



HAL
open science

Table ronde: "Des corpus linguistiques avant les corpus électroniques"

Emilie Aussant, Marc Baratin, Franck Cinato, Anne Grondeux, Cendrine Pagani, Pascale Rabault-Feuerhahn, Bernard Colombat

► To cite this version:

Emilie Aussant, Marc Baratin, Franck Cinato, Anne Grondeux, Cendrine Pagani, et al.. Table ronde: "Des corpus linguistiques avant les corpus électroniques". *Dossiers d'HEL*, 2017, *Analyse et exploitation des données de corpus linguistiques*, 11, pp.97-117. hal-01512695

HAL Id: hal-01512695

<https://hal.science/hal-01512695>

Submitted on 24 Apr 2017

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

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

TABLE RONDE : « DES CORPUS LINGUISTIQUES AVANT LES CORPUS ÉLECTRONIQUES ? »

Émilie AUSSANT, Marc BARATIN, Franck CINATO,
Anne GRONDEUX, Cendrine PAGANI-NAUDET

Compte rendu de la discussion : Pascale RABAULT-FEUERHAHN

Modérateur : Bernard COLOMBAT¹

PRÉAMBULE

Ce texte rend compte de la table ronde organisée le 30 janvier 2015. Lors d'une réunion préparatoire (23 janvier 2015), les participants avaient décidé de réduire au maximum leur première intervention pour donner largement la parole à l'assistance, de telle sorte qu'il en résulte un vrai dialogue. Pascale Rabault-Feuerhahn a bien voulu établir le compte rendu des échanges. Nous donnons ci-dessous le texte (parfois revu par l'auteur) des interventions initiales et le relevé des échanges auxquels ces dernières ont donné lieu.

BERNARD COLOMBAT.
POUR LANCER LE DÉBAT...

L'idée directrice de la table ronde est de se demander s'il existe une « linguistique de corpus » avant même la création des corpus récents, qui semblent désormais un préalable incontournable à la constitution de toute entreprise « sérieuse » de description des langues. Il semble qu'on peut répondre affirmativement, car :

– il y a des « grammaires de corpus », c'est-à-dire des grammaires qui s'appuient sur un corpus important d'exemples ou d'extraits de la langue, *vs* des ouvrages qui s'appuient sur *très* peu d'exemples, donnés comme « prototypiques » (cela va des *Turba ruunt* ou *Prata rident*, ou même le simple *curro* des médiévaux, au *Pierre aime Marie* de la grammaire générative) ; ces grammaires de corpus sont courantes par exemple au XVI^e s. (ex. Valla, Linacre, Sanctius pour le latin, Palsgrave pour le français, etc.) et s'opposent à des ouvrages à l'illustration plus parcimonieuse (comme le *De causis* de Scaliger, la *Gramerç* de 1562 de Ramus, etc.) ;

¹ Les appartenances des intervenants sont les suivantes : pour É. Aussant, F. Cinato, B. Colombat, A. Grondeux : UMR 7597 « Histoire des Théories Linguistiques », Univ. Paris Diderot/CNRS ; pour M. Baratin : UMR 8163 « Savoirs, Textes, Langage », Univ. Lille III/CNRS ; pour C. Pagani-Naudet : UMR 7320 « Bases, Corpus, Langage », Univ. de Nice-Sophia Antipolis/CNRS ; pour Pascale Rabault-Feuerhahn : UMR 8547 « Pays germaniques », ENS Ulm/CNRS.

– il y a même des grammaires qui contiennent un tel « réservoir » qu’elles peuvent se constituer comme « image » d’une langue donnée et permettre par elles-mêmes une reproduction assez complète de la langue (pensons par ex. à l’*Ars Prisciani*, opposée à l’*Ars Donati*, qui a permis aux moines irlandais d’appréhender sinon la totalité, du moins un très large échantillon de la latinité) ;

– ces ouvrages s’appuient sans doute sur les corpus préexistants, pas encore constitués comme « grammaire », mais réservoirs encore informels de données.

On peut alors se poser, entre autres, les questions suivantes :

– comment se construit le corpus ? comment se fait alors la sélection des données ? quels sont les moyens techniques qui permettent de les exploiter (par ex. la réunion en codex, etc.) ? quelle mémoire est en jeu pour l’exploitation et l’organisation de ces données ?

– ces masses de données sont-elles homogènes, ou hétérogènes ? le grammairien maîtrise-t-il toujours son corpus ? l’exhaustivité est-elle un but recherché, et pourquoi ?

– le fait de constituer une grammaire de corpus a-t-il des incidences sur la théorie mise en œuvre ? quel est le rôle de la reprise de l’exemple, que cette reprise soit simple ou au contraire dupliquée sous différentes formes ? l’utilisation massive d’exemples, opposée à une sélection rigoureuse, a-t-elle une signification ? une même donnée peut-elle servir à illustrer des faits différents ?

En quelques mots : quelle est la fonction du corpus ? quelle est sa nature ? quel est l’intérêt de l’accumulation par rapport à la sélection ? comment se construit la représentativité ?

ÉMILIE AUSSANT.

CORPUS ET GRAMMAIRE DANS L’INDE ANCIENNE : DES LIENS CONSUBSTANTIELS

L’idée que je souhaiterais faire passer, en participant à cette table ronde, c’est que la grammaire, dans l’Inde ancienne d’expression sanskrite a, dès le départ, entretenu des rapports consubstantiels avec les corpus. Mon propos s’articule en trois points.

1. La grammaire sanskrite, née d’une technique de transmission orale d’un corpus de textes sacrés

La question des corpus se pose en effet, de manière centrale, dès les plus anciens textes qui nous soient parvenus, puisque ces textes, que l’on désigne sous le nom de *Veda* (litt. « le savoir », c’est-à-dire, la connaissance sacrée), constituent le corpus qui fonde la religion védique, la forme la plus ancienne du brahmanisme, et qui constitue le point de départ de toutes les doctrines de l’Inde ancienne. Ces textes, dont les plus anciens remonteraient à 1500 avant notre ère, sont de contenu et de forme très divers et ont en commun de reposer sur une « révélation » (*śruti*). Ils se distinguent en cela d’un autre vaste ensemble de textes (dont font partie, entre autres, les traités de phonétique, de métrique, d’étymologie et de grammaire) qui reposent, quant à eux, sur la « (tradition confiée à la) mémoire » (*smṛti*). Les textes « révélés » comprennent les *saṃhitā*, des « collections » en vers rassemblant des hymnes, des prières, des formules rituelles, ainsi que des commentaires de ces *saṃhitā*.

Il y a quatre types distincts de *saṃhitā* – les strophes (*ṛc*) récitées au cours du rituel et consignées dans la *Rksamhitā* ou *Ṛgveda*, les formules rituelles (*yajus*) regroupées (avec ou sans commentaire) dans la *Yajuḥsaṃhitā* ou *Yajurveda*, les mélodies (*sāman*) rassemblées dans la *Sāmasaṃhitā* ou *Sāmaveda*, enfin les formules magiques (*atharvan*), qui forment l'*Atharvasaṃhitā* ou *Atharvaveda* – ; cette division répond, dans l'organisation définitive du culte, aux fonctions des officiants (Renou 1947, p. 3). Ainsi, les officiants réalisant tel acte cultuel se réfèrent à tel corpus² de textes védiques, les officiants réalisant tel autre acte cultuel se réfèrent à tel autre corpus, etc. La tradition védique, voire brahmanique, s'ordonne résolument autour de ces quatre Veda. On n'imagine pas de textes qui ne gravitent autour de ces masses : c'est même la seule définition plausible que nous ayons du « canon » védique³ (Renou 1947, p. 13).

La majeure partie de cette littérature védique a été conçue et conservée, donc transmise, oralement⁴. Cette pratique a profondément marqué la culture indienne, son rapport aux textes et, surtout, sa réflexion sur la langue. De très bonne heure, en effet, la transmission orale du corpus que constituent les textes sacrés a imposé la mise au point de techniques de récitation reposant sur une analyse fine de la langue. La plus ancienne⁵ de ces techniques consiste en l'établissement de la version mot-à-mot (*padapāṭha*) d'un texte, où l'on abroge les règles euphoniques de liaison en insérant des pauses entre les mots et les morphèmes, à partir de sa version « continue » (*saṃhitāpāṭha*), où les règles euphoniques sont réalisées. Cette analyse mot-à-mot des textes implique, d'une part, une connaissance très poussée de la phonétique et, d'autre part, une réflexion particulièrement élaborée concernant le découpage de la chaîne parlée (ce ne sont pas forcément les mêmes règles qui s'appliquent selon que l'on est à la frontière entre deux mots ou à la frontière entre deux morphèmes). C'est très probablement dans cette connaissance des unités sonores et dans cette réflexion relative aux unités de la chaîne parlée que la grammaire – le *Vyākaraṇa* – s'enracine. Autrement dit, la grammaire indienne d'expression sanskrite serait née d'une technique développée en vue de la transmission orale d'un corpus de textes sacrés, les *saṃhitā* védiques.

