrès Mondial de Linguistique Française (Paris, 9-12 juillet 2008) – CMLF'08*, Paris, Institut de Linguistique Française, 1251-1256, <http://www.linguistiquefrancaise.org/articles/cmlf/pdf/2008/01/cmlf08332.pdf>.

Martin, Robert / Gerner, Hiltrud / Souvay, Gilles, 2010. « Présentation de la seconde version du DMF (Dictionnaire du Moyen Français) », in: Iliescu, Maria / Danler, Paul / Siller-Runggaldier, Heidi (ed.), *Actes du XXVe Congrès International de Linguistique et de Philologie Romanes (Innsbruck, 3-8 septembre 2007)*, Berlin/New York, De Gruyter, vol. 6, 213-220.

MCVF = Martineau, France (dir.), 2005–. *Modéliser le changement : les voies du français*, Ottawa, Université d'Ottawa, <www.voies.uottawa.ca>.

Parodi, Giovanni (ed.), 2007. *Working with Spanish corpora*, Londres/New York, Continuum.

Parodi, Giovanni, 2010a. *Lingüística de corpus : de la teoría a la empiria*, Francfort, Veruert/ Iberoamericana.

Parodi, Giovanni, 2010b. « Research challenges for corpus cross-linguistics and multimodal texts », *Information Design Journal* 18/1, 69-73.

Pierrel, Jean-Marie / Buchi, Éva, 2009. « Research and Resource Enhancement in French Lexicography : the ATILF Laboratory's Computerized Resources », in : Bruti, Silvia / Cella, Roberta / Foschi Albert, Marina (ed.), *Perspectives on Lexicography in Italy and in Europe*, Newcastle upon Tyne, Cambridge Scholars Publishing, 79-118.

Pierrel, Jean-Marie / Petitjean, Étienne, 2007. « Valorisation et exploitation scientifiques de documents numériques pour la recherche en linguistique : l'exemple du CNRTL », in : *Le document numérique dans le monde de la science et de la recherche. Actes du dixième colloque international sur le document électronique (CIDE.10)*, Paris, Europia, 13-24.

Pivot, Bernard, 2013. *Les tweets sont des chats*, Paris, Albin Michel.

Pustejovsky, James, 1995. *The Generative Lexicon*, Cambridge, MIT Press.

Rastier, François, 2004. « Enjeux épistémologiques de la linguistique de corpus », in: Williams, Geoffrey (ed.), *Actes des deuxièmes journées de linguistique de corpus (Lorient, 2002)*, Rennes, Presses Universitaires de Rennes, 31-46.

Rodríguez, Horacio / Climent, Salvador / Vossen, Piek / Bloksma, Laura / Peters, Wim / Alonge, Antonietta / Bertagna, Francesca / Roventini, Adriana, 1998. «The Top-Down Strategy for Building EuroWordNet : Vocabulary Coverage, Base Concepts and Top Ontology», *Computers and the Humanities*, 32/2-3, 117-152.

Ruimy, Nilda / Piccini, Silvia / Giovannetti, Emiliano, 2012. « Les Outils Informatiques au Service de la Terminologie Saussurienne », in : Neveu, Franck / Muni Toke, Valelia / Blumenthal, Peter / Klingler, Thomas / Ligas, Pierluigi / Prévost, Sophie / Teston-Bonnard, Sandra (ed.) : *3e Congrès Mondial de Linguistique Française (Lyon, 4-7 juillet 2012). Recueil des résumés, CD-ROM des actes*, Paris, Institut de Linguistique Française, http://www.shs-conferences.org/articles/shsconf/pdf/2012/01/shsconf_cmlf12_000294.pdf, 1043-1056.

Souvay, Gilles / Pierrel, Jean-Marie, 2009. « LGeRM : lemmatisation de mots en moyen français », *Traitement Automatique des Langues* 50/2, 149-172.

Städtler, Thomas, 2010. « Die evolutive Lexikografie am Beispiel der Geschichte des *Dictionnaire du Moyen Français* », *Zeitschrift für französische Sprache und Literatur* 120, 1-13.

Steinfeld, Nadine (dir.), 2007–. *Base des mots fantômes*, Nancy, ATILF, <http://www.atilf.fr/MotsFantomes>.

Steinfeld, Nadine, 2013. « La traque des mots fantômes à travers les terres de La Curne et de Godefroy : un tableau de chasse chargé de trophées pittoresques extraits du *Livre des deduis du roy Modus et de la royne Ratio* », in : Casanova Herrero, Emili / Calvo Rigual, Cesáreo (ed.), *Actas del XXVI Congreso Internacional de Lingüística y de Filología Románicas (Valencia, 2010)*, Berlin/New York, De Gruyter, vol. 7, 411-422.

Steinfeld, Nadine / Andronache, Marta, 2011. « Quoi de neuf du côté de la lexicographie étymologique ? La méthode utilisée dans le cadre du projet TLF-Étym pour distinguer les emprunts au latin de l'Antiquité de ceux faits au latin médiéval », *Estudis romànics* 33, 151-170.

TLF = Imbs, Paul / Quemada, Bernard (dir.), 1971–1994. *Trésor de la langue française. Dictionnaire de la langue du XIXe et du XXe siècle (1789–1960)*, 16 vol., Paris, Éditions du CNRS/Gallimard.

TLF-Étym = Steinfeld, Nadine (dir.), 2005–. *Programme de recherche TLF-Étym*, Nancy, ATILF, <http://www.atilf.fr/tlf-etym>.

TLFi = ATILF, 2004. *Trésor de la langue française informatisé. Dictionnaire de la langue du XIXe et du XXe siècle (1789–1960)*, Nancy, ATILF, <http://www.atilf.fr/tlfi>.

Venegas, René, 2007. « Clasificación de textos disciplinares en función de su contenido léxico-semántico », *Revista Signos. Estudios de Lingüística* 40/63, 239-231.

Wittenburg, Peter / Bel, Nuria / Borin, Lars / Budin, Gerhard / Calzolari, Nicoletta / Hajicova, Eva / Koskenniemi, Kimmo / Lemnitzer, Lothar / Maegaard, Bente / Piasecki, Maciej / Pierrel, Jean-Marie / Piperidis, Stelios / Skadina, Inguna / Tufis, Dan / Veenendaal, Remco van / Váradi, Tamás / Wynne, Martin, 2010. « Resource and Service Centres as the Backbone for a Sustainable Service Infrastructure », in : Calzolari, Nicoletta *et al.* (ed.), *Proceeding LREC 2010. Seventh International Conference on Language Resources and Evaluation (Valletta, Malta, 17-23 May 2010)*, La Valete, European Language Resources Association, 60-63, < http://www.lrec-conf.org/proceedings/lrec2010>.

Table Ronde « Littérature, philologie, linguistique : l'unité de la romanistique »

Les cinq intervenants, Margarita Borreguero Zuloaga, Elisabeth Gülich, Barbara Frank-Job, François Rastier et Wolfgang Raible, avaient au préalable cerné quelques domaines qui pourraient servir d'éléments fédérateurs entre la linguistique et les études littéraires. Chacun des intervenants était responsable d'un de ces domaines. Le tout formera un parcours à travers la linguistique textuelle, par définition point de rencontre possible entre littérature et linguistique, à travers une autre interface, l'aspect générique ou les traditions discursives (*Diskurstraditionen*), pour choisir le terme créé par Brigitte Schlieben-Lange, ensuite à travers les possibilités offertes par les études médiévales, véritable berceau de la 'Romanistik', puis la narratologie et ses possibilités parfois inattendues ; en fin de compte on va effleurer et illustrer à l'aide d'un exemple le travail sur grands corpus accessibles et analysables par des moyens informatiques.

1. La linguistica testuale, punto di incontro della linguistica, la filologia e la letteratura

Le premier domaine est celui de la dimension textuelle, qui rapproche tout naturellement linguistique et littérature. La tâche de mettre au clair ses possibilités inhérentes revient à Margarita Borreguero Zuloaga, titulaire de philologie italienne à l'Université Complutense de Madrid. Voici sa contribution :

Gli studi testuali come li conosciamo oggi sono il risultato della convergenza di un insieme di discipline che si sono sviluppate tra gli anni '50 e gli anni '80 dello scorso secolo: la stilistica e la nuova retorica, la narratologia di stampo strutturalista, la semiotica testuale, l'analisi del discorso anglossassone e la *Textlinguistik* tedesca, tra le altre. Quest'ultima, che come la linguistica romanza è una creazione nettamente tedesca, si è costituita come un punto di incontro tra la letteratura – o più precisamente la teoria letteraria – e la linguistica.

Se pensiamo per esempio alle diverse teorie testuali che dagli anni sessanta si sono sviluppate in Germania – sia nelle università della Repubblica Federale (Konstanz, Bielefeld) sia presso l'Accademia delle Scienze di Berlino – allo scopo di stabilire il testo come l'unità basica di analisi, si possono identificare due influssi determinanti per la nascita di questa disciplina: la grammatica generativa, ancora allo stadio della teoria standard proposta in *Aspetti della teoria della sintassi* di Chomsky (1965), e

la narratologia strutturalista che ha in Vladimir Propp il suo più rinomato pioniere. Quindi la *Textlinguistik* nasce in qualche modo come un approccio metodologicamente misto, che prende spunto sia da una disciplina che in ambito letterario lavorava già con i testi, sia dalla linguistica che le forniva gli strumenti di analisi, anche se ancora limitati all'analisi frasale.

Nonostante tutti gli studiosi che si occupavano allora di linguistica testuale l'avessero concepita come una disciplina con un oggetto di studio e una metodologia a sé, cioè una disciplina che elaborava e metteva a disposizione dello studioso uno strumento adeguato per l'analisi dei testi e lo studio dei processi di costruzione e interpretazione testuale, in realtà gli approcci non erano affatto omogenei. C'erano approcci di carattere filosofico, como quello di Peter Hartmann o Siegfried J. Schmidt, più formali come quelli di János Sándor Petőfi, Teun A. van Dijk, Hannes Rieser e Roland Harweg e più pragmatici come quello di Dieter Wunderlich[1].

Ma se ci limitiamo ai lavori che hanno avuto maggiore impatto sulle lingue romanze, cioè le grammatiche testuali di tipo formale – originalmente elaborate de Manfred Bierwisch, Wolfgang Motsch, Wolfdietrich Hartung, Karl Erich Heidolph, Horst Isenberg e Renate Steinitz (membri del gruppo «Arbeitsstelle für Strukturelle Grammatik» a Berlino-Est) o da Wolfgang U. Dressler a Vienna – che dopo si sono trasformate in modelli testuali più complessi, possiamo individuare i contributi sia della teoria letteraria in ambito narratologico che della grammatica generativa e quindi concepire la linguistica testuale come un luogo di incontro tra la linguistica e la letteratura. Infatti, la linguistica testuale viene concepita come una disciplina che gioverà ugualmente agli studi linguistici e a quelli letterari perché permetterà in entrambi casi un'analisi più accurata e completa dei testi. «The textual extension of linguistics is not only a necessary development of grammar itself, but also a more adequate basis for the description of literary texts» (van Dijk 1972, 201).

1.1 I corpora e lo sviluppo degli studi testuali

Sebbene le teorie testuali emerse durante la seconda metà del Novecento abbiano stabilito i fondamenti teorici della linguistica testuale, il vero grande sviluppo degli studi testuali è avvenuto quando i linguisti si sono avvalsi dei grandi corpora testuali che erano stati compilati da grandi gruppi di ricerca formatisi negli anni novanta. I corpora testuali non solo hanno permesso di approfondire l'analisi di molti aspetti della configurazione testuale, dai meccanismi di coesione all'articolazione informativa del testo alla macrostruttura dei diversi generi discorsivi, ma hanno fatto emergere nuovi oggetti di studio.

Tra questi vogliamo qui ricordare molto brevemente la lingua orale, che non è stata studiata in modo sistematico fino alla compilazione dei corpora di lingua parlata, come

[1] Per una panoramica delle principali proposte teoriche che diedero avvio alla nascita della *Textlinguistik,* cf. Bernárdez 1982, Sáiz Noeda 1994, Hölker 2001, Borreguero Zuloaga 2006 e la bibliografia citata in questi studi.

per il francese il *Corpus de Référence du Français Parlé* (CRFP) iniziato da Claire Blanche-Benveniste e la sua équipe all'Università di Aix-en-Provence e continuato dopo dal gruppo Description Linguistique Informatisée sur Corpus (Équipe DELIC 2004), per l'italiano i *Corpora Linguistici per l'Italiano Parlato e Scritto* (CLIPS) o il *Corpus di italiano parlato* raccolto al Laboratorio Linguistico dell'Università di Firenze (Lab.L.Ita., cf. Cresti 2000), per lo spagnolo il *Corpus Valencia Español Coloquial* (Val.Es.Co.; cf. Briz / Grupo Val.Es.Co 2002) o il *Corpus del Proyecto para el Estudio Sociolingüístico del Español de España y de América* (PRESEEA; cf. Moreno Fernández 1996; 2005) e per le grandi lingue romanze il *C-Oral-Rom* (Cresti / Moneglia 2005), per fare solo qualche nome. La stessa cosa vale per l'interlingua, cioè il sistema linguistico che emerge e evolve durante il processo di apprendimento linguistico, che solo grazie ai corpora è diventato un vero e proprio oggetto di studio nell'ambito della linguistica acquisizionale, lontano ormai dagli approcci impressionistici sugli 'errori' e dai resoconti sulle divergenze riscontrate in confronto con le produzioni testuali dei madrelingua (cf. Giacalone Ramat 2001).

Per ultimo, la compilazione dei corpora dei testi storici è stata decisiva non solo per lo slancio energico che hanno sperimentato gli studi di storia della lingua e di variazione linguistica in genere, ma anche per la creazione di un concetto fondamentale – anche se ancora poco sfruttato – per chi si occupa della configurazione dei testi dal punto di vista diacronico e per chi è interessato a studiare i fenomeni di cambiamento linguistico tenendo conto del cotesto. Si tratta della nozione di tradizione discorsiva ideata dai romanisti tedeschi (Schlieben-Lange 1983; Koch 1991; Oesterreicher 2007) che ha permesso di identificare la funzione di determinate 'famiglie' testuali nel dare impulso o rallentare certi fenomeni del cambio linguistico e di creare in questo modo un legame solido tra gli studi filologici e gli studi linguistici. Tra queste 'famiglie' sono stati studiati, per esempio, i testi giuridici come propulsori del cambiamento sintattico (Kabatek 2005).

Ma oltre alla creazione di nuovi oggetti di ricerca e di nuove nozioni teoriche la possibilità di disporre di vaste quantità di testi appartenenti a diversi tipi testuali e generi discorsivi, a diversi registri e contesti di produzione, rappresentativi della variazione diatopica, diacronica e diastratica di una lingua, della distanza e dell'immediatezza comunicativa insieme alla possibilità di interrogarli grazie a sistemi che permettono di individuare una parola o struttura nel loro contesto ha contribuito ad approfondire certi aspetti della costituzione testuale che prima erano stati solo sfiorati o addirittura completamente ignorati. Tra questi vogliamo soffermarci in queste pagine sull'articolazione informativa del testo, indagata con grande successo nell'ultimo decennio grazie ai cosiddetti modelli di segmentazione discorsiva nati appunto nell'ambito della linguistica romanza.

1.2 I modelli di segmentazione discorsiva: una proposta teorica della linguistica romanza per l'analisi dei testi

Negli ultimi anni, la linguistica romanza ha dato un contributo teorico importante alla linguistica testuale proponendo diversi modelli di segmentazione discorsiva, cioè modelli che hanno cercato di individuare un'unità minima di organizzazione testuale che sia valida sia per i testi orali che scritti. Questi modelli sono stati elaborati in un primo momento dagli studiosi che si occupavano dell'analisi della lingua orale (Roulet *et al*. 1985; Blanche-Benveniste *et al*. 1990; Cresti 2000; Briz / Grupo Val.Es.Co 2003), ma sono stati poi estesi ai testi scritti (Ferrari *et al*. 2008).

L'idea fondamentale che soggiace a questi modelli è che l'unità basica del discorso non è la frase e quindi l'analisi sintattica è insufficiente per spiegare il comportamento discorsivo degli enunciati che formano un testo. In effetti, un testo non può definirsi come un insieme di frasi e la sintassi non può essere dunque la disciplina linguistica che si occupa dell'analisi dei testi né la linguistica testuale può essere concepita come una grammatica transfrastica, come avevano fatto i primi studiosi (Petőfi 1971). È necessario stabilire una nuova unità di organizzazione discorsiva che permetta di spiegare la struttura informativa, illocutiva e gerarchica dei componenti testuali[2].

I tentativi di delimitare unità basiche di analisi anche nel livello discorsivo rispondono alla necessità di segmentare un continuum per poterlo analizzare, di stabilire dei rapporti gerarchici tra i segmenti e di spiegare i fenomeni testuali inquadrati in questa gerarchia. In altre parole, rispondono alla necessità cognitiva di creare categorie per la comprensione della realtà e rispecchia quanto è già avvenuto in altri livelli di analisi linguistica, dalla fonologia alla sintassi.

Non è una casualità che siano stati gli studiosi della conversazione i primi a parlare di una sintassi del parlato (Narbona 1991) che non aveva corrispondenza con la sintassi descritta nelle grammatiche. Inoltre, quando i fonetisti che si occupavano di prosodia decisero di abbandonare l'analisi di frasi registrate in condizioni sperimentali ottime, per studiare il ruolo che l'intonazione aveva nella struttura discorsiva, giunsero alla conclusione che, nel caso della lingua parlata, l'individuazione delle unità basiche di analisi non poteva prescindere da elementi come le pause o i tonemi.

Le prime proposte sono emerse in ambito francofono, tra gli studiosi che si occupavano appunto dell'analisi della conversazione. Presenterò qui molto brevemente unicamente quelle che hanno avuto maggiore diffusione in altre lingue romanze, anche se autori come Alain Berrendonner o Liesbeth Degand hanno pure elaborato interessanti proposte (cf. Andersen / Nølke 2002; Pons 2014a).

---

[2] Evidentemente quest'idea è già presente in molti dei fondatori della linguistica testuale: ricordiamo la macrosintassi di van Dijk (1977) o gli approcci transfrastici di Stati (1990), per fare solo qualche esempio.

Da una parte, Eddy Roulet e il suo gruppo all'Università di Ginevra hanno preso come unità basica di analisi l'atto linguistico come era stato concepito da John L. Austin e John R. Searle (Roulet *et al*. 1985; 2001). Quando un parlante interviene in una conversazione compie una serie di atti linguistici (domanda, invito, affermazione, richiesta, ecc.) e la struttura conversazionale emerge dal concatenarsi di questi atti linguistici in unità più complesse fino ad arrivare all'intera conversazione, come avevano anticipato gli etnometodologi americani (Sacks / Schegloff / Jefferson 1974). Un atto linguistico può suddividersi a sua volta in sottoatti, che possono essere sostantivi o adiacenti, a seconda del loro contributo informativo. Al di sopra dell'atto ci sono altre unità monologiche come gli interventi e dialogiche come lo scambio e il dialogo. In un primo momento il modello ha stabilito unicamente un rapporto gerarchico tra le diverse unità individuate, ma negli sviluppi più recenti si è arricchito con altre 16 funzioni ed è quindi diventato un modello modulare.

Questo modello di analisi conversazionale ha avuto particolare fortuna in ambito ispanofono dove Antonio Briz e i membri del gruppo Val.Es.Co. (Valencia Español Coloquial) hanno fatto un adattamento a un tipo concreto di conversazione, la conversazione colloquiale (Briz / Grupo Val.Es.Co. 2003). Combinando l'analisi informativa con l'analisi prosodica, hanno individuato diversi tipi di sottoatti, atti, turni, scambi, dialoghi e altre unità che non erano state descritte dai ginevrini come la conversazione, unità definita grazie alla scoperta della posizione inziale assoluta (Estellés / Pons 2014). Comunque, sempre in ambito ispanico sono emerse anche altre proposte di adattamento che seguono il modello di Roulet (cf. Cortés / Camacho 2005).

Dall'altra parte, Claire Blanche-Benveniste e il suo gruppo di ricerca all'Università di Aix-en-Provence (il cosiddetto GARS, Groupe aixois de recherches en syntaxe) hanno fissato come unità basica del discorso l'enunciato (prendendo spunto dalla teoria dell'enunciazione di Benveniste). Questo approccio, a differenza di quello ginevrino, concede un ruolo maggiore all'analisi sintattica ma considera che i limiti imposti dalla sintassi si possono risolvere aggiungendo un ulteriore livello di macrosintassi. Questo livello macrosintattico è formato da tre unità all'interno dell'enunciato: (a) il 'noyau' o parte dove avviene la predicazione, (b) il 'préfixe' o parte che precede il 'noyau' e presenta le condizioni di validità della predicazione, (c) il 'suffixe' o parte che segue il 'noyau' e contiene normalmente un'informazione secondaria che specifica o precisa l'informazione nucleare (cf. Blanche-Benveniste *et al*. 1990).

A partire da questa struttura base si sono sviluppati altri modelli, prevalentemente in ambito italiano. Per l'italiano orale, Emanuela Cresti (2000) e il gruppo Lab.L.Ita a Firenze hanno adottato e raffinato la proposta di Blanche-Benveniste rinominando le categorie base (topic, comment, appendice) e creando nuove categorie di analisi (introduttore illocutivo, incipit). Questo modello, denominato Teoria della Lingua in Atto, è basato prevalentemente sull'analisi prosodica degli enunciati ed è stato applicato anche al portoghese (Raso / Mittmann 2009).

Il modello di Cresti è un buon esempio del peso acquisito dalla prosodia nell'individuazione delle unità discorsive. La centralità della componente prosodica è dovuta in gran parte alle innovazioni tecnologiche che hanno agevolato enormemente il compito dell'analisi prosodica (si pensi per esempio al software Praat). Infatti, modelli come quello della Cresti e quello di Edoardo Lombardo Vallauri hanno scelto i parametri prosodici come la base della segmentazione discorsiva. Per il francese si veda il modello di Morel che distingue, sempre su base prosodica, tra preambolo e rema.

Poco dopo, Angela Ferrari, a Basilea, ha adattato e semplificato la proposta di Cresti per poterla applicare all'analisi dei testi scritti, mettendo in evidenza che i modelli di segmentazione discorsiva non sono esclusivi della lingua parlata. Dalla prima presentazione del modello nel 2003 all'ultima e più completa nel 2008 è avvenuto un processo di approfondimento teorico e analitico che ha portato a una semplificazione delle unità informative di primo livello ridotte unicamente a tre: il nucleo, il quadro e l'appendice.

Il merito fondamentale dei modelli di segmentazione discorsiva è proporre un approccio teorico unificato per l'analisi dei testi orali e scritti che fino a pochi anni fa costituivano due campi di ricerca completamente autonomi.

1.3 Le unità informative e il cambio linguistico: un nesso di unione tra linguistica testuale e filologia

Per finire, vorremmo mettere in luce come questi modelli di segmentazione discorsiva, che sono nati e si sono sviluppati nell'ambito della linguistica romanza e che possono essere considerati uno dei grandi contributi teorici della linguistica romanza alla linguistica testuale attuale, abbiano fornito allo stesso tempo una cornice teorica per spiegare certi fenomeni dell'evoluzione storica delle lingue nel livello discorsivo. Come sappiamo, lo studio diacronico della lingua ha arricchito i suoi campi di ricerca accogliendo lo studio pragmatico e testuale insieme agli studi di fonetica, morfologia, sintassi, lessicologia e semantica storiche già consolidati e con una lunga tradizione nello studio delle lingue romanze che risale a metà dell'Ottocento.

Per mostrare l'utilità di tali modelli prenderemo il caso dei cosiddetti segnali discorsivi, elementi funzionali che agiscono a livello testuale ma che emergono a partire da altre categorie grammaticali. Alcuni studiosi avevano osservato che nel processo di grammaticalizzazione di un elemento del sistema linguistico che con l'andare del tempo diventa un connettivo testuale o assume funzioni di tipo interazionale o metadiscorsivo si possono individuare diversi fenomeni semantici e sintattici (Brinton 1996). Tra i primi possiamo menzionare la desemantizzazione – parziale o totale – del suddetto elemento che perde una parte del significato concettuale ma conserva il significato procedurale o istruttivo; tra i secondi, lo spostamento all'interno dell'enunciato da posizioni centrali verso posizioni periferiche.

Se invece di parlare di posizioni enunciative – che è un concetto alquanto vago – prendiamo come cornice teorica della nostra analisi le unità informative descritte nei

modelli menzionati, la nostra descrizione diventa molto più precisa, accurata e sistematica. Come abbiamo accennato, il modello di Basilea, che risulta il più idoneo perché concepito per l'analisi dei testi scritti, distingue tre unità informative all'interno dell'enunciato (Ferrari *et al*. 2008)[3]:

(a) il Nucleo, che determina la forza illocutiva dell'intero enunciato e contiene il rapporto predicativo centrale, e che quindi è semanticamente autonomo;
(b) il Quadro, che precede sempre il nucleo e la cui portata può limitarsi all'enunciato in cui compare o estendersi a un insieme di enunciati successivi; la sua funzione principale è indicare le circostanze spazio-temporali in cui avviene la predicazione, stabilire la connessione con enunciati precedenti o introdurre topic discorsivi;
(c) l'Appendice, che segue il Nucleo o il Quadro e ha come scopo offrire un'informazione secondaria che precisa o chiarisce quanto detto nel Nucleo o nel Quadro; occupa sempre una posizione parentetica[4].

Ecco due enunciati in cui abbiamo individuato le unità informative definite[5]:

(1) Durante l'estate scorsa,/^{Quadro} conobbi un ragazzo molto simpatico,/^{Nucleo} che poi sparì dalla mia vita/^{Appendice}//.
(2) Tre anni fa,/^{Quadro} se non ricordo male/$^{\text{Appendice del Quadro}}$ incontrai una persona che cambiò il mio modo di vedere il mondo,/^{Nucleo} per sempre/$^{\text{Appendice del Nucleo}}$//.

Queste unità sono a loro volta articolate in *Topic* e *Comment* (o Tema e Rema, che sono i termini usati in versioni precedenti del modello), unità che appartengono al livello di organizzazione tematica degli enunciati. Senza pretese di fare una descrizione completa di questo complesso modello, vorrei mostrare brevemente come l'individuazione di queste unità costituisca anche un grande contributo per gli storici della lingua e non solo per chi si occupa di linguistica sincronica.

Se prendiamo ad esempio *allora*, che è stato uno dei segnali discorsivi dell'italiano che ha ricevuto maggiore attenzione da parte dagli studiosi (Bosco / Bazzanella 2005; Bazzanella *et al*. 2007; 2008; Bazzanella / Borreguero Zuloaga 2011), possiamo osservare che già sin dai testi del Duecento si attestano diverse funzioni sia come avverbio temporale sia come connettivo intrafrasale, ma che queste funzioni sono contraddistinte dalle diverse posizioni enunciative. In effetti, quando *allora* ha valore

[3] Comunque il modello di Basilea non è stato l'unico ad avere un impatto nell'ambito filologico. Anche il modello Val.Es.Co. è usato per spiegare il processo di grammaticalizzazione dei segnali discorsivi (si veda Pons 2014b su *o sea*).

[4] Bisogna aggiungere che l'unità massima presa in considerazione dal modello è il paragrafo, che viene poi suddiviso in enunciati o unità comunicative, tra le quali esistono diversi tipi di rapporti: a) rapporti tematici, che riguardano l'introduzione e l'arricchimento di topic discorsivi grazie all'uso di meccanismi coreferenziali di ogni tipo (dalla semplice ripresa pronominale agli incapsulatori); b) rapporti logici e argomentativi, che riguardano le relazioni di causa, conseguenza, ordine temporale, finalità, ecc. tra i diversi stati del mondo rappresentati dagli enunciati; c) rapporti polifonici, relativi alla introduzione di diverse voci e punti di vista nella costruzione del discorso.

[5] La sbarra semplice indica il confine tra le unità informative, la sbarra doppia il confine tra enunciati.

semantico temporale e funzione avverbiale (modificatore del predicato) appare sempre in posizione intrapredicativa, cioè nel *Comment* dell'unità informativa. Dato che si tratta di un elemento di natura anaforica normalmente è preceduto o seguito da un'espressione temporale – con frequenza una subordinata temporale introdotta da *quando* – che precisa il riferimento e permette di situare nel tempo con esattezza l'azione espressa dal verbo. In questa funzione *allora* può apparire in un'unità informativa di Nucleo (3), di Appendice (4) o di Quadro che siano saturate da almeno un *Comment*:

(3) Et / **allora** si dimanda mala costruzione del luogo[Comment-1], *quando* i periculi si dubitano[Comment-2]/[Nucleo] (Andrea da Grossetto, 1268)

(4) Al tempo di costui nel CCXXXVIII /[Quadro] più di mille sì cominciaro li Fiorentini a lastricare Firenze,/[Nucleo] la città[Topic] che infino **allora** non era lastricata[Comment]…/[Appendice] (Paolino Pieri, *Cronica*, fior. c. 1305)

Quasi contemporaneamente si osservano casi in cui *allora* acquisisce la funzione di connettivo testuale che con un grado di desemantizzazione ancora ridotto – e quindi senza perdere completamente il valor temporale – svolge la funzione di creare una sorta di continuità temporale tra gli eventi narrati, tipica dei testi narrativi e particolarmente delle cronache medievali. Allo scopo di acquisire questa funzione, *allora* subisce un cambiamento posizionale verso la posizione periferica iniziale dell'enunciato e quindi, nei termini del modello di Basilea, satura l'unità informativa di Quadro:

(5) [...] et andaosenne a l'insola de Ponto et avendo lo pecorone retornaosende in Grecia. **Allora/**[Quadro] recordandose Iason co li sopradecti conpangi la iniuria ke li fece fare Lamendont rege [...] et entraro ne la nave e andaro de reto a Ppari, ma no lo trovaro. **Allora**/[Quadro] li greci acconçaro li navi in Athena. Agamenon mandao Acchilles e Cascas, filio de Nestore [...] Et in quella annata Accilles occise Teutras et Telefeus lo fece sepelire honoratamente. **Allora**/[Quadro] Accilles disse a Ttelofo [...] E durao la vactalgia continua .vij. dii e fo facta trieugua. **Allora**/[Quadro] dicea Accilles a li greci ke tucti devessero essere sopre a Troylus [...] Brutus et Ayrons fecero grande bactalge et gravemente feruti, occiserosse. **Allora/**[Quadro] generalemente tucti romani piangeano Bruto, sìcomo ad onne homo fossi morto oi frate oi nepote [...] vendesenne ne lo monte de Cercegi in Ytalia et demorao kello per alcuno tempo. **Allora**/[Quadro] in quella isola era una donna la quale avea nome Circe (*Storie di Troia e di Roma*, 1252/1258).

Un cambiamento parallelo è stato osservato per lo spagnolo *entonces*. In uno studio precedente (Borreguero Zuloaga 2012), basato su due grandi corpora diacronici della Real Academia Española (il CORDE e il CNDH), abbiamo mostrato che il passaggio da avverbio temporale a connettivo consecutivo è strettamente legato allo spostamento di questa unità da posizioni interne al nucleo verso la posizione di quadro.

(6) *Qvando* por auentura, el pendrado asu pendrador dixiere, que non le tollio la pendra[Comment-1], o non gelo defendio[Comment-2], et que aduxo el iuez sobrel non derechurera mente[Comment-3], **estonces** prueue el pendrador[Comment-4/Nucleo-], con testimonio de aquel uezino con el que querie pendrar/[Appendice], quel auie tollido la pendra[Comment], et quita la pendra de mano del iuez/[Nucleo] //. Si por auentura aquesto el pendrador prouar non pudiere, el mismo quitela de mano del iuez la pendra. (*Fueros de Zorita de los Canes*, 1218)

(7) Daqui adelant fablo de los reyes que avien a venir apres del e cascuno de qual poder serie, e dixo: «En essos dias destos regnos levantara el Dios de los cielos regno que por consieglo non sera crebantado e regno otro pueblo non dara, crebantara e taiara todos estos regnos e el sera firme por consieglo. Est es el suenno e es firme su soltura». **Estonz**,/Quadro el rey Nabucodonosor echos sobre sus fazes e humillos a Daniel e mandol fer sacrificio much ondrado. E dixo a Daniel: «Es de cierto que vuestro Dios es el Dios de los cielos, Rey de los Reys, e escubre las cosas cerradas». **Entonces/**Quadro el rey dio grandes dones a Daniel e diol sennoria sobre sos sabios e la cibdat de Babilonia, e fue privado del rey, e establido a Sydrac e Misaach e Abdenago que estidiesen a servicio del rey. (Almerich, *La fazienda de Ultra Mar*, c. 1200)

Questo spostamento a posizioni periferiche che permette l'acquisizione di funzioni discorsive rappresenta uno dei principali controargomenti usati contro la teoria della grammaticalizzazione quando in un primo momento sosteneva che il passaggio da un'unità lessicale libera ad un'unità grammaticale comportava necessariamente una riduzione nella portata di questo elemento (postulato che funziona nei casi di morfemi liberi che diventano affissi, ma evidentemente non nel caso di avverbio o di altri elementi che diventano connettivi).

Infatti negli ulteriori sviluppi di *allora* verso la funzione di connettivo consecutivo (9), osserviamo che conserva la posizione nell'unità informativa di Quadro, posizione che gli permette di proiettare il suo nuovo valore semantico a tutto l'enunciato e svolgere la funzione di esplicitare il legame logico-argomentativo esistente tra gli enunciati:

(8) cioè che debbia éssare il consellio ordinato o vero almeno XVJ di loro, con agionta di sei buoni uomini per terzo; e **allora**,/Quadro quello che 'l detto consellio ne ordinarà e fermarà, essendo le due parti del consellio in concordia a scruptinio a pallocte a pugno chiuso (*Statuto dell'Università ed Arte de la lana de Siena*, 1298)

Con queste brevi riflessioni abbiamo tentato di ripercorrere molto velocemente alcune delle tappe fondamentali nello sviluppo della linguistica testuale per mostrare come questa disciplina in diversi momenti della sua storia sia diventata come luogo di incontro delle tre discipline fondamentali che costituiscono la romanistica: la filologia, la letteratura e la linguistica. Linguistica testuale e studi di romanistica hanno un'origine comune nel mondo accademico tedesco a partire dal quale si sono diffuse fino a costituire tradizioni internazionali arricchite dai contributi di numerosi studiosi e discipline. Se nei suoi inizi la *Textlinguistik* tentò di combinare la ricerca linguistica più innovativa con la tradizione teorico-letteraria di stampo strutturalista, e i diversi modelli elaborati per il tedesco e per l'inglese furono presto adottati dai linguisti dell'ambito romanico, negli ultimi decenni sono stati i romanisti a offrire un contributo decisivo alla linguistica testuale che ha permesso allo stesso tempo di precisare la descrizione di fenomeni dell'evoluzione diacronica della lingua, come i processi di grammaticalizzazione che hanno dato luogo all'emergenza di nuovi elementi e alla trasformazione funzionale e categoriale di molti altri.

