

Bernard COLOMBAT

REI (01) – *Présentation du colloque*
[1]

corpus de textes linguistiques fondamentaux

Notices Bibliographie Images Textes Articles

Recherche dans les articles de l'équipe CTLF
| Actualité du CTLF | Actualité de la base d'articles | Fonctionnement du portail |

titre

auteur

rubrique

accès réservé

archive
 non oui

présentation
 abrégée détaillée

par titres par rubriques par dates

1 Pour citer cet article :

COLOMBAT Bernard. «REI (01) – Présentation du colloque». *CTLF – Articles*. Colloque 'Refonte et extension internationale du CTLF: Corpus de textes linguistiques fondamentaux', Université Paris Diderot, UMR 7597, «Histoire des théories linguistiques», Programme Action structurante 2015-2018, 31 mai et 1er juin 2018. Mise en ligne le 01/10/2018 à l'adresse: <http://ctlf.ens-lyon.fr/documents/articles/colloque/REI-01.pdf>.

Présentation du colloque

Bernard Colombat

REI (02) – [Programme](#) (l'affiche du colloque)

REI (11) – [La base CTLF aujourd'hui](#) (Bernard COLOMBAT)

Grammaires des langues européennes

REI (12) – [Les grammaires portugaises et latino-portugaises](#) (Rolf KEMMLER)

REI (13) – [Les grammaires brésiliennes](#) (Marli Quadros LEITE & Arnaud PELFRENE)

REI (17) – [Les grammaires de l'anglais](#) (Doug KIBBEE et Wilfrid ANDRIEU)

REI (15) – [Les grammaires du grec moderne](#) (Rea DELVEROUDI)

Le domaine slave

REI (19) – [La mise à jour des notices des grammaires russes](#) (Sylvie ARCHAIMBAULT)

REI (16) – [Les grammaires slovènes](#) (Mitja TROJAR)

REI (18) – [Les grammaires polonaises](#) (Patrycja KRYSIAK)

Linguistique générale

REI (14) – [Le premier grammairien – vieux norrois](#) (Olaf MIKKELSEN)

REI (27) – [Charles F. Hockett, 'A Course in Modern Linguistics'](#) (Jacqueline LEON)

REI (26) – [Problèmes liés à la rédaction de la notice sur l'œuvre d'Antoine Culioli](#) (Rémi CAMUS)

Traditions non occidentales

REI (22) – [La linguistique missionnaire dans les deux Amériques](#) (Otto ZWARTJES)

REI (21) – [Les premières grammaires du vietnamien](#) (Thi Kieu Ly PHAM)

REI (25) – [Les grammaires de l'arabe...](#) (Jean-Patrick GUILLAUME)

Questions techniques

REI (23) – [La nouvelle base CTLF](#) (Frédéric WEISS & Najib FAWZI)

REI (24) – [La nouvelle base Frantext-CTLF dans Frantext 2](#) (Gilles SOUVAY)

REI (28) – [Manuel de transTyp_M1548](#) (Chia-Hung HSUEH)

L'objet du colloque

Le colloque a eu pour objet de faire le point sur l'état du CTLF et voir ses développements futurs. Il n'a abordé que quelques points, et d'abord des questions techniques, extrêmement importantes : il faut concevoir un outil à la fois pérenne et évolutif. À ce titre les communications « techniques » dues à nos collègues de Lyon et de Nancy sont essentielles : dans le premier cas [REI (23)], la question est de savoir comment concevoir la base de demain comme un outil unifié à partir de plusieurs éléments non homogènes ? dans le deuxième cas [REI (23)], de savoir comment faire passer le CTLF de l'ancien Frantext à Frantext 2 ?

Si l'on nous autorise une métaphore automobile (?), dans le premiers cas (de la première version du CTLF à l'actuelle, en cours de réalisation), l'objectif est de remplacer une petite camionnette avec beaucoup de petites remorques derrière (façon convoi de cirque) en semi-remorque rassemblant dans sa grande benne tout le contenu de ces petites remorques. Dans le second (de Frantext à Frantext 2), il s'agirait plutôt de remplacer une Coccinelle Volkswagen progressivement gonflée en Porsche en vraie Porsche. Une troisième communication technique, celle de Chia-Hung HSUEH [REI (28)] propose des outils de type XML pour une analyse fine des textes de Louis Meigret rédigés dans une orthographe réformée : c'est le texte *Le menteur* (1548) qui est pris pour objet d'étude.