2. Les données pāṇinéennes

La grammaire (*vyākaraṇa*), la phonétique (*śikṣā*), la métrique (*chandas*) et l'étymologie (*nirukta*) font partie, avec d'autres savoirs d'érudition, des disciplines qualifiées d'« auxiliaires du Veda » (*Vedāṅga*) ; elles furent spécifiquement élaborées pour la bonne conservation des textes sacrés⁶ et le bon déroulement matériel du culte⁷. Les deux premières – grammaire et phonétique – ne sont pas rattachées à un Veda particulier, il s'agit de textes « valant pour tous les Veda » (*sarvavedasādhāraṇa*) et qui traitent d'une

² Louis Renou (1947, p. 11) emploie déjà ce terme ; on trouve une autre occurrence p. 21.

³ Louis Renou précise ailleurs (1985, § 614) : « En dépit de son caractère sacré, c'est une littérature ouverte où des textes nouveaux, des annexes aux textes anciens ont toujours pu s'accréditer. »

⁴ C'est à une date récente que les textes ont été mis par écrit (le témoignage d'Al Birûnî au XI^e s. évoque un Veda récemment consigné par écrit au Cachemire).

⁵ Soit autour de 800 av. notre ère (voir Pinault 2000, p. 223-224).

⁶ C'est le cas de la grammaire (*vyākaraṇa*), de la phonétique (*śikṣā*), de la métrique (*chandas*) et de l'étymologie (*nirukta*). Ce statut d'« auxiliaire du Veda » n'a pas empêché la grammaire de s'affranchir de la sphère religieuse ; elle s'est rapidement développée de manière autonome, en faisant de la langue – et non pas seulement ou préférentiellement des textes védiques – son centre d'intérêt.

⁷ C'est le cas de l'astronomie (*jyotiṣa*) et du rituel (*kalpa*).

matière qui est qualifiée de *sāmānya* « générale » (Renou 1947, p. 14). Les textes relevant des deux autres disciplines – la métrique et l'étymologie – sont rattachés à des textes védiques précis.

Le plus ancien traité de grammaire sanskrite qui nous soit conservé est l'*Aṣṭādhyāyī* (A) de Pāṇini (P), dont la date de composition se situerait au V^e s. avant notre ère. Ce traité grammatical, qui s'inscrit dans la lignée de travaux antérieurs aujourd'hui perdus (P cite dix prédécesseurs) est le texte fondateur de la tradition grammaticale la plus ancienne et la plus longue qui soit (V^e s. avant notre ère – XVIII^e s.). Sur un plan pratique, l'objectif de l'A est d'enseigner comment former des mots corrects à l'aide d'affixes. Ces affixes (*pratyaya*) sont directement introduits dans les quelques 4000 règles du traité Pāṇinéen, contrairement à la majorité des bases auxquelles ils s'adjoignent : les racines verbales (*dhātu*) sont listées dans un recueil qui leur est dédié (le *Dhātupāṭha*⁸) ; quant aux thèmes nominaux, ils sont soit 1) introduits de manière générale par une règle (A 1.2.45 *arthavad adhātur apratyayah prātipadikam* (« une [unité] signifiante, qui n'est ni une racine verbale ni un affixe [est] un thème nominal (*prātipadika*) »), 2) soit donnés dans des listes à part (*Gaṇapāṭha*⁹), 3) soit directement enseignés dans les règles du traité. L'A présuppose donc l'existence de listes d'items verbaux et nominaux. Les phonèmes sont également enseignés sous forme de liste (*Pratyāhārasūtra*¹⁰). Ces trois (groupes de) listes (*Dhātupāṭha*, *Gaṇapāṭha*, *Pratyāhārasūtra*) ont sans doute été compilées avant P, mais leur arrangement a été adapté à la théorie grammaticale. Le grammairien exploite ces différentes listes pour renvoyer à des groupes, plus ou moins importants, d'unités en ne citant que la première unité du groupe. C'est un procédé fort économique qui lui a permis d'alléger considérablement la forme des règles de son traité (transmis, lui aussi, oralement).

Quant aux exemples (c'est-à-dire les mots corrects auxquels les dérivations enseignées dans l'A doivent aboutir) et contre-exemples, la grammaire de Pāṇini en est totalement dépourvue ; on suppose qu'ils étaient donnés dans la glose orale des *sūtra* (on les trouve dès les tout premiers commentaires de l'A).

Maintenant, sur quels types de données P s'appuyait-il ? L'objet de son traité est une langue commune, qui englobe la langue sacrée (*chandās*), c'est-à-dire les formes des *samhitās* védiques, et la langue parlée ordinaire, non sacrée (*bhāṣā*), probablement un dialecte du Nord. Le traité est censé valoir pour les deux variétés, sans distinction chronologique : la langue sacrée ne représente pas, pour le grammairien, un stade plus ancien de la variété parlée ; il pense décrire une seule et même langue qui ne comporterait que des registres différents. L'enseignement védique de P est volontiers

⁸ Cette liste comprend près de 2000 racines verbales réparties en dix classes (*gana*), nommées d'après le premier élément de chaque classe. La répartition des différentes classes se fait selon la formation du présent et, subsidiairement, selon d'autres critères (voix, accentuation). L'objectif du recueil est d'établir la liste des racines dont l'existence est présupposée par l'A. Le nombre de racines enseignées dépasse largement celui des emplois connus, seule la moitié de ces racines environ est attestée dans la littérature.

⁹ Le *Gaṇapāṭha* comprend 261 listes ordonnées d'éléments autres que les racines, constituées notamment de thèmes nominaux, de pronoms, de numéraux, etc., arrangées par ordre alphabétique. Certaines listes sont closes, tandis que d'autres, nommées *ākṛtigāna*, sont à considérer comme des énumérations de spécimens. L'objectif de l'ouvrage est de permettre l'application d'une ou plusieurs règles de l'A à un ensemble de formes présentant certains traits de dérivation communs.

¹⁰ L'ordre est le suivant : voyelles simples, diphtongues, aspirée et semi-voyelles, nasales, sonores aspirées, sonores non aspirées, sourdes aspirées, sourdes non aspirées, sifflantes, aspirée.

approximatif : il travaille à partir d'un corpus de textes lacunaire (d'après les spécialistes), tiré de sources variées. Des traits importants de la langue védique (comme l'emploi des temps et des modes) sont passés sous silence, sans doute parce qu'ils diffèrent trop profondément de l'usage classique (Pinault 1989, p. 350). La description que le grammairien donne de la variété ordinaire montre bien qu'il s'agit d'une langue qui se pratique « dans le monde » et non « dans les livres » : une relative liberté est accordée dans la réalisation de nombreux processus. Pāṇini, en effet, admet très souvent un choix, en indiquant (ou non) une préférence (fondée sur son propre parler), par un grand nombre de règles optionnelles. Ces règles optionnelles correspondent très largement à l'usage connu et ont été testées avec succès sur trois corpus : l'usage propre à l'A, la littérature védique et la littérature classique antérieure au V^e s. La grammaire de P se fonde donc en partie sur des corpus (formes védiques tirées des textes, listes de racines, de thèmes nominaux, de phonèmes) et en partie sur les formes en usage à son époque et considérées comme correctes.

3. La grammaire de Pāṇini, le corpus du « sanskrit pāṇinéen »

La situation s'inverse avec le changement de statut du sanskrit, lorsque celui-ci devient langue seconde autour du début de notre ère : la grammaire de P constitue alors (avec deux de ses tout premiers commentaires, les *Vārttika* de Kātyāyana – III^e s. av. notre ère – et le *Mahābhāṣya* de Patañjali – II^e s. av. notre ère) l'unique référence en matière de langue correcte.

Contrainte d'être conforme au système pāṇinéen, la variété linguistique s'est inévitablement figée. Mais elle a tout de même continué sa transformation, dans le cadre des limites du traité, en exploitant certaines constructions sur lesquelles Pāṇini, par souci de généralisation, n'avait pas statué. En ne limitant pas, par exemple, le nombre de sous-membres dans les composés, le grammairien a ouvert la voie – en poésie (*kāvya*), principalement – à la formation de composés « à rallonge ». Le sanskrit, en épousant et en tirant parti des contraintes du système, a amélioré la précision du modèle descriptif, le rendant plus adéquat à son objet qu'il ne l'était au départ. Un peu comme si l'A avait, *in fine*, fondé un corpus, celui du sanskrit pāṇinéen.

MARC BARATIN.

L'ARS PRISCIANI, UNE GRAMMAIRE DE CORPUS ?

