



Corpus de textes : composer, mesurer, interpréter

Le laboratoire junior « Des nombres et des mots (N&M's) », qui réunit de jeunes chercheurs en géographie, en informatique, en linguistique, en stylistique et en sociologie, organise un colloque pour initier des réflexions interdisciplinaires sur les usages des corpus de textes. Cette manifestation s'inscrit dans la continuité de la journée d'étude du 29 mars 2012 organisée à l'ENS de Lyon sur les « Pratiques de corpus numériques : enjeux et méthodes interdisciplinaires », où des réflexions théoriques alternaient avec des études de cas. Ces deux types d'approches seront attendus, alimentés tant par des aspects méthodologiques que par des résultats.

Le corpus de textes constitue l'un des lieux les plus favorables à l'observation des réalisations de la langue. Deux conceptions du corpus coexistent et peuvent se nourrir : il est tantôt considéré comme un recueil d'exemples ou d'attestations servant à vérifier des hypothèses ou à étayer des théories linguistiques, tantôt perçu comme un outil d'exploration à part entière, susceptible de faire émerger des savoirs linguistiques.

En parallèle, les chercheurs en sciences sociales ont conscience du fait que les objets qu'ils étudient émergent discursivement. Répondant à des démarches empiriques, le corpus de textes est ainsi devenu le terrain privilégié de la mise au jour et/ou de l'observation de phénomènes historiques, géographiques, sociologiques, à partir de leurs formulations et dénominations par différents acteurs.

La construction du corpus, son analyse, son interprétation et sa mise en perspective questionnent certes le contexte d'énonciation des discours, mais aussi la place ou la posture du chercheur. Au sein de chaque discipline, la place accordée aux corpus textuels, ainsi que la question de leur investigation, de leur taille, de leur représentativité et de leur interprétation doit être éclaircie et débattue.

En outre, la situation centrale occupée par le texte interroge les cloisonnements disciplinaires et la possibilité de les dépasser. A quelles conditions peut-on faire communiquer les savoirs et les méthodes d'un champ à l'autre ? Les corpus de textes (considérés comme des construits) représentent-ils un dénominateur commun ou sont-ils le révélateur de visions segmentées qui reproduisent des clivages disciplinaires ?

Axes de réflexion

- Corpus et traitement numérique

En termes de recueil et de formalisation des corpus, la numérisation ou non des données interroge le transfert de méthodes de traitement traditionnelles sur papier, auxquelles le chercheur a été formé lors de son parcours universitaire, vers des méthodes encore parfois peu stabilisées du fait de l'émergence d'offres logicielles plus complètes. Les avantages et les inconvénients de la numérisation, de l'édition numérique, de la place faite aux métadonnées et au codage pourront notamment être abordés.

Différents axes de réflexion parmi d'autres pourront faire l'objet d'une communication. Quels sont les liens entre choix d'un logiciel et positionnement épistémologique ? Si l'utilisation de logiciels favorise la prise en compte de vastes corpus, ne risque-t-elle pas de minimiser des particularités au profit d'une homogénéisation des phénomènes ?

- Corpus et contexte

Tout corpus est « en contexte » : cette perspective modèle en effet le recueil des données. Ainsi, la composition, la mesure et l'interprétation des corpus sont des processus qui s'imbriquent selon des logiques circulatoires. Le contexte, considéré ici comme les éléments extérieurs au texte que le chercheur peut être amené à incorporer dans son analyse, constitue un élément central de la réflexion. L'utilisation de certains logiciels ne permet parfois pas d'intégrer facilement ces différents éléments contextuels.

La recherche sur de vastes corpus numériques ne risque-t-elle pas d'occulter l'influence du contexte, par exemple en observant uniquement des occurrences ? Le contexte doit-il être pris en compte dès les hypothèses (formatant/influçant alors le regard du chercheur), ou doit-il être intégré sous la forme d'un codage d'analyse de contenu, ou encore doit-il être analysé ultérieurement comme un élément d'interprétation des résultats ?

- Corpus et contraste

L'approche par corpus permet de faire apparaître des contrastes, au sens de différences jugées significatives. Les discours varient d'un locuteur à l'autre selon des valeurs, des intérêts ou des pratiques largement influencés par le contexte historique, social, économique ou culturel. Le corpus permet de cerner la manière dont un événement ponctuel bouscule des systèmes de représentations, ou comment ceux-ci évoluent progressivement de manière sous-jacente.

Les concepts de saillance ou de tendance pourront être discutés. Quels critères permettent de dégager des contrastes pertinents ? Comment mettre en place des points de comparaison ?

Public visé

Le colloque se veut résolument interdisciplinaire et ouvert aux chercheurs confirmés comme aux jeunes chercheurs. Nous souhaitons mobiliser autant les sciences humaines et sociales que les Humanités, autant les approches quantitatives (analyse de contenu, analyse de données textuelles) que qualitatives.

Conditions de soumission

Le résumé s'étendra sur une page maximum ; une deuxième page pouvant être mobilisée pour des figures. La feuille de style à utiliser est fournie : <http://nombresetmots.ens-lyon.fr/spip.php?rubrique12> .

Les propositions retenues feront l'objet d'une communication orale en français ou en anglais de vingt minutes, suivie d'un temps de questions. Une sélection de communications sera publiée sous la forme d'un recueil d'articles en français.

Mots clés

Corpus – discours – méthodologie – recueil – traitement – interprétation – réflexivité.