Les autres communications ont porté sur certains secteurs arrivés à une maturité plus ou moins grande. Certaines ont porté sur les traditions européennes.

Tout d'abord les grammaires portugaises et brésiliennes. Ce secteur constitue un cas d'école intéressant car c'est l'un de ceux qui a le plus évolué depuis l'origine du corpus. Il y avait au départ (en 1998) des notices pour 20 grammaires portugaises et 5 grammaires brésiliennes. L'action conjointe de nos collègues de l'Universidade de Trás-os-Montes e Alto Douro, à Vila Real (Carlos ASSUNÇÃO *et al.*), et de l'Université de São Paulo (Marli QUADROS LEITE *et al.*) a permis : (1) une augmentation sensible du nombre de notices : 63 au total, soit 28 grammaires portugaises, 16 traités orthographiques portugais ; 19 grammaires brésiliennes ; s'ajoutent à ces notices 14 descriptions de grammaires latines élaborées au Portugal à la Renaissance ; (2) de proposer en mode image 34 volumes ; (3) de permettre des recherches sur 56 textes numérisés. Le secteur lusitanien est donc actuellement un des mieux documentés de la base CTLF. C'est d'abord la partie portugaise qui s'est développée, dans les années 2015-2016, par la collaboration avec l'UTAD : adjonction de notices et mise à disposition des textes en mode image et sous forme numérisée. La communication de Rolf KEMMLER [REI (12)] a posé la question du rapport entre représentativité et exhaustivité. Même si la description s'est considérablement affinée, il ne peut s'agir de donner un tableau complet des ouvrages consacrés au secteur. Le séjour de Marli QUADROS LEITE au laboratoire HTL en 2017-2018 a permis de compléter la partie comportant les grammaires brésiliennes. Son texte, réalisé avec Arnaud PELFRENE [REI (13)], dresse le bilan de cette collaboration.

Du côté des secteurs documentés dès 1998, il y a le secteur des grammaires anglaises qui n'avait pas bougé depuis l'origine (27 ouvrages). Nos collègues Doug KIBBEE et Wilfrid ANDRIEU [REI (17)], ont accepté de faire une mise à jour, le premier pour les ouvrages antérieurs au XVIIIe siècle, le second pour la période XIXe-XXe siècles.

De l'autre côté de l'Europe, dans sa partie orientale, le secteur des grammaires grecques constitue un cas intéressant. Dès l'origine, le secteur de la grammaire grecque a existé dans le CTLF puisque nous avons une quinzaine de notices contenant des descriptions du grec ancien composées dans l'Antiquité ou au Moyen Age. Mais comme certains utilisateurs du CTLF nous l'ont fait remarquer, la grammaire grecque ne se restreint pas au domaine antique, voire antique et médiévale. C'est même une des traditions linguistiques les plus documentées sur une très longue période. Nous voulions donc étendre la description aux grammaires grecques moderne : notre collègue de l'université d'Athènes, Rea DELVEROUDI [REI (22)], s'en est chargée et a constitué une équipe de recherche. Elle évoque dans sa communication la genèse de ce projet, pour lequel des compétences linguistiques diverses sont requises puisque les premières grammaires du grec moderne ont été

rédigées le plus souvent en dehors du territoire national, par des étrangers, et dans une métalangue différente (allemand, italien, français, etc.). Autre fait notable : comme pour le slovène, un site est développé parallèlement à l'Université nationale et capodistrienne d'Athènes, comportant la description d'un plus grand nombre d'ouvrages et fournissant des données plus approfondies (notamment sur l'histoire éditoriale des ouvrages).