Peut-on dire que la grammaire systématique gréco-romaine est née de la constitution d'un corpus ? La fondation de la Bibliothèque d'Alexandrie, au début du III^e s. avant notre ère, a provoqué la réunion d'un nombre considérable de manuscrits, témoins souvent de traditions différentes des mêmes textes, et a ainsi entraîné la nécessité de classer ces manuscrits, d'établir les textes, de les commenter pour les rendre accessibles au public et pour en justifier les choix critiques. Ce travail, qui fut celui des philologues alexandrins, a d'abord donné naissance, au cours des III^e et II^e siècles avant notre ère, à toutes sortes d'observations ponctuelles sur tel ou tel auteur, de monographies sur des difficultés particulières, de recueils de termes spécifiques, dialectaux, poétiques, sortis de l'usage, etc. Cette dynamique a fini par donner lieu, à la charnière du II^e et du I^{er} siècle, à une réflexion sur les règles générales de la langue grecque, indépendantes de ces emplois particuliers. La grammaire comme description du fonctionnement systématique de la langue est ainsi à ses origines adossée au corpus des textes réunis dans la Bibliothèque. Ce corpus est néanmoins implicite : jamais il n'est désigné comme

tel, et les grammairiens alexandrins ne le considèrent apparemment pas comme un ensemble spécifique, préférant établir des distinctions entre les textes « à licences » (sc. poétiques) et les textes « sans licences » (de prose), ainsi qu'entre ceux-ci et l'usage commun.

Chez les grammairiens latins, le corpus implicite des auteurs classiques sert de support à la grammaire normative : les exemples sont « prélevés » (c'est le sens étymologique du mot *exemplum*) dans les textes de ces auteurs qui, à l'époque tardive, sont souvent très éloignés dans le temps des grammairiens qui les citent (il s'agit pour l'essentiel d'auteurs comme Térence, Cicéron ou Virgile, donc des deux derniers siècles avant notre ère, même chez des grammairiens tardifs des III^e, IV^e ou V^e siècle de notre ère). Par ailleurs, on trouve des recueils qui sont dans la lignée de la grammaire alexandrine : recueils lexicaux ou grammaticaux, conçus comme des séries d'exemples, ainsi des *idiomata*, ou tournures spécifiques à l'une des deux langues par rapport à l'autre, intégrés à la grammaire de Charisius, au IV^e siècle. Ces ensembles ne sont cependant pas présentés comme des corpus, mais plutôt comme des listes ouvertes.

La situation se modifie à l'extrême fin de l'Antiquité, avec Priscien, grammairien de Constantinople du début du VI^e siècle. Sa *Grammaire* ne vise pas tant à l'établissement d'une norme qu'à une présentation large des emplois attestés en latin, et rend donc compte d'un nombre considérable de formes, qu'elles soient anciennes, usitées, propres à des auteurs. Interprété par la suite comme ayant visé une certaine exhaustivité, il deviendra lui-même une sorte de corpus des formes latines. Surtout, il introduit une modification décisive dans l'emploi des exemples. Analysant le mode subjonctif dans le dernier livre de son traité, c'est-à-dire dans la partie consacrée à la syntaxe, il en présente d'abord une définition et commence à en décrire les emplois, en les illustrant d'exemples, à la manière usuelle des grammairiens de l'époque ; tout au plus le nombre des exemples fournis est-il particulièrement élevé ; mais, sans doute sous la poussée de ce nombre élevé, il se propose dans un second temps de prendre un texte particulier, le *De praetura urbana* de Cicéron (18, 109-121, *GL* 3, 258.1-264.15), qui est l'un des discours contre Verrès, et d'en examiner les occurrences du subjonctif. Toutes ces occurrences ne sont pas examinées, mais le principe est posé, novateur, de prendre un texte dans son entier pour analyser dans ce cadre une catégorie grammaticale.

L'utilisation de ce corpus entraîne-t-elle des modifications dans l'analyse ? Dans le cas de Priscien, cela paraît manifeste : les éléments de définition initialement posés se révèlent limités et étroits par rapport à la diversité d'emploi qui se dégage du corpus, pourtant lui-même peu étendu, et la cohérence de l'analyse souffre d'être soumise à cette méthode nouvelle. L'un des intérêts de cette œuvre majeure est d'assumer, au moment où se clôt l'Antiquité, les difficultés nées de cette innovation.

FRANCK CINATO.

COMPILATIONS MANUSCRITES EN TANT QUE CODEX-CORPUS

1. La grammaire est liée à la bibliothèque

En préambule, je citerai un article de M. Baratin (1996, p. 253) : « La genèse même de la grammaire est liée à la bibliothèque... » à partir de l'époque hellénistique et de la fondation de la bibliothèque d'Alexandrie (déb. III^e s. a.C.), l'objet de la grammaire n'est plus uniquement la transmission de la langue mais sa structuration en système qui se fonde sur un vaste corpus textuel issu de l'accumulation bibliographique. Les thèmes

de réflexions soulevés par cet article sont riches et nombreux, notamment à propos de la convergence entre grammaire et philologie, mais aussi sur la question de l'oscillation entre approche descriptive ou prescriptive.

Les réflexions suivantes voudraient simplement revenir sur cette problématique de la relation entre bibliothèque et grammaire dans le contexte du haut Moyen Âge. Marc Baratin a montré que la naissance de la littérature artigraphique s'appuie sur l'accumulation prise en charge par les bibliothèques, or, la perpétuation des savoirs grammaticaux, après la christianisation de la société antique, à l'inverse conditionne le contenu des bibliothèques.

2. La constitution des bibliothèques dépend des choix des grammairiens

Deux événements importants accompagnent les changements majeurs intervenus entre la fin de l'Antiquité et le début du Moyen Âge : le christianisme devenu religion d'état et le passage du *volumen* au *codex*. Le premier a opéré une *crisis* – une sélection (réalisée non sans difficulté) – dans le corpus textuel hérité de l'Antiquité et a conjointement opéré un remplacement culturel par l'introduction d'une nouvelle littérature (et l'oblitération partielle de l'ancienne) ; le second a permis la juxtaposition dans un même ensemble, ou unité codicologique, de plusieurs textes (ou livres) relevant d'un thème ou d'un domaine.

Les bibliothèques monastiques s'organisent donc autour de deux axes principaux : sacré / profane. La production de livres répond d'abord aux besoins des communautés, en termes de culte, mais aussi de formation. De ce fait, la nature des textes copiés est prescrite par les autorités de l'Église et par les grammairiens. Du point de vue de la grammaire, d'ailleurs, je dirais simplement que les grammairiens médiévaux ont puisé à un double corpus d'exemples paradigmatiques, l'ancien et le nouveau.

L'ancien corpus est défini par la tradition de l'héritage artigraphique des *Pagani* ; le nouveau, qui est constitué par toutes les productions chrétiennes, a eu une influence limitée et ponctuelle (cf. par exemple Bède, Boniface, Smaragde, etc.)¹¹. Le nouveau corpus n'a jamais supplanté l'ancien, car, disons-le simplement, il est impossible d'apprendre à lire le latin de Virgile avec celui de la Bible comme exemple.

L'exemple du catalogue de la bibliothèque de Reichenau (rédigé ca 821/822) montre que tous les livres qui n'appartiennent pas au domaine sacré (ou théologique) traitent soit d'histoire, soit de droit ou de médecine, soit encore des arts libéraux avec une composante grammaticale importante. Suivant l'ordre du catalogue, nous trouvons d'abord : 1) les textes canoniques ; 2) les Pères et les docteurs de l'Église ; 3) histoire, droit (selon les nations) et médecine ; 4) les livres liturgiques ; 5) autres docteurs ; 6) arts libéraux (suivant le programme de Boèce, Bède et Alcuin) ; 7) droit canon et liturgie ; 8) grammaire et connexes¹².

¹¹ Sur cet aspect, voir par exemple, Ellspermann, 1949 ; Hagendahl, 1958 ; Munk Olsen, 1992 ; Law, 1993 ; Perez-Rodriguez, 2001 ; Chin, 2008.

¹² La liste des livres de la bibliothèque de Reichenau (dressée en 821/822 ; Becker, 1885, § 6) présente les textes grammaticaux, en toute fin, à la suite du *De libris glossarum*, dans une section intitulée *De libris Prisciani*, où l'on notera que les poètes présents dans le programme de lecture sont presque essentiellement chrétiens, à l'exception de Virgile : <Grammatica et similia> (N° 392-414) : *De libris glossarum*. – 392. *glossarum ex diversis doctoribus excerptarum codex grandis I*. [= *Liber glossarum* (voir Grondeux 2015)] (...) *De libris Prisciani*. – 396. *De litera, de syllaba, et de octo partibus lib. XVI et de constructione lib. V [s. II] in cod. I*. – 397. *item Prisciani de litera, de syllaba et de octo partibus lib. XVI et de constructione lib. V [s. II] et Pompei tractatus in Donatum in cod. I*.

Il y aurait beaucoup à dire sur cette question du corpus-double (antique/chrétien) et des problématiques qui l'entourent, mais je souhaite surtout attirer l'attention sur le second aspect que j'ai évoqué, celui des livres manuscrits.

Le passage des *volumina* aux *codices* a permis la conception des recueils thématiques : toutes disciplines confondues, il semble que les maîtres responsables de la production des livres ont très tôt tenté de grouper les textes « thématiquement », avec comme résultat dans les faits que plusieurs textes en relation ou connexes ont bénéficié de transmissions communes ou parallèles. Sous cet aspect, les corpus textuels ont une réalité matérielle, ou du moins liée à une contingence matérielle.

