

Recherches linguistiques et corpus

Responsable Franck Neveu

Le thème *Recherches linguistiques et corpus* mis en place au sein du laboratoire STIH de l'Université Paris-Sorbonne a pour objet le développement d'une réflexion commune et croisée, d'ordre épistémologique et méthodologique, sur la notion de corpus telle qu'elle est exploitée aujourd'hui dans les **sciences du langage**, et, plus largement, dans les **sciences humaines et sociales**. On y traite des questions relatives aux notions de donnée, d'observable, d'empiricité, de théorie (lien type/occurrence), de variable contextuelle, d'annotation, de codage, de catégorisation, etc. On s'interroge sur la fonction des corpus dans l'activité de recherche. Ce thème transversal est organisé en **séminaires-ateliers** ouverts notamment aux *chercheurs*, aux *enseignants-chercheurs*, aux *ingénieurs*, aux *doctorants* et aux *étudiants de master*. Les séances sont constituées de deux conférences, suivies d'un atelier d'observation et d'application.

Mercredi 27 mars 2019, 14h-17h00
Université Paris-Sorbonne, salle des Actes

**Le décalage contextuel dans les communications médiées par écran :
à la croisée de la dématérialisation du contexte
et de la re-textualisation des échanges**

Béatrice Fracchiolla

Université de Lorraine, CREM EA 3476

**Réflexions historiques et épistémologiques
autour de la notion de corpus**

Franck Cinato

HTL - *Histoire des Théories Linguistiques*, CNRS, UMR 7597 — Université Paris Diderot

Entrée libre

1, rue Victor Cousin, 75005, Paris

Contact : franck.neveu@sorbonne-universite.fr

Le décalage contextuel dans les communications médiées par écran : à la croisée de la dématérialisation du contexte et de la re-textualisation des échanges

Béatrice Fracchiolla

Les médias de communication numérique à distance et les créations logicielles de communication entre les individus qui ont émergé ces trente dernières années ont entraîné un développement sans précédent de nouvelles technologies de l'écrit et de l'écriture ainsi qu'une retextualisation des échanges concomitante à une dématérialisation du contexte. Or, la dématérialisation du contexte propre à ces environnements complexifie de manière exponentielle sa compréhension, multipliant les risques de malentendus, de décalages, d'incompréhension. Les personnes qui communiquent n'étant en effet pas physiquement présentes l'une à l'autre, leur rencontre s'effectue et se matérialise *via* un texte. La textualité devient dès lors un nouveau mode de rencontre, dont le texte est le lieu. Le fait que nous imaginions la communication comme si elle était instantanée développe une compréhension textuelle du message, dans l'instant où nous le lisons, qui est propre à la réception. Tous ces échanges reposent sur une faille identique : ils privent l'interaction des nombreuses possibilités de régulation que permet la présence et dont nous privent les écrans, alors que la communication ne repose que sur un pourcentage très faible de verbal. Je définis ainsi le *décalage contextuel*, qui s'insinue dans ce manque, comme étant d'abord un décalage énonciatif augmenté d'une constante et nécessaire recontextualisation. Il s'agit de problématiser ce processus de *décalage contextuel* comme une notion opérationnelle pour l'analyse d'échanges textuels. Je propose d'en illustrer les manifestations à travers l'analyse d'un extrait de corpus, constitué par des réactions au texte *la liberté d'importuner* paru dans le *Monde* du 10 janvier 2018. Plus largement, dans le contexte du séminaire-atelier, il serait intéressant de réfléchir ensuite aux différentes situations où il se manifeste et aux manières que nous pouvons avoir de l'appréhender en fonction de nos corpus et de nos outils d'analyse respectifs.