Dans le secteur des langues slaves, les grammaires du slavon et du russe sont également présentes dans le corpus depuis 1998. Sylvie ARCHAIMBAULT [REI (19)], qui avait assuré la publication des premières notices (2 pour le slavon, 19 pour le russe), réfléchit à une extension. Portant un regard rétrospectif sur ses propres choix effectués en 1998, elle propose une extension relativement restreinte – placée sous le signe de la représentativité plutôt que de l'exhaustivité – qui s'accompagnera d'une mise à jour bibliographique des notices existantes. Dans ce même domaine, le slovène est également aujourd'hui bien représenté. Le secteur a été intégré plus récemment : si la notice sur Bohorič a été introduite dès 2010, la description des grammaires slovènes a été réalisée pour l'essentiel dans les années 2013-2014, par une équipe dirigée par Kozma AHACIC (Fran Ramovš Institute of the Slovenian Language, Ljubljana), qui a développé en parallèle un site slovène. Dans sa communication, Mitja TROJAR [REI (16)] – qui représentait Kozma Ahačič et qui a réalisé lui-même la traduction française de toutes les notices des grammaires slovènes –, a comparé leur rédaction dans les deux métalangues (français et slovènes) et évoqué les outils qui permettront d'exploiter les textes décrits dans les notices pour parvenir à un dictionnaire terminologique.

Toujours dans les langues slaves, un nouveau secteur a été ouvert : celui des grammaires du polonais. S'appuyant sur une liste établie par une équipe de l'Université de Varsovie dans le cadre d'un projet similaire, existant uniquement en polonais, et qui donne une description factuelle des ouvrages, Patrycja KRYSIAK [REI (18)] évoque les premières notices réalisées dans ce cadre, aux deux bouts du domaine : Statorius 1568 [notice 3445] et Szober 1923 [notice 3465].

Un texte bien connu des historiens de la phonétique et de la phonologie, le *Fyrsta málfræðiritgerðin*, relève autant de la linguistique générale que de l'histoire d'une langue particulière, en l'occurrence le vieux norrois, puisqu'il décrit, sans doute pour la première fois d'une façon aussi systématique, les paires minimales distinguées dans la recherche d'une orthographe adaptée à cette langue. Olaf MIKKELSEN [REI (14)], qui a réalisé pour le CTLF une traduction française du traité en l'accompagnant d'une étude réalisée avec Patrick Guelpa, présente le *Codex Wormianus* qui contient le texte et les principes d'orthographe proposés par le premier grammairien.

Relève également, et plus classiquement, de la linguistique générale le texte de Jacqueline LEON [REI (27)] décrivant les recherches nécessaires à l'élaboration d'une notice, celle décrivant le *Cours de linguistique générale* de Hockett (décédé en 2000) : depuis, la notice [5348] a été publiée (2018-09), mais l'intérêt de la contribution est de montrer les recherches préalables nécessaires, notamment sur la biographie et le choix du texte.

La démarche concernant Antoine Culioli, tout récemment décédé (9 février 2018), est similaire. Il s'agit de faire entrer dans le corpus des données concernant un des linguistes les plus influents de l'école française au XXe siècle. La notice, qui porte, cette fois-ci, sur toute l'œuvre d'Antoine Culioli, récemment rassemblée et publiée, a été rédigée par un nouveau descripteur, Rémi CAMUS [REI (26)], et il nous a paru intéressant – surtout pour les nouveaux collaborateurs venus et à venir dans l'entreprise – de donner un état absolument brut de sa rédaction et du dialogue qui s'établit entre le descripteur et le responsable du projet.

Dans le secteur des traditions non occidentales, Jean-Patrick GUILLAUME [REI (25)] envisage l'extension de la description de l'arabe au-delà de la « TGA », c'est-à-dire la tradition grammaticale arabe autochtone. Aux 18 notices existant depuis l'origine (2000), il propose l'adjonction de notices internes à la TGA, mais également des notices de grammaires « arabisantes », c'est-à-dire réalisées par des occidentaux. Thi Kieu Ly PHAM [REI (21)] présente, quant à elle, le secteur des grammaires vietnamiennes. Au départ la grammatisation du vietnamien a été réalisée par les Jésuites, d'abord sur le modèle latin (Alexandre de Rhodes) avant que le service des Missions Étrangères ne prenne

le relai et impose le français comme métalangue. Enfin Otto ZWARTJES [REI (22)] fait le bilan de son travail sur la linguistique missionnaire dans les deux Amériques. Si, au 1^{er} octobre 2018, 50 notices (dont 18 pour le nahuatl) ont été publiées, à terme c'est plus du double qui devraient être réalisées.