3. Exemple des œuvres grammaticales de Priscien

Prenons comme exemple l'*Ars* de Priscien. Selon L. Holtz¹³, le corpus de ses œuvres se serait composé de 3 ou 4 lots¹⁴ qui s'observent encore dans les plus anciens témoins de ces textes : 1) *Ars* – 2) trois *opuscula* (*fig.*, *Ter.*, *praeex.*) – 3-4) œuvres de pédagogie élémentaire (*nom.* et *part.*).

Dans les manuscrits l'*Ars* est fréquemment accompagnée de plusieurs textes annexes (par ex. *Ars* I-XVIII + *Carm. de pond.* + *fig.* + *Ter.* + *Praeex.* + Rufin, *metr. com.*¹⁵) : ici, les trois *opuscula* de Priscien sont encadré par le *Carmen de ponderibus et mensuris* de Remius Favinus [Ps.-Priscien] et du traité de Rufin d'Antioche *in metra comicorum*. Cependant, en de très rares occasions seulement, l'*Ars Prisciani* se trouve associée aux œuvres de pédagogie élémentaire¹⁶.

Le *De nom.* (selon Passalacqua) est la plus diffusée des œuvres de Priscien (connue par plus de 33 manuscrits du IX^e s.), souvent en association avec les *Partitiones* ou d'autres *opera minora* de Prisc. ou ps.-Prisc. (*de fig.*, *de acc.*), mais surtout l'*Ars* de Donat.

– 398-399. *item Prisciani de litera, de syllaba, de nomine et verbo lib. X in codic. II.* – 400. *Donati et Sergii et Isidori de octo partibus in cod. I.* – 401. *item Donati, Isidori de octo partibus et Sergii de differentiarum ... Isidori, et glossae spirituales Eucherii in cod. I. (...)* – 414. *metra Virgilii Maronis in vol. I.* – 415. *item Virgilii georgicorum lib. III et Aeneidos lib. VI in cod. I.*

¹³ Holtz, 2009.

¹⁴ Holtz, 2009, p. 42-44 ; le découpage des lots est établi en fonction des dédicaces qui précèdent les œuvres ; à propos des colophons de Theodore (élève de Priscien et son « éditeur ») qui rythment le texte de l'*Ars grammatica* de Priscien, voir Ballaira, 1989 ; aussi De Nonno, 2009 à propos de la structure.

¹⁵ Paris, Bnf, 7501, s. IX/3 , Corbie ou encore Paris, Bnf, 7496 : *Ars* I-XVIII + *Fig.* + *Ter.* + *Praeex.* + Rufin. *metr. com.* + *metr. or.* + *carm. de pond.* (s. IX med., Auxerre).

¹⁶ L'association *Ars* + *inst. de nom.* est rarissime, car c'est le cas seulement dans deux manuscrits dont l'organisation dénote une tentative de réagencement majeure, puisque le *de nom.* s'est vu copié entre le 7^e et 8^e livre ; cette entorse au corpus priscienique normal s'explique par le fait qu'on a voulu regrouper tout ce que Priscien avait dit à propos du *nomen*, avant de traiter du *verbum*. – Dans deux autres manuscrits, la question est un peu différente : on a associé *Ars* + *nom.* postérieurement (dans un manuscrit en écriture irlandaise, Leiden, BPL 67 (siglé *L* par Hertz), où *De nom.* a été ajouté au XI^es.) ; tandis que le manuscrit de St-Amand, Paris, BnF, lat. 7498 – qui constitue le prototype même du second volume des *opera omnia* de Prisc. selon la typologie de L. Holtz, 2009, p. 45 sq. : *Ars XVII-XVIII* + *Part.* + *Inst. de nom.* + *Eutyches*, *Ars* + Rufin. *metr. com.* + *carm. de pond.* + *Fig.* + *Ter.* + *Praeex.*) – représenterait l'unique témoignage du corpus avant son éclatement. Cf. C. Jeudy, 1972, p. 85-86.

Louis Holtz note que l'*Inst. de nom.* et les *part.* « ont pu rapidement se détacher du corpus primitif, avant même la diffusion massive des *IG* sur le continent »¹⁷.

Or cette redistribution (lot 1-2 d'un côté et 3-4 de l'autre) ne s'explique pas par la codicologie, puisque les deux œuvres élémentaires ont été remplacées par des textes – dont Priscien n'est pas l'auteur (quoique que le *carm. de pond.* lui ait été parfois attribué) –, mais qui entretiennent un rapport avec le contenu des *opuscula*. Il s'avère donc qu'une réflexion induite par les circonstances historiques et linguistiques a été portée sur la nature (thème, contenu, niveau etc.) des textes à assembler. Les traités de pédagogie élémentaire ont rejoint chacun les corpus convenant à leur contenu : le *de nom.* avec Donat et les autres « grammaires scolaires » et les *part.* dans le voisinage de Virgile.

4. Corpus grammatical tripartite

Ce classement thématique n'a pas seulement affecté les œuvres de Priscien, car à côté des traités artigraphiques ont coexisté des recueils orthographiques, qui ont rassemblé de véritables ensembles à vocation normative. Ces derniers se trouvent souvent associés aux glossaires¹⁸, troisième versant de l'outillage grammatical. Du côté de la lexicographie/glossographie, les livres manuscrits montrent cette même volonté de regrouper les glossaires entre eux (ou en compagnie d'autres textes glossographiques, surtout des *synonyma* et des *differentiae*) en des ensembles d'*instrumenta* cohérents¹⁹.

En résumé, les manuscrits contenant des recueils de *grammaticalia* se répartissent en trois groupes (avec parfois quelques recoupements) :

- traités artigraphiques
- traités orthographiques
- ensembles lexicographiques et glossographiques

À partir de cette tripartition des ouvrages grammaticaux ou afférents, il est possible d'avancer quelques constats :

a) Les bibliothèques sont élaborées sur une notion d'ensemble induite par la distinction entre le domaine sacré et le domaine profane.

b) En leur sein le groupement répond à une organisation des disciplines qui donne la première place aux textes religieux et aux auteurs chrétiens. La grande majorité des

¹⁷ Holtz, 2009, p. 44 ; voir aussi Law 1993.

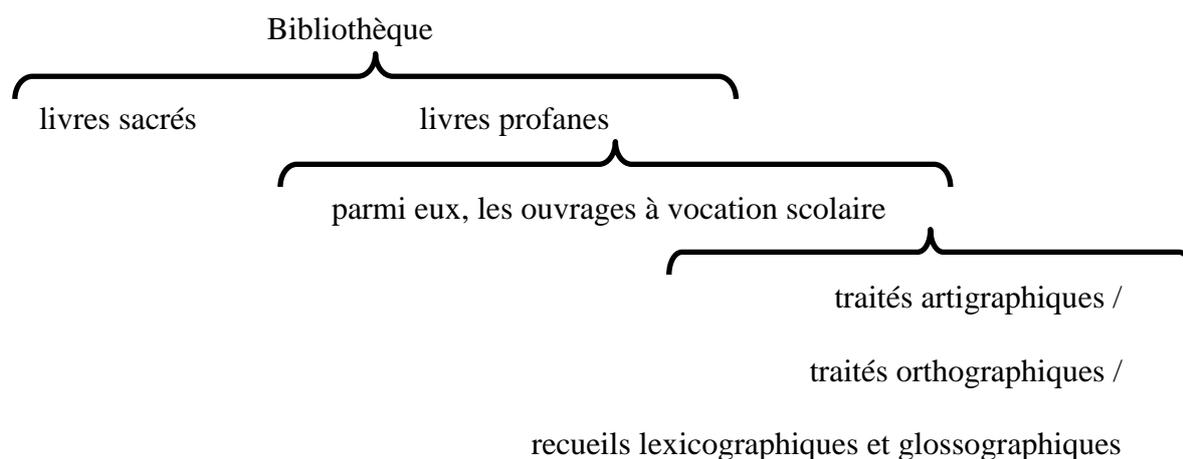
¹⁸ Par exemple, St-Gall, Stifsbibl. 249 ou Montpellier, Bibliothèque interuniversitaire, Section médecine, H 306 (voir De Paolis, 2004, spécial. p. 208-210 ; De Paolis, 2010, spéc. p. 290-291 et p. 276-277 où il distingue trois types de compilation manuscrites : a) traités artigraphiques associés aux orthographiques ; b) traités orthographiques seuls ; c) tr. orthographiques associés à des glossaires ; cf. aussi, par exemple, pour la fin du Moyen Âge, le contexte des petits textes *de nota aspirationis* et *de diphthongis*, cf. Biondi, 2011, p. 370 sq. – Voir aussi, plus généralement, Holtz, 1977.

