



HAL
open science

Les outils socio-informatiques et les recherches collectives

Patrick Trabal

► **To cite this version:**

Patrick Trabal. Les outils socio-informatiques et les recherches collectives : Une expérience de corpus partagés dans le domaine des sciences du sport. 2010. hal-03033267

HAL Id: hal-03033267

<https://hal.parisnanterre.fr/hal-03033267>

Submitted on 1 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Les outils socio-informatiques et les recherches collectives

Une expérience de corpus partagés dans le domaine des sciences du sport

Patrick Trabal

(CERSM – Univ. Paris Ouest Nanterre / GSPR – EHESS)

L'approche socio-informatique peut utiliser les capacités de calcul informatique pour tester, en les rendant réversibles à moindre coût, différentes opérations de codage. Dans cette logique, l'usage de l'ordinateur, classiquement réduit à une simple délégation des calculs dans le but de fournir une représentation unique des données, peut être réorienté vers la mise en série d'expériences sur des corpus de textes, en enrichissant progressivement l'analyse. Cette approche, inscrite dans la « [logique de l'enquête](#) » chère à Dewey (1993), ouvre sur une logique coopérative et un travail collectif. Considéré comme un mot d'ordre des nouvelles sociologies, le fonctionnement de ce « collectif » reste toutefois inabouti comme mode de production de connaissances. La figure la plus courante reste en effet celle du chercheur individuel, seul sur son terrain, recueillant ses propres données, les analysant à destination de diverses arènes (réunions de labo, séminaires, congrès), puis passant à la publication. La controverse porte généralement les résultats finaux une fois le travail accompli, et de fait, il est souvent difficile de garder du temps pour discuter finement les contenus des corpus, les cadres d'analyse et les raisonnements.

Le projet ANR [Trascinter](#) rassemble des chercheurs de trois laboratoires (CERSM- Univ. Paris Ouest Nanterre / ACP- GREHSS – Univ. Paris Est Marne La Vallée / SOI – Univ. Paul Sabatier de Toulouse) dans le but d'expérimenter un tel espace de travail collectif tout en forgeant les outils nécessaires. L'objet commun placé au cœur de cet espace coopératif est l'interdisciplinarité dans les sciences du sport. Je n'entrerai pas ici dans une discussion sur la pertinence sociologique de cet objet qui préoccupe de nombreux acteurs tant du point de vue de la sociologie des sciences que de celui des débats concernant les formations sur les métiers du sport ou encore des polémiques comme le dopage et plus généralement les relations sport-santé. Je propose plutôt de préciser trois moments décisifs pointant vers un travail collaboratif et de discuter des moyens d'outiller cette réflexion.

1. Echanges de corpus

Les expériences collectives sur les corpus sont plutôt rares. Une des raisons en est la nécessité de protéger « ses » données. Au-delà des questions de confidentialité dont on peut s'affranchir en anonymisant les informations sensibles, les chercheurs en sciences sociales livrent difficilement leurs corpus, ce que l'on peut comprendre dans un contexte de concurrence exacerbée. Du même coup est éludée la question épistémologique de la conduite d'une discussion du matériau empirique d'un collègue auquel on ne peut accéder.

Entre une position de mise à disposition publique intégrale, qui nécessiterait pour être généralisée un consensus large de la communauté épistémique concernée – ce qui est loin

d'être le cas, et un protectionnisme scientifique poussé à l'extrême – ce qui hypothèque les discussions, on peut préconiser une posture intermédiaire : partager et discuter avec des collègues qui partagent les mêmes préoccupations intellectuelles sur la base d'une confiance qui s'éprouve dans un échange réciproque des « données ».

L'outil BACH (Base d'Archives pour les Chercheurs) a été construit sur ce principe et développé par le Centre de Ressource Informatique de l'Université Paris Ouest Nanterre¹. Dans un dispositif dédié, les collègues peuvent déposer leurs corpus en précisant les communautés de chercheurs susceptibles de travailler sur leurs données. Ainsi, au-delà du projet TRASCINTER différents groupes de chercheurs utilisateur de Prospéro peuvent déclarer une mise en collaboration de corpus.