Béatrice Fracchiolla est professeure en sciences du langage à l'université de Lorraine, Centre de Recherches sur les Médiations EA 3476. Elle travaille plus particulièrement sur la violence verbale en analyse de discours, énonciation et analyse conversationnelle autour des notions d'agression, d'injure, d'attaque courtoise, de politesse et d'impolitesse ; et ce, plus particulièrement à propos du genre et en politique. Depuis 2007, elle coordonne le thème 7 de l'axe 1 « Anthropologie de la communication » à la MSH Paris Nord de Saint-Denis avec Georges Chapouthier, directeur de recherches émérite. C'est dans ce cadre qu'elle a développé un intérêt pour les approches interdisciplinaires qui visent à observer un même objet sous des angles différents, afin de mieux saisir la complexité du réel et des objets scientifiques, y compris lorsqu'ils sont linguistiques. Elle s'intéresse au langage dans sa dimension pragmatique et relationnelle, dans la mesure où toute parole est adressée à autrui, selon certaines fins, et plus particulièrement à travers l'étude de certains éléments de langage (pronoms personnels, termes d'adresse, interpellations sous diverses formes, utilisation ou non d'un féminin référentiel...). Cette perspective l'a menée ces dernières années à travailler en collaboration avec des médecins dans le cadre d'un projet (COREV) soutenu et financé par la mission interdisciplinaire du CNRS pour tenter de comprendre ce qui se produit lors de la réception d'un message, en particulier lorsque celui-ci est violent, et à l'articulation du langage en relation avec les émotions ressenties et la définition de l'altérité. Sites : beatricefracchiolla.com et <http://violenceverbale.hypotheses.org>

Réflexions historiques et épistémologiques autour de la notion de corpus

Franck Cinato

Cette présentation vise à préciser ce qu'est un corpus en replaçant cette notion dans la perspective historique d'un temps long. L'évocation d'un certain nombre de corpus, d'abord sous forme manuscrite, puis d'imprimés, et enfin, après l'apparition des outils numériques, sous forme de bases de données, offrira l'occasion d'observer en quoi l'informatique a considérablement fait évoluer la nature des corpus et ce qu'on attend d'eux, sans pour autant avoir créé de rupture avec la notion traditionnelle.

Plus que de simples parallèles, il existe une véritable continuité du point de vue épistémologique entre des pratiques attestées dans le haut Moyen Âge et celles contemporaines en ce qui a trait à l'analyse de données textuelles. Ce fil d'Ariane qui relie les grammairiens (médio-)latins aux linguistes contemporains, s'exprime à travers le phénomène que la tradition nomme 'glose'. Or, la garantie, selon nos standards scientifiques actuels, d'une analyse fiable — quantifiable et vérifiable —, repose en grande partie sur (a) la qualité d'un corpus (en termes de nature, contenu et pertinence) et (b) le soin apporté à sa préparation en vue d'analyse. Il apparaît donc, qu'outre le choix des matériaux à questionner, leur « mise en forme » (préparation préalable des données) constitue un aspect crucial d'une recherche. Et c'est précisément sur cet aspect que les pratiques anciennes et contemporaines se rencontrent : sur le terrain de l'annotation.

Dans un tel cadre, une recherche sera donc largement influencée par la nature de cette étape en amont que l'on peut qualifier de pré-analyse, ou préparation des données, qui consiste à ajouter une dimension supplémentaire aux textes. L'exposé se terminera alors par un regard « derrière » les données en considérant quelques problématiques relatives à l'encodage, aux annotations et autres métadonnées attachées aux corpus.

Franck Cinato est chargé de recherche au laboratoire *Histoire des Théories Linguistique* (HTL, CNRS, UMR 7597) [<http://htl.linguist.univ-paris-diderot.fr/laboratoire/membres/cinato>]. Ses travaux portent sur la réception de la grammaire latine de Priscien (V/VI^e s.) à l'époque carolingienne et plus largement sur la glossographie du haut Moyen Âge, à travers l'étude des glossaires [<http://liber-glossarum.huma-num.fr/index.html>] et de toutes les formes de gloses, notamment celles portant sur la syntaxe.