¹⁹ À cet égard le manuscrit Leiden, BPL 67F constitue un bel exemple : (s. VIII / IX – France, nord-est) *Affatim* (f. 1ra-54rb) ; *Ab absens* (f. 54rb-62rc) ; *Abavus* (f. 62rd-104va) ; *Gl. Vergilianae* (f. 104va-118v) ; *Proprietatum* (f. 119ra-128rb) *Proprietatum Arma bellum idest pugna...* ; *Synonyma Ciceronis* (f. 129r-140rc) ; *Item, Cycero uetere suo salutem ...* (f. 140v-141vd) ; *Glossae Nonii Marcelli* (f. 142v-147ra) ; *Gl. Abligare* (f. 147ra-148rd) ; *Glossae iuris* (f. 148rd-149rd) (cf. Keil, *Gramm. lat.* 4, 277 sqq.) ; *Glossae spirituales Eucherii* (f. 149rd-152rb) ; *Voces variarum animantium* (f. 152ra-152rb), etc.

textes antiques sont considérés du ressort des grammairiens ; lesquels ayant la charge fondamentale de l'apprentissage linguistique, déterminent les programmes de lectures.

c) Au dernier échelon de la hiérarchie, les contenus des *codices* montrent une organisation des textes rationalisée en fonction des besoins des utilisateurs.

Ces quelques réflexions montrent qu'en toile de fond de la nébuleuse des textes à vocation scolaire règne malgré tout un ordre qui s'observe non seulement à travers l'organisation des bibliothèques, mais aussi par leurs volumes conçus en tant que recueils pragmatiques (voir fig. ci-dessous) : de tels livres, imbriqués dans un ensemble plus vaste, affectent une réalité physique à la notion abstraite de corpus textuel.



Abréviations utilisées

<i>Prisc.</i>	Priscianus Caesariensis grammaticus
<i>Ars gramm.</i>	<i>Prisciani Ars grammatica (siue Institutiones grammaticae)</i> [GL 2, 1-597 et GL 3, 1-377]
<i>Inst. de nom.</i>	<i>Prisciani Institutio de nomine et pronomine et uerbo</i> [Passalacqua 1999, 5-41 (= GL 3, 443-456)]
<i>Fig.</i>	<i>Prisciani libri minores: de figuris numerorum quos antiquissimi habent codices</i> [Passalacqua 1987, 4-18 (= GL 3, 405-417)]
<i>Ter.</i>	<i>Prisciani libri minores: de metris fabularum Terentii et aliorum comicorum</i> [Passalacqua 1987, 19-32 (= GL 3, 418-429)]
<i>Praeex.</i>	<i>Prisciani libri minores: Praeexercitamina</i> [Passalacqua, 1987, 33-49 (= GL 3, 430-440)]
<i>Part.</i>	<i>Prisciani partitiones</i> [Passalacqua 1999, 45-128 (= GL 3, 459-515)]
<i>Rufin. metr. com.</i>	Rufinus (<i>grammaticus Antiochensis</i>) <i>in metra comicorum</i> [P. D'Alessandro, 2004 (= GL 6, 554-565.8 ... <i>in metra Terentiana</i>)]
<i>Rufin. metr. or.</i>	Rufinus <i>de metris oratorum</i> [P. D'Alessandro 2004 (= GL 6, 565.9-578)]
<i>Carm. de pond.</i>	Remii Favini [Ps.-Priscien] <i>Carmen de ponderibus et mensuris</i> [Baehrens, <i>Poetae Latinae Minores</i> , V, 1883, 71-82 ; Reise, <i>Anthologia Latina, pars prior</i> , n° 486, 29-37, 1906 ; voir Raïos, 1983.
[<i>Prisc.</i>] <i>acc.</i>	ps.-Priscianus, <i>de accentibus</i> [Giammona 2012, 1-67 (= GL 3, 519-528)]

ANNE GRONDEUX.
LE *LIBER GLOSSARUM* (ANTE 800) : UN CORPUS ?

Le *Liber glossarum* (*LG*) est une gigantesque encyclopédie alphabétique en 55 000 lemmes dont les plus anciennes copies ont été effectuées à l'abbaye carolingienne de Corbie vers 790. Il représente une tentative inédite, dans la mesure où les mots courants ne font pas l'objet de lexiques dans l'Antiquité. Ses origines restent mal connues mais les recherches de notre groupe de travail (ERC StG 263577) tendent à montrer que des dossiers espagnols préconstitués ont été exploités dans une aire géographique et intellectuelle située entre Saint-Gall, Reichenau et Saint-Riquier. Ses sources sont extrêmement diverses : l'encyclopédie d'Isidore de Séville (ca 636) lui a servi de base, complétée par d'autres sources, des glossaires alphabétiques (*Abstrusa*, *Abolita*), des listes de synonymes, de différences, des sources médicales, patristiques, grammaticales, etc. La masse de textes rassemblés est considérable et a servi de fondement à toute la tradition lexicographique médiévale, à commencer par l'*Elementarium* de l'Italien Papias (ca 1050).

Dans le *Liber Glossarum*, le lemme constitue la seule porte d'entrée. La glose qui suit le lemme peut être plus ou moins longue, d'un mot unique à plusieurs colonnes, en raison de l'hétérogénéité complète de l'œuvre, qui procède par collage et montage de citations, à la manière d'un patchwork : le *LG* juxtapose les dépouillements de listes de synonymes et d'œuvres de nature encyclopédique, médicale, ou encore patristique, cette dernière catégorie étant aussi exploitée dans une perspective encyclopédique (le commentaire de la Genèse d'Ambroise de Milan, ou *Hexameron*, est ainsi traité comme une source d'informations sur la Création, réservoir de notices animalières dont sont méthodiquement éliminés tous les développements spirituels).

Le résultat de ces milliers de dépouillements est la constitution d'une base fiable de mots qui existent effectivement (hormis les monstres issus de corruptions scribales), d'usages, d'extensions, le tout appuyé sur des autorités reconnues. Le problème est celui de l'étendue des dépouillements : on a autrefois considéré que le *LG* rassemblait tout ce qui était disponible à l'époque. Cette vision magnifiée doit cependant être remise en cause, car le *LG* ne procède en fait pas du dépouillement de tout ce qui était disponible dans la chrétienté, mais de ce qui était accessible à Saragosse, ce qui inclut toutefois une partie du matériel préparatoire des *Étymologies* isidorienne arrivé de Séville. Ce principe de sélection limité à la disponibilité confère au *LG* un aspect inévitablement aléatoire, tandis que sa visée mixte, à la fois lexicale et encyclopédique, en fait un instrument d'utilisation incertaine.

CENDRINE PAGANI-NAUDET.
GRAMMAIRE ET CORPUS, L'EXEMPLE DE *L'ESSAY D'UNE PARFAITE GRAMMAIRE DE*
LAURENT CHIFLET

Pour montrer que la linguistique de corpus existe bien avant les corpus récents, peut-on évoquer les ouvrages de grammaire ? Poser la question, c'est déjà proposer de nuancer un certain point de vue sur la grammaire (la figure du grammairien pédagogue, la prescription hermétique à la variété des usages, le discours figé articulant la règle à l'exemple, etc.). Y répondre suppose aussi de se mettre d'accord sur ce que l'on entend par *corpus* : à l'élasticité de la notion, liée à l'effet de mode (le moindre groupement de textes est désormais intitulé « corpus », y compris dans les exercices scolaires) répondent des définitions parfois très restrictives qui ressemblent un peu à une stratégie d'exclusion (bons linguistes, bon usage du corpus vs mauvais linguistes, conception naïve du corpus).

Il sera sans doute difficile de faire de Laurent Chiflet, père jésuite auteur de *L'Essay d'une parfaite grammaire* paru en Anvers en 1659, un linguiste présentable. Pour autant, du fait de la masse des données exhibées, du fait de sa démarche (sélection, classement, gestion des données réfractaires, reformulation de règles inadéquates, confrontation à l'usage, ajout de données complémentaires), il semble que l'on doive d'emblée récuser toute accusation de spéculation bornée sur des exemples construits. Les « exemples » (ou fragments, extraits de langue, données) ne sont pas seulement exhibés pour illustrer une règle posée *a priori*, il arrive aussi que la règle énoncée résulte de l'observation de ces données.

Le titre de l'ouvrage (*Essay d'une parfaite Grammaire de la langue française Ou le Lecteur trouvera, en bel ordre, tout ce qui est de plus nécessaire, de plus curieux, & de plus elegant, en la Pureté, en l'Orthographe, & en la Prononciation de cette Langue*) indique que l'auteur aspire à l'exhaustivité. Il est possible grâce au Grand corpus des grammaires françaises, des remarques et des traités sur la langue de retrouver les sources de la plupart des exemples et des questions abordées par Chiflet. L'ouvrage publié en 1659 s'inscrit en effet dans une chaîne de commentaires sur la langue. Il recueille la tradition grammaticale antérieure (notamment pour le XVII^e siècle Maupas, Oudin et Irson), et surtout il ajoute les Remarques. Chiflet n'est pas le premier à le faire (Irson) mais il va en exploiter la matière de façon beaucoup plus systématique.