2. Situation de la linguistique des genres

On quitte maintenant le domaine de la linguistique des textes pour aborder celui des genres textuels. Ce dernier mérite l'intérêt tant de la linguistique que des études littéraires. C'est qu'il présente au moins trois atouts : (1) avec l'évolution de traditions discursives on introduit forcément un aspect diachronique ; (2) les genres nous permettent de faire une comparaison entre littératures différentes ; (3) on peut étudier l'évolution des moyens langagiers – puisque les genres textuels établissent en partie des exigences extrêmes à cet égard (il suffit de penser aux genres juridiques). C'est François Rastier qui s'est occupé du champ générique. Sémioticien de renommée internationale, ancien directeur de recherche au CNRS, il vient en outre de publier un livre entier à ce sujet. Voici sa contribution qui, après une introduction plus générale faisant l'éloge de la *Romanistik* allemande, aborde le sujet des genres :

La romanistique, part éminente de la linguistique historique et comparée, n'a pas connu en France de développement notable. Si l'on y étudie des langues romanes, c'est généralement dans une perspective monographique (ou, au plus, contrastive) plutôt que comparative. De fait, sauf dans le domaine des langues anciennes, comparatisme et linguistique historique sont restés découplés : les questions de diachronie intéressent pour l'essentiel l'histoire de la langue française. La notion ambiguë et politisée de 'francophonie' a sans doute inhibé le développement de la romanistique.

Si la linguistique contemporaine s'est éloignée de la philologie, ce n'est pas seulement par négligence de la diachronie, mais parce que le point de vue formaliste ne peut concevoir la notion même de document (pour Chomsky, même l'écriture n'a pas d'existence linguistique).

2.1 Linguistique et littérature : comment saisir une œuvre ?

À sa constitution en tant que science, la linguistique historique et comparée était conçue notamment comme discipline auxiliaire de la lecture des textes, et elle s'est formée parallèlement à la littérature comparée. Dans le domaine littéraire, alors que la romanistique a apporté une contribution éminente (Spitzer, Auerbach), elle reste un peu négligée en France, où la séparation académique entre linguistique et littérature reste stricte, et où la stylistique, discipline de concours, évite de se confondre tant avec la littérature qu'avec la linguistique.

Comment cependant un texte devient-il une œuvre ? Cette question engage notamment les relations entre linguistique et études littéraires. Comme elle en souligne les lacunes actuelles, des propositions nouvelles sont d'autant plus nécessaires que la linguistique et les études littéraires ont beaucoup à apprendre réciproquement, pour dépasser tant la grammaire que la philosophie du langage.

Comme la littérature est un art du langage, son étude requiert une connaissance des langues que ne peuvent assurer la philosophie du langage (pour laquelle les langues ne sont pas un objet de réflexion), ni la tradition grammaticale, qui se limite pour l'essentiel à la phrase et recherche des régularités générales qui ne peuvent caractéri-

ser ni les discours, ni les genres, ni *a fortiori* les œuvres. Or, depuis un demi-siècle, les études linguistiques ont connu une restriction théorique, allant vers la formalisation, entraînant une multiplication de modèles partiels (Richard Montague secontentant par exemple d'un 'fragment de l'anglais'), traitant exclusivement de problèmes qui favorisent l'équivoque formaliste, comme la quantification en sémantique ou les 'connecteurs' en pragmatique. Cela les éloigne des sciences sociales, que des programmes réductionnistes entendent par ailleurs délégitimer. On multiplie ainsi les anecdotes néo-darwiniennes sur l'origine du langage, alors que la notion même d'histoire de la langue n'est plus comprise et que l'accès aux textes en devient difficile : j'ai par exemple vu une directrice de recherche frapper d'agrammaticalité cette modeste phrase de Balzac **Il s'aspergea d'odeur*. Bien que les littéraires 'purs' et les linguistes 'purs' soient d'apparition récente, les cours de littérature pour les linguistes ne seraient pas à présent moins utiles que jadis les cours de linguistique pour les littéraires.

Pragmatique et analyse du discours répondent à leur manière à un besoin majeur : articuler la linguistique interne (centrée de fait sur la morphosyntaxe) et la linguistique externe (qui entend tenir compte de l'incidence des situations et des contextes), d'où l'essor de la sociolinguistique et plus généralement, au sein des sciences sociales, du sociologisme. Branche de la philosophie du langage définie dans le cadre du positivisme logique par Charles Morris et Rudolph Carnap, la pragmatique traite cependant, pour l'essentiel, de l'oral et du ' langage ordinaire ', non de l'écrit et des usages peu 'ordinaires' dont témoigne la littérature[6]. Les auteurs majeurs et les plus cités, John L. Austin, John R. Searle, Ludwig Wittgenstein, Paul Grice, n'abordent pas la littérature et n'évoquent ni les discours, ni les genres, ni les styles, et comme leur sémiotique reste dualiste, ils ne mobilisent aucune théorie de l'expression (ni phonétique, ni phonologie, ni prosodie, ni métrique).

Les théories linguistiques d'inspiration logico-grammaticale n'ont pas les moyens de faire apparaître de nouveaux observables, comme des corrélations inaperçues entre paliers de complexité ou entre niveaux de l'expression et du contenu, ni *a fortiori* de poser les problèmes esthétiques que soulèvent les œuvres. Tant la philosophie du langage, trop puissante, car elle ne se soucie guère des différences entre les langues, que la grammaire, qui même étendue à la macrosyntaxe, en reste à la gamme limitée de ses observables traditionnels, n'ont la capacité ni même le projet de rendre compte des œuvres en tant que telles. Bien au contraire, elles sont mises à profit pour réduire leur complexité incompréhensible, comme on le voit avec Austin à propos de la poésie ou avec Dominique Maingueneau dans son *Contre Saint-Proust*.

---

[6] Austin écrivait à propos de la plaisanterie et de la poésie : « there are parasitic uses of language, which are <not serious>, not the <full normal use> » (Austin 1962, 104). En effet, poursuit-il, la référence peut être suspendue (*suspended*) ; il n'y a pas de « standard perlocutionary act », ni de volonté de vous faire faire quoi que ce soit (*no attempt to make you do anything*), et la théorie instrumentale du langage comme le fonctionnalisme utilitariste qu'elle inspire se trouvent déplorablement invalidés. Strawson estimait pour sa part que les poèmes homériques sont oiseux (*spurious*), car on ne peut attribuer de valeurs de vérité à leurs énoncés.

Partie de la linguistique historique et comparée, une linguistique des œuvres n'a pas vocation à se constituer en discipline autonome, mais pourrait revêtir pour la littérature une fonction analogue à celle de la musicologie pour la musique.

Les parcours génétiques et interprétatifs ouvrent et organisent la profondeur spatiale et temporelle de l'espace culturel. Par exemple, pour comprendre tel passage de *Zone* d'Apollinaire, il faut passer par le *De ave Phenice* de Lactance (cf. Rastier 1989, II, chapitre 4). Rimbaud réécrit non seulement Baudelaire, mais Horace et Virgile, et reprend Ovide ; Breton ressuscite la Grande Déesse, etc. Indépendamment des références qui pourraient aujourd'hui paraître érudites, on retrouve la matière même de toute entreprise littéraire, faite de réécritures, de rivalités avec les anciens et les étrangers, de ruptures enfin. Aussi, l'espace culturel de la littérature est-il naturellement multilingue – la notion identitaire de littérature nationale a été créée au XIXe siècle par les nationalismes européens. Or les langues de culture sont transnationales et attirent des écrivains de toute nationalité, qui à bon droit rivalisent pour enrichir leur corpus. D'autre part, les écrivains ont accès, par leur connaissance des langues comme par les traductions, à l'espace de la littérature mondiale. Enfin, la génétique littéraire fait apparaître le multilinguisme des écrivains, dont les traductions et autotraductions témoignent aussi.

Comme la pragmatique et l'analyse du discours n'ont pas de perspective diachronique ni plurilinguistique, elles ne peuvent concevoir les œuvres littéraires, qui toutes empruntent peu ou prou aux anciens comme aux étrangers. Or l'intertexte, dans toute sa dimension culturelle, reste la médiation majeure entre la linguistique interne et la linguistique externe. Alors que la philosophie du langage a démembré le champ d'investigation en disciplines presque étanches (syntaxe/sémantique/pragmatique), il revint à des romanistes comme Eugenio Coseriu le mérite d'avoir créé un cadre théorique intégrant qui permette de penser le statut propre du 'langage' littéraire et de concevoir une linguistique des œuvres, alors que la notion même d'œuvre reste étrangère à la linguistique formelle.

2.2 Les genres

Si la question des genres est traditionnellement traitée par la poétique, le nom même de cette discipline évoque toujours la littérature ; cependant, l'ensemble des normes et des usages linguistiques, oraux et écrits, artistiques ou non, relève de ce qu'on pourrait appeler une *poétique généralisée*, section importante de la linguistique historique et comparée. Ne souhaitant pas répéter des propositions descriptives (cf. Rastier 1989, chapitre 3 ; 2001, chapitre 8), nous nous limiterons à évoquer brièvement l'incidence des développements de la poétique, entendue comme linguistique des genres, sur la conception même de la textualité.

Outre le genre, il convient de distinguer le *champ générique* et le *discours*. Soit en bref trois niveaux : les *discours* (ex. juridique *vs* littéraire *vs* scientifique), les *champs*

génériques (ex. théâtre, poésie, genres narratifs)[7], les *genres* proprement dits (ex. comédie, roman 'sérieux', roman policier, nouvelle, conte, récit de voyage). Les *sous-genres* (ex. roman par lettres) constituent un niveau encore subordonné. Les différences de statut épistémologique entre ces niveaux font qu'on ne peut, sauf simplification didactique, les représenter par une simple arborescence. En matière de genres, la linguistique doit élaborer ses propres catégories descriptives, car les disciplines voisines, études littéraires et philosophie, procèdent d'autres problématiques, et leur apport reste limité.

Les études littéraires contemporaines ont presque renoncé à réfléchir sur les genres, du moins en ce qui concerne la littérature moderne, où les théories romantiques et post-romantiques à la Barthes minimisent l'intérêt d'une étude des genres, en postulant que la Littérarité d'une part, le Style de l'autre, les transcendent. Soit, mais même chez les auteurs qui affirment se défier des genres, comme Le Clézio, les méthodes de la linguistique de corpus montrent que les œuvres démentent ces propos convenus (cf. Kastberg-Sjöblom 2002, 51-55). La philosophie du langage nous dira moins encore sur les genres, car elle reste transcendante aux langues et aux textes.

2.2.1 Le genre détermine la sémiosis textuelle. – La sémiosis textuelle est une sémiose, c'est-à-dire un acte d'expression et d'interprétation. En tant qu'acte, elle relève d'une praxéologie des performances linguistiques, et non d'une ontologie ou d'une théorie des représentations. La poétique décrit des normes en action, en tenant compte que ces normes sont, par définition, socialisées et que ces actions textuelles relèvent de pratiques sociales, dont elles constituent le niveau sémiotique.

Nous définissions jadis un genre, au plan sémantique, par l'interaction de quatre composantes[8] non hiérarchisées. Par *composantes*, on entend des ensembles de normes de même type : par exemple, celles de la thématique. Les composantes sémantiques s'articulent à des composantes de l'expression : sans préjuger que la 'pensée' choisisse son expression, comme si elle s'infusait dans le langage, on peut décrire des compromis entre des contraintes sémantiques et des contraintes au plan du signifiant. Un genre définit précisément un rapport normé entre le plan du signifiant et le plan du signifié au palier textuel : par exemple, dans le genre de l'article scientifique, le premier paragraphe, sur le plan du signifiant, correspond ordinairement à une introduction, sur le plan du signifié ; dans le genre de la nouvelle, il s'agit le plus souvent d'une description. Bref, les genres déterminent voire permettent les relations entre

[7] Un champ générique est un groupe de genres qui contrastent voire rivalisent dans un champ pratique : par exemple, au sein du discours littéraire, à l'époque classique, le champ générique du théâtre se divisait en farce, comédie, comédie héroïque et tragédie.

[8] Nommément, la thématique rend compte des thèmes, décrits comme des formes sémantiques (molécules sémiques) ; la dialectique étudie la succession des intervalles dans le temps textuel, comme les états qui y prennent place et les processus qui s'y déroulent ; la dialogique étudie les relations modales entre univers et entre mondes et rend compte de l'énonciation représentée ; et la tactique rend compte de la linéarité du signifié et de la disposition des unités textuelles. Ce modèle modulaire non-hiérarchique a été repris dans son principe par diverses grammaires du texte (cf. Adam 1992).

les deux plans du langage. On voit par exemple dans les brouillons des écrivains et des penseurs, qu'il s'agisse de Flaubert ou de Saussure, le texte d'abord inclassable se conformer peu à peu à un genre. Sans qu'il soit utile d'évoquer ici une illusoire optimalité, l'énonciation établit en outre un compromis entre les contraintes de la pratique (dans laquelle se définit le genre) et la situation (c'est-à-dire, au-delà du *hic et nunc*, la position historico-culturelle de l'énonciateur et de l'interprète).

La sémiosis limitée proposée par la langue aux paliers inférieurs, du mot à la phrase, ne devient effective que si elle est compatible avec les normes de genre, voire de style, qui assurent la sémiosis textuelle[9]. Aussi, la proposition grammaticale ne fait sens que dans une période, c'est-à-dire en tant que passage d'un texte. Enfin, la sémiosis des lexies est elle aussi déterminée par le genre ou le champ générique : par exemple, *amour* en poésie a fort peu de contextes communs avec *amour* dans le roman, et ces deux mots sont donc quasi-homonymes[10].

Outre les régimes de production et d'interprétation des textes, la sémiosis textuelle détermine le mode de mimésis. En règle générale, plus les rapports entre les deux plans du texte sont normés, plus son effet de réel empirique ou transcendant est intense, comme l'attestent les textes gnomiques ou religieux. L'interaction complexe entre plans du langage, en quoi consiste l'énonciation du texte comme performance sémiotique, ne se joue pas seulement dans une langue, mais aussi dans un genre. Ainsi, le genre reste un 'chaînon manquant' des modèles énonciatifs. Ses normes remplacent avantageusement les représentations mentales intermédiaires, et prototypes divers, car elles permettent des planifications de l'action énonciative compatibles avec la pratique et la situation. Il en va de même pour l'interprétation, car le texte doit être lu selon son genre : sa littéralité et même sa prononciation en dépendent (on ne lit pas un extrait de roman comme un poème).

On dit ordinairement qu'un texte appartient à un genre. Cette proposition mériterait d'être renversée : le genre appartient au texte, qui contient des indications de son genre (dans son titre, dans son support, mais aussi dans son lexique, dans son mode compositionnel, etc.). En conséquence, nous ne retiendrons pas les notions ordinaires de péritexte ni d'architexte, qui n'apportent rien à la compréhension de la textualité.

Le genre n'est ni une classe, ni un type, mais une lignée. Au sein de cette lignée, un texte compte, pour ainsi dire, des ancêtres, des rivaux, mais point de 'patrimoine

---

9 Le système de la langue, tel que le conçoivent généralement les linguistes, ne détermine pas la sémiosis textuelle et ne contraint la sémiosis qu'au palier de complexité le plus élémentaire, celui des morphèmes : au palier des morphèmes, la langue propose voire impose des appariements entre signifiant et signifié (ex. *re-* est itératif) ; mais les morphèmes n'ont pas en eux-mêmes de signification définie, ce pourquoi le (faux) problème de la référence n'est jamais posé à leur propos. Au palier immédiatement supérieur, celui des lexies, les mots sont déjà des unités 'de discours', car la mise en relation de leurs morphèmes est réglée par une syntaxe interne : leur signification et leur mode de sémiosis dépend déjà des relations contextuelles entre les morphèmes qui les composent.

10 Ainsi, dans le roman, *amour* a pour antonyme *mariage* ou *argent*. En poésie, point de mariage, ni d'argent. Pour une analyse en corpus dans la banque Frantext, cf. Bourion (2001, 42-45).

générique' qui lui serait transcendant. Il n'est pas une occurrence d'un genre, mais un moment dans une série de transmissions, dans une tradition faite de ruptures. C'est pourquoi d'autres parlent de 'traditions discursives'.

Comme tout usage modifie et configure potentiellement la langue, chaque texte atteste et modifie son genre. Mais ces perturbations locales ne sont sensibles qu'au sein d'une stabilité globale, et en retour le genre configure le texte. Ainsi le genre et le texte, en quelque sorte, s'interprètent-ils mutuellement.

Des résultats récents confirment l'incidence du genre sur les variations morphosyntaxiques (cf. Rastier 2011, chapitre 3). À partir d'un corpus de 2 500 textes complets classés par genres et discours et étiquetés par 251 types d'étiquettes, morphosyntaxiques pour la plupart, on a retrouvé et validé les différents niveaux de classification présentés ci-dessus, en utilisant des pourcentages calculés sur les étiquettes. On a conduit des analyses univariées pour qualifier les variations selon les catégories d'étiquettes, puis une analyse multivariée utilisant des méthodes de classification automatique. Les résultats, à affiner, mais cependant probants, confirment la corrélation entre les variables globales de genre, champ générique et discours d'une part, et d'autre part les variables morphosyntaxiques, locales par définition. Ainsi, les conditions d'application de la grammaire, censée représenter la langue dans sa pureté systématique, varient-elles selon les discours, champs génériques et genres.

Au plan sémantique, le genre règle pour une part importante les afférences socialement normées. Nous avions évoqué jadis ces instanciations de sèmes : elles dépendent d'une topique, et par là d'un genre ou d'un champ générique. Par exemple, en poésie baroque, un *trait* est luisant, car c'est une métaphore récurrente du regard. C'est sa redondance dans le corpus du champ générique qui permet d'inférer ce sème.

En tant qu'unité minimale d'analyse, le texte agit comme instance globale par rapport à ses éléments, mais aussi comme instance locale par rapport à son corpus. Par le biais notamment de son genre, le texte pointe sur son corpus et sélectionne en quelque sorte les autres textes qui permettent de l'interpréter. Puisque le genre n'est pas une instance abstraite, le rapport d'un texte à ses voisins n'est pas médié par une abstraction architextuelle, mais par les parcours intertextuels propres à leur genre commun. Le premier cercle du corpus, corpus nécessaire mais pas toujours suffisant, est ainsi constitué des textes du même genre. Le genre détermine ainsi un mode de régulation du contexte interne et externe (textualité et intertextualité). Il permet enfin au texte écrit de rester compréhensible, alors même que la situation qui lui a donné naissance, ou du moins prétexte, a disparu sans retour.

2.2.2 Linguistique(s) de la langue, de la parole et des normes. – La lecture des leçons et manuscrits de Saussure confirme que Bally a gommé l'apport de Saussure à la linguistique de la parole. Les deux linguistiques, celle de la langue et celle de la parole, sont restées séparées parce que la linguistique des normes n'a pas encore été construite. On pense à tort qu'il ne peut y avoir de science des normes : elle serait une

déontologie qui échapperait par son caractère relatif et conditionné à l'imaginaire logico-grammatical des règles, voire à l'imaginaire scientiste des lois.

Le rapport entre langue et parole est tantôt pensé comme un passage du virtuel à l'actuel, tantôt comme un passage des contraintes à une liberté, et l'on peine à concilier les virtualités impératives de la langue avec les libertés actuelles de la parole. En effet, de la langue, conçue abstraitement, à la parole, on ne passe pas seulement par des degrés de systématicité décroissants, mais aussi par des statuts épistémologiques divers.

Si l'on prend la mesure des diversités effectives des discours, champs génériques et genres, le noyau invariant qu'on peut appeler *langue* se réduit drastiquement à l'inventaire des morphèmes, à des contraintes comme la structure de la syllabe, la structure du syntagme, etc. ; par exemple, les lexèmes n'en font pas partie, car ils sont déjà des phénomènes de 'discours'[11].

Aucun texte n'est écrit seulement 'dans une langue' : il est écrit dans un genre, en tenant compte des contraintes d'une langue. D'ailleurs, l'analogie des pratiques et celle des genres qui en découle permettent la traduction voire tout simplement l'intercompréhension. On note certes des régularités transgénériques et transdiscursives. Par exemple, des domaines comme la littérature et les essais sont voisins : il arrive que les mêmes auteurs y transposent des thèmes comparables. Au palier morphosyntaxique et au plan de l'expression, ces régularités relèvent de la langue ; au plan sémantique, elles relèvent de l'idéologie ou de la doxa.

Alors que la morphosyntaxe reste pour l'essentiel affaire de règles, bien qu'elle ne soit aucunement indifférente aux normes, la sémantique est pour l'essentiel affaire de normes. Même si l'on définit, à bon droit, les règles comme des normes invétérées, il faut souligner les différences de leurs modes d'évolution diachronique. Le lien problématique entre langues et représentations collectives reste crucial : une langue contraint mais ne dicte pas une vision du monde, comme le prétendent, après certains linguistes nazis, certains culturologues russes d'aujourd'hui. Pour mieux saisir ce lien, il faut poursuivre le programme d'une poétique généralisée. L'étude des genres a manqué à l'école française d'Analyse du discours pour pouvoir lier effectivement la linguistique à l'étude des idéologies. Avec les méthodes de la linguistique de corpus, on dispose à présent de moyens nouveaux pour tester les hypothèses sur le rapport entre normes et règles, comme sur le rapport entre les deux plans du langage.

Le problème du genre dépasse enfin les sciences du langage, car l'on retrouve des problèmes analogues à propos des autres sémiotiques, qui connaissent elles aussi des genres. En outre, alors que les diverses sémiotiques (langage, musique, danses) sont ordinairement décrites comme des systèmes isolés, c'est dans certains genres plurisémiotiques, comme l'opéra, le cinéma, le site interactif, qu'elles trouvent des modes réglés d'interaction ; en d'autres termes, l'étude des genres commande celle de

[11] C'est pourquoi le lexique, du moins celui des lexies, n'appartient pas à la langue.

l'intersémioticité. Ainsi la poétique occupe-t-elle sans doute une place éminente au sein de la linguistique, mais aussi de la sémiotique.

2.2.3 La linguistique de corpus. — D'abord limitée à des questions d'attribution, la *philologie numérique* (cf. Rastier 2001, chapitre 2) s'est développée avec les corpus, pour traiter notamment de leur codage, au sein de la *Text Encoding Initiative* (TEI), en particulier. Elle donne à présent les moyens empiriques et expérimentaux de tester, valider et invalider des hypothèses, comme d'articuler les recherches nomothétiques (description des normes) et idiographiques (description des singularités). Appliquée à la littérature, elle s'attache à la typologie des genres et des discours, à la description de formes et de fonds sémantiques, au repérage de thèmes, à la caractérisation de leur évolution, à l'étude des corrélations jamais observées entre les plans du contenu et de l'expression. Épistémologie, méthodologie, descriptions empiriques et applications vont ici de pair et permettront d'affermir, dans le développement des humanités numériques, le statut propre des sciences de la culture.

3. Études médiévales

Après avoir traité de l'interface entre linguistique et littérature qu'est le domaine des genres, nous nous attaquons à présent aux études médiévales. Ce berceau de la romanistique au sens propre du mot forme un troisième domaine fédérateur entre linguistique et littérature ; il s'y ajoute les travaux de lexicographie historique. C'est Barbara Frank-Job, titulaire de philologie romane à l'Université de Bielefeld, qui va approfondir ce point. Voici sa contribution :

3.1 Tradition et objets d'étude de la philologie romane médiévale

Depuis sa constitution, la philologie romane médiévale est une 'interdiscipline' qui réunit les connaissances philologiques traditionnelles incluant la paléographie et la codicologie des manuscrits médiévaux, la critique textuelle, mais aussi la littérature, la linguistique historique et l'histoire. C'est donc elle qui représente peut-être le mieux l'idée de l'unité de la romanistique. Ceci vaut aussi pour sa conception de la Romania médiévale, unité à la fois linguistique et culturelle qui existait déjà comme telle pour les acteurs historiques. Brunetto Latini en témoigne dans la préface de son *Livre dou Trésor* : « Et se aucuns demandoit por quoi cis livres es *escrit en roumanç selonc le raison de France*, puisque nos somes italien, je diroie que c'est pour .ii. raisons, l'une ke nous somes en France, l'autre pour çou que la parleure est plus delitable et plus commune a tous langages [var. *gens*] »[12].

Le passage à l'écrit des langues romanes au Moyen Âge s'effectue dans une situation sociale et culturelle marquée aussi bien par la culture orale des personnes illettrées que par la culture littéraire latine du clergé. En s'appropriant successivement les acquis de l'écriture, les langues vernaculaires entrent dans le domaine culturel du

[12] Brunetto Latini, *Li Livre dou Tresor*, ca 1260-1266, cité d'après Carmody (1948, 18).

latin, où elles rencontrent un système de traditions discursives tant orales qu'écrites. Les pratiques communicatives dans lesquelles se placent les textes latins du Moyen Âge et dont ils sont, en même temps, l'expression, sont à peu près les mêmes dans les régions de langue romane et dans celles d'autres langues (Frank-Job / Selig 2015).

Le latin représente donc pour les auteurs des premiers documents romans le cadre de référence commun tant sur le plan grammatical – c'est pourquoi on appelle *grammatica* la langue latine au Moyen Âge – que sur le plan culturel. C'est pour cela que la philologie médiévale romane ne s'est jamais contentée de l'analyse de textes vernaculaires seuls, mais en a étudié aussi les cotextes et contextes latins. Un exemple très connu en est constitué par le cas des Serments de Strasbourg. Ainsi l'usage répété de techniques syntaxiques latines[13] dans le texte roman des Serments reflète-t-il l'ancrage de ce même texte dans les traditions discursives juridiques latines, en particulier les serments féodaux (cf. Gärtner / Holtus 1995, 121-122). Le choix du registre ou de la langue en fonction de la tradition discursive est également essentiel pour l'expansion d'innovations linguistiques. Un autre exemple en est fourni par la haute fréquence de spécificateurs nominaux dans les chartes privées (cf. Raible 1985) ; l'exemple le plus illustre en est celui des Serments de Strasbourg.

Quant aux traditions de textes, la philologie médiévale traditionnelle a eu de bons arguments pour ne pas réaliser une séparation disciplinaire entre littérature et linguistique. Comme Paul Zumthor (1972 ; 1983 ; 1986 ; 1987 ; 1990 ; 1997) nous l'a montré, le Moyen Âge ne connaît pas notre distinction moderne entre texte littéraire, texte instructif ou texte religieux, entre narration fictive et récit historique, mais il crée ses différenciations propres, dont la reconstruction constitue l'un des travaux les plus importants du philologue.

Ce qui caractérise la philologie médiévale traditionnelle est donc essentiellement le travail de recontextualisation des documents anciens, qui est faite à la base d'une analyse

- des contextes de transmission écrite (analyse codicologique et paléographique approfondie) ;
- du contexte socio-pragmatique (pratique communicative et tradition discursive du document) ;
- du rôle des traditions écrites préexistantes (le plus souvent latines) et
- des pratiques plurilingues de l'époque[14].

Bien sûr, un tel programme de recherches sans bornes linguistiques et hautement interdisciplinaire est difficilement réalisable à l'heure actuelle, où les Instituts de

[13] Cf. Raible (1994, 117) : « Whereas the whole morphology of the oaths is already Old French, all the important syntactic hinges are borrowed from the underlying Latin original worked out by the royal chancellery ». – Pour une analyse exhaustive des phénomènes de 'relatinisation' dans les Serments de Strasbourg (cf. Raible 1996, 122-123).

[14] Tous ces aspects ont été pris en compte dans la conception de l'*Inventaire systématique des premiers documents des langues romanes* (Frank / Hartmann 1997) ; cf. aussi Selig (2006 ; 2008). Pour l'aspect des pratiques multilingues dans la Romania médiévale, cf. Frank-Job (2010) et Frank-Job / Selig (à paraître).

romanistique même en Allemagne se différencient de plus en plus en rassemblant des spécialistes d'une seule langue romane, d'un seul siècle, voire d'une seule méthode de recherche. De plus, les chaires de philologie médiévale semblent menacées partout en Europe, l'histoire de la langue ne figurant même souvent plus dans les programmes de formation universitaire.

Et pourtant, la philologie romane médiévale existe toujours, et elle est même de plus en plus présente, grâce aux possibilités techniques des nouveaux médias qu'elle a adoptées pour constituer de grands corpus de textes anciens. Ceci vaut tant pour les langues vernaculaires[15] que pour les textes latins du Moyen Âge français : et les chartes et les autres types de textes latins sont aujourd'hui accessibles dans des éditions électroniques[16]. La constitution et l'annotation de ces corpus ont été soutenues par les institutions nationales ou régionales respectives, car on a compris qu'elles servent à la sauvegarde du patrimoine culturel. En s'ouvrant à des techniques de travail nouvelles et à des méthodes de recherche nouvelles – en intégrant au travail interprétatif et descriptif traditionnel des méthodes quantitatives déjà bien établies dans les sciences empiriques voisines –, la philologie romane actuelle peut donc regagner du terrain sans être forcée de nier ses traditions.

3.2 Le travail à base de grands corpus: l'exemple du projet Computational Historical Semantics

Le groupe de recherche dont je fais partie représente une coopération de romanistes, historiens et informaticiens de quatre universités allemandes[17]. Notre groupe se propose de faire des analyses qualitatives et quantitatives sur un ensemble de collections de textes du latin tardif. Le but des linguistes est de trouver les contextes qui ont favorisé la diffusion de changements grammaticaux et sémantiques dans les textes écrits entre le IIe et le XIIe siècle. En même temps, nos recherchent doivent contribuer à la perfection d'un portail internet déjà existant et qui met à la disposition des philologues les instruments électroniques nécessaires pour leurs analyses respectives sur grand corpus (*e-humanities desktop*, cf. Gleim et al. 2009 ; 2010).

---

15 Cf. le 'Corpus de la littérature médiévale' des Classiques Garnier, le projet des 'Plus anciens documents linguistiques de la France' de l'Université de Zurich (www.mediaevistik.uzh.ch/docling/), la 'Base de Français Médiéval' de l'Université de Lyon (bfm.ens-lyon.fr) ou encore les 'Anglo-Norman Source Texts' des universités d'Aberystwyth et de Swansea (www.anglo-norman.net/texts).

16 Cf. TELMA 'Chartes originales antérieures à 1121 conservées en France' (www.cn-telma.fr/originaux/), la 'Patrologia Latina Database' (pld.chadwyck.co.uk) ou encore les 'Monumenta Germaniae Historica' électroniques (http://www.mgh.de/dmgh).

17 Il s'agit des universités de Bielefeld (Barbara Frank-Job, linguistique des langues romanes), Francfort (Bernhard Jussen, histoire, et Alexander Mehler, informatique), Tübingen (Peter Koch, linguistique des langues romanes) et Ratisbonne (Maria Selig, linguistique des langues romanes). Pour une description du projet *Computational Historical Semantics* (CompHistSem), qui se réalise entre 2013 et 2015, cf. http://www.comphistsem.org/project.htm.

Notre travail actuel consiste en une analyse semi-automatisée des occurrences et cooccurrences d'innovations langagières (telles que l'apparition de verbes auxiliaires) dans des corpus électroniques du latin (dont la *Patrologia latina* et des collections de chartes). Par ces analyses, nous cherchons une vérification ou une falsification qualitative et quantitative de maintes hypothèses anciennes sur le développement du latin aux langues romanes (Frank-Job 2011 ; Mehler et al. 2011).

3.3 Conclusion

Pour effectuer un tel programme de recherches, une philologie médiéviste moderne doit donc d'une part se fonder sur une conception traditionnelle de la romanistique qui embrasse et littérature et linguistique ainsi que tous les idiomes romans pratiqués dans une région précise et dans une époque précise, pour ouvrir à un travail interdisciplinaire avec les historiens et les informaticiens de la langue (les représentants de ce que l'on nomme en allemand *geisteswissenschaftliche Fachinformatik*), qui contribuent tous à la constitution et à l'analyse de grands corpus. Les évolutions de ces dernières décennies me portent donc à voir le futur de l'unité romaniste – en ce qui concerne le champ de la philologie romane médiéviste – plutôt positif, pourvu que cette discipline s'ouvre au travail interdisciplinaire.

4. Narratologie – analyse de récit – récit oral – récit conversationnel – reconstruction narrative : la diversité des approches théoriques et méthodologiques

C'est à Elisabeth Gülich, prédécesseure de Barbara Frank-Job sur la chaire de philologie romane à Bielefeld, que revenait le rôle de montrer, d'une part, que, sur la base des acquis d'un savoir de philologue romaniste, il est possible de transgresser les frontières entre langues romanes : les phénomènes de production orale sont assez comparables entre différentes langues, et que, d'autre part, fort d'un savoir narratologique, on peut beaucoup mieux juger des particularités de la narration de personnes épileptiques, pour ne citer que deux exemples. Voici sa contribution :

4.1 Narratologie

Parmi les sujets qui nous ont été proposés, j'ai choisi la narratologie, qui semble être un domaine idéal pour la thématique de cette table ronde, car elle appartient – ou : on peut la rattacher – aux trois disciplines que nous discutons ici sous l'aspect de l'unité de la romanistique : à la littérature, à la philologie et à la linguistique. La narratologie pourrait ainsi ouvrir une perspective sur cette unité ou du moins aider à préciser ce que l'on veut entendre par cette notion.