Rappel des objectifs initiaux

L'objectif initial du CTLF était de proposer un panorama des textes essentiels qui ont marqué le développement de la réflexion sur le langage et les langues. Il s'agissait d'établir, selon l'expression de Sylvain Auroux, le « noyau théorique du long terme », qui, dans un projet initial de *Dictionnaire de la terminologie linguistique*, devait occuper 700 000 caractères sur les 2,4 millions. Le projet suppose d'admettre que la réflexion sur le langage et les langues n'est pas née de rien, qu'elle « se réinscrit » dans un long passé qui (cf. les notions de taux de réinscription » et d'« horizon de rétrospection »).

D'emblée, le projet était centré sur les *textes*, non sur les *auteurs*. Les études en histoire des théories linguistiques¹ ont montré qu'on pouvait dégager à la fois des « textes fondateurs » – ceux qui refondent une vision du domaine (la *Grammaire générale et raisonnée* de Port-Royal, le *Cours de linguistique générale* de F. de Saussure) – et des « textes piliers » – ceux autour desquels s'articule la réflexion linguistique, parfois pour des années, voire des siècles (la *Tekhnê* de Denys le Thrace, l'*Ars Donati* ou l'*Ars Prisciani*).

Le CTLF était donc centrage sur *quelques* grandes traditions, bien identifiées comme telles : par exemple la tradition de la grammaire indienne, la tradition de la grammaire arabe, etc.

Des notices descriptives ont donc été établies, dans le but de fournir deux types de données :

- des données factuelles, concernant la taille de l'ouvrage, le nombre d'éditions et de rééditions, etc.
 - des éléments qui relèvent plus de l'évaluation: intérêt général, influence subie, influence exercée.
- Ces derniers sont évidemment les plus importants : on ne conçoit guère la description d'un ouvrage qui, restant un hapax, serait absolument hors de la transmission.

Quels ouvrages choisir ? les *outils* pour explorer les langues sont innombrables et de nature très diverse. D'emblée a été établie la priorité pour les ouvrages qui fournissent le plus de matériel métalinguistique, à savoir les grammaires. Évidemment les textes généraux traitant du langage, qu'il s'agisse des compilations linguistiques, des textes de phonétique et de phonologie, ou encore de linguistique « générale » s'imposaient d'emblée dans le corpus. Mais d'autres types de textes se sont ajoutés, comme les remarques sur la langue française, ou encore les traités orthographiques portugais. Il en est de même pour certains matériaux si difficiles à séparer du contexte religieux, à l'intérieur duquel ils ont été élaborés et souvent au service desquels ils se trouvent (ajout dans les mot clés de « texte religieux », dans le cadre de la linguistique « missionnaire »).

Les difficultés

L'adjonction incessante de nouveaux textes nous a obligés à changer la « granularité » dans la description. Il s'agissait surtout au départ d'explorer quelques grandes tendances dans leurs grandes lignes. Il s'agit maintenant d'explorer de façon beaucoup plus systématique certains domaines.

Mais cela dépend des secteurs. Comparons le secteur des grammaires françaises et celui des grammaires des langues d'Amérique. Le premier secteur n'a pratiquement pas bougé actuellement ; nous sommes passés de 100 à 116 notices. Le second contient actuellement 50 notice et devrait dépasser à terme les 120. Dans le premier cas, il n'y a que les textes les plus représentatifs. La production de grammaires françaises devient exponentielle à partir du XIXe siècle, comme l'ont montré par exemple les répertoires d'André Chervel. Dans le second cas, Otto Zwartjes répertorie

¹ Cf. par ex. Colombat, B., Fournier, J.-M. et Puech, C., *Histoire des réflexions sur le langage et les langues*, Paris, Klincksieck, p. 4-6 et 21.

tous les ouvrages décrivant les grammaires amérindiennes au fur et à mesure qu'ils parviennent à sa connaissance. Dans un cas, on a affaire à un domaine très documenté et relativement bien exploré. Dans l'autre, à un domaine vierge, peu documenté, et où les items se réduisent parfois à *un* manuscrit. Les formes d'investigation des deux domaines ne peuvent être que différentes. Il est à noter cependant que le nombre d'exemplaires n'est pas un critère en soi : le dernier ouvrage entré dans les grammaires françaises est justement aussi une grammaire manuscrite, *De L'Eloquence ou Regles pour parler correctement*, d'Antoine de Courtin [notice 2313], retenue pour l'originalité de la perspective adoptée par l'auteur.