Peut-on considérer que cet ensemble constitue un corpus (c'est-à-dire autre chose qu'un réservoir d'exemples) ? On pourrait au moins le dire de Vaugelas, dont Chiflet déclare avoir « transporté dans *tout ce qu'elles contiennent de plus beau* ». Les *Remarques* rassemblent pour Chiflet des données (= ce qui est connu et admis, et qui sert de base au raisonnement). C'est un ensemble clos, stable, public et déjà largement discuté au moment où Chiflet s'en empare. Vaugelas propose des énoncés qui ne sont pas des exemples forgés, mais recueillis et contextualisés. Chiflet les emprunte littéralement ou en les modifiant à la marge, valide ou discute ses analyses. Il exploite assez librement les données : le reclassement de certains exemples montre qu'il conserve son autonomie dans la manière d'évaluer l'intérêt d'un fragment de langue en l'assignant à une nouvelle problématique linguistique. Inversement un énoncé peut devenir prototypique, voire résumer à lui seul une règle. Chiflet épuise le texte des *Remarques*, jusqu'à citer la prose même de Vaugelas : le chapitre huitième *De la syntaxe et du style* contient ainsi plusieurs extraits de la préface des *Remarques*. Chiflet est en outre attentif à la manière

dont Vaugelas met en pratique ses propres préceptes. Ainsi au sujet des emplois respectifs de *avec* et *avecques* : il note que « luy-mesme en ses œuvres ne s'est point obligé à la rigueur de cette regle » (p. 120). Autrement dit, c'est bien la démarche de Chiflet qui constitue les *Remarques* en corpus : Chiflet conduit son analyse à partir de données rassemblées par un autre que lui, mais dont l'autorité offre donc les garanties de représentativité et de qualité nécessaires à son projet linguistique. Chiflet toutefois ne se contente pas d'enregistrer les faits et de procéder à des classements : il maintient le corpus ouvert, « décision », qui « a pour corollaire l'implication plus franche du linguiste dans le modelage de celui-ci » (Dalbera 2002).

Chiflet enrichit le corpus initial, dont les limites ont conduit Vaugelas à énoncer des règles erronées. Il s'ouvre à d'autres langues (notamment le flamand), il ajoute le résultat de ses propres observations. Ce travail sur les données ne fait pas l'objet de déclarations très longues sur sa méthode mais le texte en conserve les traces. La forme du discours manifeste les tensions entre les attendus du genre (une grammaire) et l'ambition de décrire la langue. Chiflet est bien conscient du dilemme : il faut formuler des règles, car renvoyer le lecteur à l'usage serait un aveu d'échec, mais ces règles doivent correspondre à la réalité de la langue. Il entend donc se démarquer des grammairiens qui ont pêché par l'un ou l'autre excès, ceux qui « après s'estre bien debatus, desesperant d'en voir le fond, (...) vous renvoient à l'usage » et ceux qui formulent des règles fausses car fondées sur « petit nombre d'exemples ». Le traité qu'il consacre à la place des adjectifs illustre bien cette prétention :

Quand il est question de former une regle de Grammaire, ils s'arrestent à dix ou douze mots, qui se presentent à leur memoire (...) & là dessus, sans examiner plus avant, ils prononcent leurs arrests de l'usage de la langue. Je n'en ay pas usé de la sorte: car avant que de former les regles de ce Traité, j'ay parcouru & examiné tous les adjectifs de la langue Frãoise: & si quelcun m'est eschapé, ce n'est pas faute de diligence. (1659, p. 244)

Les règles qu'il s'efforce de formuler ensuite sont donc fondées sur l'observation d'un corpus qu'il veut exhaustif. Pour autant, elles ne lui semblent jamais assez respectueuses des données, il n'y a pas de règle « certaine » qui ne soit agrémentée d'« exceptions ». Il y a toujours des restes, qui viennent remplir des listes. Pour la prononciation de *s*, il entend donner les règles, « pour abbreger le catalogue des mots, où l'*s* doit estre prononcée » (p. 226), mais il doit finalement se résoudre à présenter une « table » qui contient près de 400 mots (p. 228). On assiste ainsi à l'altération de la matrice d'origine (d'où l'impression de fouillis et d'émiettement) : tentatives de reformulation des règles, discussion ouverte (notamment dans les sections intitulées « Observations » où il expose, sans toujours vouloir trancher, les points de conflit entre Vaugelas et la censure), listes où sont regroupées les données réfractaires. Les exemples inconfortables obligent le grammairien à dévoiler sinon son incapacité, du moins les possibilités de doute. Il concède au sujet du genre des noms de villes « qu'on n'en sçauroit donner une regle certaine, à cause de leur grand nombre », et suggère une stratégie d'évitement : « vous ne sçauriez faillir si vous les faites tous feminins, en y ajoutant ces mots, *la ville* : comme, *la ville de Paris est grande* » (p. 234).

Le corpus est à la mode et constitue désormais un gage de scientificité. Comme on s'en doute, ce dont s'enorgueillit la linguistique contemporaine n'a rien de bien nouveau. Et au risque de l'anachronisme, on peut affirmer le grammairien jésuite Chiflet entrainé de manière consciente dans l'ensemble des questionnements qui agitent la communauté linguistique autour des corpus : clôture, sélection des données, représentativité, variation, authenticité, exhaustivité, possibilités d'extrapolations théoriques. En citant la

source de ses données, Chiflet remplit la condition d'un corpus publiquement partagé, qui permet la cumulativité et la discussion.

DISCUSSION

Q Bernard Colombat

Pour un moderne un corpus est un construit, qui l'est à des fins précises, mais sans qu'on soit obligé d'explicitement lesquelles. On ne se pose plus la question de la finalité parce qu'elle est évidente. Anne Grondeux s'étonnait d'être invitée à une table ronde sur les corpus, alors même que le *Liber glossarum* apparaît d'emblée comme un corpus, qui s'impose en tant que tel dans sa constitution comme pour la suite. Idem dans l'exposé d'Émilie Aussant : le corpus préalable est extérieur à la grammaire, et c'est à partir de là que le grammairien constitue son propre corpus. Est apparue la notion de « corpus lacunaire », expression qui traduit par elle-même une *évaluation* du corpus ; que veut-elle dire exactement ?

Dans la tradition on a une sorte d'enchaînement de corpus. À un moment le corpus est devenu tellement complexe qu'il faut inventer une manière simplifiée et adroite pour le présenter. Donc : le corpus est-il un construit ou non ?

Dans le cas de Priscien, les exemples sont construits. Mais ce n'est plus le cas quand on prend un texte dans son entier, ce que fait Priscien au livre 18. Évidemment il y a un choix, une sélection du texte, mais ensuite on est confronté à ce qu'il contient.

R Anne Grondeux

L'expression de « corpus lacunaire » s'applique bien au *LG*, dans la mesure où l'aspect *corpus* est fortement limité par le recours à ce qui était disponible sur place. Au moment de la mise au net du *LG*, peut-être dans le nord de la France, le stock de définitions n'a pas été augmenté par de nouveaux dépouillements, en sorte que la source la plus récente est de 690 environ. Les responsables de la mise au net du *LG*, dont l'histoire débute en Espagne au VII^e s. et s'achève comme un monument de la culture carolingienne, paraissent avoir considéré que ce qu'ils avaient sous la main était suffisamment représentatif pour ne pas requérir de dépouillements supplémentaires. L'absence de revendication éditoriale ou de préface constitue ici un handicap évident pour l'historien, mais la méthodologie transparait au moins dans l'apposition d'étiquettes (ou tags) qui permettent de donner une base réelle à une citation, de l'authentifier pour employer la terminologie de Papias, qui, héritier du *LG* vers 1050, justifie ainsi, dans sa préface, le système qu'il reprend à son compte et étend en y insérant de nouveaux tags adaptés à ses nouvelles autorités.

R Franck Cinato

Le *LG* est devenu un corpus pour tous les continuateurs (les lexicographes occidentaux, pour qui il est la base de données de références). Mais une base de données est-elle un corpus ?

R Émilie Aussant

À propos de l'expression « corpus lacunaire » : Pāṇini décrit une variété commune qui englobe :

- la variété linguistique sacrée (*chandās*, i.e. sanskrit védique) ;
- la langue parlée ordinaire (*bhāṣā*), probablement un dialecte du nord dont il était locuteur.

Le corpus lacunaire concerne la partie védique. Les textes védiques ont été transmis dans des recensions variées à travers des écoles et Pāṇini n'a probablement pas eu accès à toutes celles dont on dispose actuellement.

Par ailleurs si sa grammaire couvre ces deux variétés, l'essentiel porte sur la grammaire ordinaire et les faits védiques ne sont convoqués que de manière marginale. Il est possible qu'il ait fait un tri dans les données et qu'il ait cité les sources védiques qu'il pensait être suffisamment proches des formes parlées.

Donc la lacune s'explique soit par un problème d'accès aux textes védiques, soit par un choix délibéré d'écarter des formes qui s'éloignaient trop de ce qu'il souhaitait décrire.

Q *Bernard Colombat*

Les 4000 règles pāṇinéennes constituent-elles un corpus ? Cela devient tellement complexe que cela semble difficile à considérer comme un ensemble de règles séparées. Comment dominer une telle matière ? Cela implique-t-il de recourir à des procédés mnémotechniques qui pourraient s'assimiler à un corpus bien rangé ?