Mais qu'est-ce que la narratologie ? Tzvetan Todorov introduit ce terme dans sa *Grammaire du Décaméron* (Todorov 1969, 10) : « cet ouvrage relève d'une science qui n'existe pas encore, disons la NARRATOLOGIE, la science du récit ». À l'époque,

il avait raison : dans le *Dictionnaire encyclopédique des sciences du langage* qu'il a publié lui-même avec Oswald Ducrot (Ducrot / Todorov 1972), un article *narratologie* fait effectivement défaut. Cette notion ne figure même pas dans l'index des termes, où on ne trouve que *analyse narrative*, terme qui est mentionné dans l'article *texte* (un des articles faisant partie de la section « Les concepts descriptifs »). Cependant, Gérard Genette, également en 1972, utilise cette notion dans l'avant-propos de son *Discours du récit* : il parle de « la théorie du récit ou *narratologie* » (Genette 1972, 68). Dans le *Nouveau dictionnaire encyclopédique des sciences du langage* (Ducrot / Schaeffer 1995), la narratologie existe : on y trouve un article *narratologie*, écrit par Marielle Abrioux (1995), qui résume l'histoire et le développement de cette science, dont l'existence n'est plus mise en question. Dans le *Dictionnaire d'analyse du discours* par Charaudeau / Maingueneau (2002), il y a un article *récit* de Jean-Michel Adam, dans lequel l'auteur constate deux défauts des théories narratologiques : « trop exclusivement littéraires » ou « beaucoup trop générales ». Il propose de reprendre la distinction de Genette entre 'acte de narration', 'histoire racontée' et 'mise en texte'. Il plaide donc pour une différenciation terminologique et un « concept de *narration* […] réintégré dans le phénomène linguistique plus large de l'énonciation » (Adam 2002, 484-485). Un dernier exemple de la terminologie au niveau des encyclopédies : dans une encyclopédie allemande des études de français, *Handbuch Französisch* (Kolboom / Kotschi / Reichel 2008), on ne trouve ni *narratologie*, ni *récit*, mais *discours narratif*. Et on trouve ce terme dans un contexte tout à fait différent, à savoir dans un article de Laurent Filliettaz (2008) consacré aux types de discours, qui fait partie de la section : *Le français dans l'interaction verbale* (« Das Französische in der verbalen Interaktion »). Dans cet article, il n'est cependant guère question d'interaction ; l'auteur reste au contraire dans une tradition qui définit le récit comme un texte monologique et qui privilégie la description des structures narratives, en l'occurrence d'après le modèle de Jean-Michel Adam. D'autre part, l'auteur parle déjà de variations selon le contexte situationnel et de négociation des formes narratives entre les interlocuteurs (Filliettaz 2008, 335).

En conclusion, si la narratologie existe et si nous nous fions à la façon dont les encyclopédies en tiennent compte, c'est une science du récit écrit, en principe monologique, même si les structuralistes français se sont déjà intéressés au récit oral (littéraire). Aujourd'hui le concept de narratologie me semble être utilisé plutôt dans le contexte des études littéraires qu'en linguistique. Ainsi son importance pour l'unité de la romanistique me paraît-elle assez réduite.

J'aimerais proposer une autre approche, linguistique, qui tient compte du fait que le récit est quasiment omniprésent dans les interactions quotidiennes, familières ou professionnelles. C'est une pratique communicative qui remplit des fonctions multiples dans des contextes divers. Elle est située dans un contexte conversationnel, et on doit l'analyser systématiquement en tenant compte de ce contexte. Ce point de vue a déjà été développé par Lorenza Mondada et moi-même dans un article du *Lexikon der Romanistischen Linguistik* : c'était l'article 48 sur l'analyse conversationnelle, dont

un sous-chapitre est consacré aux séquences narratives (Gülich / Mondada 2001). Cette approche a été reprise plus tard dans un livre (Gülich / Mondada 2008).

4.2 Réflexions théoriques et méthodologiques : l'approche de l'analyse conversationnelle

Le système de l'alternance des tours de parole sert de point de départ à Sacks (1992) pour étudier la production de tours 'longs', par exemple des séquences narratives émergeant dans les conversations. L'organisation conversationnelle devient ainsi pertinente pour la production d'un récit. Cette réorientation de l'analyse narrative se distingue par les caractéristiques suivantes d'une narratologie au sens de "science du récit littéraire" :

- La reconstruction narrative est considérée comme une activité conversationnelle. D'après les sociologues allemands Bergmann / Luckmann (1995), ce qu'on a l'habitude d'appeler *récit* compte parmi les genres reconstructifs (*reconstructive genres*).
- La reconstruction narrative est une 'méthode' au sens ethnométhodologique du terme, faisant partie de la méthodologie quotidienne dont disposent les membres d'une société ou d'un groupe social pour résoudre leurs tâches communicatives ; en disant *reconstruction narrative*, je mets l'accent sur l'activité et le processus de son accomplissement et non sur son résultat ou sur le produit : le récit.
- Dans l'analyse de ce processus ou de cet accomplissement, il faut décrire l'organisation séquentielle et prendre en considération non seulement les ressources verbales, mais toutes les ressources communicatives auxquelles les interactants ont recours.
- Le résultat, le 'récit', est une co-construction, c'est-à-dire que l'interlocuteur (ou les interlocuteurs) y contribue(nt) aussi ; ce qui est pertinent, ce qui est 'racontable', est également une construction interactive. La 'racontabilité' n'est donc plus considérée (ou : pas seulement) comme une qualité du texte narratif.

Cette conception du récit ou de l'activité narrative a des conséquences pour la pratique de l'analyse :

- il faut décrire le déroulement de la conversation, le développement thématique, le déclenchement (auto- ou hétéro- déclenchement) de l'activité narrative ;
- il faut tenir compte des activités des interlocuteurs, de la co-construction du récit ;
- il faut inclure dans l'analyse les tentatives de raconter même si elles échouent et les fragments de récits.

4.3 Perspectives concernant l'unité de la romanistique

L'étude du récit oral dans le cadre de l'analyse conversationnelle que je viens d'esquisser peut-elle contribuer à l'unité de la romanistique ? Spontanément on dirait sans doute : non, parce que le cadre ne semble pas être approprié par exemple à l'analyse de textes littéraires : on ne peut pas retracer le processus de la reconstruction narrative d'un événement, on ne peut pas tenir compte des aspects interactifs, de la co-construction d'un récit – et on pourrait avancer d'autres contre-arguments du même type. Mais d'un autre côté, on trouve souvent dans les textes littéraires

la description – ou mieux encore : le récit – de ce processus et de cette interaction. L'auteur/le narrateur raconte la manière dont les personnages interagissent, comment ils parlent entre eux et comment ils se racontent mutuellement leurs expériences. Ainsi l'étude de la reconstruction narrative d'un événement dans la communication orale peut-elle tout de même servir de modèle pour l'analyse d'un texte littéraire. En disant cela, je pense à un double apprentissage : l'analyse littéraire pourrait partager certaines techniques avec l'analyse linguistique, et les linguistes qui s'intéressent à la communication pourraient beaucoup apprendre des textes littéraires, par exemple des conversations dans le Salon Verdurin, mises en scène dans *Un amour de Swann* de Marcel Proust (cf. Gülich 1990).

D'une façon plus générale, l'étude de l'oral pourrait être un élément très fort, même constitutif de l'unité de la romanistique : dès que l'on se place au niveau de la communication orale, on trouve de nombreux procédés communicatifs ou interactifs typiques de l'oralité qui se prêtent parfaitement aux études comparatives entre les langues romanes. Les tâches communicatives des membres d'une société sont largement les mêmes ou se ressemblent beaucoup au niveau de l'organisation et de la structuration de l'interaction, par exemple des tâches comme l'alternance des locuteurs, ou l'organisation de la production verbale spontanée avec les différentes techniques qu'on utilise pour surveiller sa production, revenir en arrière (auto-corrections, ruptures), les changements de construction ou encore les différents procédés de formulation ou de reformulation.

Aujourd'hui, les recherches sur l'oral profitent, elles aussi, des nouvelles techniques de documentation (enregistrements audio et vidéo, transcription) et de travail sur grands corpus numérisés. Ces techniques pourraient aider les romanistes à reprendre ou à poursuivre la tradition des études comparées qui est propre à la romanistique. Il me semble par exemple que dans le domaine des marqueurs discursifs, l'intérêt pour l'aspect comparatif se développe et se répand (cf. par exemple Drescher / Frank-Job 2006 et Rodríguez Somolinos 2011). Et c'est un exemple parmi d'autres.

4.4 Un exemple tiré de de mes recherches actuelles

Je travaille depuis quelques années dans le domaine de la communication médicale, plus précisément, sur les entretiens médecin-patient. C'est une coopération avec un neurologue, Martin Schöndienst, spécialisé dans le domaine de l'épilepsie. Pour comprendre l'intérêt de cette recherche, il est utile de se faire une idée de la situation de départ : le diagnostic dans ce domaine est extrêmement difficile. Les crises épileptiques (causées par une lésion dans le cerveau) et les crises non-épileptiques (qu'on appelle souvent psychogènes) sont difficiles à distinguer sous plusieurs aspects. Un patient souffrant de crises non-épileptiques met en moyenne sept ans avant de recevoir le diagnostic exact.

Mon partenaire dans cette recherche, un épileptologue expérimenté, avait observé que les descriptions que les patients donnent de leurs crises étaient très variées – au point qu'il formulait l'hypothèse qu'il y a un rapport entre type de description

et type de crise. Nous avons enregistré et transcrit une centaine d'entretiens, dont l'analyse a révélé différents modèles de description chez les patients : différents dans les stratégies de formulation, le recours aux métaphores, les types de reconstruction narrative des crises, etc., ce qui nous a permis de confirmer l'hypothèse de départ. Le résultat de ce travail est donc un moyen de diagnostic supplémentaire (cf. pour une présentation en français : Furchner & Gülich 2002 ; Gülich 2010).

Comme c'est un travail sur des corpus allemands, on pourrait penser que cela n'a rien à voir avec la romanistique. Cependant, nos observations et nos résultats ne semblent pas se limiter à l'analyse d'entretiens avec des patients allemands. Un neurologue à Sheffield, Markus Reuber, a refait la même recherche avec des patients anglais – et ses recherches ont abouti au même résultat (pour une présentation en français, cf. Reuber *et al*. 2014). En Italie, à Milan, il y a un groupe d'épileptologues qui a également pu vérifier certaines de nos observations (Cornaggia *et al*. 2012). En français, j'ai trouvé des exemples intéressants dans une émission télévisée, mais je n'ai pas de corpus qui permettrait une étude sérieuse. Si j'en avais un, je n'hésiterais pas à me mettre au travail.

Conclusion : l'étude du récit oral dans un contexte professionnel permet aussi, à condition d'adopter une orientation conversationnelle ou interactive, une orientation vers d'autres disciplines. Et l'interdisciplinarité est un travail particulièrement approprié aux romanistes.

5. L'apport fédérateur des humanités numériques

Nous disposons aujourd'hui, grâce à la numérisation, de grands corpus de textes qui représentent un défi pour la linguistique (*Text Encoding Initiative* – TEI ; linguistique textuelle) de même que pour les études littéraires et comparatistes. On avait prévu un intervenant spécialement pour ce domaine. Comme notre spécialiste a eu un contretemps, c'est Wolfgang Raible, professeur émérite de philologie romane et de linguistique générale à l'Université de Fribourg-en-Brisgau et président de cette table ronde, qui s'est chargé d'approfondir ce sujet. Voici sa contribution :

J'aimerais profiter de l'occasion pour présenter, à titre d'exemple, le travail d'une de mes doctorantes, Susanne Mocken, terminé avec grand succès fin 2012 : *Informatik erschließt Literatur. Der 'discours citant' in der französischen Romanliteratur zwischen 1750 und 1920* (Mocken 2013). Cette analyse englobe 70 romans avec environ 24 500 pages imprimées, donc un corpus vraiment grand, voire gigantesque. Le discours citant (et son importance) est un phénomène qui peut être illustré par la citation suivante, provenant du *Colonel Chabert* de Balzac :

> – « Oh ! » *répondit froidement le colonel en relevant la tête par un mouvement de fierté*, « si je succombe, je saurai mourir, mais en compagnie. » Là, le vieillard avait disparu. Les yeux de l'homme énergique brillaient rallumés aux feux du désir et de la vengeance.

Le travail de Susanne Mocken est une analyse tant qualitative que quantitative. Analyse qualitative veut dire qu'il faut analyser la structure syntaxique avec un instrument comme le *tree-tagger* qui marque forcément aussi les parties du discours. Cela permet de distinguer toute une série de types de discours introducteur (*dit-il* + gérondif, + adverbe, etc., au total une dizaine). Le logiciel créé par la thésarde comporte un menu avec une série d'options : pour un texte donné, on peut se faire donner une synopse de tous les tours tout court, de tous les tours après un retour à la ligne, etc. Dans le cas de *Madame Bovary*, par exemple, cette dernière option nous livrerait les 966 tours observés, qu'on pourrait étudier, le cas échéant, l'un après l'autre, à l'aide d'une longue liste imprimée.

Analysant les 70 romans en question, on constatera qu'on peut distinguer plusieurs phases. Dans un premier temps, entre Voltaire et Bernardin de Saint-Pierre, il n'y a que peu de dialogues, tous très stéréotypés, donc avec peu de variation, le nombre des verbes communicatifs lui-aussi étant très restreint. Dans une deuxième phase, qui va de *Han d'Islande* de Victor Hugo jusqu'à Flaubert, le nombre de verbes communicatifs et l'envergure du discours citant augmentent considérablement. Un cas extrême est le *Juif errant* d'Eugène Sue, où nous trouvons, tout au long du texte, 85 types différents de verbes communicatifs. En même temps, le discours citant comporte des informations sur la qualité de la voix, l'expression du visage, les gestes, les mouvements du corps, le regard : voir l'exemple cité plus haut. Lors d'une troisième étape, qui va de Flaubert au Zola tardif, le discours direct perd de l'importance (c'est entre autres à cause de l'avènement du style indirect libre). À la fin de l'époque qui a été analysée par Susanne Mocken, Proust est un écrivain plutôt traditionnel (mais, comme vient de nous dire Elisabeth Gülich, hautement intéressant pour l'étude linguistique de la conversation), tandis que Colette aime encore une fois la variation, avec 67 types différents de verbes communicatifs dans le seul texte relativement bref de son ouvrage *Chéri*. À partir de la deuxième phase, on observe un énorme essor dans le nombre des verbes qu'on peut utiliser comme verbes communicatifs : le français en possède maintenant plus de 450 (donc des verbes comme *balbutier*, *bégayer*, *chuchoter*, *glapir*, *grommeler*, *hurler*, *jurer*, *murmurer*, *nasiller*, *souffler*, *soupirer* et *vociférer*).

Outre la distinction de différentes phases dans l'histoire littéraire du roman, on peut observer, dans le genre du roman, un changement profond de la conceptualisation, par les écrivains, d'une situation de communication et de ses aspects pragmatiques – voir encore une fois l'exemple cité de Balzac. C'est là que se trouve le point de rencontre entre une analyse comme celle faite par Susanne Mocken et la 'reconstruction narrative' évoquée tout à l'heure par Elisabeth Gülich.

On observera en outre, en ami de la littérature comparative, que cette conceptualisation nouvelle d'une situation de communication n'est pas un phénomène français, mais plutôt européen. Il a été découvert et bien décrit, pour des auteurs allemands et russes (des XIXe et XXe siècles), par les formalistes russes (Ejchenbaum, Vinogradov, Tynjanov), qui lui ont donné le nom de *skaz*, technique dont le but est d'engendrer une oralité fictionnelle.

Bref : l'exemple du travail de Susanne Mocken fait voir combien une telle analyse, et forcément les humanités numériques, peuvent intéresser tant l'historien et le théoricien de la littérature que le comparatiste, le linguiste, le lexicologue et même celle ou celui qui s'intéresse à la psychologie de la perception.

6. Conclusion

En guise de conclusion, on aura vu que, pour la survie de la conception d'une romanistique intégrale, à savoir une romanistique comprenant les langues et littératures romanes, un rôle central revient aux humanités numériques. Tous les participants de cette table ronde, Margarita Borreguero Zuloaga, François Rastier, Barbara Frank-Job, Elisabeth Gülich et Wolfgang Raible, ont souligné, d'une façon ou d'une autre, l'importance cruciale de cet outil de travail. Il nous ouvre, entre autres, la chance de ne plus être trop submergés par le nombre immense de données, de textes, de détails. En même temps, nous profitons des grand corpus numérisés de textes qui existent déjà et dont le nombre va croissant.

Université de Fribourg-en-Brisgau	Wolfgang RAIBLE (président)
Université Complutense de Madrid	Margarita BORREGUERO ZULOAGA
Université de Bielefeld	Barbara FRANK-JOB
Université de Bielefeld	Elisabeth GÜLICH
CNRS	François RASTIER

Références bibliographiques

Abrioux, Marielle, 1995. «Narratologie», in : Ducrot, Oswald / Schaeffer, Jean-Marie (ed.), *Nouveau dictionnaire encyclopédique des sciences du langage*, Paris, Seuil, 191-201.

Adam, Jean-Michel, 1992. *Les textes : types et prototypes*, Paris, Nathan.

Adam, Jean-Michel, 2002. «Récit», in : Charaudeau, Patrick / Maingueneau, Dominique (ed.), *Dictionnaire d'analyse du discours*, Paris, Seuil, 484-487.

Andersen, Hanne Leth / Nølke, Henning (ed.), 2002. *Macro-syntaxe et macro-sémantique. Actes du colloque international d'Århus, 17-19 mai 2001*, Berne, Lang.

Austin, John, 1962. *How to do things with words*, Oxford, Oxford University Press.

Bazzanella, Carla / Bosco, Cristina / Tini Brunozzi, Federica / Gili Fivela, Barbara / Miecznikowski, Johanna, 2008. «Polifunzionalità dei segnali discorsivi, sviluppo conversazionale e ruolo dei tratti fonetici e fonologici», in : Pettorino, Massimo / Giannini, Antonella / Vallone, Marianna / Savy, Renata (ed.), *La comunicazione parlata. Atti del congresso internazionale. Napoli 23-25 febbraio 2006*, Naples, Liguori, vol. 2, 934-963.

Bazzanella, Carla / Borreguero Zuloaga, Margarita, 2011. «*Allora* e *entonces* : problemi teorici e dati empirici». *Oslo Studies in Language* 3/1, 7-45.

Bazzanella, Carla / Garcea, Alessandro / Bosco, Cristina / Gili Fivela, Barbara / Miecznikowski, Johanna / Tini Brunozzi, Francesca, 2007. «Italian *allora*, French *alors* : functions, convergences, and divergences». *Catalan Journal of Linguistics, Special issue : Contrastive Perspectives on Discourse Markers* 6, 9-30.

Bergmann, Jörg R. / Luckmann, Thomas, 1995. «Reconstructive genres of everyday communication», in : Quasthoff, Uta M. (ed.). *Aspects of oral communication*, Berlin/New York, De Gruyter, 289-304.

Bernárdez, Enrique, 1982. *Lingüística del texto*, Madrid, Espasa-Calpe.

Blanche-Benveniste, Claire *et al.*, 1990. *Le français parlé : étude grammaticale*, Paris, Éditions du CNRS.

Borreguero Zuloaga, Margarita, 2006. «La lingüística textual alemana : nacimiento y primeros desarrollos de una nueva disciplina lingüística (1960–1975)», in : Roldán Pérez, Antonio (ed.), *Caminos actuales de la Historiografía Lingüística. Actas del V Congreso Internacial de la SEHL*, Murcia, Universidad de Murcia, 279-301.

Borreguero Zuloaga, Margarita, 2012. «Análisis del discurso. La consolidación de las funciones discursivas de los marcadores en el s. XIX : el caso de *entonces*», in : Zamorano Aguilar, Alfonso (ed.), *Reflexión lingüística y lengua en la España del siglo XIX : marcos, panoramas y nuevas aportaciones*, Munich, Lincom, 301-337.

Bosco, Cristina / Bazzanella, Carla, 2005. «Corpus linguistics and the modal shift in Old and Present-Day Italian : Temporal pragmatic markers and the case of 'allora'». In : Pusch, Claus D. / Raible, Wolfgang (ed.), *Corpora and Historical Linguistics*, Tübingen, Narr, 443-453.

Bourion, Évelyne, 2001. *L'aide à l'interprétation des textes électroniques*, thèse, Nancy, Université de Nancy II.

Brinton, Laurel J., 1996. *Pragmatic Markers in English. Grammaticalization and Discourse Functions*, Berlin, De Gruyter.

Briz, Antonio / Grupo Val.Es.Co., 2002. *Corpus de español coloquial*, Madrid, Arco Libros, Oralia. Anejos 2.

Briz, Antonio / Grupo Val.Es.Co., 2003. «Un sistema de unidades para el estudio del lenguaje coloquial». *Oralia* 6, 7-61.

Carmody, Francis J. (1948). *Li livres dou tresor*, Berkeley/Los Angeles, University of California Press.

Charaudeau, Patrick / Maingueneau, Dominique (ed.), 2002. *Dictionnaire d'analyse du discours*, Paris, Seuil.

Chomsky, Noam, 1965. *Aspects of the Theory of Syntax*, Cambridge, MIT Press.

Cornaggia, Cesare Maria *et al.*, 2012. «Conversation analysis in the differential diagnosis of Italian patients with epileptic or psychogenic non-epileptic seizures. A blind prospective study», *Epilepsy and Behavior* 25, 598-604.

Cortés, Luis / Camacho, Maria Matilde, 2005. *Unidades de segmentación y marcadores del discurso*, Madrid, Arco Libros.

Cresti, Emanuela, 2000. *Corpus di italiano parlato*, Florence, Accademia della Crusca.

Cresti, Emanuela / Moneglia, Massimo, 2005. *C-Oral-Rom. Integrated Reference Corpora for Spoken Romance Language*, Amsterdam, Benjamins.

Dijk, Teun A. van, 1972. *Some Aspects of Text Grammars. A Study in Theoretical Linguistics and Poetics*, La Haye, Mouton.

Dijk, Teun A. van, 1977. *Text and Context. Explorations in the Pragmatics and Semantics of Discourse*, Londres, Longman.

Drescher, Martina / Frank-Job, Barbara (ed.), 2006. *Les marqueurs discursifs dans les langues romanes. Approches théoriques et méthodologiques*, Frankfort, Lang.

Ducrot, Oswald / Schaeffer, Jean-Marie (ed), 1995. *Nouveau dictionnaire encyclopédique des sciences du langage*, Paris, Seuil.

Ducrot, Oswald / Todorov, Tzvetan, 1972. *Dictionnaire encyclopédique des sciences du langage*, Paris, Seuil.

Équipe DELIC, 2004. «Présentation du Corpus de référence du français parlé», *Recherches sur le français parlé* 18, 11-42.

Estellés, María / Pons, Salvador, 2014. «Absolute initial position», in : Pons, Salvador (ed.), *Discourse Segmentation in Romance Languages*, Amsterdam, Benjamins, 121-155.

Ferrari, Angela *et al*., 2008. *L'interfaccia lingua-testo. Natura e funzioni dell'articolazione informativa dell'enunciato*, Alexandrie, Edizioni dell'Orso.

Filliettaz, Laurent, 2008. «Diskurstypen», in : Kolboom, Ingo / Kotschi, Thomas / Reichel, Edward (ed.), *Handbuch Französisch. Sprache, Literatur, Kultur, Gesellschaft ; für Studium, Lehre, Praxis*, Berlin, Schmidt, 332-340.

Frank, Barbara / Hartmann, Jörg, 1997. *Inventaire systématique des premiers documents des langues romanes*, Tübingen, Narr, ScriptOralia 100, I-V.

Frank-Job, Barbara, 2010. «Traditions discursives et élaboration écrite des langues romanes au Moyen Âge», *Aemilianense : Revista Internacional sobre la Génesis y los Orígenes Históricos de las Lenguas Romances* 2, 13-36.

Frank-Job, Barbara, 2011. «Zu den Leistungen eines netzwerkanalytischen Ansatzes für die empirische Linguistik», in : Dessì Schmid, Sarah / Detges, Ulrich / Gévaudan, Paul / Mihatsch, Wiltrud / Waltereit, Richard (ed.), *Rahmen des Sprechens. Beiträge zu Valenztheorie, Varietätenlinguistik, Kreolistik, Kognitiver und Historischer Semantik*, Tübingen, Narr Francke Attempto, 3-16.

Frank-Job, Barbara / Selig, Maria (à paraître). «Early Evidences and Sources», in : Ledgeway, Adam / Maiden, Martin (ed.), *Oxford Handbook of The Romance Languages*, Oxford, Oxford University Press.

Furchner, Ingrid / Gülich, Elisabeth, 2002. «L'expertise des patients dans l'élaboration d'un diagnostic médical – Analyse linguistique d'entretiens avec des patients souffrant de crises épileptiques ou non-épileptiques», *Bulletin suisse de linguistique appliquée* 74, 83-107.

Gärtner, Kurt / Holtus, Günter, 1995. « Die erste deutschfranzösische 'Parallelurkunde'. Zur Überlieferung und Sprache der Straßburger Eide», in : Gärtner, Kurt / Holtus, Günter (ed.), *Beiträge zum Sprachkontakt und zu den Urkundensprachen zwischen Maas und Rhein*, Trèves, Kliomedia, 97-127.

Genette, Gérard, 1972. «Discours du récit», in : *Figures III*, Paris, Seuil, 67-282.

Giacalone Ramat, Anna, 2001. *Verso l'italiano*, Rome, Carocci.

Gleim, Rüdiger / Mehler, Alexander / Waltinger, Uli / Menke, Peter, 2009. «eHumanities Desktop - an extensible online system for corpus management and analysis», in : *5th Corpus Linguistics Conference*, Liverpool, University of Liverpool, http://pub.uni-bielefeld.de/publication/1918907.

Gleim, Rüdiger / Warner, Paul / Mehler, Alexander, 2010. «eHumanities Desktop - an architecture for flexible annotation in iconographic research», in : *Proceedings of the 6th International Conference on Web Information Systems and Technologies (WEBIST '10), April 7-10 (2010)*, Valence, http://www.researchgate.net/publication/220724277_eHumanities_Desktop_-_An_Architecture_for_Flexible_Annotation_in_Iconographic_Research.

Gülich, Elisabeth, 1990. «Erzählte Gespräche in Marcel Prousts ‹Un amour de Swann›», *Zeitschrift für französische Sprache und Literatur* 100, 89-108.

Gülich, Elisabeth, 2010. «Le rôle du corpus dans l'élaboration pluridisciplinaire d'un instrument de diagnostic linguistique : l'exemple de l'épilepsie», *Pratiques* 147/148, 173-197.

Gülich, Elisabeth / Mondada, Lorenza, 2001. «Analyse conversationnelle», in : Holtus, Günter / Metzeltin, Michael / Schmitt, Christian (ed.), *Lexikon der Romanistischen Linguistik* (LRL), Tübingen, Niemeyer, vol. I/2, 196-250.

Gülich, Elisabeth / Mondada, Lorenza, 2008. *Konversationsanalyse : eine Einführung am Beispiel des Französischen*, Tübingen, Niemeyer.

Hölker, Klaus, 2001. «Per ricostruire un contesto : gli inizi della 'Textlinguistik'», in : Prandi, Michele / Ramat, Paolo (ed.), *Semiotica e linguistica. Per ricordare Maria Elisabeth* Conte, Milan, Franco Angeli, 63-79.

Kabatek, Johannes, 2005. «Tradiciones discursivas y cambio lingüístico», *Lexis* 29/2, 151-177.

Kastberg-Sjöblom, Margareta, 2002. *L'écriture de J.M.G. Le Clézio – Une approche lexicométrique*, thèse, Nice, Université de Nice.

Koch, Peter, 1991. «Diskurstraditionen : zu ihrem sprachtheoretischen Status und ihrer Dynamik», in : Frank, Barbara / Haye, Thomas / Tophinke, Doris (ed.), *Gattungen mittelalterlicher Schiftlichkeit*, Tübingen, Narr, 43-79.

Kolboom, Ingo / Kotschi, Thomas / Reichel, Edward (ed.), 2008. *Handbuch Französisch. Sprache, Literatur, Kultur, Gesellschaft ; für Studium, Lehre, Praxis*, Berlin, Schmidt.

Mehler, Alexander / Diewald, Nils / Waltinger, Ulli / Gleim, Rüdiger / Esch, Dietmar / Job, Barbara / Küchelmann, Thomas / Pustylnikov, Olga / Blanchard, Philippe, 2011. «Network Analysis of Late Latin and Early Romance Corpora», *Leonardo* 44/3, 244–245.

Mocken, Susanne, 2013. *Informatik erschließt Literatur. Der 'discours citant' im französischen Roman zwischen 1750 und 1920*, thèse, Fribourg-en-Brisgau, Albert-Ludwigs-Universität Freiburg.

Moreno Fernández, Francisco, 1996. «Metodología del 'Proyecto para el Estudio Socio-lingüístico del Español de España y de América'», *Lingüística* 8, 257-287.

Moreno Fernández, Francisco, 2005. «Project for the Sociolinguistic Study of Spanish from Spain and America (PRESEEA) – A Corpus with a Grammar and Discourse Bias», in : Takagaki, Toshihiro / Zaima, Susumu / Tsuruga, Yoichiro / Moreno-Fernández, Francisco / Kawaguchi, Yuji (ed.), *Corpus-Based Approaches to Sentence Structures*, Amsterdam, Benjamins, 265-288.

Narbona, Antonio, 1991. «Sintaxis coloquial y análisis del discurso», *Revista española de lingüística* 21/2, 187-204.

Oesterreicher, Wulf, 2007. «Gramática histórica, tradiciones discursivas y variedades lingüísticas – esbozo programático», *Revista de Historia de la Lengua Española* 2, 109-128.

Petőfi, János S., 1971. *Transformationsgrammatiken und eine ko-textuelle Texttheorie. Grundfragen und Konzeptionen*, Frankfort, Athenäum.

Pons, Salvador (ed.), 2014a. *Discourse segmentation in Romance languages*, Amsterdam, Benjamins.

Pons, Salvador, 2014b. «Paths of grammaticalization in Spanish 'o sea'», in : Ghezzi, Chiara / Molinelli, Piera (ed.), *Discourse and Pragmatic Markers from Latin to Romance Languages*, Oxford, Oxford University Press, 108-135.

Raible, Wolfgang, 1985. «Nominale Spezifikatoren ('Artikel') in der Tradition lateinischer Juristen oder Vom Nutzen einer ganzheitlichen Textbetrachtung für die Sprachgeschichte», *Romanistisches Jahrbuch* 36, 44-67.

Raible, Wolfgang, 1994. «Literacy and Language Change», in : Cmejrková, Svetla / Daneš, František / Havlová, Eva (ed.), *Writing vs Speaking. Language, Text, Discourse, Communication. Proceedings of the Conference held at the Czech Language Institute of the Academy of Sciences of the Czech Republic, Prague, October 14-16, 1992*, Tübingen, Narr, 111-125.

Raible, Wolfgang, 1996. «Relatinisierungstendenzen», in : Holtus, Günter / Metzeltin, Michael / Schmitt, Christian (ed.), *Lexikon der Romanistischen Linguistik (LRL)*, Tübingen, Niemeyer, vol. II/1, 120-134.

Raso, Tommaso / Mittmann, Maryualê Malvessi, 2009. «Validação estatística dos critérios de segmentação da fala espontânea no corpus C-ORAL-BRASIL», *Revista de Estudos da Linguagem* 17/2, 73-91.

Rastier, François, 1989. *Sens et textualité*, Paris, Hachette.

Rastier, François, 2001. *Arts et sciences du texte*, Paris, PUF.

Rastier, François, 2011. *La mesure et le grain. Sémantique de corpus*, Paris, Champion.

Reuber, Markus *et al.*, 2014. «Comment ce que disent les patients peut nous renseigner sur leurs crises non-épileptiques psychogènes», *Neurophysiologie Clinique / Clinical Neurophysiology* 44/4, 375-388.

Rodríguez Somolinos, Amalia (ed.), 2011. *Les marqueurs du discours : approches contrastives*, *Langages* 184.

Roulet, Eddy *et al.*, 1985. *L'articulation du discours en français contemporain*, Berne, Lang.

Roulet, Eddy *et al.*, 2001. *Un modèle et un instrument d'analyse de l'organisation du discours*, Berne, Lang.

Sacks, Harvey / Schegloff, Emanuel A. / Jefferson, Gail, 1974. «A Simplest Systematics for the Organisation of Turn-Taking for Conversation», *Language* 50, 696-735.

Sacks, Harvey, 1992. *Lectures on Conversation*, Oxford, Blackwell, 2 vol. (ed. Gail Jefferson).

Sáiz Noeda, Belén, 1994. *De la gramática transfrástica a la lingüística comunicativa : nacimiento y desarrollo de la lingüística del texto*, Alicante, Universidad de Alicante.

Schlieben-Lange, Brigitte, 1983. *Traditionen des Sprechens*, Stuttgart, Kohlhammer.

Selig, Maria, 2006. «Die Anfänge der Überlieferung der romanischen Sprachen : Quellentypen und Verschriftungsprinzipien», in : Ernst, Gerhard / Gleßgen, Martin-Dietrich / Schmitt, Christian / Schweickard, Wolfgang (ed.), *Romanische Sprachgeschichte.. Ein internationales Handbuch zur Geschichte der romanischen Sprachen*, Berlin/New York, De Gruyter, vol. 2, 1924-1944.

Selig, Maria, 2008. *La Naissance des langues romanes*, Avignon, Éditions universitaires d'Avignon.

Stati, Sorin, 1990. *Le transphrastique*, Paris, PUF.

Todorov, Tzvetan, 1969. *Grammaire du Décaméron*, La Haye/Paris, Mouton.

Zumthor, Paul, 1972. *Essai de poétique médiévale*, Paris, Seuil.

Zumthor, Paul, 1983. *Introduction à la poésie orale*, Paris, Seuil.

Zumthor, Paul, 1986. «Y a-t-il une <littérature> médiévale ?», *Poétique* 66, 131-141.

Zumthor, Paul, 1987. *La Lettre et la voix. De la <littérature> médiévale*, Paris, Seuil.

Zumthor, Paul, 1990. *Écriture et nomadisme. Entretiens et essais*, Montréal, L'Hexagone.

Zumthor, Paul, 1997. *Babel ou L'inachèvement*, Paris, Seuil.

Le faible rôle de l'État dans l'évolution de la langue française

1.0 Introduction

Lorsque nous contemplons le paysage linguistique de l'Europe actuelle, une des choses qui frappent – par rapport à ce que l'on voit dans les Amériques, par exemple – c'est sa diversité. On y trouve une trentaine d'États avec une vingtaine de langues officielles, et une cinquantaine de langues « minoritaires », et partout les rapports entre *langue*, *nation* et *État* sont différents : il existe en Europe des nations multilingues, des États multinationaux, des États sans nation et des nations sans État.

	État	Langues officielles	Nations
Royaume-Uni	1	3	4 ?
Espagne	1	4	3 ?
Belgique	1	2	2 ?
Suisse	1	4	1 ?
Norvège	1	2	1
France	1	1	1

Deux choses frappent dans ce tableau :

1° Les points d'interrogation : tout le monde sait ce que c'est qu'un *État* et ce que c'est qu'une *langue officielle*, mais quand il s'agit d'une *nation*, on est dans l'incertitude – ce qui est *nation* pour l'un n'est que *minorité culturelle / linguistique* pour l'autre. Une *nation* n'est pas une entité objective, c'est une chose qui se pense, qui dépend essentiellement du point de vue de l'observateur. C'est une construction sociale.