Cocher les cases correspondant aux groupes de chercheurs pouvant accéder à vos fichiers

Informations sur les fichiers à copier de votre ordinateur sur le serveur

Abdel TAG Test <input type="checkbox"/>	APS <input type="checkbox"/>	Chimères <input type="checkbox"/>	Dopage <input type="checkbox"/>	DOXA <input type="checkbox"/>	GSPR <input type="checkbox"/>	Perso <input type="checkbox"/>	ssd <input type="checkbox"/>	Test <input type="checkbox"/>	Trascinter <input checked="" type="checkbox"/>
Source	<input type="text" value="Photocopie"/>								
Original	<input type="text" value="Archives du CIO"/>								
Etat	<input type="text" value="Bon"/>								
Descripteurs	<input type="text" value="dopage olympisme"/>								
Langue	<input type="text" value="Français"/>								
nom fichier compressé		<input type="text"/>		<input type="button" value="Parcourir..."/>					
<input type="button" value="Valider"/>									

Figure 1 : Fenêtre de dépôt de corpus

Le dépôt de textes se fait par l'envoi d'une archive contenant les textes, les fichiers livrant des éléments de contexte et les ressources associées.

Les fichiers texte sont alors déposés dans une base de données à laquelle on accède par des champs de deux types : ceux renseignés lors du dépôt (cf. figure ci-dessus) et ceux présents dans les fichiers contextuels.

Il est alors possible d'accéder aux différents fichiers avec la possibilité de les télécharger individuellement, d'éditer et de modifier les champs (Figure 2), de corriger éventuellement le texte (notamment en cas de coquilles).

—mise à jour du contexte

identifiant	906	Titre	<input type="text" value="PostFace"/>	Auteur	<input type="text" value="Anquetil"/>
Nom fichier	1966Anquetil05	Destinataire	<input type="text"/>	Date	<input type="text" value="01/09/1976"/> <input type="button" value="Calendrier"/>
Nom support	<input type="text" value="En brulant les étapes"/>	Type support	<input type="text" value="Ouvrage"/>	Observation	<input type="text" value="p 189-193"/>
Statut auteur	<input type="text" value="ancien cycliste"/>	Lieu emission	<input type="text" value="Cyclisme"/>	Calcul narrateur	<input type="text" value="n"/>
Calcul destinataire	<input type="text" value="n"/>	Heure	<input type="text" value="00h00"/>	Résumé	<input type="text"/>
KeyWord	<input type="text" value="test"/>	Remarques	<input type="text"/>	Champs libre 1	<input type="text"/>
Champs libre 2	<input type="text"/>	Champs libre 3	<input type="text"/>	Champs libre 4	<input type="text"/>
Champs libre 5	<input type="text"/>	Champs libre 6	<input type="text"/>	Champs libre 7	<input type="text"/>
Champs libre 8	<input type="text"/>				

Liste des fichiers

—Liste des ressources externes

Référence Externe	Action
-------------------	--------

Figure 2 : Edition et modification du contexte

A un premier niveau, ce système fonctionne sur un principe d'économie. On ne corrige le texte qu'une seule fois (et les différents collègues qui utiliseront ce texte bénéficieront de la version corrigée). La construction d'un corpus collectif s'en trouve facilitée. On peut en effet regarder les documents disponibles avant de se lancer dans une coûteuse numérisation et organiser le rassemblement des textes en se répartissant la charge de travail. Enfin, l'échange de corpus s'en trouve simplifié puisqu'il peut se dérouler sans encombrer le serveur mail qui parfois d'ailleurs refuse les pièces jointes un peu trop volumineuses.

Mais à un second niveau, l'outil permet une reconstruction de corpus grâce à des fonctions de recherche.

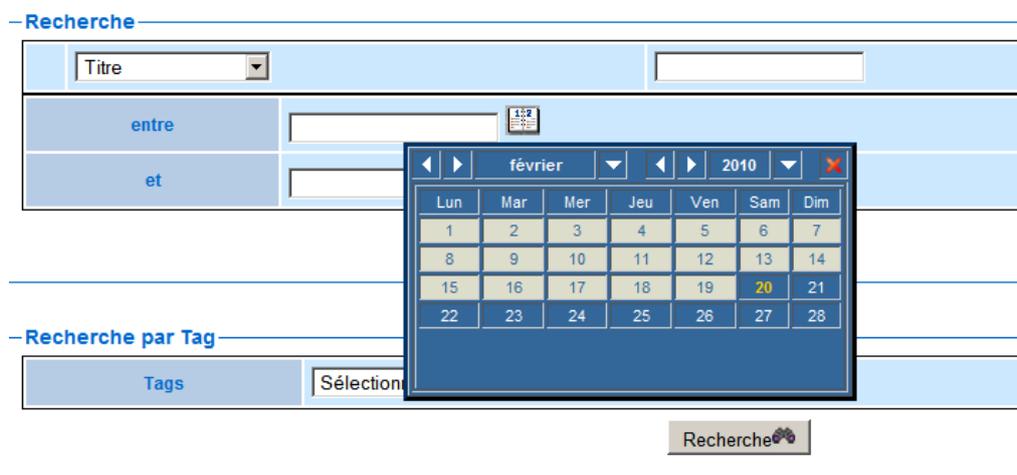


Figure 3 : Fenêtre permettant les recherches multicritères

Une fois les critères saisis, l'outil permet la génération du corpus (les textes, leurs contextes, les ressources associées, et le fichier PRC qui rassemble ces informations) et bien sûr son téléchargement.