R *Émilie Aussant*

L'objectif de Pāṇini n'était pas de construire un corpus. Les 4000 règles de son traité constituent une grammaire générative, l'objectif principal étant de permettre de dériver n'importe quelle phrase correcte de la variété linguistique dite *bhāṣā*. Certes, Pāṇini se fonde en partie sur un corpus de textes pour les données védiques, ainsi que sur des listes (dont quelques-unes sont ouvertes) pour certaines données de la *bhāṣā*, mais pour le reste, il s'en remet à l'usage qu'il observe et dont il note parfois les variantes possibles. On suppose que la glose orale du traité grammatical fournissait exemples et contre-exemples ; nous les retrouvons dans les tout premiers commentaires de la grammaire de Pāṇini.

Q *Bernard Colombat*

Mais comment se fait le travail de commentaire ?

R *Émilie Aussant*

Il y a eu plusieurs types de commentaires : des gloses explicatives, qui reconstruisent la forme complète des règles en réintégrant des éléments tacitement reconduits de règle en règle et qui donnent exemples et contre-exemples (ex. *Kāśikāvṛtti*, VII^e s.), mais également des commentaires qui mêlent gloses grammaticales et développements philosophiques et qui proposent parfois des amendements dans la formulation des règles pāṇinéennes (ex. *Mahābhāṣya*, II^e s. av. notre ère).

Q *Gabriel Bergougnieux*

On semble avoir affaire moins à un corpus qu'à une réinterprétation d'un certain nombre de données. On a l'impression d'avoir un choix qui a été fait dans de l'écrit, qui est déjà préexistant, qui n'était pas destiné à cela. Aujourd'hui, quand on fait un corpus, on se dit que ça servira toujours, donc à un peu tout ; aussi ne fixe-t-on pas trop les

choses. Alors que dans les exemples cités, les corpus semblaient très resserrés sur une fin précise.

Par ailleurs dans ces exemples de la table ronde, on a affaire, très largement, à de l'oral. Il s'agit de corpus nourris par un préalable oral.

R *Marc Baratin*

Sur la question du corpus donné ou construit : oui, il y a des éléments de donné. La littérature latine classique peut être considérée comme un donné. Mais un discours de Cicéron dans le cas précis de l'analyse de Priscien n'est pas un donné. Priscien le construit comme corpus. Chronologiquement, il procède d'abord comme les autres grammairiens, en s'adossant à un donné (la littérature latine), dont il se sert comme réservoir d'exemples. Quand il prend le *De praetura urbana* de Cicéron pour l'analyser, il inverse le mécanisme habituel en n'envisageant plus de poser une règle et de l'illustrer ensuite par du donné : au contraire, il prend tous les exemples du texte pour les confronter à la théorie. Il se rend compte que ça marche mal, il patauge un peu, mais continue bravement, au risque de démolir sa théorie de l'intérieur. S'il construit le discours de Cicéron comme corpus, c'est ainsi dans la mesure où ce n'est plus un réservoir d'exemples, mais quelque chose qui s'impose à la théorie qu'il construit, et qui en montre les faiblesses.

Cette construction me paraît poser le problème de la finalité du corpus. Aujourd'hui, la construction d'un corpus se fait de A à Z en fonction de sa finalité. Mais l'intention initiale ne suffit-elle pas, à elle seule, à construire le corpus ?

Q *Francine Mazière*

À propos de la modification méthodologique opérée par Priscien, en rapport avec la définition du corpus : l'analyse du discours a été citée comme une linguistique de corpus ; or elle n'a jamais travaillé sur un corpus textuel défini, mais toujours sur une modification de corpus.

Cela ne pose plus la question ni de l'exhaustivité ni de la restriction, mais de la restriction en tant qu'on ne va pas travailler forcément sur du lexique, du réel, de l'attesté, mais qu'on va réécrire cet attesté et ce réel.

Par exemple Vaugelas travaille dans l'optique d'une restriction par rapport à l'usage du XVII^e s. et toute une série d'analyses ne prennent pas le corpus comme une donnée.

Q *Bernard Colombat*

On peut aussi poser la question du rapport entre corpus et auteur. Priscien a écrit l'*ars* mais aussi des traités mineurs. Il est intéressant de savoir que l'on mettait sous l'étiquette « Priscien » des grammairiens divers dans les bibliothèques. Qu'est-ce qu'un auteur dans ces conditions, à partir du moment où on redistribue son œuvre ?

R *Franck Cinato*

Pour les médiévistes, il est difficile de définir un auteur puisque les textes sont des montages à très forte intertextualité. Priscien est utilisé comme tête de liste parce qu'il fait figure de grammairien encyclopédique, presque exhaustif, qui compare systématiquement le latin avec le grec.

Au début du IX^e s. il est considéré comme la nouveauté. L'appui de la relance culturelle carolingienne permet la diffusion de l'*Ars Prisciani*. Les catalogues montrent ce qu'il faudrait qu'il y ait dans les bibliothèques, plus que leur contenu réel.

Q Bernard Colombat

L'auteur s'efface mais il devient en même temps emblématique ?

R Franck Cinato

Cela pose la question de la limitation des corpus. Les corpus ne cherchent pas une exhaustivité. La limite est matérielle. On utilise ce qu'on a sous la main. On cherche à obtenir et compléter mais on fait avec la limite des ressources.

R Anne Grondeux

Un corpus est représentatif d'une époque, d'un lieu : en matière de bibliothèques, des politiques d'acquisition visent à une sorte de canon homogène, car les centres intellectuels forment des réseaux, liés par des raisons historiques : Murbach ayant été fondée comme Reichenau par Pirmin, elle détient une copie du catalogue de Reichenau, établi par Reginbert en 821/822. En ce qui concerne le *Liber glossarum*, l'influence du contexte historique se traduit par une surreprésentation de tout ce qui tourne autour des hérésies, conséquence d'une élaboration qui puise à pleines mains dans les traités polémiques de l'Espagne du VII^e s.

Q Bernard Colombat

La notion d'« exemple inconfortable », évoquée par Cendrine Pagani-Naudet, pose la question de la construction d'un corpus. On veut donner une vision de la langue, donc envisager celle-ci de manière *complète* : aussi est-on forcément confronté à l'exemple inconfortable.

R Cendrine Pagani-Naudet

La deuxième partie de la grammaire de Chiflet est une sorte de traité qu'il développe à part. Il reproche à des grammairiens d'avoir formulé des règles, ou d'avoir simplement renvoyé à l'usage sans aller au-delà, ce qui n'est guère utile à l'usager.

Q Bernard Colombat

L'idée n'est-elle pas qu'au fond c'est la démarche de Chiflet qui constitue l'approche de Vaugelas encore plus comme corpus ?

R Cendrine Pagani-Naudet

Les *Remarques* de Vaugelas ne sont pas simplement un réservoir d'exemples. C'est la manière même dont Vaugelas s'est exprimé qui devient un témoignage intéressant. Dans la partie stylistique de l'*Essay d'une parfaite Grammaire* de Chiflet, Vaugelas est présent non pour ses préconisations, mais comme représentant du bon usage.

Q Bernard Colombat

Cela pose la question de la représentativité. Pour nous, un contenu, c'est un construit. Cendrine envisage la démarche de Chiflet précisément sous cet angle-là.

Il est difficile de savoir pourquoi on se fixe sur tel ou tel auteur, par exemple pour le choix des grammaires françaises du XVIII^e s. à retenir pour le corpus que nous sommes en train de constituer en complément du *Grand Corpus des grammaires françaises, des remarques et des traités sur la langue*. Il s'agit d'une décision collective dans le cadre de discussions (laborieuses) sur la constitution d'un corpus, qui doit être hétérogène.

Il faut donc documenter ce que l'on fait pour se rappeler comment on y est arrivé ! Quand on fait de la linguistique de corpus, c'est pour échapper à une linguistique qui prend pour exemple *Pierre aime Marie*.

Q *Danielle Candel*

On s'accorde généralement à reconnaître qu'un corpus doit être : (i) homogène ; (ii) exhaustif ; (iii) représentatif. Or la représentativité se décrète et c'est elle qui induit l'homogénéité et l'exhaustivité.

Je serais donc moins sévère qu'Émilie et irais plus dans le sens de Bernard. On peut délimiter un corpus en assumant et explicitant ses choix. On peut se donner des règles à soi-même.

Les intervenants sont-ils d'accord pour dire que leurs corpus sont « des terrains » ?

R *Émilie Aussant*

Il y a bien une dimension anthropologique, lorsque l'on travaille sur les textes techniques (grammaticaux, mais pas seulement) de l'Inde ancienne : on a la chance de pouvoir échanger avec des paṇḍits, des brâhmanes lettrés formés de manière traditionnelle (apprentissage par cœur des textes auprès d'un maître pendant de longues années) et de les interroger sur le contenu de ces textes, souvent abscons. Sans ces informateurs « à double titre » (maîtrise de la langue sanskrite et des textes techniques qui la décrivent), notre connaissance de ces disciplines d'érudition ne serait pas ce qu'elle est, même si beaucoup reste à faire.

R *Marc Baratin*

Marc Baratin présente le mode opératoire dans la traduction en cours de l'*Ars Prisciani* par le Groupe Ars Grammatica : il effectue un travail de préparation de la traduction en amont de séances de travail collectives ; ce travail est presque toujours modifié, parfois profondément, par les autres participants pendant les séances de travail en commun, parce que l'expérience prouve toujours que la confrontation des différents points de vue fait sortir des choses plus intéressantes, ce qui ne laisse d'être surprenant pour le préparateur qui a une grande familiarité avec le texte. Autrement dit, il y a une dimension imprévisible !