2° S'il y a un pays où les rapports entre *langue*, *nation* et *État* sont clairs et sans équivoque, c'est la France, où ces trois idées sont liées entre elles d'une façon inextricable. Plus que dans les autres pays, ces trois mots portent en France sur trois aspects d'une seule et même chose. Voici pourquoi la langue occupe une place tellement centrale dans le sentiment d'identité nationale, et pourquoi les historiens du français associent avec tant de facilité la langue à l'État.

Dans cette intervention, qui va survoler plus de mille ans d'histoire de la langue française, nous ne mettrons pas les pieds dans le territoire périlleux des rapports entre la *langue* et la *nation*, mais nous allons néanmoins examiner le rôle de l'*État* dans cette affaire. Nous allons regarder, dans une première partie, la version 'offi-

cielle' de l'histoire du français et le rôle central qu'elle attribue aux activités de l'État. Dans la deuxième partie, nous allons voir que cette vision des choses repose sur une idée trop étroite de ce que c'est qu'une langue, qu'elle sous-estime le rôle des premiers intéressés – les locuteurs ordinaires – et qu'elle surestime, par conséquent, le rôle de l'État dans l'histoire du français.

2.0 Les versions officielles de l'histoire de la langue

Les historiens de la langue, dont je fais partie, sont premièrement des linguistes et seulement plus tard des historiens. Ce fait a des conséquences néfastes pour l'historiographie des langues, car l'histoire de France, telle qu'elle transparaît dans les histoires de la langue, est souvent une histoire simplifiée que les historiens spécialistes ont du mal à reconnaître.

2.1 Histoire de la Nation

Dans les versions 'officielles', l'histoire de la Langue est liée d'une manière indissoluble à l'histoire de la Nation. La Nation française, comme la Langue, ne se fit pas d'elle-même, elle fut construite au cours de mille ans par une succession d'hommes d'État, agissant – sans nécessairement le savoir – dans les intérêts à long terme du peuple français. L'histoire de France se résume en quatre périodes, le point culminant étant la Révolution de 1789, qui donna à la Nation sa forme définitive :

I Formation de la France (IXe - XIIe s.)

II Expansion du royaume (XIIIe - XVe s.)

III Consolidation du royaume (XVIe - XVIIIe s.)

IV Triomphe de la Nation (XIXe - XXe s.)

2.1.1 Formation de la France (IXe - XIIe s.)

La nation prend sa naissance dans le démantèlement de l'Empire de Charlemagne, l'empire des *Francs*, au milieu du IXe siècle. La Gaule est coupée en deux : le roi de France prend la partie occidentale et centrale, et l'Empereur germanique se réserve la partie orientale. Sous les derniers Carolingiens et les premiers Capétiens, il se constitue progressivement un noyau de pouvoir dans l'Ile-de-France autour de Paris.

2.1.2 Expansion du royaume (XIIIe - XVe s.)

Au début du XIIIe siècle, sous Philippe Auguste, le royaume de France s'agrandit : les Anglais sont boutés hors de la Normandie (après la bataille de Bouvines) et la Croisade des Albigeois permet la réunification à la Couronne du centre de la France (en particulier l'Auvergne) et de toute la partie du Midi qui relevait des comtes de Toulouse. La Guerre de Cent Ans enraya pendant un certain temps l'expansion du royaume, mais le mouvement reprit au XVe siècle avec l'expulsion définitive des Anglais du sud-ouest du pays.

2.1.3 Consolidation du royaume (XVIe - XVIIIe s.)

Le XVIe siècle et la période de l'Ancien Régime virent l'extension du pouvoir du roi de France dans l'est de la Gaule, dans le domaine relevant jadis de l'Empire germanique. Cette époque vit également la consolidation du pouvoir monarchique dans un royaume de France devenu maintenant très vaste, sous les auspices de François Ier, Henri IV, Richelieu, Louis XIV, Colbert, etc.

2.1.4 Triomphe de la Nation (XIXe - XXe s.)

La monarchie avait œuvré pendant huit siècles à l'expansion du royaume, mais ce fut la tâche de la Révolution, de Napoléon et de la IIIe République de faire de ce royaume une véritable Nation - une société qui offrait à tous les Français Liberté, Égalité et Fraternité. C'est la dernière des grandes périodes de l'histoire de France : l'époque actuelle n'appartient pas au domaine des historiens de la langue.

Qu'en est-il maintenant de l'histoire de la langue, dans ses versions officielles ? La construction de la Langue emboîte le pas à la construction de la Nation.

2.2 Les étapes de la langue française

Le français appartient, comme chacun le sait, à la famille des langues romanes. L'histoire du français commence donc avec le latin et la désintégration de l'Empire romain au Ve siècle. Ce cataclysme enchaîne sur quatre longs siècles durant lesquels la langue latine - en Italie, en Espagne et en Gaule - se fragmente en une multitude de dialectes. Au niveau de l'Église et des administrations politico-juridiques, la communication est assurée, longtemps après l'effondrement de l'Empire romain, par le latin. En revanche, au niveau de la population générale, composée à 95% de paysans incultes, c'est une véritable catastrophe : leur latin se fragmente en une multitude de patois, rendant la communication en-dehors de la province de plus en plus difficile. La masse des Français se voient ainsi condamnés à une vie d'ignorance et d'immobilité, qui va durer mille ans jusqu'à la Révolution,

L'histoire de la langue, dans sa version officielle, raconte essentiellement la lente rectification de cette catastrophe linguistique et sociale : c'est l'histoire d'un long processus d'uniformisation / de standardisation de la langue. Une vraie langue est une langue unie, sans variation, « une et la même pour tous ». Tout comme ce fut la tâche des rois et d'hommes d'État de construire une Nation unie à partir d'une bigarrure de provinces, il incombait de la même façon à l'État, secondé par les grands écrivains et les cercles de gens instruits, d'œuvrer à la construction d'une grande langue *nationale* (ou *standard*).

La standardisation des langues est un sujet qui intéresse non seulement les linguistes universitaires, mais aussi les planificateurs de langues, administrateurs s'efforçant de régler les problèmes linguistiques, par exemple, dans les nouveaux pays d'Afrique et d'Asie. Parmi les linguistes, c'est le Norvégien Einar Haugen qui nous a

fourni l'analyse la plus élégante du phénomène de standardisation. Haugen distingue dans ce processus quatre phases :

Les processus de standardisation		
	Forme	Fonction
Société	*Sélection*	*Diffusion*
Langue	*Codification*	*Élaboration fonctionnelle*

Au niveau de la société, le point de départ c'est un pays sans langue commune divisé en plusieurs zones dialectales. Le premier pas à franchir c'est de *sélectionner* l'un des dialectes en présence pour former la base de la future langue nationale (ou *standard*). Ensuite, il sera nécessaire de *diffuser* ce dialecte privilégié et de le faire accepter par l'ensemble de la population. Au niveau de la langue, le dialecte de base sera, à l'origine, un simple parler oral comme tous les autres. Pour qu'il devienne une langue standard utilisable dans l'administration publique, dans la justice, dans l'éducation, etc., il faudra la doter d'une forme écrite, et augmenter son vocabulaire et sa syntaxe (*élaboration fonctionnelle*). Il sera nécessaire également que ce dialecte, qui sera proposé comme le modèle à imiter par l'ensemble de la population, soit uniforme et invariable. Il convient donc que le dialecte standard obéisse strictement à des règles, *codifiées* dans des dictionnaires et dans des livres de grammaire.

2.3 Langue et Nation

La construction de la Langue emboîte le pas de la construction de la Nation :

	Construction de la Nation	*Construction de la Langue*
Ve-VIIIe s.	Les Francs	Latin vulgaire / gallo-roman
IXe-XIIIe s.	Formation de la France Capétiens (la chancellerie royale)	*Sélection* Ancien français
XIVe-XVIe s.	Expansion du royaume Valois (Villers-Cotterêts 1539)	*Élaboration* Moyen français
XVIe-XVIIIe s.	Consolidation du royaume Ancien Régime (Académie française 1635)	*Codification* Français classique

XIXe-XXe s.	Triomphe de la Nation IIIe République (Jules Ferry 1880)	*Acceptation* Français moderne

L'État intervient, à chaque stade de l'évolution de la Nation, pour promouvoir la standardisation, l'unification de la langue. Passons en revue les quatre phases.

2.3.1 Les Capétiens : sélection du dialecte de base

L'État français commence à se constituer dans le nord de la Gaule, dès la partition de l'empire de Charlemagne au IXe siècle. Et c'est justement dans un acte politique que nous trouvons la première tentative de mettre par écrit la nouvelle langue qu'était le français, les Serments de Strasbourg (842) :

Pro Deo amur et pro christian poblo et nostro commun saluament, d'ist di in auant, in quant Deus sauir et podir me dunat, si saluarai eo cist meon fradre Karlo, et in adiudha et in cadhuna cosa si cum om per dreit son fradra saluar dift, in o quid il mi altresi fazet. Et ab Ludher nul plaid nunquam prindrai qui meon uol cist meon fradre Karle in damno sit.

Ces paroles, insérées dans un texte latin, représenteraient la première tentative de forger un français capable de traiter les affaires d'État.

A cette époque, il n'y avait pas de langue commune. L'ancien français était une langue vernaculaire, c'est à dire une langue qui existait uniquement sous ses multiples formes dialectales – le picard, le normand, le champenois, le lorrain, le francien (dialecte de l'Ile de France). Pour standardiser une langue, la première chose à faire c'est de *sélectionner* l'un des dialectes pour fournir la base de la standardisation. Dans le cas du français, le dialecte sélectionné était le *francien*, le dialecte de l'Ile-de-France et de Paris. Pourquoi ce dialecte et non pas le picard, le normand, le lorrain ? C'est que les rois capétiens avaient décidé d'établir à Paris le siège de l'administration royale. Commence alors la lente ascension du français de Paris comme instrument du pouvoir.

2.3.2 Les Valois : élaboration des fonctions du français

Étant donné qu'au Moyen Age la langue de l'écrit était le latin, le français était alors une langue surtout orale. Pour que le français puisse devenir une langue standard, il fallait qu'il se dote d'une forme écrite. Pour qu'il soit utilisable dans l'administration, dans la diplomatie, dans l'éducation, il fallait qu'il s'équipe des outils linguistiques nécessaires à la tâche – un vocabulaire savant (en plus du vocabulaire de tous les jours), une syntaxe complexe (en plus de la syntaxe orale). C'est ce que l'on voit arriver, en effet, dans les textes français de la fin du Moyen Age et de la Renaissance. Il s'agit d'une grande expansion fonctionnelle de la langue.

Au XVIe siècle, la monarchie reconnaît les capacités de la langue française et entame le processus de remplacement des autres langues et dialectes de France – à

commencer par le latin, mais en enchaînant rapidement sur les langues régionales, notamment l'occitan. Pour ce faire, il a fallu, naturellement, un texte de loi, et c'est ce que François Ier a entrepris dans les ordonnances de Villers-Cotterêts de 1539 :

Et pour ce que telles choses sont souvent advenues sur l'intelligence des mots latins contenus esdits arrests, nous voulons d'oresnavant que tous arrests, ensemble toutes autres procédures, soient de nos cours souveraines et autres subalternes et inférieures, soient de registres, enquestes, contrats, commissions, sentences, testaments, et autres quelconques, actes et exploicts de justice, ou qui en dépendent, soient prononcé, enregistré et délivré aux parties *en langage maternel français et non autrement*.

2.3.3 L'Ancien régime : codification de la norme

Vouloir imposer à tous les Français le français du roi, c'était très bien, encore fallait-il que ce français du roi soit de la qualité nécessaire – une langue 'pure' gouvernée par des règles. Dans la première moitié du XVIIe siècle, après toutes les extravagances de la Renaissance et des Guerres de Religion, l'État s'efforçait de mettre de l'ordre. En matière de langue il fallait une langue commune, une norme fixe et strictement codifiée. Heureusement, l'heure a produit l'homme qu'il fallait en la personne de Richelieu. Cet homme a su reprendre le bâton de la standardisation en fondant en 1635 l'Académie française pour codifier les règles de la langue. La France des XVIIe - XVIIIe s. eut la chance de voir naître une succession éblouissante de grands auteurs qui portèrent la belle langue classique, codifiée par les grammairiens, à un point de perfection.

2.3.4 Révolution et République : extension du français en France.

Le français du roi avait maintenant parcouru les trois premiers stades de la standardisation. Il n'en restait qu'un, celui de la *diffusion* : en-dehors des cercles de gens cultivés à Paris, très peu de Français parlaient français, surtout en province dans les campagnes. C'est la Révolution qui mena à bien le processus de standardisation en promouvant l'extension du français en France. Persuadé que « la langue d'un peuple libre doit être une et la même pour tous », l'État s'engage au XIXe siècle dans une immense campagne d'alphabétisation, visant la francisation des provinces dialectophones et l'élimination des patois. Si tous les Français parlent aujourd'hui français, c'est grâce surtout à l'action de l'État, à l'Éducation nationale, à l'école primaire de Jules Ferry, et aux hussards noirs de la République.Voici donc, très abrégée, la version officielle de l'histoire du français : l'évolution de la Langue fait partie de l'évolution de la Nation. C'est essentiellement un mouvement d'uniformisation, de standardisation, dont le grand moteur a toujours été l'État – secondé par les grands écrivains et l'élite intellectuelle du pays.

3.0 Tout cela serait bien beau, si tout était vrai.

La version de l'histoire du français que je viens d'esquisser est cohérente, rassurante : c'est une vision dirigiste / volontariste de l'histoire, qui fait de la langue quelque

chose de gérable. Mais, c'est au prix de maintes distorsions et simplifications. Prenons trois ou quatre exemples, portant sur des moments clefs de l'histoire de la langue

3.1 Serments de Strasbourg

Est-il possible que la monarchie ait pensé, dès les Serments de Strasbourg (IXe siècle), faire du français la langue du royaume ? C'est cinq siècles trop tôt. Le premier État européen à utiliser le français comme langue administrative fut, non pas la France mais l'Angleterre. Ayant conquis l'Angleterre au XIe siècle, les rois anglo-normands gouvernèrent le pays non pas dans la langue du peuple conquis, mais en français. A Paris, en revanche, les rois ont attendu le XIVe siècle pour mener systématiquement leur administration en français. Pourquoi ce retard ? C'est qu'il existait à Paris une armée de fonctionnaires qui tenaient à conserver le latin dont ils tenaient le monopole.

Il est indéniable, toutefois, que le dialecte de Paris et de l'Ile de France a pris une place prépondérante en France dès le XIIe siècle – deux siècles avant que les rois ne viennent s'occuper de ces choses-là. Comment l'expliquer, si ce n'est pas par le choix de l'État ? C'est vraisemblablement pour des raisons plutôt démographiques que politiques : Paris est devenue, au cours du XIIe siècle, la plus grande ville de l'Occident :

Figure 1

1000		**1400**		**1700**		**1900**	
City	Population	City	Population	City	Population	City	Population
Constantinople	450	Paris	275	Constantinople	700	Londre	6,480
Cordoue	450	Milan	125	Londre	550	Paris	3,330
Séville	90	Bruges	125	Paris	530	Berlin	2,424
Palerme	75	Venise	110	Naples	207	Vienne	1,662
Kiev	45	Grenade	100	Lisbonne	188	St. Pétersbourg	1,439
Venise	45	Gênes	100	Amsterdam	172	Manchester	1,255
Ratisbonne	40	Prague	95	Rome	149	Birmingham	1,248
Thessalonique	40	Théodosie	85	Venise	144	Moscou	1,120
Amalfi	35	Séville	70	Moscou	130	Glasgow	1,072
Rome	35	Gand	70	Milan	124	Liverpool	940

(Source : Hohenberg & Lees 1985, 11)

3.2 Villers-Cotterêts 1539

Les Ordonnances de Villers-Cotterêts jouent un rôle extrêmement important dans l'histoire officielle: elles constituent l'acte de naissance de la langue. Cependant, lorsque l'on examine le texte de ces Ordonnances, on découvre qu'il comporte 192 articles, dont seulement deux portent sur des questions de langue, et que ces deux articles sont loin de figurer en tête de la liste - ils occupent les places 110 et 111.

Art. 111. - Et pour ce que telles choses sont souvent advenues *sur l'intelligence des mots latins contenus esdits arrests*, nous voulons d'oresnavant que tous arrests, ensemble toutes autres procédures, soient de nos cours souveraines et autres subalternes et inférieures, soient de registres, enquestes, contrats, commissions, sentences, testaments, et autres quelconques, actes et exploicts de justice, ou qui en dépendent, soient prononcé, enregistré et délivré aux parties en langage maternel français et non autrement.

François Ier ne cherchait aucunement dans cet article à lancer une grande action de planification linguistique visant la destruction des dialectes et des langues régionales. Il s'attaque avant tout au latin, dans l'optique de miner l'autorité des tribunaux ecclésiastiques. On se rappellera que le grand rival de François Ier - Henri VIII, roi d'Angleterre - venait tout juste de se proclamer « chef unique et suprême de l'Église d'Angleterre ». C'est seulement par implication que les articles 110 et 111 ont été appliqués aux autres langues et dialectes de France. Si le roi a voulu que l'on se serve de sa langue dans les parlements du royaume, notamment ceux de langue d'oc, ce n'était pas pour imposer le français, c'était surtout pour donner à ses officiers l'avantage linguistique sur les gens du pays.

3.3 Académie française

Lorsque Richelieu fonda l'Académie française en 1635, ce ne fut pas dans l'optique de créer une langue commune accessible à tout le monde : il ne cherchait nullement à établir une norme linguistique que l'ensemble de la population française devrait suivre. Au contraire, la norme codifiée par les grammairiens fut destinée exclusivement aux élites de la Cour et de la Ville qui faisaient tout pour se distinguer, pour se distancier de la masse des Français. Voici le projet de l'*Académie française* (1635), tel qu'il fut dressé par Nicolas Faret :

> (i) nettoyer la langue des ordures qu'elle avoit contractées ou dans la bouche du peuple ou dans la foule du Palais, ou dans les impuretez de la chicane, ou par le mauvais usage des courtisans ignorants
>
> (ii) notre langue plus parfaite déjà que pas une des autres vivantes, pourroit bien enfin succéder à la latine, comme la latine à la grecque, si l'on prenoit plus de soin [...] de l'élocution.

L'Académie cherchait à créer, il est vrai, une variété 'parfaite' du français, mais c'était pour exclure du pouvoir la masse du peuple et pour donner au français des

élites un prestige international. Si l'État a joué un rôle dans la *codification* du français au XVIIe siècle, ce rôle a souvent été mal compris.

3.4 Jules Ferry

Nous arrivons enfin à la quatrième phase du processus de standardisation – la diffusion de la langue 'nationale' dans les provinces dialectophones au XIXe siècle. Les versions 'officielles' de l'histoire attribuent ce mouvement principalement à l'action de l'État, sous la forme de l'école primaire, gratuite et obligatoire de Jules Ferry. La mission des hussards noirs de la République était d'enseigner à lire et à écrire à tous les petits Français en un bon français, approuvé par le Ministre de l'enseignement public. Et il est certain qu'au cours du XXe siècle les pauvres patois et langues régionales ont disparu de la face de la terre, mais peut-on être sûr que ce soit dû principalement à l'action de l'État, par l'intermédiaire des maîtres d'école ? Si cela était le cas, comment se fait-il, après tant d'efforts, que les Français parlent aujourd'hui si mal leur langue ? Si le français vient des maîtres d'école, pourquoi les Français ne parlent-ils pas tous comme un livre ? D'où vient ce 'mauvais français' que tout le monde parle ?

4.0 Le français ordinaire à Clermont-Ferrand

Je n'oublierai jamais le choc qui me fut administré, il y a 50 ans, lorsque je suis venu pour la première fois faire un long séjour en France, comme assistant d'anglais dans un lycée de Clermont-Ferrand. J'étais fier de mon français, dont tous les résultats d'examen me disaient qu'il était bon. Mais, ce que j'ai entendu au cours de mon premier repas dans la cantine de l'école m'a brusquement désabusé :

- Hé l'Anglais, la flotte.

- La bouffe dans ton pays, elle est dégueulasse, non ?

- Elle avait pas l'air bien cuite ta nana d'hier soir.

Puisqu'il existe chez nous, à côté de l'anglais de la Reine, un anglais relâché, je savais vaguement qu'il devait exister la même chose en France. Mais rien ne m'avait préparé au spectacle de tout un langage parallèle, comportant des milliers de mots étranges qui doublait le français correct que j'avais appris à Manchester :

travailler → bosser
manger → bouffer
pleurer → chialer
boire → picoler
sentir → piffer

Et j'ai vite fait de découvrir que ces mots bizarres n'étaient pas l'apanage des « pions », mais que tous les « potaches » les utilisaient, et même les « profs ».

Étudiant consciencieux, j'ai dressé une liste de ces expressions bizarres que j'ai présentée au professeur de français du lycée : « Ah ! Monsieur Lodge, me dit-il, ne

vous tracassez pas. Le français est une langue logique, claire et élégante. Ces mots de votre liste, ce n'est pas du français, c'est de l'argot. » Cette réplique ne réussit pas à me faire reprendre mes esprits : me voici venu en France pour perfectionner mon français et – pas de chance ! – je tombe sur des quantités de Français qui baragouinent un langage qui n'est pas français du tout. J'aurais mieux fait de rester à Manchester.

Depuis cette époque, l'écart qui sépare le 'bon' français (officiel, académique) du 'mauvais' français de tous les jours ne cesse de m'intriguer. Je ne suis pas seul, d'ailleurs, car la question fascine les Français eux-mêmes, à en juger par la quantité d'articles qui paraissent à ce sujet dans les journaux :

Figure 2

A chacun son français

On comprend aisément, grâce à ce tableau, à quelles dégradations s'expose notre langue en cette période de banalisation audiovisuelle.

Précieux (snob, poétique, désuet)	**Soutenu** (littéraire)	**Courant** (commercial, public, administratif)	**Familier** (privé, populaire)	**Argotique** (snob, jeune, vulgaire)
Le chef	La tête	La figure	La bobine, la bouille, la binette, la caboche, la bille	La gueule, la tronche, la trombine, la margoulette
Un mortel	Un homme	Un individu	Un type, un gars, un pékin, un zèbre	Un mec, un gonze, un zig, un gazier
Cocasse	Amusant	Drôle	Rigolo, tordant, gondolant, fendant	Marrant, bidonnant, poilant, à s'tap'.
Le véhicule	L'automobile	La voiture	L'auto, la bagnole	La tire, la caisse
Il chuta.	Il tomba.	Il est tombé.	Il s'est cassé la figure, il a pris un billet de parterre.	Il s'est cassé la gueule.
Qu'il manque d'intelligence !	Qu'il est sot !	Qu'il est bête !	Quel idiot ! Quel crétin !	Quelle andouille ! Quel couillon ! Quel cul ! Quel con !
Il est béni des dieux.	Il est né sous une bonne étoile.	Il a de la chance.	Il a de la veine, il a du pot.	Il a du bol, du cul, il est beurré, il l'a bordé de nouilles.
Il fait preuve de pusillanimité.	Il est rempli de crainte.	Il a peur.	Il a la frousse, il se dégonfle, il panique.	Il a la trouille, les chocottes, les jetons, les grelots, le trouillomètre à zéro.
Tu m'agaces.	Tu me fatigues.	Tu m'ennuies.	Tu m'embêtes, tu me fais suer, tu m'enquiquines.	Fais chier, tu me les casses, tu m'emmerdes.
Suffit !	Assez !	Par-dessus la tête !	Plein le dos !	Ras-le-bol ! Plein le cul ! Y en a marre !

TABLEAU TIRÉ DU LIVRE DE MICHEL MASSIAN ET SI L'ON ECRIVAIT CORRECTEMENT LE FRANÇAIS ?

Ma première réaction fut de penser qu'au Lycée Blaise Pascal de Clermont j'étais particulièrement mal tombé : ces gens-là n'étaient pas typiques. J'ai décidé de tester cette hypothèse en menant une enquête sur une population plus large, plus représentative. Je voulais découvrir si c'était seulement les gens du lycée qui parlaient si mal, ou si, au contraire, tout le monde parlait mal, selon les circonstances. J'ai trouvé d'abord à Clermont-Ferrand 70 informateurs, échelonnés à égalité sur toutes les tranches d'âge, avec autant d'hommes que de femmes, et avec une répartition assez égale entre les différentes classes sociales. A ces 70 personnes j'ai donné à remplir un questionnaire écrit. Le questionnaire comportait 50 paires de mots, dont le premier appartenait au 'bon français', au français officiel, et le second au 'mauvais français', au français de tous les jours (étiqueté *fam., pop., vulg.* et *arg.* dans le dictionnaire *Petit Robert*) : type *médecin ~ toubib, argent ~ fric, désordre ~ merdier, voiture ~ tire.* A propos de chaque mot j'ai demandé à l'informateur de répondre à deux questions, la première portant sur les comportements en situations d'intimité et la seconde sur les comportements en public :

Figure 3

1. Pensez-vous avoir utilisé ce mot avec vos proches / amis etc. au cours des trois derniers mois?

Jamais	Rarement	Souvent	Très souvent
0	1	2	3

2. Dans quelle mesure seriez-vous capable d'utiliser ce mot avec quelqu'un que vous ne connaissez pas?

Jamais	Rarement	Souvent	Très souvent
0	1	2	3

Imaginez-vous seul dans un compartiment de chemin de fer avec, en face de vous une personne de même âge et du même sexe que vous, que vous ne connaissez pas.

Il convient de signaler que cette enquête portait non pas sur l'usage réel de Clermontois, mais seulement sur leurs perceptions d'usage. L'analyse des réponses m'a permis, d'une part, de donner un score à chaque informateur, indiquant sa propension à utiliser les mots en question, d'autre part, de donner un score à chaque mot, indiquant son niveau d'acceptabilité (sa 'valeur sociale')[1]. Lorsque j'ai calculé les résultats, ce qui m'a surpris c'est que ceux-ci n'étaient pas aléatoires : A. Qui utilise le plus les mots du 'mauvais français'?

[1] Toutes les langues comportent une grande part de variabilité. Les locuteurs ne restent pas neutres devant les variantes en présence – elles sont toutes soumises à des évaluations sociales, tacites la plupart du temps. Les variantes à valeur sociale haute sont affectées à des situations plutôt formelles et solennelles et les variantes à valeur basse à des situations familières et intimes.

Figure 4 : *Âge / sexe*

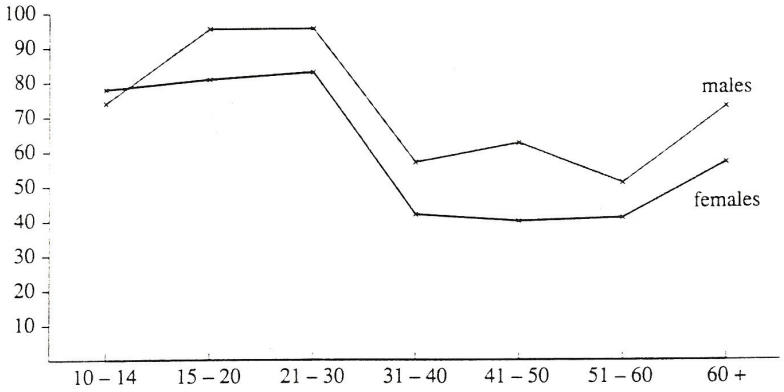


Figure 5 : *Catégorie sociale*

	Catégorie professionnelle	*Avec ses proches*	*Avec des inconnus*
I	Cadres supérieurs / Professions libérales	72.1	57.2
II	Cadres moyens	65.3	43.7
III	Employés	88.0	85.1
IV	Ouvriers / Personnel de service	74.3	69.8

B. Quels mots sont considérés comme les plus 'mauvais' ?

Figure 6

bouquin	167	fam.	roupiller	84	fam.
balader	152	fam.	bouffer	79	fam.
marre	152	fam.	con	79	fam./vulg.
marrant	147	pop.	mec	78	pop.
blague	139	fam.	bahut	76	arg. des écoles
flic	134	pop.	baffe	75	pop.
boulot	132	fam.	dégueulasse	75	vulg.
chouette	131	pop.	pognon	74	pop.
moche	124	fam.	pompes	74	pop.
rigolo	119	fam.	frangin	71	pop.
trouille	119	pop.	godasse	71	pop.
dingue	116	fam.	piaule	61	pop.
bosser	115	pop.	foutre	60	vulg.
costaud	115	fam.	nana	60	pop.
pagaie	113	fam.	pif	55	pop.
gaffe	110	pop.	merdier	51	vulg.
esquinter	107	fam.	pieu	51	pop.
toubib	106	fam.	tifs	50	pop.
frousse	103	pop.	chiottes	40	pop.
bagnole	94	fam./pop.	clope	43	pop.
fric	94	pop.	couillon	43	très fam.
fringues	91	fam.	gonzesse	39	vulg.
bousiller	86	fam.	pouffiasse	36	vulg.
baratin	85	pop.	pèze	33	arg.
cramer	84	pop.	tire	26	arg.

Ce qui peut surprendre dans les résultats de ce sondage, c'est que l'ensemble des informateurs, tout en parlant différemment, tout en possédant chacun sa façon personnelle de s'exprimer, était d'accord sur la valeur *sociale* à attribuer à chaque mot relativement aux autres - un mot jugé très mauvais par les uns était également jugé très mauvais par les autres. J'en tire la conclusion que tous les informateurs, d'une manière tacite et inconsciente, jugeaient la valeur de chaque mot selon les mêmes critères, et qu'ils participaient tous aux mêmes normes évaluatives.

Mais qu'est-ce que tout cela peut avoir à faire avec l'histoire de la langue ? Eh bien, si la diffusion du français à Clermont-Ferrand était due principalement à l'action de l'État - à l'école de la IIIe République - comment se fait-il que les Clermontois, comme sans doute les Nancéens, parlent si mal la langue ? Ce n'est pas l'école d'État qui leur a appris les mots *fam., pop., vulg.* et *arg.* Il a dû y avoir d'autres facteurs en jeu.

5.0 Registre H versus Registre B

Lorsqu'un linguiste examine une grande langue comme le français, il y voit non pas un seul type de français mais deux : d'un côté, un français normé, académique,

utilisé à l'écrit et dans les situations plus ou moins solennelles, approuvé par l'école et par l'État, et, de l'autre, un français ordinaire, quotidien, employé par tout le monde dans les situations familières, désapprouvé par l'école. Toutes les grandes langues vivantes possèdent nécessairement deux registres analogues : une langue qui ne dispose que d'un registre familier, oral, est une sorte de patois ; une langue qui ne dispose que d'une forme écrite, normée, est une langue morte, comme le latin.

Les linguistes appellent la langue normée le registre haut (H) ou le registre *standard*, et la langue de tous les jours le registre bas (B) ou le registre *vernaculaire*. Les deux niveaux de langue sont complémentaires l'un de l'autre :

Standard (H) ⟷ *Vernaculaire (B)*
grand prestige ⟷ aucun prestige
langue de distance ⟷ langue de proximité
à l'extérieur du groupe ⟷ à l'intérieur du groupe
centré sur le message ⟷ centré sur l'interlocuteur
objectivité ⟷ subjectivité
discours planifié ⟷ discours spontané

Dans certaines communautés linguistiques, que l'on appelle *diglossiques*, les deux fonctions H et B sont remplies par deux langues différentes - c'est le cas des pays maghrébins, par exemple, où les fonctions H sont remplies par l'arabe classique et les fonctions B par les arabes régionaux. En français, comme dans la majorité des langues, les deux types de fonctions sont remplis par deux registres différents de la même langue.

Dans la vie normale, dans la vie de tous les jours, c'est le registre B (vernaculaire) qui prédomine. Cela dépend du métier que nous exerçons : dans la vie des cadres supérieurs, des professions libérales, le registre H (standard) occupe une place vraisemblablement plus importante que dans celle des femmes au foyer, des employés, du personnel de service, etc. Mais, cela n'empêche pas que, pour tous les Français, le registre qu'ils utilisent la plupart du temps c'est le registre B (vernaculaire).

L'on cherche à nous faire croire que le registre B est une forme paresseuse, dégénérée du registre H. Ceci n'est pas le cas : le registre B est la forme la plus basique de la langue, sur laquelle vient se greffer plus tard la forme standard, avec l'école. Vu sous cet angle, dire de telle ou telle tournure familière ou populaire qu'elle n'est pas française, comme le faisait autrefois M. Anglade, c'est un non-sens. Ne serait-il pas préférable d'adopter une vision holistique de la langue, et d'inclure sous le parapluie *français* toutes les variétés du français ? – les formes dévalorisées au même titre que les formes les plus valorisées – en réservant à chacune la fonction qui lui revient ?

6.0 La partie invisible de l'iceberg

Lorsque l'on fait l'histoire d'une langue, il importe de ne pas abandonner cette vision holistique, inclusive, des choses, à la différence de ce que font les histoires offi-

cielles du français. Nous avons vu que celles-ci se préoccupent peu de l'évolution de la langue ordinaire. Est-il légitime de réduire l'histoire d'une grande langue comme le français à l'histoire du seul registre standard ? Nous venons de voir qu'il existe en même temps un registre B, vernaculaire. Pour celui qui veut remonter l'histoire de la langue, le registre B c'est la partie invisible de l'iceberg :

Figure 7

Plus nous remontons dans le passé, plus cette immense partie de l'iceberg cachée sous l'eau devient difficile à voir. Mais son invisibilité n'autorise pas à penser - comme l'a fait il y a cent ans le capitaine du *Titanic* – qu'elle n'était pas là. Le centre de gravité se trouve loin au-dessous de la surface de l'eau, et il en va de même pour le passé d'une langue.

La langue parlée, le registre B, n'a jamais cessé d'exister, mais elle est très différente du registre H – elle est réfractaire à toutes les autorités. Elle échappe au contrôle institutionnel. Si nous donnons au français le sens large qui inclut les deux registres H et B – les deux parties de l'iceberg – la capacité de l'État d'infléchir le mouvement général de la langue se réduit considérablement. Dans l'évolution de la langue prise dans ce sens large, le rôle de l'État se voit beaucoup diminuer.

Si ce n'est pas l'État qui fait bouger la langue, quelles sont les grandes forces motrices ? Dans le système grammatical de l'ancien français comme dans celui du latin, quand il s'agissait de verbes et de substantifs, on exprimait d'abord le sens du mot (en majuscules) et l'on ajoutait ensuite une terminaison pour marquer sa fonction grammaticale (en minuscules soulignées) :

Ancien français : Iluec MANJUent POIREs.

(Là-bas mangent poires.)

