

# **Création d'une banque de textes socio-politiques en Fédération Wallonie- Bruxelles. « Belgofrancotext »-BFT**

## **Demande d'un financement de recherche d'initiative ministérielle.**

### **Introduction**

L'intérêt pour l'analyse des discours sociopolitiques et médiatiques est en pleine croissance dans les sciences humaines. Les nouvelles techniques dites « de l'information et de la communication » accroissent les capacités de production et de vitesse de diffusion de ces types de discours tandis que le perfectionnement des logiciels d'analyse de données textuelles incitent de plus en plus de chercheurs et d'étudiants à placer l'étude des discours au cœur de leurs stratégies d'apprentissage et de recherche.

Un travail sérieux tant sur le plan de la recherche que de l'enseignement à tous les échelons de l'enseignement supérieur ne peut plus faire l'impasse de l'impact qu'a cette démultiplication de la production discursive sur les dynamiques de la transformation sociale et sur le développement de la connaissance des processus sociaux et politiques.

Or, pour alimenter les nouveaux besoins de la recherche ou de l'enseignement dans le domaine de l'analyse des discours, en Belgique francophone, de nombreux enseignants ou chercheurs, et c'est sensiblement le cas pour de nombreux doctorants, se trouvent obligés de recourir à des données de source étrangère car il n'existe aucune collecte systématique et institutionnellement organisée des données textuelles, sous format numérisé, accessibles gratuitement, pour couvrir l'ensemble des principales productions discursives socio-politiques ou médiatiques de la Belgique francophone, hormis le cas d'une partie de l'activité parlementaire.

De très nombreux étudiants se penchent sur l'étude des discours socio-politiques américains ou français car ils sont accessibles sur internet alors qu'ils pourraient s'intéresser avec fruit à l'étude de la vie socio-politique belge francophone.

Au sein des nombreuses institutions d'enseignement supérieur francophone du pays nous sommes face à une urgence : la nécessité de mettre en place un appareil de collecte systématique et fiable, doublé d'une infrastructure publique d'accès à ces données textuelles, pour sortir du bricolage et présenter à l'étranger des études de qualité comparable à ce qui se fait ailleurs.

Le besoin est clair et simple : pouvoir disposer de séries chronologiques systématiques et exhaustives de données textuelles qui reflètent le déroulement de l'activité socio-politique et médiatique de notre propre système politique.

### **Dimensions techniques**

La question soulève évidemment des questions techniques. Celles-ci ne soulèvent pas de difficultés majeures. Elles nécessitent cependant la consultation d'un comité d'utilisateurs à mettre en place.

Dépôt : Imposer des normes de standardisation archivistiques, informatiques ou documentologiques risquerait bien d'entraver les dépôts. Pour éviter tout frein et faciliter lesdits dépôts, la banque autorisera une variété de formats courants, indépendamment des besoins des chercheurs (PDF, traitement de textes, format de composition utilisé par les

graphistes).

À partir de ces formats de dépôt, la plateforme permettra d'ajouter des versions éditées par les chercheurs, sous des formats différents, utilisables par les logiciels de traitement de données textuelles. La version originale sera toujours conservée permettant l'existence de différentes révisions et les comparaisons de version. A terme, et en fonction du développement de la science informatique, il sera discuté des possibilités visant à mettre l'ensemble des textes sous un format numérisé identique, facilement reconvertible pour les besoins des analyses et recherches particulières.

La plateforme prévoit de vérifier l'identité du déposant, pour éviter que la plateforme puisse être détournée pour servir des stratégies de désinformation croisée (entre partis politiques, par exemple). La mise en œuvre de cette certification pourra par exemple passer par l'envoi automatique d'un code secret à l'adresse de tout déposant bénéficiant d'une adresse institutionnelle (administration, ministères, partis politiques...).