Plusieurs usages peuvent être a priori distingués. D'une part, il est possible de faire des expériences sur les choix d'un collègue. Il prétend que les textes sur telle controverse, que les documents sur tel débat n'apporte rien. Dans la mesure où ces fichiers sont disponibles, pourquoi ne pas en faire l'expérience ? D'autre part, le dispositif permet d'étendre des corpus, de prolonger des réflexions. Prenons un exemple : les débats sur la pertinence de l'[électrostimulation](#) a conduit une partie des protagonistes à défendre l'idée que cette technique s'apparentait à du dopage (Collinet et Terral, 2006). Les sportifs en effet augmenteraient ainsi leurs performances en restant allongés et branchés sur ces appareils, ce qui est contraire à une valeur du sport qui exige que ce soit le travail et la force de caractère du champion qui soient consacrés. A la marge de cet argument qui ici vise surtout à défendre l'efficacité très controversée de l'électrostimulation, les collègues qui travaillent sur les questions de dopage peuvent facilement récupérer, non pas la totalité du corpus sur l'électrostimulation (ce qui n'aurait pas de sens) mais les textes qui abordent cette qualification – tout aussi discuté – de « dopage ».

Encore faut-il pouvoir repérer ces textes. Tous les champs – y compris les mots-clés saisis lors du dépôt – peuvent être mobilisés comme critère et une recherche *full text* est prévue. Mais lorsque l'on a travaillé sur la langue, on sait qu'il y a bien des manières de désigner une thématique sans utiliser les mots qui la désignent directement. Aussi est-il prévu d'implanter un champ « résumé » sur lequel on pourra bien sûr faire une recherche. Ces résumés seront générés automatiquement par Marlowe et mentionneront les principales propriétés du texte, tout en faisant référence au corpus dans lequel il faisait initialement sens et aux concepts dominants. On pourra alors utiliser le travail de codage des collègues pour effectuer des recherches.

2. Echanges de concepts

Encore faut-il connaître les stratégies d'indexation des différents chercheurs ! La variété des codages est une ressource (c'est même ce sur quoi la socio-informatique est fondée) mais s'apparente à un obstacle à partir du moment où il s'agit de travailler collectivement. Deux personnes peuvent créer des objets proches, oublier des représentants, coder des expressions, formaliser des recherches d'expression... et parfois les oublier eux-mêmes !

Sur le même principe que précédemment, l'outil CAMBRA (Concepts pour l'Analyse Multiple, Pragmatique et RAisonnée) visera à mettre à disposition :

- Les dictionnaires ;
- Les concepts
 - mais aussi des éléments des concepts (sous-concepts)...
 - et des représentants,
- Des formules ;
- Des configurations discursives.

Le cahier des charges et l'analyse des tables sont en cours. L'outil sera disponible dans quelques mois et fera l'objet d'un autre billet.

3. Echanges de raisonnements

Comme cela a été plusieurs fois souligné (Chateauraynaud 2003, Torny et Trabal, 2006), le logiciel Prospéro est moins conçu pour administrer directement une preuve via à une représentation unique du corpus (graphique, tableau, énoncé...) que comme un moyen de tracer des raisonnements qui mobilisent eux-mêmes plusieurs stratégies. Le logiciel [Marlowe](#), qui est venu doubler Prospéro, a été réalisé dans ce but : formaliser, expliciter et confronter plusieurs formes de raisonnements. Fondé sur cet investigateur sociologique, [Marloweb](#) mis en place dans le cadre d'un observatoire informatisé de veille sociologique, apparaît comme un complément décisif pour le travail collectif.

Dans le cadre du projet TRASCINTER, nous ne sommes pas partis d'un corpus unique, d'une série d'hypothèses fixées *a priori* suite à la définition d'une approche donnée. Les différentes équipes ont construit des corpus évolutifs placés dans un espace de variation puisqu'ils déploient des débats dans des arènes plus ou moins confinées (depuis des congrès de sociétés savantes jusqu'à la construction publique de « questions de société » sur les relations sport-santé). S'ils partagent une posture commune, les chercheurs impliqués dans le projet ont essayé de raisonner en prenant l'objet « interdisciplinarité » par plusieurs bouts.