En français moderne on fait l'inverse, en marquant d'abord la fonction grammaticale, et ensuite le sens du mot :

Français moderne : la'bɑ i mɑ̃ʒ de 'pwaʀ

(Là-bas ils MANGENT des POIRES.)

On se rappellera que les terminaisons ont presque totalement disparu, sauf à l'écrit (qui est toujours à la traîne). Qui est-ce qui a pu organiser une restructuration si profonde de la syntaxe française ? Ce ne sont pas les rois, ni les grands auteurs, ni tout l'appareil de l'État, ce sont les locuteurs ordinaires du français, agissant collectivement et inconsciemment. Ce n'est pas l'État qui a dirigé l'évolution de la langue française, telle qu'elle est parlée la plupart du temps par la plupart des gens, ce sont les locuteurs eux-mêmes. D'où vient le 'mauvais français' familier et populaire ? Loin d'être une forme corrompue du français normé, le français ordinaire lui a préexisté de presque mille ans. C'est le successeur du latin ordinaire parlé par les Romains et par leurs peuples colonisés, que les habitants de la Gaule n'ont jamais cessé d'utiliser et qu'ils ont modifié continuellement et sans rupture pour qu'il réponde continuellement à leurs besoins de communication.

Pour la période médiévale et préindustrielle, nous avons trop longtemps exagéré, d'une part, les fossés linguistiques censés empêcher la communication entre les différentes provinces du royaume, d'autre part, l'immobilisme de la société française à cette époque. Personne ne peut nier l'importance de la variabilité dialectale de la France, surtout avant la période industrielle. Mais on aurait vu la même chose dans tous les autres pays d'Europe – en Angleterre, en Italie, en Allemagne. Les nombreux dialectes et langues de la Gallo-Romania remontent tous au latin et restent fortement apparentés les uns aux autres. La France médiévale n'avait ni trains ni automobiles, mais cela n'empêche pas que ce fut un des pays les plus dynamiques de l'Europe : les gens se déplaçaient continuellement d'un bout à l'autre du royaume. Et pas seulement les élites – je pense aux marchands qui sillonnaient constamment le pays, je pense aux migrations saisonnières d'ouvriers agricoles, je pense aux compagnons apprenant leur métier en faisant leur « tour de France ». Tout porte à croire que, grâce à de tels déplacements, il se créa une *Umgangsprache*, une langue commune orale indépendamment du français standard parachuté de Paris par les élites.

Pour la période moderne, industrielle, si ce n'est pas Jules Ferry qui a diffusé dans tous les coins de la République le français familier et populaire tellement déprécié par l'école, il est nécessaire d'en chercher les causes ailleurs. Il faut les chercher, je

crois, dans les transformations technologiques et démographiques apportées par la révolution industrielle. Paris est devenue, comme Londres, plus monstrueuse que jamais, attirant, grâce alors au chemin de fer, des foules de travailleurs de toutes les provinces de France. Ce sont surtout ces gens-là qui ont diffusé la langue commune orale partout dans l'Hexagone. Ce sont les grands mouvements d'industrialisation et d'urbanisation qui ont sonné le glas des parlers vernaculaires de la campagne - les patois :

Figure 8

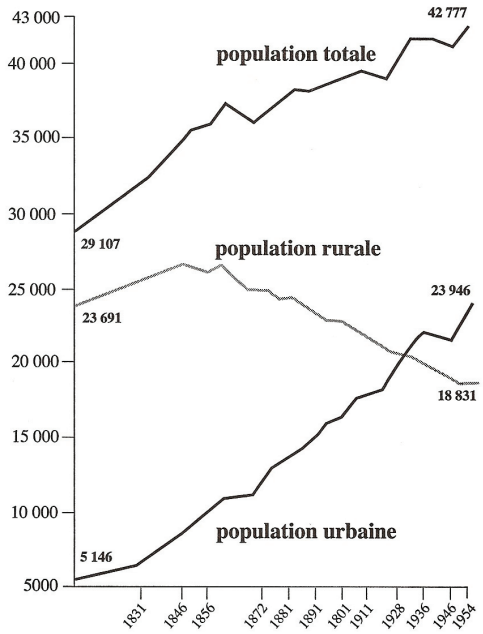


Figure 8. Population rurale et population urbaine en France (d'après Braudel, 1986).

7.0 Conclusions

J'ai essayé de montrer, dans cette intervention, que les versions officielles de l'histoire du français donnent à l'État un rôle bien plus important qu'il n'en a eu en réalité. Les histoires en question ciblent presqu'exclusivement la langue standard. Or, le processus de standardisation est un processus éminemment politique. La langue standard – le registre H – est le registre adopté par les grandes institutions de l'État, et elle a, de ce fait, toujours été assujettie aux ingérences du pouvoir. Il est donc difficile de nier le poids de l'État dans l'évolution de cette partie de la langue. Mais, la grande faiblesse de cette approche, c'est qu'elle s'occupe uniquement de la partie visible de

l'iceberg, et ne réserve aucune place à la langue parlée par la plupart des Français la plupart du temps. Elle laisse ainsi dans les limbes les principaux intéressés – les millions de locuteurs ordinaires du français – qu'elle semble considérer comme des veaux. Le rôle directeur dans l'évolution des langues a toujours été, non pas les élites, encore moins les suppôts de l'État, mais la collectivité des locuteurs.

Université de Saint-Andrews Anthony LODGE

Ce que la philologie / linguistique romane doit aux savants allemands[1]

En 1898, Gaston Paris, au sommet de la gloire dont il jouit en tant que chef de file de la philologie romane universitaire de France, écrit à propos des Normands et de leur littérature le paragraphe suivant :

> On peut croire que dès avant l'établissement des Danois les habitants de la Neustrie [= Normandie] avaient développé plus fortement que ceux d'autres régions les caractères qui les distinguèrent plus tard, l'esprit positif et clair, le sens pratique, l'amour de l'ordre et de la règle. Les conquérants du Nord [...] y ajoutèrent ce qui leur était propre, la hardiesse, l'esprit d'entreprise, le besoin d'expansion qui se traduisit dans les faits par ces conquêtes extraordinaires où revit la hasardeuse audace des Vikings, dans les idées par une curiosité ouverte de toutes parts; à l'amour de la règle et de l'ordre, qui répondait à leur forte conception du droit, ils joignirent le sentiment d'indépendance personnelle qui leur était inné.[2]

Si la littérature de Normandie mérite sans l'ombre d'un doute toute l'attention des savants d'hier et d'aujourd'hui puisque c'est la plus ancienne en langue d'oïl et puisque c'est depuis la Normandie que la langue et la littérature de France sont parties à la conquête de l'Angleterre, on peut être étonné par la caractérisation des habitants de Neustrie : hasardeuse audace et amour de la règle et de l'ordre, esprit positif et clair et sentiment d'indépendance personnelle inné, c'est une véritable synthèse de toutes les qualités que peut comporter une civilisation moderne. La raison de cette présentation plutôt élogieuse des Normands est toute simple : Gaston Paris parlait devant la Société des Antiquaires de Normandie. Il s'agit d'un travail de commande, tout comme peut l'être une conférence sur l'héritage des savants allemands délivrée à l'Institut Goethe. Les deux sujets sont néanmoins de vrais bons sujets.

En effet, si la littérature de la vieille France a essaimé à partir de la Normandie, la Romanistique a essaimé depuis l'Allemagne. Avec l'Allemagne, nous sommes, sinon au cœur, au moins à l'origine de la Romanistique, à l'époque où les fondations se mettent en place et les *topoi* de l'histoire de la philologie romane s'inventent. C'est le moment rare, unique, où l'universitaire découvre un terrain vierge qu'il se dépêchera de quadriller au point de le transformer en lieu commun. C'est le moment où le savoir s'organise, où l'on instaure de l'ordre grâce aux lois phonétiques, au classement des

---

1 Ce texte, présenté le 17 juillet 2013 à l'Institut Goethe de Nancy, en conserve l'allure orale. Les références bibliographiques et certains développements ont toutefois été ajoutés pour cette version écrite.

2 Paris (1912, 82-83).

langues en familles, aux stemmata qui servent à établir des éditions critiques. C'est le moment où se mettent en place les canons d'une nouvelle discipline universitaire avec ses toutes premières certitudes qui perdurent partiellement jusqu'au jour d'hui et font partie des fondations de notre discipline[3].

Dans cette grande quête du savoir, l'Allemagne joue un rôle important puisque c'est de Berlin que part une nouvelle conception de l'université, celle qu'on désigne aujourd'hui simplement par le nom de son initiateur Wilhelm von Humboldt, dont l'université de Berlin honore le souvenir. La conception humboldtienne inclut tout type de savoir et, surtout, n'a d'autre visée que de permettre à l'Homme de mieux comprendre qui il est. D'où la large place qui est faite aux Lettres et Beaux-Arts. L'histoire de la Romanistique n'est donc qu'un fragment de cette histoire plus grande qu'est l'histoire de l'université moderne qui, depuis l'Allemagne, touchera l'ensemble de l'Europe et de l'Amérique du Nord.

Ici, il ne pourra pas être question de retracer ce rayonnement allemand depuis les origines jusqu'à nos jours. Une liste de tout ce qui a été accompli par la Romanistique allemande n'aurait d'ailleurs qu'un intérêt limité et se transformerait vite en tableau de chasse. Il semble à la fois plus raisonnable et plus intéressant de se borner au moment initial quand naît le concept même de la romanistique universitaire et d'examiner ponctuellement quelques points cruciaux dans les ramifications européennes de cette histoire qui, à l'origine, est donc une histoire allemande.

Toute histoire a un début et, souvent, les histoires se nouent autour d'une figure fondatrice. Pour la Romanistique, c'est naturellement Friedrich Diez qui occupe ce rôle à l'université de Bonn, petite ville universitaire où se rendront le Suisse Adolf Tobler, le Français Gaston Paris et l'Italien Ugo Angelo Canello, sans compter ceux venus d'Allemagne. Mais nous sommes là déjà dans les marges de l'histoire, là où elle se mélange à la légende : cette histoire est construite autour d'une source où coule la parole fondatrice du maître et des pèlerinages de disciples variés qui essaiment ensuite à travers l'Europe pour répandre le nouvel enseignement chez eux. Une configuration presque évangélique ou, au moins, un modèle de *translatio studii*, qui présente, toutefois, une faille, au moins pour ce qui concerne la France. Tobler était naturellement de langue allemande, et Canello, citoyen de l'empire autrichien originaire de Trévise, était lui aussi assez germanophone[4]. Ils ont effectivement suivi, et avec profit, l'enseignement de Diez. Gaston Paris, par contre, lors de son séjour à Bonn, était plus jeune, trop jeune peut-être, et ne maîtrisait pas, à l'époque, la langue de Goethe, c'est-à-dire la langue de Diez. En effet, il avait été envoyé en Allemagne par son père Paulin, professeur au Collège de France, précisément pour qu'il apprenne cette langue indispensable à quelqu'un qui aspirait à s'occuper sérieusement de choses académiques. C'était en 1856-1857, et Gaston Paris avait 17 ans.

---

[3] Les études sur l'émergence de la romanistique en Allemagne constituent désormais un champ très ample. On peut consulter Hirdt (1993) et Christmann (1985). De nombreuses informations de détail sont rassemblées aussi par Kalkhoff (2010).

[4] Canello y séjourna en 1870-71 grâce à une bourse d'étude autrichienne, cf. Formisano (2014).

Tobler, qui était à Bonn au même moment que Gaston Paris, résume la situation comme suit :

> Jeder von uns beiden [Paris et Tobler] hat zwei Semester in Bonn studiert und da selbst neben anderen vortrefflichen Männern auch Diez gehört, ich allerdings insofern im Vorteil, als ich die Landessprache nicht erst zu erlernen brauchte, vier Jahre älter war, vier Semester akademischen Studiums an meiner Heimatuniversität hinter mir und Diezens bis dahin erschienene Werke fleissig durchgearbeitet hatte.[5]

Il y aurait donc bien eu pèlerinage, mais pas de relation particulièrement privilégiée du type maître à élève, et pour des raisons objectives. Si Tobler revient, en 1876, soit 20 ans après les événements, sur ces années de jeunesse, les siennes et celles de Gaston Paris, avec lequel il entretenait d'ailleurs d'excellents rapports, c'est qu'il y avait naturellement derrière la relation à Diez l'enjeu de la filiation de la discipline universitaire entière. Gaston Paris, en effet, revendiquait volontiers pour lui-même le statut d'élève de Diez, mais plus par piété à l'égard de celui dont il s'efforça de faire connaître les travaux en France que pour s'accaparer d'un héritage spirituel. D'autres, en France et en Allemagne, s'en chargèrent, conférant ainsi au voyage allemand de Gaston Paris le but précis de s'initier à la Philologie romane et plaçant du même coup toute l'existence du savant français sous des auspices quasi téléologiques. Ce faisant, ils reconnaissaient à l'université allemande le statut de modèle que les Paris, père et fils, avaient eu la clairvoyance d'identifier, mais pour se l'approprier. Frédéric Masson, successeur de Gaston Paris à l'Académie française, écrit sans ambages : « Il nationalisa cette science qu'il apportait d'Allemagne »[6]. Gaston Paris, nouveau Prométhée, voleur de feu.

D'autres, surtout en France, trouvèrent un autre moyen pour secouer l'influence allemande, en défendant un mythe fondateur alternatif, construit non pas autour de la personne de Diez, mais autour de celle de Raynouard, qui effectivement précéda Diez du point de vue chronologique. François Raynouard était en grand homme en France, qui avant – ou : en plus – de s'occuper de philologie, était surtout un auteur dramatique à succès et un homme politique, emprisonné sous la Terreur, mais fait député sous le Consulat et l'Empire. C'est d'ailleurs en tant qu'écrivain, non en tant que savant, qu'il est élu, dès 1807, à l'Académie française[7]. En tant que savant, né dans le Midi, il se consacra surtout à la poésie lyrique des troubadours, dont il publia, un des premiers, des poèmes.

Mais le fait le plus important est qu'il a eu l'idée d'ouvrir son édition des poèmes des troubadours par une *Grammaire comparée des langues de l'Europe latine dans leurs rapports avec la langue des troubadours*, parue en 1816, soit une vingtaine

---

5 Tobler (1912, 445-446). Le passage est cité également par Bähler (2004, 74) qui fournit aussi un exposé très nuancé des années allemandes de Gaston Paris et de sa relation avec Diez, aux pages 38-48 et 73-88. Sur son séjour en Allemagne, voir aussi Trachsler (2011).

6 Masson (1904, 14), cité d'après Bähler (2004, 70).

7 Par contre, lorsqu'il intègre l'Académie des Inscriptions et des Belles-Lettres, en 1816, c'est bien en tant que savant, il avait abandonné la scène pour la bibliothèque.

d'années avant la *Grammatik der romanischen Sprachen* de Diez, qui reconnaissait volontiers sa dette à l'égard de Raynouard. Dans cette grammaire comparée – donc effectivement écrite dans une perspective de philologie romane – Raynouard estime que la langue occitane conservait l'image la plus authentique d'une sorte de proto-roman, une forme primitive de la langue romane, dont seraient issues toutes les langues romanes modernes, thèse totalement erronée qui a assez vite été abandonnée par la suite[8]. Mais toujours est-il que la tentative d'envisager toutes les langues romanes dans leur relation au latin a pu constituer un titre pour transférer sur Raynouard la gloire dont jouissait Diez. En 1872, le combatif et patriotique Léon Gautier, par exemple, écrit à propos de l'Académicien dans la préface à son édition de la *Chanson de Roland*, qu'il

> [...] avait ressuscité ou plutôt créé, non-seulement en France, mais en Europe, la science de la littérature provençale: c'était le grand Raynouard, que l'imperfection de sa science ne nous empêchera jamais d'admirer comme un des pères de l'érudition française et même allemande.[9]

Voilà qui inverse, un an après la guerre franco-prussienne et dans la préface d'une œuvre que Gautier qualifie de « poème national », les prérogatives, pour faire revenir en France les débuts de la nouvelle discipline. Sous la plume de Gautier, Raynouard devient le père de l'érudition allemande.

Il est vrai que Raynouard est le plus ancien. Il aurait pu être le père de quelque chose, le premier d'une lignée, mais, au moins au sens traditionnel, il n'a pas eu d'élève, si ce n'est qu'il a rencontré, dans l'omnibus entre Paris et Passy, le jeune François Guessard qu'il initia aux principes de la philologie et fit arriver à l'Ecole des Chartes. Guessard allait devenir professeur à l'Ecole Nationale des Chartes, où il ne brilla pas, mais eut pour élève les plus grands romanistes français de cette époque, Gaston Paris et, surtout, Paul Meyer. C'est dans ce sens que Raynouard, indirectement, se trouve tout de même à la tête d'une filiation française[10].

Voilà donc pour les mythes fondateurs concernant la Philologie romane dont les différentes versions montrent assez que le sujet est sensible. Il est sensible parce qu'il met en concurrence deux pays qui se trouvent, après 1870-71, dans un rapport d'émulation, et que l'ordre chronologique instaure un ordre de préséance et un rapport de maître à élève.

Mais laissons ici les mythes fondateurs et tournons-nous vers l'histoire. Contrairement au mythe, qui la fonde et explique, toute véritable histoire a une préhistoire. C'est essentiellement ce qu'il eut avant Raynouard et Diez qui explique que le visage de la première romanistique est allemand et non français. L'universitaire Diez, en effet, vers 1820, travaillait dans un cadre institutionnel très différent que celui que

---

8 Voir, pour le parcours assez surprenant de cet abandon, Rettig (1976).

9 *La chanson de Roland* (1872, clxii).

10 Guessard est évoqué sous un jour sympathique par Monfrin (1958), réimprimé dans Monfrin (2001).

connut l'Académicien Raynouard quelques années auparavant de l'autre côté du Rhin. La Philologie romane en tant que discipline universitaire est allemande parce que l'université moderne toute entière est allemande et que, quand le nouveau modèle de l'enseignement universitaire s'est exporté, certes avec de nombreux ajustements, la romanistique a fait partie du voyage. La clé réside donc dans l'histoire des institutions.

A Bonn, en 1830, Friedrich Diez, se voit installé sur une chaire de *Geschichte der mittleren und neueren Literatur*[11]. Avant d'accéder à sa chaire, Friedrich Diez était *Lektor der Neueren Sprachen*, "lecteur de langues modernes", ce qui, en soi, n'est rien de nouveau. Autrefois, au XVIIIe siècle, cette fonction s'appelait *Sprachmeister*, "maître de langues", comme il y avait des maîtres de danse ou d'équitation, d'ailleurs bien mieux rémunérés que les professeurs de langue. Il y en avait dans presque toutes les universités au XVIIIe siècle, et l'enseignement de la langue, en réalité de plusieurs langues étrangères, leur incombait. Mais cet enseignement était totalement orienté vers la pratique et ne comportait aucune dimension théorique ou historique. Il était en général assuré par des personnes sans diplôme universitaire que les professeurs universitaires regardaient avec condescendance[12]. C'est cela qui allait changer avec Diez et les lecteurs nouveau style.

Sous l'impulsion de Humboldt, des chaires dédiées aux philologies modernes arrivent dans les universités, bien que sous un angle particulier et typiquement humboldtien : pour qu'un tel enseignement ait un sens, il fallait qu'un seul homme puisse le prendre en charge. Cet homme expliquerait et étudierait alors de façon transversale toute la production littéraire d'un point de vue à la fois historico-philologique et esthétique. Face à la philologie classique, face à la philologie orientale, il devait y avoir une philologie moderne, une seule. Une chaire consacrée à une seule langue ne servirait à rien. Au début du XIXe siècle, apparaissent donc les premières chaires de philologie moderne, embrassant, concrètement, la philologie allemande, anglaise et romane. C'est la *Neuphilologie*, la Romanistique est encore loin. En effet, les arguments contre une dissociation des trois entités ne manquaient pas : si on remonte assez haut dans le temps, l'anglais n'existe pas, puisqu'on a affaire à de l'anglo-saxon, donc à du germanique, et les langues romanes accusent elles aussi l'influence des Francs. Par ailleurs, seule une vue comparatiste permet de dégager les spécificités dans chacune des philologies. Éclater la chaire en trois serait donc une erreur. Pas de spécialisation selon les différentes philologies, donc, pour l'instant, mais avènement graduel d'une *Neuphilologie* modelée sur l'*Altphilologie*. Ce qui change par rapport à l'époque des *Sprachmeister*, c'est l'arrivée, à l'université, d'une approche des langues et littératures modernes axée non plus sur une connaissance empirique, mais proprement scientifique, philologique.

---

11 Hirdt (1993). Voir aussi Kalkhoff (2010, 277-285) et passim.

12 Voir, pour l'exemple de Göttingen, Krapoth (2001). Certains éléments sont repris, en français, dans Krapoth (2014). Pour un contexte plus large, voir *L'enseignement du français en Europe autour du XIXe siècle. Histoire professionnelle et sociale, numéro thématique de Documents pour l'histoire du français langue étrangère ou seconde*, 33/34.

C'est ainsi que s'explique l'intitulé de la chaire de Diez évoqué à l'instant : *Geschichte der mittleren und neueren Literatur*. Diez était romaniste par goût, par vocation, par intérêt, mais son poste lui imposa d'assurer un enseignement de *toutes* les philologies, obligation dont il s'acquitta d'ailleurs toujours sans difficulté. Quant à la première chaire de philologie romane, elle sera créée à Berlin, en 1867, et sera occupée par Adolf Tobler[13]. C'est alors seulement, presque quarante ans après la création de la chaire de Diez, la reconnaissance, par le monde académique, de la spécificité, au sein des différentes philologies, anciennes et nouvelles, des langues et littératures romanes. C'est donc en 1867 que la Philologie romane s'affranchit avec succès à la fois de l'emprise de la philologie classique et de celle de la philologie anglaise et germanique, pour accéder à l'autonomie institutionnelle.

Le cheminement de la triade néophilologique – philologie anglaise, germanique, romane – analogue à l'*Altphilologie*, vers l'éventail plus différencié qui est, au fond, encore en vigueur aujourd'hui, passe par plusieurs étapes, dont la plus visible est constituée par la défection de la philologie allemande, qui quitte le trio quand Georg Friedrich Benecke est nommé à Göttingen en 1805 sur la première chaire de philologie germanique, ouvrant ainsi la voie à l'autonomie graduelle de la germanistique et inaugurant, du même coup, l'ère des *Seminare für Englische und Romanische Philologie*. Ces deux philologies, étrangères sur le sol allemand, resteront assez liées pendant quelques décennies encore et prennent le relais des chaires de philologies modernes. Entre 1860 et 1875, une douzaine de ces *Doppelprofessuren* est créée et pourvue en Allemagne[14]. En effet, à cette époque, la plupart des savants que nous romanistes considérons comme étant des nôtres, appartiennent en réalité aussi à nos collègues anglicistes et, plus rarement, germanistes : Karl Bartsch, à Rostock puis à Heidelberg, est germaniste et romaniste, d'autres, comme Edmund Stengel à Marbourg, Eduard Mall à Münster, Gustav Gröber à Breslau, couvrent la philologie anglaise et la philologie romane[15].

Cette évolution conduisant à la différentiation des philologies se lit, par exemple, dans la création des revues scientifiques. On passe ainsi des *Göttingischen Gelehrte Anzeigen* (1739 et 1753) où philologies classique et moderne cohabitaient et côtoyaient des contributions touchant aussi d'autres disciplines des sciences humaines, à l'*Archiv für das Studium der Neueren Sprachen und Literaturen* (1846), puis au *Jahrbuch für romanische und englische Literatur* (1859)[16]. C'est l'émergence de la *Neuphilologie*, puis l'ère des *Doppelprofessuren*. En 1867, à Berlin, Tobler occupe, on l'a dit, la pre-

---

13 Swiggers (2000, 1279). Voir aussi Briesemeister (2001, 565). Sur les circonstances de l'arrivée de Tobler à Berlin, voir Lebsanft (2009, 70-72).

14 Le cas de Marbourg est bien documenté, Finkenstaedt / Haenicke (1994).

15 Outre Christmann (1985), voir le très utile répertoire de Storost (2000). Voir aussi Swiggers (2000) ainsi que les tableaux chez Kalkhoff (2010).

16 Certaines de ces publications de la première heure ont fait l'objet d'études précises au moment de leurs anniversaires respectifs. Voir, par exemple, Ringleben (2002).Voir aussi les publications citées *infra*, à la note suivante.

mière chaire de Philologie Romane, par voie de conséquence, quelques années plus tard, sont créées la *Zeitschrift für romanische Philologie* (1877) et les *Romanische Forschungen* (1883), qui documentent l'indépendance de notre discipline par rapport aux autres philologies modernes, tout comme les collections spécifiques pour les éditions de textes qui apparaissent au même moment : la *Bibliotheca Normannica* (Halle, 1879), l'*Altfranzösische Bibliothek* (Heilbronn, 1881) et la *Romanische Bibliothek* (Halle, 1888)[17]. Après la chaire de Tobler, d'autres se créent, souvent, d'ailleurs, par le biais de la scission d'une *Doppelprofessur* existante, laissant au professeur en poste son domaine de prédilection et recrutant un nouveau collègue, en général avec rang d'*Extraordinarius*, au profil complémentaire : parmi les premières chaires dédiées exclusivement à philologie romane figurent Munich (Konrad Hoffman 1869), Zurich (Gustav Gröber 1872), Strasbourg (Eduard Böhmer 1872) Leipzig (Adolf Ebert 1873). Il vaut la peine de s'arrêter un instant sur le cas de Strasbourg, où a s'est produit un malentendu, peut-être volontaire, qui traduit très précisément le changement en cours. Les instances strasbourgeoises avaient en effet obtenu la permission des ministères berlinois pour entrer en négociation pour un poste d'*eines Lehrers für die romanischen Sprachen und für die englische Sprache*, "un professeur de langues romanes et de langue anglaise", clairement compris, de la part du ministère, comme une *Doppelprofessur*, ce que l'université de Strasbourg fit en toute bonne foi, proposant un poste à Eduard Boehmer pour la philologie romane et un autre à Bernhard ten Brink pour l'anglais[18]. C'était la voie de la modernité que ce passage de la *Doppelprofessur* à la chaire dédiée à une seule philologie, et c'est probablement pour cela que Bismarck, à Berlin, a suivi le raisonnement et a choisi d'implanter, dans cette ville sensible qu'était Strasbourg, une chaire consacrée exclusivement à la philologie romane.

En l'espace de quelques décennies, tout a donc été mis en place. A partir de 1870, en gros, on commence à avoir les chaires, les revues suivent dans les années 1880 et, en dernier lieu, on procède aussi à la création d'instituts dédiés plus ou moins aux philologies et aux disciplines qui existent encore aujourd'hui. C'est la scission administrative, permettant le passage des *Seminare für Englische und Romanische Philologie* à la *Einzelphilologie*, qui a été le plus long à réaliser – car il fallait dissocier aussi les bibliothèques et trouver de nouveaux bâtiments –, mais en gros, l'évolution est accomplie durant les deux premières décennies du XXe siècle, et la Philologie

[17] Sur le rôle des revues scientifiques, voir le volume *Tra Ecdotica e Comparatistica : Le riviste e la fondazione della Filologia Romanza* (Meneghetti / Tagliani (ed.) 2009), contenant des contributions sur la *Romania*, la *Zeitschrift für romanische Philologie* et plusieurs grandes revues italiennes. On notera d'ailleurs que cette évolution est assez générale parce qu'elle atteint aussi, avec un temps de retard, les États-Unis, où *The Romanic Review* (fondée en 1910), première revue consacrée à la philologie romane, fait suite aux *Modern Language Notes* (1884), *Publications of the Modern Language Association of America* (1885), *Modern Philology* (1903). Sur ce point, voir Malkiel (1972) et sur la *Romanic Review*, voir le fascicule jubilaire, 100/1-2 (Jan-March 2010), contenant notamment la contribution de Cerquiglini-Toulet (2010).

[18] Voir la contribution de Rudolf (2014).

Romane jouit depuis lors non seulement d'un statut reconnu dans l'enseignement universitaire, mais aussi d'une existence juridique, administrative, financière et physique, avec des chaires, du personnel, des crédits, des locaux, des bibliothèques et des textes, voire des masses de textes, régissant jusqu'au dernier détail son fonctionnement. Pour preuve, ce passage, tiré du règlement interne du *Seminar für Romanische Philologie* de Göttingen, qui date de 1913 : *Wer das Zimmer als Letzter verläßt, hat das Gaslicht auszudrehen*. Si l'administration a pensé à cela pour vous, cela veut dire que vous existez officiellement, et elle aussi.

Voilà, donc, pour la Romanistique à la l'allemande telle qu'elle existe au sein de la nouvelle université humboldtienne. C'est globalement sous cette forme qu'elle sera importée, avec le système universitaire dont elle fait partie, par les pays voisins. Il ne sera pas possible d'évoquer ici tous les cas de figure que peut présenter ce processus d'importation et je me limiterai surtout au cas de la France. D'une part, parce que c'est le domaine que je connais le moins mal et d'autre part parce qu'on voit là le plus nettement les limites qui, dès le début, sont inhérentes au concept même de la Romanistique du modèle allemand.

Au premier abord, un regard sur la situation française dans la décennie de 1870, pourrait inviter à démentir l'influence allemande. Avec la *Collection des Poètes Français du Moyen Âge* (Paris, 1859), celle des *Poètes de Champagne antérieurs au XVIe siècle* (Paris, 1847), et les *Anciens Poètes de France* (Paris, 1858), la France dispose de collections d'éditions de textes spécifiques pour l'ancienne littérature de France bien avant l'Allemagne, de même que l'enseignement de la philologie médiévale se pratique, à l'École des Chartes dès 1829, et qu'au Collège de France, Paulin Paris enseigne la Littérature Médiévale Française depuis 1853. En réalité, il s'agit là de collections de textes méritoires, mais totalement préscientifiques, et ni l'Ecole des Chartes ni le Collège de France n'offrait un enseignement universitaire humboldtien à proprement parler. C'est précisément lorsqu'on se rend compte, pour reprendre le mot célèbre de Léon Gautier, que la seule ville de Marbourg compte plus de spécialistes de la chanson de geste que toute la France, que les choses se débloquent : devant l'impossibilité de réformer la Sorbonne, on crée de toutes pièces l'Ecole Pratique des Hautes Etudes qui dispense un enseignement explicitement calqué sur le modèle allemand où Gaston Paris, en 1868, commence ses séminaires de Philologie romane. En 1872, le même Gaston Paris fonde avec son ami Paul Meyer la revue *Romania*, et les deux sont aussi à l'initiative de la création de la Société des Anciens Textes Français en 1875, la première collection vraiment scientifique pour des éditions de textes. C'est donc, comme en Allemagne, à partir de 1870 que se fait l'introduction de la philologie romane dans le paysage universitaire français. Le fait que l'on soit presque en synchronie avec les créations des premières chaires allemandes, alors qu'il n'y avait, en France, aucune base comparable à la *Neuphilologie* allemande, s'explique précisément par ce retard : quand on comprend brusquement, côté français, les enjeux, on prend le train en marche sans se soucier des diverses mues qu'a connues la *Neuphilologie* en Allemagne, pour adopter directement le système dans son dernier état, celui

des *Einzelphilologien*, sans passer par la case néophilologique – philologie anglaise, allemande, romane – et gagnant ainsi une bonne cinquantaine d'années.

Il faut tout de même remarquer que l'introduction de la Philologie Romane en France donne lieu à un léger ajustement par rapport à la pratique dans les universités d'outre-Rhin : en France, on cultive beaucoup plus l'étude de la Philologie Française que Romane. En effet, dans l'Hexagone, aujourd'hui comme alors, on n'enseigne pas, selon le modèle allemand, la Philologie romane, mais la langue et la littérature françaises, sauf à Paris et dans de rares universités du Sud[19]. Les deux revues de philologie romane de France, la *Romania* (fondée en 1872) et la *Revue des Langues Romanes* (1870) à Montpellier figent précisément l'état programmatique des débuts, qui ne s'est jamais vraiment concrétisé au niveau institutionnel, mais reflète avec force le modèle allemand[20].

On touche ici à quelque chose de très fondamental de la conception de la Romanistique à l'allemande. Dans l'université allemande, la *Romanische Philologie* comprenait et comprend la science de la langue et de la littérature, *Sprach- und Literaturwissenschaft*, dans toutes les langues et dialectes romans, du portugais au roumain, du picard au sicilien. Ailleurs, même quand la désignation paraît identique – *filología románica*, *filologia romanza*, philologie romane – le contenu n'est pas exactement le même. En général, il y a une sorte de rétrécissement qui concerne soit la période – en général, on se limite au Moyen Age – ou les langues concernées – en général, on privilégie celle du pays. Ce qui a fait dire à Fritz Nies lorsqu'il présidait le *Deutsche Romanistenverband*, l'Association des romanistes allemands :

> Romanistik müßte als Kuriosität in das Artenschutz-Abkommen aufgenommen werden: in freier Wildbahn lebt sie einzig im deutschsprachigen Raum, hat aber mit diesem nichts zu tun; sie ist in ihrem Bestand gefährdet; und nach ihren prächtigsten Platzhirschen zu schließen, würde durch ihr Verschwinden die geisteswissenschaftliche Fauna merklich verarmen[21].

> La romanistique devrait être intégrée à titre de curiosité dans la Convention sur le commerce international des espèces de faune et de flore sauvages: dans la nature, elle ne vit que dans l'espace germanophone, mais n'a rien à voir avec ce dernier; sa population est menacée;

---

19 Cette tendance est ancienne et s'observe particulièrement bien dans le domaine de la langue, comme le montrent les dépouillements de Bergounioux (1991).

20 À vrai dire, les deux revues ne sont pas tout à fait équivalentes : contrairement à la *Romania*, qui se veut internationale, universitaire et professionnelle, la *Revue des Langues Romanes* des débuts est fortement enracinée dans la culture occitane et fait volontiers une place à des poètes contemporains et des érudits locaux, et ne regarde pas sans hostilité vers Paris, d'où les grands professeurs contemplent, et commentent, ce qui se publie dans la *Revue des Langues Romanes* avec un mépris occasionnel. Le premier volume de la *Revue des Langues Romanes* devait d'ailleurs lui aussi s'ouvrir par quelques pages un peu programmatiques, mais comme les rédacteurs – de façon symptomatique – les avaient demandées à M. Saint-René Taillandier, une éminente figure politique du Midi, elles n'ont jamais été livrées : en 1871, dans la confusion qui a suivi la guerre franco-prussienne, M. Saint-René Taillandier a été appelé à Paris comme conseiller d'état.

21 Nies (1988).

> et, à en juger par ses mâles dominants les plus impressionnants, la faune intellectuelle se trouverait considérablement appauvrie si elle venait à disparaître.