Les risques

Sous cet afflux de nouveaux textes, l'entreprise ne courait-elle pas un nouveau risque, celui d'éparpillement, de submersion sous la variété des langues, le nombre de grammaires, et de ce fait, une perte de « représentativité » ? Le projet initial, qui avait donné lieu à une édition papier en 1998-2000, s'intitulait « Corps *représentatif* des grammaires et des traditions linguistiques ». Il était en quelque sorte contenu naturellement par des contraintes d'ordre matériel. Lors de la publication papier, il fallait nécessairement le restreindre à un certain nombre de pages. Avec la numérisation et l'édition numérique, le verrou a sauté, et d'une certaine façon, nous avons dérivés :

- tendance à l'exhaustivité plutôt qu'à la représentativité : l'objectif d'Otto Zwartjes est bien de décrire *toutes* les grammaires des langues amérindiennes, communes et à découvrir. En fait la couverture dépend des secteurs : pour les domaines bien documentés, comme celui des grandes langues européennes, nous retiendrons toujours le principe de la sélectivité ; pour des « niches », des documents rares, atypiques, mais importants en ce qui concerne le développement des idées linguistiques – le *Fyrsta málfraeðiritgerðin* mentionné *supra* est un bon exemple –, nous serons beaucoup plus accueillants ;

- inflation des notices : celles-ci devaient être contenues dans 8000 car. Or beaucoup atteignent 15 000, voire 20 000 caractères. Nous nous sommes récemment imposé – et nous avons imposé à nos collaborateurs – de nous limiter à un chiffre raisonnable (15 000 caractères).

Les demandes

Parmi les demandes qui nous ont été adressées, l'une des plus fréquentes a été celle de rendre la base plus accessible. Quel que soit le portail sur lequel on se positionne, la base a parfois été estimée « muette » puisqu'on arrive sur un écran vide et l'internaute ne sait pas a priori ce qu'il faut faire pour la « faire parler ». Il ne comprend pas forcément qu'il suffit de cliquer sur « envoyer », sans sélection préalable, pour récupérer toutes les données : les 743 notices, les 4733 références bibliographiques, les 194 images ou les 804 textes numérisés. En fait, lors de la conception de la base, la volonté initiale était une volonté de « dépouillement » : il s'agissait, et il s'agit toujours, de ne pas encombrer le site avec des informations de tous ordres mal hiérarchisées. De même, le site est apparu parfois (trop ?) discret sur ses concepteurs et ses réalisateurs.

Nous sommes sensibles à ces remarques ou à ces critiques. Nous pouvons répondre : (1) qu'il y a un certain nombre d'aides à l'utilisation du site, comme sur le site d'accueil, par la mention « Informations », l'article « Fonctionnement du portail », accessible aussi par la base d'articles (site jaune) ; (2) que le nombre de fonctionnalités utilisables (par exemple les possibilités de tri) est important, au point même que... nous en oublions nous-mêmes quelques-uns ; (3) que nous travaillons constamment à l'amélioration de l'ergonomie de la base. Donnons-en deux exemples :

– la gestion de la bibliographie : dans chaque notice, les renvois bibliographiques ne permettaient auparavant de renvoyer – en cliquant sur le nom concerné – qu'à la liste des auteurs de bibliographie secondaire (site bleu) ; il permet maintenant (en cliquant sur « Références ») de renvoyer uniquement aux « Références » de la notice (ce qui peut être intéressant, par ex. pour les notices les plus chargées en références, comme celle de Vološinov [notice 5321] qui comprend 103 références) ;

– le perfectionnement du tableau chronologique. Statique au départ, mais de plus en plus chargé en informations au point de devenir difficilement lisible, il est devenu interactif et « dépliable », permettant ainsi d'accéder au détail de certains secteurs. En fait, il en vient peu à peu à être le moteur central de la base des notices, au point que nous envisageons de donner accès par son intermédiaire à toutes les notices, même si celles-ci resteront également accessibles par les formulaires de recherche.

Bref, le CTLF est et restera en constant développement, comme le montre la page « [Actualité du CTLF](#) » qui elle-même avait fait son apparition au 2^e semestre de 2014. Et la meilleure preuve en est... que nous avons toutes les peines du monde à actualiser les chiffres dans chacun des portails, tant ceux-ci changent souvent !