A côté de ces dimensions techniques (nécessaires mais néanmoins aisées à régler), la durabilité et la régularité de la récolte et de l'accès nécessite une intervention forte du monde politique pour imposer et organiser ceux-ci.

### **Pourquoi une banque de textes en Fédération Wallonie-Bruxelles?**

Non seulement de nombreux documents socio-politiques essentiels pour l'analyse ne sont pas numérisés et se perdent très facilement (comme les tracts électoraux toute-boîte et les affiches électorales), l'internet crée une illusion de conservation des données alors que les « sites » et les « blogs » changent très vite – de très nombreuses données disparaissant sans laisser de trace –, l'intérêt du stockage de données d'une institution peut être lié à des raisons très pratiques qui peuvent réduire son cycle de conservation alors que le chercheur a besoin de données continues à très long terme, des initiatives de stockage peuvent exister sans être vraiment connues de tous et les multiples formats actuels de données « sauvegardées », notamment le format pdf, rendent souvent l'analyse difficile voire impossible, car ils ne se prêtent pas à un traitement avec les logiciels d'analyse disponibles.

La disparition des données, leur éparpillement, leur conservation dans des formats inadéquats... tout plaide pour la création d'une telle banque.

### **L'infrastructure à mettre en place**

Nous proposons de créer une base de données textuelles socio-politiques à implanter sur le site du FNRS sur la toile mondiale. Cette base serait en libre accès.

Le travail de rubriques et de catalogues sera essentiel pour organiser les données en type d'acteurs et en séries chronologiques de façon aisément manipulable.

### **Création d'une obligation légale de dépôt**

Afin d'assurer la systématisme et l'étendue de la récolte, il serait nécessaire d'imposer légalement l'envoi sous format numérisé des « données textuelles structurelles » de toutes les organisations socio-politiques et médiatiques recevant une aide financière publique.

Par « données structurelles », nous entendons les documents publics destinés à faire connaître l'action entreprise par ces organismes : rapports annuels d'activités, communiqués de presse, périodiques, déclarations publiques, tracts électoraux,...

De nos jours, tous ces documents, même s'ils sont destinés à être imprimé sur papier, passent par

un stade premier de numérisation. Ce qui facilitera leur dépôt sans coût.

### **Publicité publique pour inciter à dépôt**

Dans le cas où les organisations socio-politiques et médiatiques ne seraient pas directement dépendantes du financement public, une incitation publique permanente sera organisée pour susciter l'envie du dépôt à des fins d'aide à la recherche et à l'enseignement.

### **Concentration de données existantes et indication claire de leur accès et maniement pour les institutions administratives publiques ou para-publiques disposant déjà de bases de données numérisées**

Dans le cas où les administrations disposeraient déjà de données numérisées et stockées sur un site *ad hoc*, il serait nécessaire de créer une page d'information particulière sur le futur site de la banque belgofrancotext afin d'y indiquer le lien d'accès via l'internet (adresse URL), le type de données qu'on peut y trouver et leur format.

Par « organisations socio-politiques », nous entendons les partis politiques, les ONG, les mutuelles, les organisations syndicales et patronales,...

Par « organisations médiatiques », la radio et télévision publiques francophones. Dans ce cas particulier, dans un premier temps, l'effort de collecte portera sur les seuls « journaux parlés » portant sur l'actualité socio-politique.

### **Budget nécessaire pour l'opération.**

Engagement de deux spécialistes en base de données informatisées intégrés au personnel permanent du FNRS, idéalement aux profils complémentaires (un documentaliste et un informaticien, par exemple). La consultation d'un juriste spécialiste en droit de la propriété intellectuelle, voire sur les licences documentaires publiques (« *commons* ») pourrait être utile lors de la mise sur pied de ladite plateforme. Ces personnes travailleraient en interaction avec un comité d'utilisateurs à mettre en place. Ce comité d'utilisateur pourrait être formé au sein du Groupe de contact scientifique « Langue(s) et Politique(s) » déjà créé au sein du FNRS.