Le point essentiel, ici, est le simple constat que la Romanistique ainsi comprise n'a rien à voir avec le pays qui la cultive. Contrairement aux pays qui l'entourent, l'Allemagne ne parle aucune langue romane et la discipline ne bénéficie donc d'aucun ancrage naturel. Par conséquent, elle ne dispose pas non plus d'un appui politique, à la différence des pays de langue romane où la discipline peut toujours s'adosser à la langue – ou une langue – nationale et se trouve, de ce fait, protégée par un ensemble, que l'on ne peut pas défaire sans toucher du même coup la langue nationale. Cette conception des langues romanes en tant que système protège, par exemple, l'ancien français, qui tend à disparaître même des études de français, dès lors qu'on l'envisage comme indissociable de l'ancien italien et l'ancien espagnol auxquels on peut plus difficilement toucher en Italie ou en Espagne. Il est d'autant plus frappant de constater que c'est précisément en Allemagne qu'on continue à cultiver la conception originelle d'une famille de langues unitaire et solidaire et d'une approche scientifique qui ne conçoit que depuis peu le partage entre *Sprach-* et *Literaturwissenschaft*. Cette conviction concernant l'unité fondamentale de leur discipline a permis aux romanistes allemands de concevoir, et souvent de réaliser, des grandes entreprises comme le *Romanisches Etymologisches Wörterbuch* de Wilhelm Meyer-Lübke ou le *Grundriss der romanischen Philologie* de Gustav Gröber, qui fait, à l'aube du XXe siècle, déjà, le bilan des premières décennies de sa discipline. Ces monuments, et d'autres, comme le dictionnaire de l'ancien français de Tobler ou le *Französisches Etymologisches Wörterbuch* de Walther von Wartburg, étaient dès leurs débuts tellement indispensables à tous que le jeune Mario Wandruszka, à sa grande surprise, à la rentrée universitaire 1928 en Sorbonne, entend son enseignant français recommander à tous les étudiants l'apprentissage de la langue allemande[22].

Pourtant, l'unité fondamentale de la discipline, ou, mieux, la *possibilité* de pouvoir l'embrasser dans sa totalité, a périodiquement été questionnée par les chercheurs les plus autorisés, en dehors de l'Allemagne et en Allemagne même. Il est frappant de voir que ce sentiment d'impuissance est apparu très tôt. Jacques Monfrin, dans sa leçon inaugurale à l'Ecole des Chartes en 1958, situe l'éclatement de la discipline peu après sa naissance, au moment où s'éteint la première génération de romanistes, la seule qui ait pu réellement prétendre embrasser toute la romanistique. Faisant le bilan de l'activité scientifique de Paul Meyer, il écrit :

> En 1869, Paul Meyer était la philologie romane; il n'avait qu'à aller de l'avant, de toute sa force. En 1916, des choix s'imposaient. La philologie romane était une science complexe, aux multiples liaisons.

[22] Wandruszka rappelle l'anecdote en ouverture de son article (Wandruszka 1988, 27).

> Depuis longtemps, nul ne pouvait plus songer à embrasser d'un seul regard l'ensemble du domaine roman. Désormais, un seul homme n'avait guère la possibilité d'en dominer qu'une province ou deux[23].

Dès Paul Meyer, « depuis longtemps », la Romanistique se serait donc complexifiée et ramifiée au point de devenir indomptable. Il est clair que ce sentiment est largement partagé au sein de la communauté scientifique allemande elle-même où l'on plaide pour une spécialisation. Non seulement les chercheurs que nous sommes ne pouvons plus tout suivre, mais, surtout, les enseignants que nous sommes ne pouvons plus rien proposer de cohérent à nos étudiants, qui doivent balayer, en quelques heures d'enseignement hebdomadaire, une discipline aux dimensions amazoniennes. Du point de vue didactique, la Romanistique, comme le rappelle avec force Hans Rheinfelder en 1960, est en effet une exception, une anomalie :

> Die deutschen Universitäten haben Lehrstühle für Deutsche Philologie; an den Höheren Schulen Deutschlands gibt es Deutsch-Unterricht. Die deutschen Universitäten haben Lehrstühle für Englische Philologie; an den Höheren Schulen gibt es Englisch-Unterricht. Die Universitäten haben Lehrstühle für Romanische Philologie, d. h. für die Geschichte der neun romanischen Sprachen und Literaturen; an den Höheren Schulen gibt es Unterricht im Französischen, Spanischen und Italienischen[24].

L'approche des langues et littératures romanes en tant que système qui ne livre sa cohérence qu'à condition d'être envisagé dans sa globalité, c'est-à-dire ce qui faisait au moment de la création de la discipline sa force, se révèlerait désormais inapte aux exigences de l'université moderne. Il convient, dit-on, d'acter enfin la transformation du romaniste à l'ancienne en « romanistischer Einzelsprachler »[25].

Il ne faut pas en déduire que la Romanistique, le legs que l'Allemagne de la fin du XIXe siècle a fait au monde, a vécu. La *Vollromanistik* existe bel et bien, et les grandes entreprises pan-romanistes qui se sont récemment réalisées sous impulsion allemande, comme le *Lexikon der Romanischen Linguistik* ou les trois gros volumes de la *Romanische Sprachgeschichte*, témoignent de la fécondité du concept. Sans doute la *Romania* semble plus cohérente quand on la regarde du point de vue du linguiste plutôt qu'en historien de la littérature, mais l'enlisement du nouveau *Grundriss*

---

23 Monfrin (1958, 186). Mario Wandruszka (1988, 29) lui fait écho : « Länger als andere haben die deutschen Romanisten an der Idee der Gesamtromania als einer umfassenden Einheit, als eines einzigen großen Hauses mit vielen Wohnungen festgehalten. In den meisten Ländern hat die immer stärkere Beschränkung auf die Erforschung einer einzigen romanischen Sprache, ihrer Geschichte und Gegenwart, ihrer verschiedenen regionalen, sozialen, kulturellen Varietäten und Variationen dazugeführt, daß es dort kaum noch Romanisten in unserem Sinn gibt. » Stempel, dans le même volume, compare la Philologie romane à l'actrice principale d'une pièce de théâtre absurde qui lui serait dédiée, alors qu'elle aurait disparu depuis longtemps : « hier ist die romanische Philologie […] alten Zuschnitts tatsächlich im Gespräch, aber sie ist als Lehrinhalt wie durch die damit beschäftigten Protagonisten der Lächerlichkeit preisgegeben und taugt nur noch für (allerdings sehr tiefgründige) Nonsens-Veranstaltungen. » (Stempel 1988, 41).

24 Rheinfelder (1960, 201).

25 Le terme est de Wandruszka (1988, 29).

der romanischen Literaturen des Mittelalters, ouvrage qui devait porter sur l'époque où les différentes littératures étaient encore censées refléter une unité première, ne suffit pas pour prouver que le concept de Romanistique, en littérature, est devenu inopérant. Il est clair, toutefois, que la Romania, par rapport à l'époque de Diez et de Tobler, s'est agrandie : essentiellement européenne au début du XXe siècle, elle est aujourd'hui mondiale. L'Amérique latine a détrôné la Péninsule ibérique, la francophonie a pris le relais du classicisme français et les *Colonial Studies* invitent à regarder les textes non plus depuis le centre, mais depuis la marge[26]. Après le structuralisme, les courants post-modernes, le féminisme, les *gender studies*, les *queer studies*, ont apporté d'autres paradigmes interprétatifs à la palette des romanistes. Il est difficile, en effet, d'y voir encore l'unité centripète que dégageait le *Grundriss* de Gröber, voire celle que pouvaient encore incarner ces personnalités d'exception qu'étaient Spitzer et Auerbach, sans doute les derniers qui remplissaient la discipline de l'intérieur, au moment où ils apportaient aux Etats-Unis non seulement une certaine forme de Romanistique, mais une certaine forme de philologie tout court. La Romanistique allemande était alors le modèle pour les études de lettres en général et c'est donc à travers la Romanistique que le Continent nord-américain a rencontré – dans le cas de Spitzer, on devrait dire heurté – une nouvelle façon de lire les textes. C'était il y a longtemps, c'était hier.

L'erreur la plus grave que la Romanistique puisse faire aujourd'hui, si elle entend préserver ce qui lui reste de son identité, serait de se chercher quelque part, guidée par une théorie littéraire exogène, un îlot encore inoccupé pour s'y installer un temps avant de passer au prochain îlot, plus exotique encore. Là ne peut pas être l'avenir d'une discipline qui, si elle doit avoir un sens, a plutôt besoin de se recentrer que de se diversifier. D'ailleurs, la romanistique allemande a déjà trouvé le moyen de se reforger une identité commune en s'interrogeant sur son histoire. Les études sur l'histoire de la discipline, y compris pendant les périodes difficiles et douloureuses, prouvent, si besoin était, que la romanistique est bien une histoire allemande. En même temps, elles attestent aux romanistes une conscience particulière concernant leur propre faire et les conséquences de leurs choix idéologiques qui peut bien servir de modèle à d'autres pays.

A bien y réfléchir peut-être même qu'à son insu, à l'insu de nous tous, la romanistique allemande n'a jamais cessé d'être en phase avec ses origines : c'est à Heidelberg que se fait le *Dictionnaire Etymologique de l'Ancien Français*, et le *Lessico Etimologico Italiano* se fait à Sarrebruck, tous deux financés dans le cadre de projets d'Académies, et ce depuis les années 1970. La romanistique allemande n'en tire que la gloire. *Ars gratia artis*, comme toute université qui se respecte, même si elle ne se situe ni en Grèce ni en Italie, entretient un département d'archéologie classique. Aucune retombée politique positive n'est à attendre puisqu'on s'occupe du patrimoine des autres, non du sien. A la limite, précisément pour cette raison-là, les porteurs des projets

[26] Pour une présentation très accessible, voir von Stackelberg (1991).

essuient les critiques des instances politiques qui leur reprochent un mauvais usage de l'argent public. Un contribuable allemand peut comprendre pourquoi il faut financer l'étude étymologique de l'allemand, mais pas du français ou de l'italien. Pourquoi donc le faire?

La romanistique allemande s'occupe d'étymologie française et italienne, avec lesquelles elle n'a, pour reprendre l'expression de Fritz Nies, *nichts zu tun*, parce qu'elle perpétue ici la tradition humboldtienne, où la connaissance est une fin en soi et le savoir est universel. C'est peut-être l'enseignement le plus éclatant que la romanistique allemande délivre, depuis 150 ans, aux romanistes du monde.

Université de Zurich Richard TRACHSLER

Bibliographie

Bähler, Ursula, 2004. *Gaston Paris et la Philologie Romane*, Genève, Droz, Publications Romanes et Françaises 234.

Bergounioux, Gabriel, 1991. « L'Introduction de l'ancien français dans l'université française (1870-1900) », *Romania* 112, 243-258.

Briesemeister, Dieter, 2001. « Geschichte des Faches Romanistik an den Hochschulen. Deutschland », in: Holtus, Günter / Metzeltin, Michael / Schmitt, Christian (ed.), *Lexikon der Romanistischen Linguistik (LRL)*, vol. I, 1, Tübingen, Niemeyer, 562-574.

Cerquiglini-Toulet, Jacqueline, 2010. « La littérature médiévale dans l'histoire littéraire et la théorie : l'exemple de la *Romanic Review* », *Romanic Review : 1910-2010. A hundred years of romance studies*. Special issue : *Romanic Review* 100/1-2 (Jan-March 2010), 91-100.

Christmann, Hans Helmut, 1985. *Romanistik und Anglistik an der deutschen Universität im 19. Jahrhundert. Ihre Herausbildung als Fächer und ihr Verhältnis zu Germanistik und klassischer Philologie*, Mainz-Stuttgart, Steiner, Akademie der Wissenschaften und der Literatur, Abhandlungen des Geistes- und Sozialwissenschaftlichen Klasse, année 1985, n°1.

Finkenstaedt, Thomas / Haenicke, Gunta, 1994. « Gutachten des Prof. Tobler in Berlin betreffend die Vorschläge des Prof. Stengel in Marburg wegen Gründung von Universitätsseminarien für romanisch-englische Philologie », in: Baum, Richard / Böckle, Klaus / Hausmann, Frank Rutger / Lebsanft, Franz (ed.), *Lingua et Traditio. Geschichte der Sprachwissenschaft und der neueren Philologien. Festschrift für Hans-Helmut Christmann zum 65. Geburtstag*, Tübingen, Narr, 481-488.

Formisano, Luciano, 2014. « La romanistique allemande et l'Italie au XIXe siècle », in: Trachsler, Richard (ed.), *Bartsch, Foerster et Cie. La première romanistique allemande et son influence en Europe*, Paris, Classiques Garnier, Rencontres. Secteur Moyen Age. Civilisation Médiévale, 129-143.

Hirdt, Willi (ed.), 1993. *Romanistik. Eine Bonner Erfindung,* Bonn, Bouvier, 2 vol. (I : Présentation ; II : Documentation).

Kalkhoff, Alexander, 2010. *Romanische Philologie im 19. und frühen 20. Jahrhundert. Institutionengeschichtliche Perspektiven*, Tübingen, Narr, Romania Monacensia 78, 277-285.

Krapoth, Hermann, 2001. « Die Beschäftigung mit romanischen Sprachen und Literaturen an der Universität Göttingen im 18. und frühen 19. Jahrhundert », in: Lauer, Reinhard (ed.), *Philologie in Göttingen. Sprach- und Literaturwissenschaft an der Georgia Augusta im 18. und beginnenden 19. Jahrhundert*, Göttingen, Vandenhoeck & Ruprecht, Göttinger Universitätsschriften, Serie A : Schriften, 18, 57-90.

Krapoth, Hermann, 2014. « Les premiers séminaires de philologie romane. L'exemple de Göttingen », in: Trachsler, Richard (ed.), *Bartsch, Foerster et Cie. La première romanistique allemande et son influence en Europe*, Paris, Classiques Garnier, Rencontres. Secteur Moyen Age. Civilisation Médiévale, 23-36.

La chanson de Roland, 1872. *Texte critique, accompagné d'une traduction nouvelle et précédé d'une introduction historique*, par Léon Gautier, Tours, A. Mame.

Lebsanft, Franz, 2009. « Adolf Tobler (1835-1910). I : ‹ Der gesamte Reichtum der Menschennatur › », in: Bähler, Ursula / Trachsler, Richard (ed.), *Portraits de médiévistes suisses (1850-2000). Une profession au fil du temps*, Genève, Droz, Publications Romanes et Françaises 246, 61-95.

L'enseignement du français en Europe autour du XIXe siècle. Histoire professionnelle et sociale, numéro thématique : *Documents pour l'histoire du français langue étrangère ou seconde*,

33/34 (2005) version mise en ligne le 01 janvier 2012, consultée le 13 juillet 2013. <http://dhfles.revues.org/1647>

Malkiel, Yakov, 1972. « The First Quarter-Century (and Some Antecedents) », *Romance Philology* 26, 3-15.

Masson, Frédéric, 1904. *Discours prononcé dans la scéance publique tenue par l'Académie française pour la réception de M. Frédéric Masson, le jeudi 28 janvier 1904*, Institut de France, Académie française, Paris, Firmin-Didot.

Meneghetti, Maria Luisa / Tagliani, Roberto (ed.), 2009. *Tra Ecdotica e Comparatistica : Le riviste e la fondazione della Filologia Romanza,* Firenze, Edizioni del Galluzzo per la Fondazione Ezio Franceschini.

Monfrin, Jacques, 1958. « Leçon d'ouverture du cours de philologie romane à l'École des chartes (6 novembre 1958) », *Bibliothèque de l'École des chartes* 116, 170-193.

Monfrin, Jacques, 2001. *Études de philologie romane*, Genève, Droz, Publications Romanes et Françaises 230, 3-20.

Nies, Fritz, 1988. « Die Zukunft eines 'unmöglichen Fachs' », in: Nies, Fritz / Grimm, Reinhold R. (ed.), *Ein 'unmögliches Fach': Bilanz und Perspektiven der Romanistik*, Tübingen, Narr, 9-12.

Paris, Gaston, 1912. « La Littérature normande avant l'annexion (912-1204) », *Mélanges de Littérature française du Moyen Age*, publiés par Mario Roques, Paris, Champion. [Discours lu à la séance publique de la Société des Antiquaires de Normandie, le 1er décembre 1898].

Rettig, Wolfgang, 1976. « Raynouard, Diez und die romanische Ursprache », in : Niederehe, Hans-Josef / Haarmann, Harald (ed.), *In Memoriam Friedrich Diez, Akten des Kolloquiums zur Wissenschaftsgeschichte der Romanistik (Actes du Colloque sur l'Histoire des Etudes Romanes / Proceedings of the Colloquium for the History of Romance Studies), Trier, 2.-4. Okt. 1975*, Amsterdam, John Benjamins, 247-273.

Rheinfelder, Hans, 1960. « Vom Notstand der romanischen Philologie », *Die Neueren Sprachen*, neue Folge 9, 201-208.

Ringleben, Joachim, 2002. « Über die Anfänge der Göttingischen Gelehrten Anzeigen », in: Smend, Rudolf (ed.), *Die Wissenschaften in der Akademie. Vorträge beim Jubiläumskolloquium der Akademie der Wissenschaften zu Göttingen im Juni 2000*, Göttingen, Vandenhoeck & Ruprecht, 345–355.

Romanic Review : 1910-2010. A hundred years of romance studies. Special issue : *Romanic Review* 100/1-2 (Jan-March 2010).

Rudolf, Winfried, 2014. « Bernhard ten Brink and Edmund Max Stengel – Two Pupils of Friedrich Diez and Their Contributions to the Study of Medieval English in the National Discourse of the Nineteenth Century », in: Trachsler, Richard (ed.), *Bartsch, Foerster et Cie. La première romanistique allemande et son influence en Europe*, Paris, Classiques Garnier, Rencontres. Secteur Moyen Age. Civilisation Médiévale, 223-246.

Stempel, Wolf-Dieter, 1988. « Die schwierige Einheit der romanischen Philologie », in: Nies, Fritz / Grimm, Reinhold R. (ed.), *Ein 'unmögliches Fach': Bilanz und Perspektiven der Romanistik*, Tübingen, Narr, 41-58.

Storost, Jürgen, 2000. « Die 'neuen Philologien', ihre Institutionen und Periodica : eine Übersicht », in: Auroux, Sylvain / Koerner, E.F.K. / Niederehe, Hans-Josef / Versteegh, Kees (ed.), *History of the Language Sciences*, Berlin-New York, De Gruyter, Handbücher zur Sprach- und Kommunikationswissenschaft, 18.2, 1240-1272.

Swiggers, Pierre, 2000. « Les débuts et l'évolution de la philologie romane au XIXe siècle, surtout en Allemagne », in: Auroux, Sylvain / Koerner, E. F. K. / Niederehe, Hans-Josef / Versteegh,

Kees (ed.), *History of the Language Sciences*, Berlin-New York, De Gruyter, Handbücher zur Sprach-und Kommunikationswissenschaft, 18.2, 1272-1285.

Tobler, Adolf, 1912 [1876[1]]. « Briefe von Gaston Paris an Friedrich Diez », *Vermischte Beiträge*, Leipzig, Hirzel, vol. 5, 443-475.

Trachsler, Richard, 2011. « Portrait des fernen Freundes. Die Briefe von Léopold Jacottet an Gaston Paris », in: Overbeck, Anja / Schweickard, Wolfgang / Völker, Harald (ed.), *Lexikon, Varietät, Philologie. Romanistische Studien zum 65. Geburtstag von Günter Holtus*, Berlin, De Gruyter, 733-740.

von Stackelberg, Jürgen, 1991. « Zu neuen Ufern… Die französischsprachige Literatur Afrikas, der Antillen und Kanadas als romanistischer Studiengegenstand », in: Engelbert, Manfred / Knauf, Michael (ed.), *Romanistik und Französische Revolution : 200 Jahre Französische Revolution — 100 Jahre Seminar für Romanische Philologie Göttingen*, Bochum, AKS Verlag, Fremdsprachen in Lehre und Forschung 9, 29-42.

Wandruszka, Mario, 1988. « Deutsche Romanistik: Kritische Bilanz und Perspektive », in: Nies, Fritz / Grimm, Reinhold R. (ed.), *Ein 'unmögliches Fach' : Bilanz und Perspektiven der Romanistik*, Tübingen, Narr, 27-39.

Les noms de famille en Lorraine

Quelques mots tout d'abord sur ce que sera et ne sera pas cette présentation. Contrairement à ce que son titre pouvait laisser craindre, elle ne prendra pas la forme d'une longue liste monotone de noms de famille suivis de leur étymologie. Son ambition n'est en effet pas de donner un tableau des noms de famille en Lorraine le plus exhaustif possible, mais de faire découvrir la méthode de l'étude de ces noms, d'exposer la pratique pour établir leur origine. Son objectif est également de proposer une réflexion sur cette pratique et sur ces noms eux-mêmes. Cette explication méthodologique s'appuiera sur de nombreux exemples lorrains ; elle sera suivie d'un tour d'horizon des noms de famille sur lesquels nous avons été amenée à travailler (Wirth 2007)[1].

Les noms de famille en question sont essentiellement des noms de famille d'origine romane ou de la Lorraine romane – par opposition aux noms de famille de la zone germanophone – pour deux raisons : d'une part, parce que cette conférence prend place dans le cadre d'un congrès axé sur l'étude des langues romanes, d'autre part, parce que, n'étant pas germaniste, nos compétences pour l'étude des noms trouvant leur origine dans cette partie de la Lorraine est très réduite. Nous éviterons donc de nous avancer sur un terrain qui n'est pas le nôtre.

1. De la méthode avant toute chose

L'entrée dans le vif du sujet se fera à travers le nom de famille (ci-après NF) *Poirot*, particulièrement fréquent dans les Vosges et en Meurthe-et-Moselle (ci-après MM). Pour tenter de connaître son origine, le non spécialiste pourra consulter principalement deux ouvrages destinés au grand public, le *Dictionnaire étymologique des noms de famille et prénoms de France* d'Albert Dauzat (1978[6]), dont la première édition date de 1951 mais qui a été réédité plusieurs fois depuis cette date ; cet ouvrage se trouve aisément dans les librairies. Le second est le *Dictionnaire étymologique des noms de famille* de Marie-Thérèse Morlet (1997[2]), dont la première édition date de 1991. C'est une version revue et augmentée du dictionnaire de Dauzat, Morlet ayant été son élève. Voici donc ce que l'on trouve respectivement dans ces deux dictionnaires pour *Poirot* :

> **Poirier++, Poirrier** (nord d'Abbeville ; **Poirriez** depuis 1785) et **Dupoirier** : arbre caractéristique de la propriété. V. Périer. – **Poirot**, " petite poire ", surnom probable de marchand de poires.

---

[1] Nos remerciements vont aux organisateurs du Congrès international de linguistique et de philologie romanes pour leur invitation.

> **Poirier**, var. **Poirrier**, **-iez** (Picardie), **Poirié** (S.O.), avec l'art. contr. **Dupoirier**, arbre caractéristique de la propriété. DIMIN. : (de *poire*), **Poirot**, surnom de fruitier, composé **Poircuitte**, surnom de marchand de poires cuites. DIMIN. : [de *poirier*], **Poirion**.

Que peut-on dire de ces deux articles ? Il est possible, en premier lieu, d'observer quelques différences : Dauzat glose ainsi le NF *Poirot* par « petite poire », en précisant qu'il s'agit d'un « surnom probable de marchand de poires », tandis que Morlet en fait un diminutif de *poire* et un « surnom de fruitier ». Si on se place sur le seul plan de la vraisemblance, on peut imaginer un surnom de fruitier ; il est en revanche beaucoup plus difficile d'accepter l'idée que des marchands aient pu être spécialisés dans certains fruits uniquement, comme le suggère Dauzat avec son « surnom probable de marchand de poires ». Des doutes peuvent donc déjà être émis sur cette première explication, rien qu'en tentant de se représenter ce métier ; Morlet a sans doute eu les mêmes, puisqu'elle élargit ce surnom à celui d'un fruitier en général. On ne peut donc pas dire que ces explications soient vraiment convaincantes.

Un autre problème peut être souligné : celui de la façon dont sont présentées ces notices étymologiques. Dans le dictionnaire de Dauzat, le NF *Poirot* est suivi de ce qui ressemble à une définition, « petite poire », comme pour un substantif. Or *Poirot* n'est évidemment pas un substantif, mais un nom propre ; ce qui pourrait ou aurait pu avoir le sens de « petite poire », c'est le surnom initial, à l'origine du NF, pas le NF lui-même qui en est la cristallisation. Pour être exacte et si on suit la logique de Dauzat, la formulation aurait dû être du type « le nom de famille *Poirot* a pour étymon le nom commun *poirot*, qui signifie 'petite poire' ». La même remarque peut être faite pour le dictionnaire de Morlet : *Poirot* ne peut pas être issu directement de *poire*.

D'autres points sont eux aussi sujets à caution. Parmi ceux-ci, l'absence d'éléments permettant d'appuyer les motivations avancées, comme des arguments logiques, des citations de textes ou de dictionnaires attestant l'existence d'un substantif *poirot*, des renvois à des travaux préalables ayant été menés sur ce nom, des attestations anciennes, des formations parallèles, etc. En résumé, l'étymologie proposée par ces deux articles n'est pas convaincante. Alors comment procéder pour établir l'étymologie de ce NF sur des bases solides ?

Pour y parvenir, nous allons nous appuyer sur la méthodologie développée dans le cadre d'un projet de recherche européen intitulé PatRom (*Patronymica Romanica. Dictionnaire historique de l'anthroponymie romane*), fondé en 1987 par Dieter Kremer, professeur à l'Université de Trèves. Les grands principes qui sous-tendent cette entreprise ont été élaborés à partir de cette même année lors d'un premier congrès international. Comme son nom l'indique en partie, ce projet a pour but l'élaboration d'un dictionnaire étymologique du noyau commun des noms de famille romans délexicaux (d'origine lexicale : substantifs, adjectifs ou verbes), c'est-à-dire qu'il n'étudie ni ceux issus d'un nom de personne (déanthroponymiques, comme *Gauthier*, *Martin* ou *Michelet*) ni ceux issus d'un nom de lieu (détoponymiques, tels *Fressancourt*, *Grammont* et *Ligny*). Chaque pays ou région représenté dans le projet possède sa propre équipe de chercheurs (portugais, espagnols, français, belges, suisses, italiens et

roumains), regroupés dans des centres nationaux et régionaux et travaillant en étroite collaboration les uns avec les autres.

Les différents lemmes de la nomenclature de ce dictionnaire sont regroupés par thème, avec, entre autres, des noms d'animaux (ARANEA, CABALLUS, CAPRA) et de parties du corps (AURICULA, BARBA, CAPILLUS), des adjectifs de qualité (ALACER, AMARUS, FIDELIS) et de couleurs (ALBUS, *BLANK, *BLUND) et des ethniques (ALAMANNUS, BRITTUS, CATALANUS). L'article CABALLUS (" cheval " en latin) traite ainsi les noms de personnes européens de langue romane issus de ce substantif parmi lesquels on peut citer, entre autres, les noms de famille *Cheval*, *Cheveau* ou *Chival* pour la France, *Caballa* et *Cavalletta* en Italie, *Cavalinho* au Portugal, *Cabalin* en Espagne, etc. La publication de ce dictionnaire est en cours (Cano González / Germain / Kremer 2004, 2007, 2010) ; les actes des colloques annuels tenus à partir de 1987 et un volume de présentation générale du projet (Kremer 1997) sont déjà disponibles.

Dans son traitement des noms de famille, le premier principe sur lequel s'appuie le projet PatRom est le recours en premier lieu à la documentation contemporaine (pour la France, au moment de l'établissement de ces données, l'annuaire téléphonique, alors disponible sur Minitel), non à la documentation historique. Deux raisons motivent ce choix, l'une liée à l'anthroponymie elle-même, l'autre aux possibilités offertes par les progrès technologiques. La première tient au fait qu'entre le Moyen Âge et l'époque contemporaine, le stock anthroponymique s'est considérablement amoindri, que ce soit par la non-transmission de ces surnoms (donc non encore noms de famille), ou par l'extinction naturelle de leurs porteurs. La seconde est que les moyens actuellement à la disposition des chercheurs permettent des relevés systématiques et rapides sur l'ensemble du territoire français. La cartographie des données ainsi obtenues autorise des comparaisons précises entre les régions (fréquence plus ou moins grande d'un nom) que le caractère fragmentaire des sources historiques rend très hasardeuses. Ces cartes permettent également une lecture des mouvements de population récurrents : attraction exercée par les villes sur les campagnes, par la plaine sur les habitants des montagnes, par la région parisienne sur la province mais aussi traînées le long des cours d'eau, etc. La géographie des noms de famille est également mise en relation constante avec les données de la géographie linguistique, c'est-à-dire avec la répartition dans l'espace des phénomènes phonétiques (exemple : les représentants du suffixe diminutif -ITTU sont différents selon les régions : *-et*, *-at*, *-ot*), mais aussi morphologiques (exemple : prédilection pour un suffixe) ou lexicaux (emploi d'un lexème caractéristique d'une région précise), etc. Un NF du Pas-de-Calais ne sera ainsi normalement pas rapproché d'un nom commun occitan ou n'apparaissant qu'au XIXe siècle. La documentation historique vient alors à l'appui de cette première analyse en attestant de l'existence du NF étudié à date ancienne dans la région (sur la méthodologie PatRom, cf. également Buchi 2001). Elle est regroupée sur la base de données PatRom, accessible par mot de passe sur internet, et est composée d'attestations issues de trois types de textes : textes en latin médiéval, textes en langues vernaculaires et listes nominatives (rôles de taille, bans

de tréfonds, etc.). Grâce à ces données anciennes, l'anthroponymie ainsi pratiquée contribue de façon indirecte à l'antédatation du lexique : la présence en anthroponymie d'un lexème à une date antérieure à celle relevée dans le lexique prouve son existence à cette époque et permet de reculer d'autant sa première attestation.

Ce sont les grands principes de cette méthodologie qui vont être appliqués au NF *Poirot*, qu'il soit délexical ou non : ceux-ci peuvent en effet être adaptés pour l'étude de l'ensemble des noms de famille (Wirth 2003).

La première étape est celle de la localisation du nom. Pour l'établir, les anthroponymistes du projet PatRom utilisaient l'annuaire téléphonique informatisé. Depuis, grâce au développement d'internet et de la généalogie en ligne, le chercheur a accès à des données issues des travaux de l'INSEE ; il suffit de se rendre sur le site internet Géopatronyme et l'on obtient pour tous les noms de famille français une carte de répartition des naissances pour différentes périodes (1891–1915, 1916–1940, 1941–1965, 1966–1990) s'appuyant sur des données presque exhaustives ; elles remplacent donc très avantageusement celles du Minitel. Pour *Poirot*, entre 1891 et 1915 (http://www.geopatronyme.com/cgi-bin/carte/nomcarte.cgi?numero=0004405&periode=1), ce sont donc 1054 naissances de porteurs qui sont enregistrées ; les départements les mieux représentés sont, par ordre décroissant, les Vosges (548 porteurs, soit plus d'un sur deux), la Meurthe-et-Moselle (162) et la Haute-Saône (99). Paris vient ensuite (50), suivi de la Meuse (29) et de la Moselle (26). C'est donc une zone d'un seul tenant qui se dessine autour des Vosges, avec une émigration secondaire vers la capitale qui s'explique aisément ; ce NF est donc clairement lorrain.

Mettons à présent cette localisation en relation avec les phénomènes phonétiques de la région. Dans son ouvrage sur la formation diminutive, B. Hasselrot (1957, 57) présente une carte localisant les différents aboutissements du suffixe diminutif -ITTU dans le nord-est de la France ; la presque totalité des départements des Vosges, de la Meurthe-et-Moselle et de la Haute-Saône, ainsi qu'une grande partie de celui de la Meuse se trouvent dans la zone où il aboutit à *-ot*. Le NF *Poirot* a donc très certainement pour étymon un dérivé formé avec ce suffixe. Mais lequel ? *Poire* étant un substantif féminin, on s'attendrait à une forme diminutive elle aussi féminine, du type **poirotte* ; cette explication doit donc être écartée. Ce sont les attestations anciennes qui vont nous guider sur le chemin de la solution :

Colot *Poirot*, Boviolles (Meuse), 1496 (archives départementales de MM, B 6595, f°174v)

Claude *Poirot*, Épinal (Vosges), 1499 (archives départementales de MM, B 299, 2, f°5v)

Claudon *Poirot*, Châtel-sur-Moselle (Vosges), 1578 (archives départementales de MM, B 4215, f°5r)

Claude Poirot le viel, Châtel-sur-Moselle, 1601 (archives départementales de MM, B 4242, f°3r)

Estienne *Poirot*, Moriville (Vosges), 1626 (archives départementales de MM, B 4290, f°33v)

Celles-ci nous prouvent que ce NF est bien originaire de la région ; d'autres, plus anciennes encore, indiquent l'origine de ce nom :

Poirot, Martigny-les-Bains (Vosges), 1403 (archives départementales de la Meuse, B 2406, f°23v)

Poirot le Receveur, Châtel-sur-Moselle, 1425 (archives départementales de MM, B 608, 9, 1, f°1v)

Poirot Noblet, Clézentaine (Vosges), 1425 (archives départementales de MM, B 608, 9, 1, f°5r)

Poirot filz Callot, Clézentaine, 1443 (archives départementales de MM, B 608, 9, 2, f°2r)

Poirot Loillier, Épinal, 1499 (archives départementales de MM, B 299, 2, f°5r)

Poirot, Auzainvilliers (Vosges), 1499 (archives départementales de MM, B 4388, f°26r)

Poirot Cognel, Châtel-sur-Moselle, 1509 (archives départementales de MM, B 4172, f°1v)

Poirot Estienne, Moriville, 1578 (archives départementales de MM, B 4215, f°120r)

Poirot Goery, Châtel-sur-Moselle, 1601 (archives départementales de MM, B 4242, f°2r)

La position occupée par *Poirot* dans la chaîne anthroponymique indique clairement qu'il s'agit, dans ces différents exemples, d'un nom de baptême (ci-après NB), en l'occurrence lié à *Pierre*. D'autres dérivés formés sur *Pierre* sont d'ailleurs attestés en Lorraine, comme *Poirel* (à comparer à *Pierrel*, lui aussi lorrain) et *Poirson* (*Pierson* est lui aussi lorrain) :

Poirel fil le gros Mongin, Horville-en-Ornois (Meuse), 1400/1402 (archives départementales de la Meuse, B 1429, f°32r)

Poirel de Clisentenne, (Loro)montzey (MM), 1425 (archives départementales de MM, B 608, 9, 1, f°5v)

Poirel Joly, Gouécourt (Vosges), 1499 (archives départementales de MM, B 4388, f°37r)

Demongin *Poirel*, Houdelaincourt (Meuse), 1401/1402 (archives départementales de la Meuse, B 1429, f°21r)

Jehan *Poirel*, Laxou (MM), 1488 (archives départementales de MM, B 7559, 7, f°1v)

Jehan *Poirel*, Clézentaine, 1507 (archives départementales de MM, B 4170, f°28r)

Poireçons fiz Vautherin, Bleurville (Vosges), 1327 (archives départementales des Vosges, 5 H 5, f°16r)

Poiresson Holriot, Gondrecourt-le-Château (Meuse), 1400 (archives départementales de la Meuse, B 1429, f°3v)

Poiresson le Doyen, Clézentaine, 1425 (archives départementales de MM, B 608, 9, 1, f°5r)

Girart *Poiresson*, Châtel-sur-Moselle, 1425 (archives départementales de MM, B 608, 9, 1, f°1r)

Demengin *Poiresson*, Nançois-le-Grand (Meuse), 1495 (archives départementales de MM, B 6595, f°133r)

Anthoine *Poirson*, Portieux (Vosges), 1626 (archives départementales de MM, B 4290, f°10v)

Poirot (ainsi que les autres dérivés cités) n'a donc absolument rien à voir avec la poire et un surnom en rapport avec celle-ci. Un simple coup d'œil dans la documentation historique lorraine aurait permis à nos deux auteurs de s'en rendre rapidement compte.