Pour ce qui est de la définition du corpus selon les trois critères, proposée par Danielle Candel : malgré leur caractère très verrouillé, le corpus est bien une affaire de points de vue. C'est pour cela qu'on arrive aux conceptions étranges qui ont été abordées par Franck Cinato et par d'autres, à savoir l'enchâssement de corpus, qui devrait être paradoxal mais qui ne l'est finalement pas... justement du fait de cette question du point de vue.

R *Franck Cinato*

Un corpus n'est-il pas tout simplement un ensemble de données auquel on donne un sens ?

R Marc Baratin

Il faut différencier entre des auteurs qui cherchent dans la littérature latine des réservoirs d'exemples et des auteurs comme Priscien qui cherchent dans un discours des contre-exemples.

Q Pascale Rabault-Feuerhahn

Ne serait-il pas utile de confronter la notion de « corpus » telle qu'elle est mobilisée en linguistique et en philologie à son emploi dans d'autres disciplines scientifiques ? Qu'est-ce, par exemple, qu'un « corpus » pour un historien ? Ses sources primaires ou bien la délimitation de son domaine ?

Même si l'expression « linguistique de corpus » vise le fait que les corpus connaissent une croissance exponentielle notamment à l'heure d'Internet, on peut relever qu'il n'existe pas d'expression parallèle dans d'autres disciplines.

Cette question de la place, de la formation et de l'emploi du corpus selon les domaines scientifiques met en jeu un rapport à l'empiricité (induction/déduction, pluralité ou unicité du corpus, valeur paradigmatique, voire canonique qui lui est accordée...) et offre donc une bonne porte d'entrée à la confrontation épistémologique de diverses sciences humaines.

RÉFÉRENCES BIBLIOGRAPHIQUES

- BALLAIRA, Guglielmo, 1989. *Prisciano e i suoi amici*, Torino, Giappichelli.
- BARATIN, Marc, 1996. « De la bibliothèque à la grammaire : le paradigme de l'accumulation », in Ch. Jacob et M. Baratin (éd.), *Le pouvoir des bibliothèques : la mémoire des livres en Occident*, Paris, Albin Michel, 253-260.
- BARATIN, Marc, COLOMBAT, Bernard, HOLTZ, Louis (éd.), *Priscien. Transmission et refondation de la grammaire de l'Antiquité aux Modernes*, Turnhout, Brepols (Studia Artistarum 21).
- BECKER, Gustav Heinrich, 1885. *Catalogi bibliothecarum antiqui*, Bonn, M. Cohen et fils.
- BIONDI, Laura, 2011. *Recta scriptura. Orthografia ed etimologia nei trattati mediolatini del grammatico Apuleio*, Milano, LED Edizioni Universitarie.
- CHIFLET, Laurent, 1659. *Essay d'une parfaite Grammaire de la langue française*, Anvers, Jacques Van Meurs.
- CHIN, Catherine M., 2008. *Grammar and Christianity in the Late Roman World*, Philadelphia, University of Pennsylvania Press.
- COLOMBAT, Bernard, FOURNIER, Jean-Marie, AYRES-BENNETT, Wendy (éd.), 2011. *Grand Corpus des grammaires françaises, des remarques et des traités sur la langue (XIV^e-XVII^e siècles)*, Classiques Garnier Numérique.
- DALBERA, Jean-Philippe, 2002. « Le corpus entre données, analyse et théorie », *Corpus* 1, 89-104.
- DE NONNO, Mario, 2009. « *Ars Prisciani*: problemi di tipologia e di composizione », in M. Baratin et al., *Priscien. Transmission...*, Turnhout, Brepols, 249-278.

- DE PAOLIS, Paolo, 2004. «I codici miscellanei grammaticali altomedievali. Caratteristiche, funzione, destinazione », in E. Crisci – O. Pecere (ed.), *Il codice miscellaneo. Tipologie e funzioni. Atti del Convegno internazionale, Cassino 14-17 maggio 2003*, Cassino, Edizioni dell'Università, 183-211.
- DE PAOLIS, Paolo, 2010. « L'insegnamento dell'ortografia latina fra Tardoantico e alto Medioevo: teorie e manuali », in L. Del Corso – O. Pecere (ed.), *Libri di scuola e pratiche didattiche. Dall'Antichità al Rinascimento. Atti del Convegno Internazionale di Studi Cassino, 7-10 maggio 2008*, 2 vol., Cassino, Edizioni dell'Università, I, 229-291.
- ELLSPERMANN, Gerald L., 1949. *The Attitude of the Early Christian Latin Writers toward Pagan Literature and Learning*, Washington, DC, The Catholic University of America Press.
- GRONDEUX, Anne, 2015. « Le rôle de Reichenau dans la diffusion du *Liber glossarum* », *Dossier HEL 8, L'activité lexicographique dans le haut moyen âge latin. Rencontre autour du Liber Glossarum (suite)*, 79-93.
- GROUPE ARS GRAMMATICA, 2010. *Priscien, Grammaire, livre XVII, Syntaxe, 1*, texte latin, traduction introduite et annotée, Paris, Vrin (Histoire des doctrines de l'antiquité classique, 41).
- GROUPE ARS GRAMMATICA, 2017. *Priscien, Grammaire, livre XVIII, Syntaxe, 2*, texte latin, traduction introduite et annotée, Paris, Vrin (Histoire des doctrines de l'antiquité classique).
- HAGENDAHL, Harold, 1958. *Latin Fathers and the classics: a study on the apologists, Jerome and other Christian writers*, Göteborg, Studia Graeca et Latina Gothoburgensia (6). Distrib. Stockholm, Almqvist & Wiksell.
- HOLTZ, Louis, 1977. « La typologie des manuscrits grammaticaux latins », *Revue d'histoire des textes* 7, 247-269 (+ pl. II à X).
- HOLTZ, Louis, 2009. « L'émergence de l'œuvre grammaticale de Priscien et la chronologie de sa diffusion », in M. Baratin et al., *Priscien. Transmission...*, Turnhout, Brepols, 37-55.
- JEUDY, Colette, 1972. « L'*Institutio de nomine, pronomine et verbo* de Priscien. Manuscrits et commentaires médiévaux », *Revue d'histoire des textes* 2, 73-144.
- LAW, Vivien, 1993. "Erchanbert and the interpolator. A Christian *Ars minor* at freising (CLM 6414)", in Vivien Law (ed.) *History of Linguistic Thought in the Early Middle Ages*, Amsterdam, Benjamins (SiHoLS 71), 223-238.
- MUNK OLSEN, Birger, 1992. « Les poètes classiques dans les écoles au IX^e siècle », in L. Holtz, J.-C. Fredouille, M.-H. Jullien (éd.), *De Tertullien aux Mozarabes. Mélanges offerts à Jacques Fontaine*, 2 vol. + tables et index, vol. 2, *Antiquité tardive et christianisme ancien (VI^e-IX^e siècles)*, Paris, Institut d'études augustiniennes, 197-210.
- PEREZ-RODRIGUEZ, Estrella, 2001. « La christianización de la gramática latina (ss. V-IX) », in C. Macías Villalobos & A. Alberte González (ed.), *Actas del Congreso Internacional "Cristianismo y Tradición Latina" (Málaga, 25 a 28 de abril de 2000)*, Madrid, Ediciones del laberinto, 49-74.

- PINAULT, Georges-Jean, 1989. « Pāṇini et l'enseignement grammatical », in S. Auroux éd., *Histoire des idées linguistiques*, 3 vol., Liège-Bruxelles, Mardaga, vol. 1, 331-353.
- PINAULT, Georges-Jean, 2000. « Śākalya (notice 4301) », in B. Colombat éd., *Corpus représentatif des grammaires et des traditions linguistiques*, tome 2, *Histoire Épistémologie Langage*, hors-série, 3, 223-226.
En ligne : http://ctlf.ens-lyon.fr/n_fiche.asp?num=4301
- PRISCIEN [Priscianus Caesariensis], 1855-1859. *Institutionum grammaticarum libri XVIII*, éd. Martin Hertz, *Grammatici Latini*, vol. 2 et 3, Leipzig, Teubner.
- RAÏOS, Dimitris K., 1983. *Recherches sur le « Carmen de ponderibus et mensuris »*, Jannina, Dédale.
- RENOU, Louis & FILLIOZAT, Jean, 1985-1996. *L'Inde classique. Manuel des études indiennes*, 2 volumes, vol. I : Paris, Adrien Maisonneuve, 1985 [1947], vol. II : Paris, Presses de l'École française d'Extrême-Orient, 1996 [1953].
- RENOU, Louis, 1947. *Les écoles védiques et la formation du Veda*, Paris, Imprimerie nationale (Cahiers de la Société Asiatique 9).
- VAUGELAS, Claude Favre DE, 1647. *Remarques sur la langue françoise*, Paris, Veuve Jean Camusat et Pierre Le Petit.