

Lexicometrica : Appel à contribution

<http://www.cavi.univ-paris3.fr/lexicometrica/>

Thème : Topographie et topologie textuelles

Responsables : Sylvie Mellet et André Salem

Appel à contribution :

Depuis ses débuts la statistique linguistique, y compris lorsqu'elle s'applique à l'étude des textes et des discours, a principalement recouru à des modèles qui tendent à négliger ce fait majeur qu'un texte est une structure ordonnée ; les dénombrements, les relevés de fréquences, les calculs de spécificités reposent tous sur le fameux schéma d'urne et renoncent à prendre en compte le positionnement dans le texte des unités dénombrées. Certes, les résultats ainsi obtenus sont généralement intéressants et bien interprétables, et ils ont largement contribué au développement et aux succès de la discipline.

Mais ils se pourraient qu'ils soient en train d'atteindre leurs limites. Ou, du moins, de ne plus suffire pour donner entière satisfaction au chercheur. De plus en plus souvent en effet, ceux-ci souhaitent pouvoir établir, à côté de la dimension paradigmatique appréhendée par ce type de calculs statistiques traditionnels, la dimension syntagmatique des données textuelles, saisies à courte ou à longue portée : distribution régulière ou non d'une entité linguistique (mot ou catégorie grammaticale) susceptible d'arriver à intervalles à peu près égaux ou, au contraire, en paquets plus ou moins denses ; répartition d'un élément au fil du texte, selon la structure globale de celui-ci et ses parties constituantes ; phénomènes d'échos et d'alignements dans la mise en parallèle de deux textes ou deux portions de textes ; etc.

Bien sûr, des travaux, dont certains sont déjà anciens, ont abordé ces questions : parmi les plus connus citons tous ceux d'A. Salem qui ont établi la pertinence de la fameuse notion de « segment répété » et qui ont mis en place les outils pour les repérer et les analyser ; citons aussi les travaux de P. Lafon sur les « rafales » et son article "Statistique des localisations des formes d'un texte" paru en 1984 dans la revue *Mots* ; ou encore l'article de D. Sérant et Ph. Thoirion sur la « topographie des formes répétées » (*Revue Informatique et Statistique dans les Sciences humaines* 24, pp. 333-343) ; etc.

Actuellement, cette question reprend de l'acuité et les études, ainsi que les développements logiciels afférents, se multiplient. Le moment nous semble donc venu de faire le point.

Ce numéro de *Lexicometrica* accueillera donc exclusivement des contributions consacrées aux notions de topographie et topologie textuelles, c'est-à-dire à la prise en compte, dans les exploitations automatiques des textes numérisés et dans leur traitement quantitatif de la linéarité intrinsèque du texte, voire de sa structure en réseau avec d'autres textes au sein d'un corpus fortement cohérent (cas des recueils par exemple).

Les articles proposés pourront se centrer soit sur des problèmes méthodologiques généraux, soit sur des applications particulières, soit sur des développements logiciels.

La soumission se fera sous la forme d'un texte de présentation synthétique (entre 3000 et 5000 caractères), accompagné d'une bibliographie de référence. Ce texte sera envoyé sous format Word ou PDF, simultanément à André Salem (salem@msh-paris.fr) et à Sylvie Mellet (mellet@unice.fr). Après avis d'acceptation, les textes longs définitifs devront être fournis **avant le 8 septembre 2008**. Sous format Word, ils ne devront pas excéder 35000 caractères et devront respecter la feuille de style de la revue, disponible à l'adresse : <http://www.cavi.univ-paris3.fr/lexicometrica/soumis.html>