Au moment où j'écris ces lignes (soit environ un an après le début du projet d'une durée de 3 ans), il est possible de recenser plusieurs stratégies pour repérer les disciplines et, du même coup, les formes de l'interdisciplinarité.

- Certains ont cherché à repérer une interdisciplinarité revendiquée. Parfois, dans un rapport de laboratoire, dans un article ou dans une communication, les auteurs-acteurs affirment que leur recherche est « pluri » ou « inter » voire « transdisciplinaire ». A d'autres moments, on repère des énumérations de disciplines. Cela invite à étudier ces manifestations, ces intentions ou ces promesses de cette interdisciplinarité ;
- D'autres collègues ont construit, grâce à des dispositifs d'enquête, des lexiques disciplinaires pointant tout aussi bien sur des objets, des concepts ou des méthodes. A ce titre, c'est la cohabitation de lexiques qui fait l'objet d'analyse ;
- On recense aussi des tentatives sur des sous-corpus disciplinaires, lesquelles semblent heuristiques notamment lorsqu'il s'agit de repérer de façon croisée les formes de présence des autres disciplines ;
- Des chercheurs se sont quant à eux employés à reconstruire des séquences significatives de collectifs à partir d'indices d'appartenance, de marques utilisées en bibliométrie, de trajectoires scientifiques...

Dans la mesure où d'autres tentatives sont en cours, directement liés au travail de terrain qui nous conduit à interroger des scientifiques pour leur faire expliciter leur ancrage disciplinaire, leur pratique de l'interdisciplinarité et les tensions imputables à la pluralité de disciplines dans les sciences du sport, il s'agit de considérer cette série de raisonnements comme une première tentative. Mais il est possible dès maintenant, d'envisager le partage de ces expériences de pensée.

On peut envisager de les formaliser et de réaliser quatre séries de scripts pour Marloweb :

- l'un pourrait répondre à des requêtes du type « Recherche toutes les formes d'interdisciplinarité revendiquée ». La machine pourrait alors travailler sur les énumérations de disciplines et sur les usages du lexique autour de l'interdisciplinarité ;
- un deuxième inviterait à travailler sur les sous-corpus disciplinaires en croisant les formes de présence des autres disciplines ;
- un troisième travaillerait les objets frontières, à l'instar du travail mené par Matthieu Delalandre dans sa thèse sur les sciences de la performance (Delalandre, 2009). Le script fonctionnerait sur la co-présence de lexiques disciplinaires et le croisement des réseaux des catégories pointant les différentes disciplines ;
- un quatrième viserait à repérer ce que partagent des communautés épistémiques en travaillant sur les façons dont circulent les objets, les dispositifs et les trajectoires.

Dans tous les cas, l'idée consiste à conserver les raisonnements et à les confronter à d'autres corpus. Expliciter, discuter, comparer. Des activités au cœur des pratiques scientifiques et que l'on peut envisager désormais de réaliser collectivement dans nos disciplines de sciences sociales.

Bibliographie

CHATEAURAYNAUD F., 2003, *Prospero : Une technologie littéraire pour les sciences humaines*, Paris, CNRS.

COLLINET C., TERRAL P., 2006, « Une controverse scientifico-technique dans le monde des sciences du sport : le cas de l'électrostimulation », *Sociologies contemporaines*, n°64, pp.67-87.

DELALANDRE M., 2009, *Sociologie des sciences de la performance sportive en France*, Thèse NR, Univ. Paris Est Marne la Vallée.

DEWEY J., 1993, *Logique – La théorie de l'enquête*, [1ère éd : 1939], Paris, PUF.

TORNY D., TRABAL P., 2006, « Le résumé de communication comme objet sociologique – Une analyse thématique, ontologique et littéraire à l'aide du logiciel Prospéro » in D. Demaziere, C. Brousseau, P. Trabal, K. Von Meter (Dir), *Analyses textuelles en sociologie*, Rennes, Presses Universitaires de Rennes, pp. 23-80.

1. Nous tenons à remercier Abdel Ramdani pour son travail. Au-delà de la qualité de l'outil qui reste encore en développement, nous tenons à exprimer le plaisir de travailler avec un informaticien qui s'efforce de saisir les contraintes des chercheurs.