Mais comment expliquer cette erreur (et d'autres) chez des chercheurs ? Il n'y a pas de réponse toute faite (les auteurs ne se sont bien évidemment jamais exprimés sur le sujet), mais on peut tout de même émettre quelques hypothèses. En premier, il est évident que la méthodologie de ces auteurs n'était pas la bonne ; on peut même dire, pour *Poirot*, qu'ils n'ont pas suivi du tout de méthodologie. J.-P. Chambon (2002, 12-13) l'a souligné : dans ces ouvrages, la vulgarisation a précédé la recherche :

> Contrairement à la première idée qui vient à l'esprit, nos dictionnaires *ne* relèvent *pas* de la vulgarisation : la connaissance *vulgaire* (au sens technique de ce terme) est tout à fait distincte de la *vulgarisation*. *Qui dit vulgarisation dit production préalable d'un savoir scientifique*, et dit surtout qu'il s'agit d'amener le lecteur à rompre avec sa soif naïve de certitudes, et à se poser les 'bonnes' questions.

Ces auteurs ont très certainement voulu publier des résultats sans avoir fait tout le travail nécessaire en amont : de telles erreurs n'auraient pas été commises autrement. Morlet ayant, en outre, été l'élève de Dauzat et ayant repris son dictionnaire pour le continuer, on ne s'étonnera guère de retrouver dans son ouvrage les mêmes faiblesses que celles de son maître. Une autre explication à envisager tient aux contraintes qu'ont pu faire peser leurs éditeurs sur Dauzat et Morlet afin qu'ils fournissent un maximum d'articles et donc de NF traités, au détriment de la qualité des recherches. Pressés par le temps, nos deux auteurs ont probablement dû faire vite et beaucoup, et ils ont fait, parfois, comme dans notre exemple, vite et mal. On doit regretter qu'un dictionnaire étymologique à prétentions scientifiques entérine sans sourciller ce que les linguistes appellent des étymologies populaires, c'est-à-dire les étymologies des simples locuteurs non érudits. Il est donc particulièrement important de ne pas prendre pour assurées les informations présentes dans ces deux ouvrages.

Cette méthodologie PatRom, nous avons également voulu la tester sur une sélection de NF rares (Wirth 2003). Parmi ceux-ci, *Delubin*, dont un seul porteur est né durant la période 1891–1915, en Meurthe-et-Moselle (Géopatronyme). Il est également recensé trois fois dans l'annuaire téléphonique au début des années 2000, deux fois dans la Meurthe-et-Moselle (Gerbéviller), une fois dans les Vosges (Cornimont). Spontanément, on pourrait imaginer que l'étymon de ce patronyme est le NB *Lubin*, à l'image des formations du type *Degeorges*, *Debernard* ou *Derobert* ; il existe en effet un saint Lubin, évêque de Chartres au milieu du VIe siècle. Deux réserves importantes s'élèvent cependant très rapidement pour cette première tentative d'explication : d'une part, le NB *Lubin* est très peu représenté en Lorraine, d'autre part, les

formations du type préposition *de* + NB sont, à notre connaissance, absentes de Lorraine.

Sans piste sérieuse, c'est vers les attestations anciennes que nous nous tournons. Le nom apparaît dans la région de Domèvre-sur-Vezouze (MM), mais pas à époque très ancienne :

> Joseph *Delubin*, Mignéville (MM), 1831 (FamilySearch)
>
> Jeanne Claire *Delubin*, Domèvre-sur-Vezouze, 1783 (FamilySearch)

Comme, à cette étape de la recherche, il est impossible de trouver des mentions antérieures à la seconde moitié du XVIIIe siècle, la filière généalogique s'impose, grâce aux registres paroissiaux. On remonte alors jusqu'au mariage de Catherine *Delubine,* fille majeure de feu Jean *Delubine* et de défunte Madelaine Dubas, à Saint-Martin (MM), en 1758. Le nom de la commune de Lubine, dans les Vosges, pourrait convenir comme étymon : située à une trentaine de km, elle apparaît dans les archives dès 1334 sous la forme *Leubines* et en 1480 sous celle de *Lubinne* (Marichal 1941). De plus amples recherches vont cependant très rapidement mettre à mal cette explication séduisante, puisque les attestations plus anciennes dans les registres paroissiaux présentent ces formes :

> en 1747, à Pexonne (MM) : mariage de « Joseph ~~*Duli*~~ [sic] *Delubine*, fils de déft. [défunt] Jean *Delubine* et de Madelaine Dubas de la paroisse de Saint-Martin d'Ouze de Metz », avec Marie Biétry ; signature *Dulibine*
>
> en 1750, à Leintrey (MM) : mariage de Joseph *Dulibine*, veuf de Marie Biétry, avec Margueritte Masson
>
> en 1753, à Saint-Martin : naissance de Joseph *Delubine*, signature du père : *Dulibine* [fils des précédents]
>
> en 1756, à Saint-Martin : naissance de Jean François *Delubin*, signature du père : *Dulibine* [frère du précédent]
>
> en 1761, à Mignéville : mariage de Jean Claude *Delubin*, fils de Jean *Delubin* [peut-être frère du premier] et de Marie Lavaut, signatures : J. C. *Delubin,* Jean *Dulibine*.

On constate donc que *Delubin* comme *Delubine* sont des variantes d'un NF initial *Dulibine*, donc évidemment pas une formation *de* + *Lubin* ou *de* + *Lubine*. Malheureusement, ce stade dans les attestations anciennes ne semble pas pouvoir être dépassé. Dans la première attestation, les parents de Joseph Delubine sont dits « de la paroisse de Saint-Martin d'Ouze les Metz ». Il existe bien une paroisse de Saint-Martin, mais dans la ville même de Metz, et l'on n'y trouve aucune mention de ce NF (rien dans les tables de mariages) ; aucune localité n'est relevée sous le nom d'*Ouze les Metz*. Les généalogistes ne disposent pas d'informations supplémentaires et sont eux aussi arrêtés à cette date.

Deux points peuvent être soulignés : d'une part, la longévité de la forme initiale *Dulibine* dans la signature de certains porteurs de ce nom, alors que dans le texte de

l'acte, la forme est différente ; d'autre part, le rattrapage du scribe dans la toute première attestation de ce nom et qui, ayant commencé à écrire *Duli*, barre ces quatre lettres pour inscrire *Delubine*, témoignant ainsi de son interprétation d'un nom qu'il entend très certainement pour la première fois et rapproche d'un item linguistique connu. Et cette nouvelle forme s'est maintenue un certain temps, plusieurs personnes l'ayant portée, au moins dans les actes officiels.

Mais quelle est l'origine de ce nom *Dulibine* ? La réponse est loin d'être sûre, les données historiques sur les premiers porteurs faisant totalement défaut. Il pourrait s'agir d'un nom d'origine polonaise, résultat de la dérivation du NF polonais *Duleba* (de *dul-* adj. « enflé, épais » ; Hoffman 1998, 233) ou du NF *Dulęba* (de *dulęba* adj. « maladroit, fruste » ; *ibid.*) au moyen du suffixe patronymique polonais *-in* (*ibid.*, 34). La présence de Polonais en Lorraine dans l'entourage de Stanislas, devenu duc de Lorraine en 1737, rend l'hypothèse séduisante. Mais quelle que soit l'origine de ce nom, on est de toute façon bien loin du NB *Lubin* et du nom de lieu *Lubine*.

Un autre exemple intéressant nous est fourni par le NF *Palpart*, représenté, entre 1891 et 1915, par un seul porteur né en Meurthe-et-Moselle (Géopatronyme). On ne peut pas dire qu'une hypothèse sur l'origine de ce nom vient facilement à l'esprit, autant donc voir tout de suite ses attestations anciennes, classées dans l'ordre chronologique. La première, un acte de naissance à Nancy, date du 18 juin 1808, jour où :

> […] est comparu le sieur Jean Jacques Schneider commissaire de police en cette ville, y demeurant, lequel nous a déclaré que le dix-huit du courant, sur les neuf heures du soir, d'après l'avis qui lui a été donné, qu'un enfant étoit exposé près la porte d'entrée de l'hospice des orphelins de cette ville, il s'y est à l'instant transporté, et a trouvé un enfant tel qu'il nous le présente, enveloppé de trois braies, d'une bande et d'une chemise, la tête couverte d'un bonnet de soie rouge garni de dentelle, et d'un béguin, après l'avoir visité, avons reconnu qu'il étoit du sexe masculin, paroissant âgé d'un jour, qu'il n'avoit aucune remarque, et de suite l'avons inscrit sous les prénom et nom de François *Palpart,* et ordonné qu'il fut remis à l'hospice des enfans trouvés de cette ville dequoi nous avons dressé le présent procès-verbal […].

Le suivant date de 1845 : c'est celui du mariage de François *Palpart,* « âgé de trente-sept ans, domicilié à Ville-au-Val [MM, à quatre kilomètres de Sainte-Geneviève], fils majeur de père et mère inconnus, enfant trouvé à l'hospice de Nancy âgé d'un jour le dix-huit juin 1808 », Sainte-Geneviève (MM) ; il est suivi des actes de naissances de Jean François (1847), Marie (1848), Georges François (1849), Jeanne Eugénie (1851), Pierre Charles (1854) et Marie Joséphine *Palpart* (1858), à Sainte-Geneviève (FamilySearch).

Nous avons donc là l'acte de naissance du premier porteur de ce nom, mais également l'acte de naissance de ce NF, puisque celui-ci a été créé à ce moment. Malheureusement, la ou les motivations de cette création ne sont pas mentionnées : il ne s'agit pas du nom d'une personne présente, du saint du jour ou d'un objet trouvé en même temps que l'enfant, d'une marque de reconnaissance, etc. Comment l'interpréter ? On peut émettre l'hypothèse qu'il s'agit d'un dérivé formé sur le verbe français *palper* qui a le sens de " toucher avec la main pour connaître " (FEW 7, 518a, PALPARE), dérivé

formé avec le suffixe français d'origine germanique *-art* (Nyrop 1936, 173-175), de toute évidence dénué ici de connotation péjorative. Un substantif phonétiquement proche comme *poupard*, qui a le sens de " enfant au maillot " (dep. env. 1220, FEW 9, 602a, *PŬPPA) en tant que substantif et, en tant qu'adjectif, de " gros et joufflu " (dep. 1869, *ibid*.), a pu consciemment ou non avoir une influence dans le processus créatif, surtout que le sens de ce terme est parfaitement adapté au contexte de création. Le but était peut-être de donner à l'enfant de parents inconnus un nom lui aussi inconnu jusqu'alors.

Quels sont les principaux enseignements que l'on peut tirer de ces deux exemples ? Tout d'abord, c'est la prudence que le chercheur doit observer dans ses analyses ; dans le cas contraire, il risque de se tromper rapidement et lourdement. Cette prudence, elle passe par la prise de conscience de l'importance de la documentation historique, déjà bien soulignée dans le cadre du projet PatRom, mais également de celle, nos deux exemples le prouvent, de ce que l'on peut appeler la « filière » généalogique. Chercher des attestations anciennes de *Delubin*, par exemple, ne pouvait pas suffire pour établir l'origine première de ce nom. Le problème se pose si le chercheur se contente de consulter les bases de données informatiques existantes, comme GeneaNet : ce n'est pas la recherche de *Delubin* qui peut permettre de saisir l'origine de ce nom, mais la consultation ou l'établissement de la généalogie du premier porteur et des descendants. Dans des cas de ce type, il faut donc absolument sortir de l'idée qu'une interrogation toute bête des bases de données va donner une information suffisante pour établir l'origine du nom. Le chercheur ne doit pas oublier que tous les NF n'ont pas connu une histoire linéaire qui peut se résumer et se limiter à des évolutions phonétiques et graphiques comme le lexique a pu en connaître : certains, comme on l'a vu, ont pu connaître des variations, plus ou moins importantes, même à des époques récentes.

L'exemple de *Delubin* est particulièrement intéressant, parce qu'il montre le rôle des scribes et leur influence non seulement dans la graphie, phénomène bien connu des généalogistes confrontés aux nombreuses variantes d'un même nom, mais aussi dans la phonétique des noms, *Dulibine* étant bien différent de *Delubin*. Pour de tels noms, la généalogie n'est donc pas importante, mais essentielle. Dans le cas des NF rares, presque systématiquement, on avance qu'il est fort probable que l'étymon soit X, mais on en est rarement absolument sûr, contrairement à ce qui se passe pour le lexique, partagé par une population plus ou moins nombreuse. Dans le premier cas, la variation peut être le fait d'un seul individu, et donc être extrêmement importante et imprévisible ; dans le second, le partage du lexique par un groupe restreint fortement cette fourchette de variation. Les noms fréquents se rapprochent davantage du second cas : reconnus par les scribes, installés dans la langue, partagés par un grand nombre de porteurs, ils risquent beaucoup moins d'être l'objet de telles mutations. *Prudence* doit donc rester un maître-mot lorsque l'on aborde les NF, surtout s'ils sont peu répandus. Il est en outre à souligner que nombre de dépôts d'archives en France ont procédé à la numérisation de leurs registres paroissiaux et de leur état civil et que ceux-ci sont

presque toujours accessibles gratuitement en ligne ; le chercheur ne peut pas ou plus se réfugier dans la difficulté d'accès de ces documents pour ne pas en tenir compte.

Plusieurs autres enseignements peuvent être tirés de ces deux exemples. Ainsi, tous les NF n'ont pas été créés au Moyen Âge ou même à l'époque moderne, ce qui finalement est une chance pour nous, puisque la masse de documents conservés pour l'époque contemporaine est bien plus importante ; on peut donc supposer et espérer que figure dans ceux-ci une partie au moins des réponses à nos interrogations quant à l'origine de ces noms. Ce point est particulièrement important pour *Palpart*, puisque le verbe *palper* n'est attesté qu'en 1488 et à partir de 1770 ; si on n'avait pas cherché ces attestations anciennes, cet étymon potentiel aurait très probablement été écarté car employé de façon régulière à une époque bien plus tardive que celle de création des NF en général.

L'absence d'attestations anciennes d'un nom en Lorraine avant le début du XVIIe siècle et le début de la guerre de Trente Ans doit mettre la puce à l'oreille du chercheur : il est alors fort probable que le nom se soit implanté en Lorraine lors des repeuplements avec des populations issues d'autres régions (Picardie, Suisse, Tyrol) ; il est également possible que le NF en question ait connu des changements comme ceux de *Dulibine*, soit celui d'un enfant abandonné, d'un immigrant plus récent, etc. La généalogie peut alors jouer un rôle essentiel, même si elle ne permet évidemment pas d'écarter toutes les incertitudes. Dans le cas de *Delubin*, elle n'a ainsi permis que de remonter à la première forme du nom en France, sans permettre de connaître son origine géographique plus lointaine ; encore cette information est-elle assurée : en remontant de la sorte le plus haut possible, il reste encore beaucoup d'incertitudes, mais des incertitudes dont on est certain puisqu'il n'est pas possible d'en savoir davantage.

Tous les NF actuels ne sont bien sûr pas aussi rares ni des cas spéciaux comme cet exemple ; tous ne devraient pas nécessiter des recherches généalogiques. Il nous semble cependant que pour être assuré de ce point et pour être sûr qu'une démarche généalogique n'est pas nécessaire pour l'ensemble des noms, il faudrait déjà prendre un échantillon de noms contemporains représentatif, établir la généalogie de leurs porteurs pour voir les éventuelles variations, modifications ou changements et, à partir des résultats d'une telle étude, voir si effectivement les noms comme nos exemples sont très isolés ou au contraire sont plus fréquents qu'on n'a tendance à l'imaginer. Une autre question est liée à ce point : lorsque deux familles portent le même nom, peut-on vraiment dire qu'il s'agit justement du même NF ? Et cela surtout si on n'étudie pas en détail la généalogie de tous ces porteurs ? C'est ce qu'ont tenté de faire P.-H. Billy et M.-R. Sauvadet dans leur *Dictionnaire historique des noms de famille du Puy-de-Dôme* (1998–2001). L'idée mérite réflexion, parce qu'il est évident qu'un éventuel NF *Delubine* bien issu, lui, du toponyme *Lubine*, n'a absolument rien à voir avec les noms *Delubine*, *Delubin* et *Dulibine* qui eux, au contraire, sont apparentés malgré leurs différences formelles. S'agit-il là du même nom ? Cela paraît difficile à

soutenir, pourtant l'origine est, là, bien la même, avec en outre l'apparition concomitante de ces formes dans le même acte, pour désigner la même personne.

2. Panorama des noms de famille lorrains

Voyons à présent une sélection de NF lorrains et de surnoms qui en sont à l'origine, relevés dans des documents datant du XIVe au XVIIe siècle (donc avant la guerre de Trente Ans) et étudiés dans le cadre de notre thèse (Wirth 2007). Les trois grands types étymologiques sont représentés : détoponymiques, délexicaux, déanthroponymiques. Ce sont cependant ces derniers qui sont, et de loin, les plus nombreux.

2.1 Les déanthroponymiques

Parmi ceux-ci, on peut relever :

Le nom de famille *Adam*, qui a pour étymon un NB d'origine hébraïque *Adam*. L'un et l'autre sont attestés à de nombreuses reprises dans la documentation lorraine : *maistre Adans Maistre Massons* (Varennes-en-Argonne, Meuse, 1341/1343), *Adam le Moingne* (Brillon-en-Barrois, Meuse, 1473), *Adam Mercier* (Châtel-sur-Moselle, 1552), *Jehan Adam* (Moriville, 1552), *Claudon Adam* (Froville, MM, 1601).

Les 216 porteurs du nom *Adnet* nés en France entre 1891 et 1915 (Géopatronyme) se concentrent dans l'Est du pays (Marne : 60 ; Ardennes : 37 ; Meuse : 28 ; Meurthe-et-Moselle : 26) ; à date ancienne, on relève : *Adenet Malegoule* (Varennes-en-Argonne, 1341/1343), *Adenet* (Neufchef, Moselle, 1480/1481), *Nicolas Adinet* (autre attestation : *Niclos Adenet*, Neufchef, 1480/1481). C'est, à l'origine, un dérivé, formé avec l'aboutissement du suffixe déjà cité -ITTU, sur le nom de personne (ci-après NP) d'origine germanique *Adin*, lui aussi attesté dans la documentation lorraine : *Adin de Boinville* (prévôté d'Étain, Meuse, 1324).

Le NB *Amet*, attesté à époque ancienne (*Amey*, Épinal, 1499 et *Amey Drouot*, Longchamp-sous-Châtenois, Vosges, 1499), a pour étymon le NP latin *Amatus*. Il est à l'origine du NF *Amet*, dont 144 des 241 porteurs nés en France entre 1891 et 1915 ont vu le jour dans les Vosges. C'est ce même NP que l'on retrouve dans le nom de la commune vosgienne de *Saint-Amé*.

Les NF *Ancel* et *Ansel* trouvent leur origine dans le NB *Ancel*, lui-même emprunt du NP germanique *Ansehelm* : *Anceil de Boulay* (prévôté de Saint-Dié, Vosges, 1379/1380), *le maire Ancel et Martin de La Fosse consorciers de Sainte Margueree* (grande prévôté de Saint-Dié, 1416), *Ancel Vaulthier* (Trieux, MM, 1480/1481), *Jehan Ansel* (= *Jehan Ancel* = *Jehan Anseil de Mes*, prévôté d'Étain, 1324), *Mengin Ancel* (Biencourt-sur-Orge, Meuse, 1495).

Le NF *Anchier* est issu du NB *Anchier*, emprunt du NP germanique *Ansigar* : *Anchiers de Sainte Geneviere prevos de Condey sor Moselle* (prévôté de Custines, MM, 1345/1346), *monseigneur Anchier* (Pont-à-Mousson, MM, 1358/1359) ; dérivé d'*Anchier*, le NB *Ancherin*, attesté dès le XIVe siècle dans la Meuse (*Ancherin prevost*

dou Pont, prévôté d'Étain, 1324), est à l'origine du NF meusien *Encherin* (Wirth 2003, 106-112). Le NB *Eulry* (*Eulry Ferry*, Balléville, Vosges, 1499), à l'origine emprunt du NP germanique *Odalric*, est l'étymon des NF *Eulry*, *Eury* et *Olry* ; ses dérivés ont abouti aux NF *Euriat*, *Euriot*, *Eulriet* et *Eulriot*.

Le NB *Franquin*, emprunt du NP germanique *Franchin*, est à l'origine des attestations suivantes : *Francquin* (Montauville, MM, 1479), *Francquin Darmoix* (Sancy, MM, 1480/1481), *Mengin Francquin* (Mainville, MM, 1480/1481) et du NF *Franquin*, dont les 110 porteurs nés en France entre 1891 et 1915 sont très majoritairement lorrains (Meurthe-et-Moselle : 38 ; Vosges : 22 ; Moselle : 13). *Franquinet*, relevé par exemple pour *Martin Franquinet* (Bazincourt-sur-Saulx, Meuse, 1473), en est un dérivé.

Un autre NB d'origine germanique (du NP *Idulf*) est à l'origine de ces deux attestations : *monseigneur Jehan Hydoul maitre concheinoigne* (grande prévôté de Saint-Dié, 1416), *Jehan Ydou* (Sandaucourt, Vosges, 1499). Ce nom est à mettre en relation avec le culte de saint Hidulphe, fondateur, vers 704, de l'abbaye de Moyenmoutier (Vosges), à 10 km à peine de la ville de notre première attestation. Il est à l'origine des NF contemporains *Idoux* et *Ydoux*, dont les 254 et 4 porteurs respectifs nés en France entre 1891 et 1915 se concentrent, pour le premier, dans les Vosges (87) et dans le Bas-Rhin (58), département voisin, pour le second, en Meurthe-et-Moselle (3 sur 4, à Nancy et Jarville-la-Malgrange). Le petit nombre de ceux-ci et leur localisation dans une grande ville et sa région incitent à y voir un NF de même souche que le NF *Idoux* mais de forme graphique différente : à époque ancienne, on relève également de façon ponctuelle cette variante graphique dans le département des Vosges, à Denipaire (1654), Fraize (1654), Le Valtin (1688) et Biffontaine (1690), ainsi que dans le Bas-Rhin, à La Broque (1665).

Les porteurs contemporains du NF *Idoux* présentent quant à eux une micro-répartition remarquable, puisque les communes où les naissances ont été les plus nombreuses, dans les Vosges comme dans le Bas-Rhin, ne se trouvent qu'à quelques kilomètres de la commune de Moyenmoutier où le saint de ce nom a fondé une abbaye. La plus représentée dans les Vosges, Nompatelize, petite commune de 522 habitants, se situe à 7 km seulement ; viennent ensuite Saint-Dié (10 km), Senones (5 km), La Salle (9 km) et Saint-Michel-sur-Meurthe (7 km ; liste non exhaustive).

Mais c'est dans le Bas-Rhin que se trouve la commune où les naissances ont été les plus nombreuses pour l'ensemble de la France : Saales, à proximité immédiate de la limite avec le département des Vosges et à 14 km de Moyenmoutier, dans la partie de langue romane de l'Alsace tout comme les communes de Saint-Blaise-la-Roche (19,5 km) et de Ranrupt (21 km), au deuxième et au troisième rang (liste non exhaustive). Pour les siècles antérieurs (GeneaNet), les localités les plus anciennement représentées se trouvent également à une distance assez réduite de Moyenmoutier ; ce sont, dans les Vosges : Celles-sur-Plaine (1510, 9 km), Belmont-sur-Buttant (1610, à 20 km), Étival-Clairefontaine (1610, à 4 km) et Nompatelize (1620, à 7 km) ; dans le Bas-Rhin : Schirmeck (1580, à 24 km) et Russ (1610, à 29 km).

Dans *Saufroi Fuzeli* (Houdelaincourt, 1327/1330), *Saffroy Pernet* (Sornéville, MM, 1477), *Colart Saffroy* (Sornéville, 1477) et George *Saffroy* (Portieux, 1626), c'est un NB issu du NP germ. *Salafrid* que l'on trouve ; il est à l'origine du NF *Sauffroy* ainsi que du foyer lorrain du NF *Saffroy*. Les 17 porteurs du premier nés en France entre 1891 et 1915 se localisent exclusivement en Lorraine (Vosges : 16 ; Meurthe-et-Moselle : 1) ; 13 des 55 porteurs du NF *Saffroy* sont également de la région (Meurthe-et-Moselle : 8 ; Moselle : 5).

La documentation ancienne atteste aussi un NB *Symay* : *Symay Waultrenel* (Gerbécourt, Moselle, 1477 ; variante : *Xymart Waultrenel*), *Le maire Xemay* (Damas-aux-Bois, Vosges, 1534). En tant que surnom ou NF, on relève : *Colin Ximart de Pareirs* (prévôté d'Amel-sur-l'Étang, Meuse, 1321/1328), *Jehan Symart* (Louppy-le-Château, Meuse, 1406 ; variante : *Jehan Simart*, 1407), *Jehan Ximart* (Châtel-sur-Moselle, 1425), *Jehan Symay* (Damas-aux-Bois, 1507), le même que : *Jehan Xemay* (Damas-aux-Bois, 1534), *Bastien Xemay nouveau marié* (Ortoncourt, Vosges, 1601). Il a pour étymon le NP germ. *Sigimar* et est à l'origine du NF *Ximay*, dont l'unique porteur né en France entre 1891 et 1915 a vu le jour dans le département de la Moselle, à Hampont ; on trouve ce NF dès 1551/1600 à Vic-sur-Seille et dès 1697 à Jallaucourt (Moselle), en 1676/1700 à Nomeny (MM), en 1718 à Saint-Médard, en 1720 à Haraucourt-sur-Seille et en 1701/1725 à Destry (Moselle ; GeneaNet).

Les NB *Arnould*, *Aubert*, *Aubry*, *Baudoin*/*Baudouin*, *Bertrand*, *Gerard*, *Louis* viennent respectivement des NP germaniques *Arnulf*, *Athalbraht*, *Albirich*, *Baldavin*, *Berahthraban*, *Gairard*, *Hlodwic*. Ils sont à l'origine des NF de mêmes formes. *Aubert* est également à l'origine d'*Aubertin*, par dérivation au moyen du suffixe *-in* ; *Aubry*, d'*Aubriat*, *Aubriet* et *Aubriot*, tous formés avec les aboutissements du suffixe -ITTU : le NF *Aubriat* est vosgien (la totalité de ses 7 porteurs nés en France entre 1891 et 1915 ont vu le jour dans ce département), *Aubriet*, ardennais et meusien (Ardennes : 26 porteurs sur 95 ; Meuse : 16) ; le patronyme *Aubriot* est un peu plus représenté (184 porteurs) et se rencontre surtout en Haute-Marne (44), dans la Meuse (38) et en Meurthe-et-Moselle (23). Sur *Aubry* a également été formé *Aubrion*, à l'origine du NF contemporain *Aubrion*, dont 29 des 36 porteurs nés en France entre 1891 et 1915 ont vu le jour en Meurthe-et-Moselle.

Parmi les autres déanthroponymiques issus de NP germaniques, on peut citer :

- *Bazin*, du NP *Basin* : *Bazin* (Uruffe, MM, 1327/1330), *Demengeot Bazin* (Darney-aux-Chênes, Vosges, 1499)
- *Basselin*, du NP *Bezilin* : *Basselin Coquibus* (grande prévôté de Saint-Dié, 1416)
- *Liébault*, du NP *Leudbald* : *Liebault Bouchier* (Portieux, 1509), *Mengin Liebault* (Houécourt, Vosges, 1499), *Jehan Liebault* (Portieux, 1509). Il est à l'origine du NF *Liébaut*
- *Thierry*, du NP *Theoderic*, est à l'origine des NF *Thiery*, *Thiry* ; *Thirion*, *Therion*, *Thierion*, *Thiriat*, *Thiria*, *Thiriet*, *Thiriot*, *Thyriot* en sont des dérivés
- *Wichard*, du NP germ. *Wighard*, a donné les NF contemporains *Vichard*, *Vichart* et *Wichard*
- *Wiry*, du NP germ. *Wigirich*, a abouti aux NF *Viry* et *Wiry* ; ses dérivés, à *Viriat*, *Viriot*, *Wiriot*, *Virion*
- *Wiart*, du NP germ. *Withard*, au NF *Viard*.

Parmi les NF lorrains issus de NP d'origine latine figurent *Augustin* ; le NB *Curien* vient quant à lui du NP latin *Quirinus* : *Currien Barbier* (Châtel-sur-Moselle, 1537), *Quirien durant* (Clézentaine, 1601) ; la forme savante du NB *Quirin* se retrouve dans le NF *Quirin*, patronyme caractéristique de l'Alsace (des 277 porteurs nés en France entre 1891 et 1915, 137 ont vu le jour dans le Bas-Rhin), tandis que la forme héréditaire s'est fixée dans le NF vosgien *Curien* (113 des 129 porteurs français sont nés dans ce seul département).

Saintin est un emprunt du NP latin *Sanctinus* : *Sainctin* (Hadonville-lès-Lachaussée, Meuse, 1345/1346), *Saintin* (Saint-André-en-Barrois, Meuse, 1469/1471). Il est à l'origine du foyer meusien du NF *Saintin* qui regroupe 7 des 60 porteurs du NF *Saintin* nés en France entre 1891 et 1915. *Saintignon*, qui se rencontre par exemple dans *Sanctignon d'Estain* (= *Sainctignon de la val d'Estain*, prévôté d'Étain, Meuse, 1324), en est un dérivé, à l'origine du NF *Saintignon*, dont les quatre porteurs nés en France entre 1891 et 1915 sont tous meusiens.

Le NB *Urbain*, du NP latin *Urbanus*, a également laissé sa trace par le foyer lorrain du NF contemporain *Urbain*, qui regroupe 133 des 607 porteurs de ce NF nés en France entre 1891 et 1915 (MM : 66 ; Meuse : 31 ; Vosges : 31 ; Moselle : 5). Le NB lui-même n'apparaît que peu dans la documentation ancienne : *le petit Vrbain* (Nant-le-Grand, Meuse, 1496), *Vrbain Pollet maieur* (Triconville, Meuse, 1495), tout comme le surnom/NF : *Jehan Vrbain* (prévôté de Lamarche, Vosges, 1333/1334), *Didier Vrbain* (Nant-le-Petit, Meuse, 1495), *Denis Urbain* (Damas-aux-Bois, Vosges, 1626).

Les NF formés sur un NB issu de *Nicolas* (< NP latin d'origine grecque *Nicolaus*) sont particulièrement nombreux : *Colas*, *Colotte*, *Collotte*, *Colin*, *Colignon*, *Colnat*, *Collenat*, *Colnot*, *Collenot*, *Collinet*, *Collesson*, *Colson*, *Collot*, *Colet*, *Collet*, *Nicole* ; le même phénomène peut être observé pour :

- *Dominique* (< NP latin *Dominicus*) : *Demange*, *Demenge*, *Demengeon*, *Demangin*, *Demengin*, *Demenginot*, *Demangeat*, *Demangeot*, *Demangel*, *Demengel*, mais aussi *Mangel*, *Mangeolle*, *Mangeat*, *Mangeot*, *Mangin*, *Mengin*, *Petitmengin*, *Mangenot*, *Manginot*
- *Didier* (< NP latin *Dēsīderius*) est à l'origine de *Didellot*, *Didelot*, *Didion*, *Didot*, *Didon*
- *Étienne* (< NP latin d'origine grecque *Stephanus*), de *Thouvenel*, *Thouvenin*, *Thouvenot*
- *Georges* (< NP latin d'origine grecque *Georgius*), de *Georgin*, *Georgeon*, *Georgeat*, *Georgeot*
- *Jacques* (< NP latin d'origine hébraïque *Jacob*), de *Jacquel*, *Jacquin*, *Jacquinel*, *Jacquemin*, *Jaquemin*, *Jacquotte*, *Jacquinot*, *Jaquinot*, *Jacquat*, *Jacquot*, *Jaquot*
- *Poince* (< NP latin *Pontius*), de *Poincelet*, *Poincelot*, *Poincignon*, *Poinsignon*, *Poinsard*, *Poinsart*.

Parmi ces déanthroponymiques, certains sont formés sur des NB n'apparaissant que tardivement (seconde moitié du XVe siècle) dans la documentation, mais connaissant ensuite, très rapidement, pour la plupart, un emploi important. Les NF *Adrian* et *Ambroise* en font partie. Le premier apparaît dans ces attestations : *Geraird Adrian* (Clézentaine, 1552), *Parisot Adrian* (Damas-aux-Bois, Vosges, 1626) ; en tant que NB, issu du NP d'origine latine (*H*)*adrianus* : *Adrian Regnalt* (Mercy-le-Haut, MM, 1480/1481), *Andrian Jehan Loys* (Portieux, 1534) = *Adrian* Jehan Loys (Portieux,

1552) ; le second se rencontre dans les suivantes : *Perrin Ambroize* (Neufchâteau, 1499), *Mansuit Ambroise* (Portieux, 1578) ; il a pour étymon le NB *Ambroise*, à l'origine NP latin m. d'origine grecque *Ambrosius*.

Sans surprise, le NF *Anthoine* a pour étymon le NB *Antoine*, qui lui-même a pour origine le NP latin *Antonius* ; le NB apparaît durant la seconde moitié du XIVe siècle, connaît un frémissement dans son emploi durant le XVe siècle pour finalement connaître une véritable explosion au XVIe siècle. Faut-il y voir un lien avec le nom d'Antoine, duc de Lorraine et de Bar (1508–1544) ?

L'apparition du NB *Balthazar* est elle aussi tardive : *Baltazar* (Commercy, Meuse, 1478/1480), *Baltezar* (Châtel-sur-Moselle, 1509), *Baltasar Ferry* (Portieux, 1534). Ayant pour étymon le NP latin *Baltasar*, ce NB est à l'origine de plusieurs NF bien attestés en Lorraine. Les 14 personnes nommées *Balthasar* nés en France entre 1891 et 1915 ont pour la plupart vu le jour dans une zone englobant la Moselle (6) et le Bas-Rhin (4), tout comme les 2 porteurs du NF *Balthasard* (Bas-Rhin : 2) et une partie des 86 porteurs du NF *Balthazar* (Moselle : 7 sur 86). La localisation du NF *Balthazard* est voisine de la précédente (Haut-Rhin : 78 sur 292 ; Vosges : 70), tandis que celle du patronyme *Baltazard* est un peu plus éloignée (Meuse : 12 sur 34), de même que celle de *Balthazart* (Marne : 5 sur 6 ; Ardennes : 1) et de *Baltazart* (Ardennes : 39 sur 49 ; Marne : 7). Mis en relation, les différents foyers de ces variantes graphiques dessinent une zone compacte incluant la moitié nord de la région Champagne-Ardenne, la Lorraine et l'Alsace.

Le NF *Bastien* présente un cas comparable : son étymon, le NB *Bastien* (à l'origine, *Bastien* est une aphérèse du NB *Sébastien*), n'apparaît qu'à la fin du XVe siècle dans la documentation lorraine : *Bastien Robinot* (Resson, Meuse, 1495), *Bastien* (Morelmaison, Vosges, 1499), *Bastien Jehan Marischal* (Hadigny-les-Verrières, Vosges, 1534) ; l'apparition du NF est encore plus tardive : Claudon *Bastien* (Saint-Remy-aux-Bois, MM, 1601), *Jean Bastien* (Moriville, Vosges, 1626).

Plusieurs autres NB appartiennent également à cette catégorie : *Blaise*, du NP latin *Blasius* : *Blaise* (Anoux/Mancieulles, MM, 1480/1481), *Blaize Cardot* (Neufchâteau, 1499), *Mengin Blaize* (Neufchâteau, 1499), *Gaspar Blaise* (Rozelieures, MM, 1626) ; *Christophe*, à l'origine emprunt du NP latin d'origine grecque *Christophorus*, n'apparaît dans la documentation lorraine que durant la seconde moitié du XVe siècle, tout comme *Claude* (du NP latin *Claudius*) et ses nombreux dérivés *Claudel*, *Claudin*, *Claudinel*, *Claudinot* et *Claudon* (sur ce NB, cf. Corbet/Wirth 2011) ; *Ligier*, du NP germ. *Liutheri*, qui peut être relevé dans ces attestations : *Liegier* (Rambervillers, Vosges, 1489/1490), *Ligier Husson* (Resson, Meuse, 1495), *George Ligier* (Saint-Remy-aux-Bois, MM, 1664). Il est à l'origine du foyer lorrain du NF *Ligier*, qui regroupe 41 des 230 porteurs de ce NF nés en France entre 1891 et 1915 (Meuse : 21 ; Meurthe-et-Moselle : 16).

D'autres noms de famille lorrains ont pour étymon un NB féminin : c'est le cas pour *Alison*, attesté dans *Alizon de Hageyville* (Arraye-et-Han, MM, 1358/1359),

Alixon vesve (Neufchâteau, Vosges, 1499), *Mathieu Alixon* (Viocourt, Vosges, 1499). 20 des 47 porteurs du nom de famille *Alison* nés en France entre 1891 et 1915 sont meurthe-et-mosellans, 6 sont vosgiens. Ce nom a pour étymon un NB, lui-même formé sur le NB *Alix* (d'origine germanique) avec le suffixe *-on*.

Appartiennent également à cette catégorie les NF *Biétrix* et *Biétry*, du NB *Biétry*, emprunt du NP latin *Beatrix* : *Bietrix femme Jacot Mareschal*, (Clézentaine, 1443), *Betrix vesve* (Mercy-le-Haut, MM, 1480/1481), *la grosse Beatrix vefve* (Varangéville, MM, 1488), *Jehan Bietry* (Thaon-les-Vosges, Vosges, 1438), *Mengin Bietrix* (ban de Biécourt, Vosges, 1499). Il en est de même pour le NF *Helluy*, du NB *Heluy*, lui-même emprunt du NP germ. fém. *Halawit* : *Heluy fille Jennet le Besgue* (Villotte-devant-Louppy, Meuse, 1405), *Heluy La Couseresse vesve de Perresson Nauette* (Les Bordes, vill. ruiné près de Louppy-le-Château, Meuse, 1406), *Jehan Heluy* (Villotte-devant-Louppy, 1405).

Dérivé du NP *Ide* formé à l'aide du suffixe *-ette* et de sa variante lorraine *-atte*, le NB *Idette*/*Idatte* apparaît ponctuellement dans la documentation historique lorraine : *Ydatte la Cousiere famme Ferry Rougieville* (grande prévôté de Saint-Dié, 1416), *Ydette vesve* (Neufchef, 1480/1481), *Ydatte d'Aultreval vesve* (Moriville, 1534). Il est à l'origine du NF *Idatte*, dont les 14 porteurs nés entre 1891 et 1915 sont tous lorrains (Meurthe-et-Moselle : 7 ; Vosges : 4 ; Moselle : 3).

Les noms de famille *Zabé* et *Zabel* ont quant à eux pour étymon le NB *Zabel*, issu par aphérèse du NB *Isabelle*. Dans les sources, on trouve : *Zabel Colignon* (Ramber-villers,1489/1490) ou encore *Jehan Zabel* (Bazien, Vosges, 1489/1490). Les 80 porteurs du NF *Zabé* nés en France entre 1891 et 1915 se localisent majoritairement dans les Vosges (32) et en Meurthe-et-Moselle (25) ; les 49 de *Zabel* sont surtout meurthe-et-mosellans (16) et mosellans (6).

2.2 Les délexicaux

La catégorie des NF ayant pour étymon un substantif, un groupe nominal, un adjectif, etc. constitue probablement la catégorie dont l'étude est la plus délicate, pour deux raisons principalement. Il est souvent malaisé, d'une part, d'établir avec certitude le sens de cet étymon au moment où il est passé dans la nomination, ce moment et son contexte nous échappant le plus souvent ; il est encore plus difficile, d'autre part, de saisir les motivations à l'origine de l'attribution d'un surnom. Les résultats assurés de l'étude de tels noms peuvent donc paraître minces.

Parmi les NF lorrains appartenant à cette catégorie des délexicaux figure *Bailly*, qui a pour origine *bailli* n. m. "représentant du roi ou d'un seigneur, dans une circonscription où il exerce par délégation un pouvoir administratif et militaire, et surtout des attributions judiciaires, soit en première instance, soit comme juge d'appel des prévôts et des hauts-justiciers" (dep. av. 1160, TLF ; FEW 1, 207a, BAJULUS) : *Pieresson dou Pont balli de Saint Mihiel* (prévôté d'Étain, 1324), *Symonnet le Bailli* (Dainville, Meuse, 1400/1404), *Nicolas Bailly* (Épinal, 1499).

L'attestation *Jehan Blanc Cheveulx* (Sandaucourt, Vosges, 1499), à la composition évidente, a une motivation originelle qui l'est également, tout comme *Thiriat Bien Dance* (grande prévôté de Saint-Dié, 1416) et *Claude Bien Dance* (Épinal, 1499), *Jehan Demengot dit le Boyteux* (Willeroncourt, Meuse, 1495) et *Thomas le Boiteux* (prévôté d'Étain, 1473), *Jehan Bon Frere* (Borville, MM, 1448), *La vefve Brise Escuelle* (Châtel-sur-Moselle, 1601), *Berteran Grant Ventre* (Varennes-en-Argonne, Meuse, 1341/1343), *Thiriat Gros Cul* (Lunéville, MM, 1484), *Jehan Larmurier* (Épinal, 1499), *Jacommin Tainturier* (Houdelaincourt, Meuse, 1401/1402), *Jehan Tisserant* (Moriville, 1534), *Jehan Tonnellier* (Château-Salins, 1492).

Le sens originel d'autres formations est un peu moins évident : *Anthoine le Bon Homme* (Nançois-sur-Ornain, Meuse, 1495) et *Jehan Bon Homme* (Vouthon-Bas, Meuse, 1400) ont ainsi pour étymon *bonhomme* n. m. "homme bon, homme de bien" (dp. fin XIIe s., TLF ; FEW 4, 455a, HOMO), *Garin Bonnetier* (Gimécourt, Meuse, 1495) et *Anthoine Bonnetier* (Neufchâteau, 1499), *bonnetier* n. m. "celui qui fait ou qui vend des bonnets, des bas et d'autres objets de tricot" (dp. 1469, FEW 25, 41b, ABONNIS), substantif à l'origine du NF *Bonnetier*, dont les 43 porteurs nés en France entre 1891 et 1915 ont majoritairement vu le jour en Meurthe-et-Moselle (17) et en Moselle (12).

Le NF *Borgnot*, s'il n'est pas répertorié entre 1891 et 1990, est attesté dans les Vosges en 1647 et 1665 à Soncourt, et plusieurs fois à Aroffe entre 1637 et 1688 (GeneaNet) ; ces attestations doivent être rapprochées de celle de *Didier Borgnot* (Resson, Meuse, 1495) et d'afr. mfr. *borgnet* adj. "borgne, louche" (encore attesté en 1613 et 1627 ; FEW 1, 569a, *BRŬNNA).

Les noms de métiers ne sont pas rares. *Flabaix le Bosquillon* (prévôté d'Étain, 1324) avait pour ancêtre ou était bûcheron (fr. *boquillon* n. m. "bûcheron", dep. 1210/1240 [*bosquillon*] ; FEW 15/1, 195b, *BOSK- ; TLF) ; dans le cas de *la vesve Aubry le Bourlier* (Vignot, Meuse, 1478/1480) et *Jehan Bourelier* (Neufchâteau, 1499), il devait fabriquer des colliers pour animaux (fr. *bourrelier* n. m. "celui qui fait les colliers des bêtes de trait ou de somme", dep. 1268/71, TLF ; FEW 1, 639b, BURRA), tandis que pour *Husson le Bourgon* (Vigneulles-lès-Hattonchâtel, Meuse, 1477) et *Jehan Bourgon* (Moriville, 1425), il était chef de charretiers (anc. lorr. *bourgon* n. m., Meuse 1318, FEW 23, 66b, charretier). *Brulefers* (Clézentaine, 1443) et *Gerard Brulefer* (Commercy, Meuse, 1478/1480) doivent être rapprochés de fr. *brusle-fer* n. m. "surnom du forgeron" (1611, FEW 14, 80b, USTULARE).

Le patronyme de *Aubriat Le Cherrier demeurant a Viller devant Luneville* (Lunéville, 1484), *Jehan Cherrier* (Vannecourt, Moselle, 1477) et *Jacot Cherrier* (Épinal, 1499) doit quant à lui être rapproché d'anc. lorr. *cherrier* n. m. "charron, celui qui fabrique les voitures" (1483, FEW 2, 433a, CARRUS). Dans le cas de *Francoy le Corvisier* (Ollières, Meuse, 1480/1481) et de *Bastien Corvisier* (Châtel-sur-Moselle, 1552), l'étymon est afr. mfr. *corvisier* n. m. "cordonnier" (Lorraine XIIIe–XIVe s., FEW 2, 1182a, CORDUBA) ; pour *Gerardin Fournier* (Clézentaine, 1451), c'est *fournier* n. m. "celui qui tient un four banal, boulanger" (dep. 1153, FEW 3, 902a, FŬRNARIUS). Dans *Denys le Papellier* (Épinal, 1499) et *Waultrin Papellier* (Nançois-sur-Ornain, Meuse,

1495), c'est le type lexical représenté par anc. lorr. n. m. "fabricant de papier", *papelier* (Meuse 1383), *pauppellier* (Metz 1451), *pappellier* (Épinal XVIe s.), Moselle [pō'pli], Fiménil (Vosges) [pa'plɛy] " ouvrier papetier " (FEW 7, 591b, PAPYRUS) que l'on trouve. Il est à l'origine du NF *Papelier*, dont 32 des 40 porteurs français nés entre 1891 et 1915 sont lorrains (Vosges : 17 ; Moselle : 8 ; Meurthe-et-Moselle : 6 ; Meuse : 1).

Jehan Sagaire de Harvafaing (grande prévôté de Saint-Dié, 1416) constitue la première attestation de lorrain (Vosges) *ségard* n. m. "scieur, ouvrier qui débite le bois en planche, dans les scieries des Vosges" (dp. 1860, TLF), *sagard* (dp. 1876, TLF), *ségare*, attesté, dans les Vosges, sous la forme *ségaire* à Fraize et *saugard* " scieur de long " à Uriménil (FEW 17, 2b, *SÄGER ; TLF). Bien connu des généalogistes travaillant sur cette partie de la Lorraine, ce substantif n'apparaît que de façon tardive dans les sources littéraires (1860) ; notre documentation permet de l'antédater de plus de quatre siècles et demi. Il est à l'origine des NF contemporains *Sagaire* et *Sagard* : c'est ce que montre la localisation de leurs porteurs nés entre 1891 et 1915 : 13 des 16 personnes nommées *Sagaire* sont vosgiennes ; sur 23 nommées *Sagard*, 11 sont nées dans les Vosges et 6 en Meurthe-et-Moselle.

Le Cahu (Removille, Vosges, 1499) et *Didier le Cahu* (Removille, 1456) sont à rapprocher de lorrain (Cumières, Meuse) *cahu* adj. "entêté" (FEW 4, 502a, HŪ-), Girard *Camart* (Brillon-en-Barrois, Meuse, 1473) de mfr. frm. *camard* adj. "qui a le nez court et plat" (dep. 1534, FEW 6/3, 276b, MŪSUS, TLF), *Didier Cayemant* (Neufchâteau, 1499) du type lexical représenté par mfr. frm. *cayment* n. m. "mendiant" (1382–1er tiers du XVIe s.), *cayemant* (1432 ; 1ère moitié XVe s. ; FEW 22/2, 303b, mendiant ; mendier), et *Thiebault Champy* (passage de Saint-Nicolas-de-Port, MM, 1487) de fr. *champi* n. et adj. "enfant conçu ou trouvé dans les champs, bâtard" (dep. 1390 [*champiz*] ; FEW 2, 157b, CAMPUS ; TLF).

L'étymon du nom de *Jehan des Chandoillez* (grande prévôté de Saint-Dié, 1416) est le substantif féminin *chandelle* (dep. 1119 [*chandeile*], TLF ; FEW 2/1, 177b, CANDELA) ; la personne ainsi surnommée à l'origine devait avoir pour charge l'entretien des chandelles à l'église. *Jehan du Chaucheu* (Ollainville, Vosges, 1499) ou son ancêtre devaient avoir en charge un pressoir : anc. lorr. *chaucheur* n. m. "pressoir" (FEW 2, 67b, CALCATORIUM).

Chaucoullon (Vandœuvre-lès-Nancy, MM, 1488) a quant à lui pour étymon un syntagme composé de fr. *chaud* adj. (dep. *ca* 1100, TLF ; FEW 2/1, 87b, CALIDUS) et fr. *couillon* n. m. "testicule" (dep. XIIIe s., FEW 2, 889a, COLEUS, TLF). Il est assurément à l'origine du NF *Chaucouillon* ; la seule personne portant ce NF née entre 1891 et 1915 l'est dans le même département que celui de cette première attestation, la Meurthe-et-Moselle. Le même substantif est à l'origine du nom de *Gillet Couillon* (prévôté de Louppy-le-Château, Meuse, 1405) et *Andreu Coillon* (Chenicourt, MM, 1478).

Dans le même esprit, le surnom ou NF de *Demenget Vit de Boix* (Rouvres-la-Chétive, Vosges, 1458) a pour origine le syntagme nominal *vit de bois*, composé de fr. *vit* n. m. "membre viril" (dep. env. 1195, FEW 14, 211a, VECTIS ; TLF), de la préposition

de et de fr. *bois* n. m. “substance dure des arbres, des arbrisseaux” (dp. *ca* 1100, FEW 15/1, 204b, *BOSK ; TLF). La personne ainsi surnommée à l’origine devait avoir une raideur au niveau de cette partie de son anatomie (priapisme ?).

Enfin, l’étymon du NF de *Claudon Trompette* (Saint-Germain, MM, 1552) est le nom de l’instrument de musique. Ce surnom a pu être donné, à l’origine, à une personne jouant de cet instrument (surnom métonymique), ou à une personne dont le nez avait une forme rappelant celui-ci (surnom métaphorique) ; il est à l’origine du NF *Trompette*, dont les 157 porteurs nés en France entre 1891 et 1915 sont majoritairement lorrains (Meurthe-et-Moselle : 25 ; Vosges : 23 ; Meuse : 23).

2.3 Les détoponymiques

Les noms de famille issus d’un toponyme constituent la catégorie la moins bien représentée. Parmi ces NF apparaît *Bruyères*, dans les Vosges, à mettre en relation avec une attestation comme *Doron de Bruyeres* (passage de Bruyères, 1482) ; son étymon est le nom de lieu (ci-après NL) vosgien *Bruyères*, relevé dès 1138 (Marichal 1941). Le NL *Ceintrey* (MM), attesté dès 960 (Buchmüller-Pfaff 1990, 137), est également l’étymon du NF *Ceintrey*, dont les trois porteurs nés entre 1891 et 1915 sont meurthe-et-mosellans (Damelevières et Neuves-Maisons) ; Gerart *Ceintrey* apparaît quant à lui à Circourt-sur-Mouzon (Vosges) en 1499.

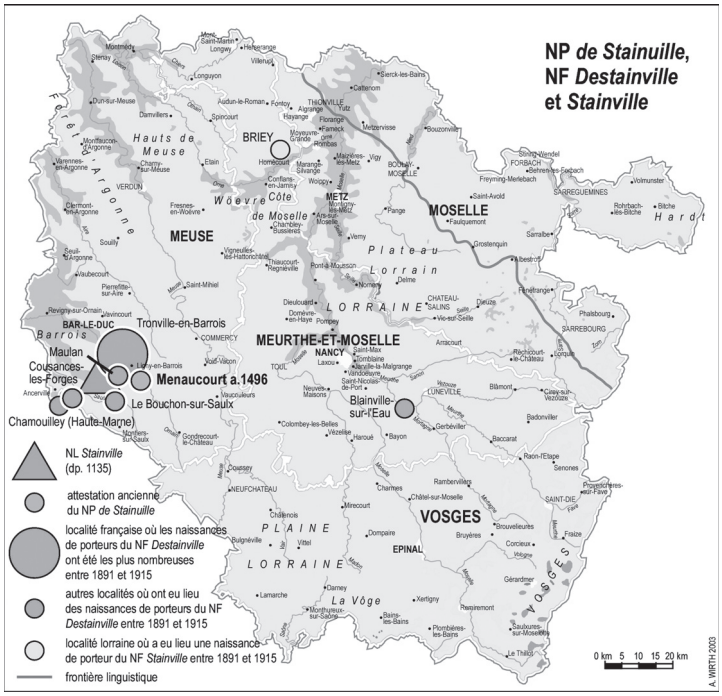
Le NF contemporain *Chamagne* a pour étymon le nom de la commune vosgienne *Chamagne* : c’est ce qu’indique la localisation de ses 67 porteurs nés en France entre 1891 et 1915 dans deux foyers, l’un au nord des Vosges (MM : 21), l’autre au sud (Haute-Saône : 19 ; Côte-d’Or : 9). Ce NL, attesté dès 1164 (Marichal 1941), est aussi à l’origine des attestations : *Nicolas de Chamaigne* (Châtel-sur-Moselle, 1578) et *Nicolas Chamagne* (Châtel-sur-Moselle, 1601), désignant la même personne. Dans *Colin de Girompaire* (grande prévôté de Saint-Dié, 1416), c’est le NL *Girompaire*, nom d’un hameau des communes de Saint-Léonard et de Saulcy-sur-Meurthe (Vosges) attesté depuis 1311 (Marichal 1941), que l’on retrouve. Il est à l’origine du NF *Girompaire*, dont les 11 porteurs nés entre 1891 et 1915 se localisent tous dans le département des Vosges, à quelques kilomètres seulement, dans leur grande majorité, du hameau qui leur a donné leur nom.

Deux auteurs se sont intéressés au NF *Destainville* : Thouvenot (1953, 64) et Morlet (1997). Le premier se contente de le citer parmi les NF détoponymiques meusiens, sans en préciser l’étymon ; la seconde, en revanche, mentionne explicitement le nom de la commune meusienne qui en constitue l’étymon. Cette étymologie est confirmée par la répartition à grande échelle des porteurs de ce NF nés en France entre 1891 et 1915, principalement dans la Meuse (14 sur 24), avec rayonnement dans la Meurthe-et-Moselle, à Blainville-sur-l’Eau (3 porteurs), et dans la Haute-Marne, à Chamouilley (2 porteurs), soit 19 porteurs sur 24 en tout dans ces trois départements.

Leur répartition à petite échelle est plus remarquable encore (cf. carte 1) : on y constate en effet la position épicentrique du NL éponyme[2] ainsi que la très faible distance entre la localité de notre attestation de 1496, Menaucourt, et celle où les naissances ont été les plus nombreuses en France entre 1891 et 1915, Tronville-en-Barrois : 9,5 km à peine les séparent. On en conclut que les porteurs, entre 1496 et 1891, se sont rapprochés de deux villes : Bar-le-Duc dans la Meuse (Tronville-en-Barrois et Maulan), et Saint-Dizier, en Haute-Marne (Chamouilley et Cousances-les-Forges n'en sont distantes respectivement que de 7 et 10 km).

Le NL *Stainville* est à l'origine d'un autre NF, *Stainville*, mais la localisation de ses porteurs nés entre 1891 et 1915 est moins parlante : seul un sur neuf est lorrain, de Briey plus précisément ; les autres se situent surtout à Paris et en région parisienne (4 dans la capitale, 1 dans les Hauts-de-Seine et 1 en Seine-Saint-Denis), ainsi que dans le Nord (1) et le Rhône (1). L'étymon de ce NF ne fait cependant aucun doute : à notre connaissance en effet, aucun NL homonyme n'existe en France ; c'est donc le même étymon qui est à l'origine de *Destainville* et de *Stainville*. La faible représentation de ce dernier et sa localisation dans le nord de la Lorraine, à bonne distance du foyer meusien du NF *Destainville*, ne nous autorise cependant pas à en faire une variante du premier (avec suppression tardive de la préposition) ou au contraire à y voir deux NF de créations distinctes. Seules des recherches généalogiques permettraient peut-être davantage de précision.

[2] « La cartographie à petite échelle fait apparaître des configurations d'aire typiques, incontestablement parlantes. Neuf fois sur dix, le NL éponyme s'inscrit à l'intérieur de l'aire d'intensité maximale du NF. L'éponyme peut être dit alors en position 'épicentrique' » (Chambon 1992, 72).



Carte 1. Le NL *Stainville* et les anthroponymes associés.

Le NF *Thiaville*, présent dans *Anthonne de Thiaville* (Thiaville/Lachapelle, MM, 1489/1490) et *Thouuenat Thiaville* (Rambervillers, 1489/1490), a pour étymon le NL *Thiaville* (MM), attesté dès 962 (Lepage 1862).

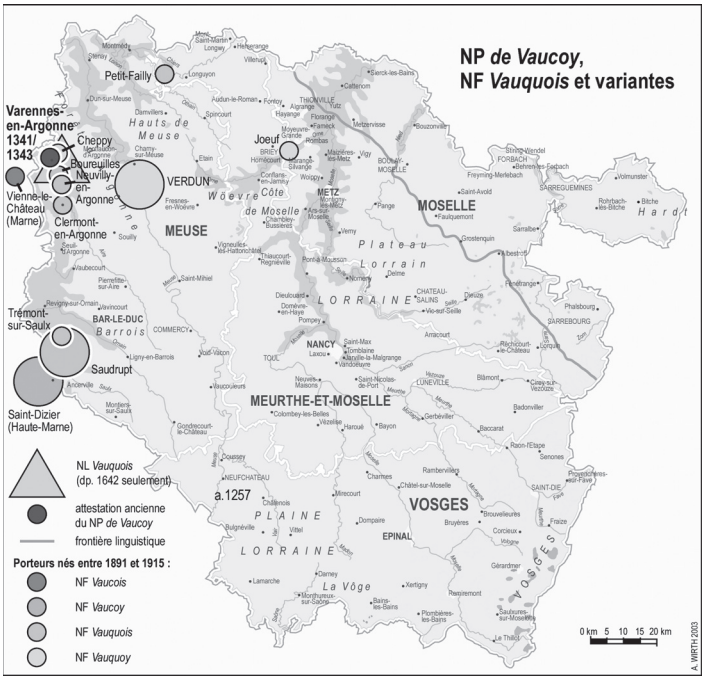
Enfin, le NL meusien *Vauquois* est à l'origine du NF meusien *Vauquois*, dont 9 des 12 porteurs nés entre 1891 et 1915 le sont dans la Meuse, essentiellement à Saudrupt et dans l'Argonne, à proximité du NL éponyme. La première mention de ce dernier relevée dans le dictionnaire topographique date seulement de 1642 (Liénard 1872) ; la mention anthroponymique *Tabar de Vaucoy* (Varennes-en-Argonne, 1341/1343) permet indirectement d'antédater de trois siècles environ l'apparition de ce NL.

Mais d'autres NF modernes doivent également être rattachés au NL *Vauquois*. Pour cette même période 1891–1915, 3 des 5 personnes nommées *Vaucoy* ont vu le jour à Saint-Dizier, en Haute-Marne ; un rapide coup d'œil sur la carte nous apprend que cette commune se situe à proximité de la frontière avec la Meuse, à quelques kilomètres seulement de Saudrupt : ce NF est donc sans aucun doute une variante assez récente du NF *Vauquois*. La même remarque peut être faite pour le patronyme *Vaucois*, dont le seul porteur qui ne soit pas né en Île-de-France est marnais (Vienne-le-Château) ; là encore, la localisation de sa commune d'origine, non loin de la Meuse, montre clairement que ce NF doit être lui aussi rattaché au NL *Vauquois*. C'est également le

cas pour le NF *Vauquoy*, qui connaît sa plus grande fréquence à Verdun, importante cité épiscopale qui a pu très tôt exercer son attraction sur les porteurs de ce NF.

Notre porteur du XIVe siècle est peut-être l'ancêtre de ces NF modernes aux graphies différentes mais à l'origine commune : c'est ce que semble indiquer la confrontation de leur localisation. Trois zones se dessinent : la première, à proximité du NL éponyme et de notre attestation ancienne, pourrait être à l'origine des deux autres, situées dans ou à proximité de villes plus importantes, Verdun pour l'une, et Bar-le-Duc et Saint-Dizier pour l'autre.

Aucun rapport n'a en revanche pu être établi entre ce foyer meusien et le NF moderne *Vauquoit* : le seul porteur de ce patronyme né entre 1891 et 1990 a en effet vu le jour dans un département éloigné, la Sarthe. Seules des recherches généalogiques poussées permettraient peut-être de le rattacher à nos autres NF.



Carte 2. Répartition du NP *de Vaucoy*, du NF *Vauquois* et de ses variantes et du NL *Vauquois*.

2.4 NF dont l'origine est obscure

Dans cette catégorie figure le NF *Challemand*, dont les 5 porteurs nés en France entre 1891 et 1915 sont tous vosgiens (Saint-Nabord et Remiremont) ; *Nicolas Challe-*

mant (Châtel-sur-Moselle, 1601) est peut-être leur ancêtre. Ce nom est peut-être une adaptation française d'un NP allemand non identifié se terminant par *-mann*.

Le NF *Chapoulot*, dont le seul porteur né en France entre 1891 et 1915 est vosgien (Moyemont), se rencontre en Meurthe-et-Moselle et dans les Vosges au XVIIe et au XVIIIe siècle (FamilySearch) ; la première attestation fournie par notre documentation (*Chapoulet*) date de 1488 et est localisée à Maron (MM).

Les NF *Chatourel* et *Chatourelle* sont quant à eux meusiens : entre 1891 et 1915, 4 porteurs du premier sur 5 au total sont nés dans la Meuse (Loxéville, Ligny-en-Barrois, Cousances-au-Bois, Brauvilliers), tout comme les 3 du second, à Willeroncourt et Velaines. Quatre siècles auparavant, dans le même département, en 1495, on trouve les mentions de : *Didette vefve de Chatourel* (Morlaincourt), *Anthoine Francois dit Chatourel sergent* (id.), *Colot Chatourel* (Willeroncourt), *Demengin Chatourel* (Nançois-le-Grand) et *Didier Chatourel* (Salmagne/Dagonville).

Le nom de *Mengin Chaufflaire* (Certilleux, 1499) pourrait être une adaptation française du NF allemand *Schaufler*, *Schäufler* (Gottschald 1982 *s.v. Schaufel*).

Ces quelques exemples montrent, s'il en était encore besoin, que l'étude des noms de famille lorrains et de leur origine et histoire possède encore un beau potentiel de travail et de découvertes…

Université Catholique de Louvain Aude WIRTH-JAILLARD

3. Références bibliographiques

Billy, Pierre-Henri / Sauvadet, Marie-Renée, 1998–2001. *Dictionnaire historique des noms de famille du Puy-de-Dôme,* Clermont-Ferrand, A.R.G.H.A., 2 vol.

Buchi, Éva, 2001. «La méthodologie de l'étymologie des noms de famille (domaine français et galloroman)», *RIOn* 7, 105-127.

Buchmüller-Pfaff, Monika, 1990. *Siedlungsnamen zwischen Spätantike und frühem Mittelalter. Die* -(i)acum-*Namen der römischen Provinz Belgica Prima*, Tübingen, Niemeyer.

Cano González, Ana Maria / Germain, Jean / Kremer, Dieter (ed.), 2004. *Dictionnaire historique de l'anthroponymie romane. Patronymica Romanica* (PatRom). *Volume II/1. L'homme et les parties du corps humain (première partie)*, Tübingen, Niemeyer.

Cano González, Ana Maria / Germain, Jean / Kremer, Dieter (ed.), 2007. *Dictionnaire historique de l'anthroponymie romane. Patronymica Romanica* (PatRom). *Volume I/1. Introductions. Cahier des normes rédactionnelles. Morphologie. Bibliographies*, Tübingen, Niemeyer.

Cano González, Ana Maria / Germain, Jean / Kremer, Dieter (ed.), 2010. *Dictionnaire historique de l'anthroponymie romane. Patronymica Romanica* (PatRom). *Volume I/2. Bibliographie des sources historiques*, Tübingen, Niemeyer.

Cano González, Ana Maria / Germain, Jean / Kremer, Dieter (ed.), 2004. *Dictionnaire historique de l'anthroponymie romane. Patronymica Romanica* (PatRom). *Volume II/1. L'homme et les parties du corps humain (première partie)*, Tübingen, Niemeyer.

Chambon, Jean-Pierre, 1992. «Quelques noms de famille français issus de noms de lieux. Essai de traitement lexicographique», in: Taverdet, Gérard (ed.), *Dictionnaire historique des noms de*

famille romans. Actes du Colloque IV (Dijon, 24–26 septembre 1990), Tübingen, Niemeyer (*Patronymica Romanica* 6), 69-92.

Chambon, Jean-Pierre, 2002. «D'une linguistique populaire écrite par des savants : notes sur les dictionnaires français d'anthroponymie», AICOS 18/5, Niemeyer, 7-13.

Corbet, Patrick / Wirth, Aude, 2011. «Claude… le lorrain. Le culte de saint Claude en Lorraine des origines au XVIIe siècle : premiers résultats», in: Roth, François (ed.), *Lorraine, Bourgogne et Franche-Comté, mille ans d'histoire. Actes du colloque tenu les 13 et 14 novembre 2009 au Conseil régional de Lorraine*, Moyenmoutier, EDHISTO, 115-134.

Dauzat, Albert, 1978[6] [1951[1]]. *Dictionnaire étymologique des noms de famille et prénoms de France*, édition revue et augmentée par M.-Th. Morlet, Paris, Larousse.

FamilySearch, Église de Jésus-Christ des Saints des derniers jours. <www.familysearch.org/search>.

GeneaNet. <www.geneanet.org>.

Géopatronyme. <www.geopatronyme.com/>.

Gottschald, Max, 1982[5]. *Deutsche Namenkunde. Unsere Familiennamen*, Berlin/New York, Walter de Gruyter.

Hasselrot, Bengt, 1957. *Étude sur la formation diminutive dans les langues romanes*, Uppsala/Wiesbaden, Lundequistska Bokhandeln-Harrassowitz.

Hoffman, William F., 1998. *Polish Surnames : Origins and Meanings,* second edition, revised, Chicago, Polish Genealogical Society of America.

Kremer, Dieter (ed.), 1997. *Dictionnaire historique de l'anthroponymie romane* (PatRom). *Présentation d'un projet*, Tübingen, Niemeyer.

Lepage, Henri, 1862. *Dictionnaire topographique du département de la Meurthe*, Paris, Imprimerie Impériale.

Liénard, Félix, 1872. *Dictionnaire topographique du département de la Meuse*, Paris, Imprimerie Nationale.

Marichal, Paul, 1941. *Dictionnaire topographique du département des Vosges comprenant les noms de lieu anciens et modernes*, Paris, Imprimerie Nationale.

Morlet, Marie-Thérèse, 1997[2] [1991[1]]. *Dictionnaire étymologique des noms de famille*, nouvelle édition revue et augmentée, Paris, Perrin.

Nyrop, Kristoffer, 1936[2]. *Grammaire historique de la langue française. Tome troisième : formation des mots*, Copenhague, Gyldendalske Boghandel.

Thouvenot, Jean, 1953. «Sur une méthode d'investigation anthroponymique», *Le Pays lorrain* 34, 63-68.

Wirth, Aude, 2003. *Éléments d'anthroponymie lorraine. Application de la méthodologie PatRom à une sélection de noms de famille*, Dijon, Association bourguignonne d'études linguistiques et littéraires.

Wirth, Aude, 2007. *L'anthroponymie de la Lorraine romane du XIe au XVIIe siècle*, Nancy, thèse inédite de sciences du langage.