

*LES CORPUS  
VUS PAR  
LE CNRS*

CORLI, Paris, 3 octobre 2019

# *Le plan*

L'institution face aux données  
(linguistiques)

La politique (linguistique) du CNRS  
depuis sa création

Les évolutions du CNRS

Que signifie le nom de CNRS par  
rapport aux corpus ?

Bilan et perspectives

---

*L'institution  
face aux  
données  
(linguistiques)*

---

La linguistique et ses données

La grammaire comparée des  
langues indo-européennes et les  
données philologiques : le  
*Corpus* de Berlin

La variation synchronique et la  
collecte des données : les  
enquêtes des frères Grimm



*L'institution  
face aux  
données  
(linguistiques)*

Un dispositif préexistant

Les collections des Bénédictins de Saint-Maur et les « corpus inscriptionum » de l'Académie des Inscriptions et Belles-Lettres

De l'ESLOV à l'Ecole des Chartes, de l'EPHE à l'EFEO, entre enseignement et recherche

(sans les universités)

---

*L'institution  
face aux  
données  
(linguistiques)*

---

Quelques tentatives

Les « missions » du Ministère de l'Instruction Publique et les politiques d'acquisition des institutions.

Les carnets d'enquête d'Edmont pour *l'Atlas Linguistique de la France* et les fiches lexicales de Mario Roques  
(mais l'Afrique ?)

*L'institution  
face aux  
données  
(linguistiques)*

Les lacunes du système :

Les données comme matériaux d'un chercheur (ex. Tesnière)

La gestion des ressources patrimoniales (ex. dépôt légal du son)

Les domaines sectorisés (l'I.E.S. et les langues slaves) ou non pris en compte (les langues amérindiennes)


---

*La politique  
(linguistique)  
du CNRS  
depuis sa  
création*

---

Après les tentatives du Front Populaire, la création du CNRS dans l'effort de reconstruction : des objectifs scientifiques et patriotiques

Deux directions initiales : la préservation du patrimoine (les données) et la reconduction de l'ALF (les « patois »)



*La politique  
(linguistique)  
du CNRS  
depuis sa  
création*

---

Le CNRS a été dès le départ conçu comme complémentaire : il devait assumer ce que les autres institutions, en particulier l'université et les grands établissements, ne pouvaient prendre en charge

L'exemple de la phonétique comme institut de la Sorbonne





*La politique  
(linguistique)  
du CNRS  
depuis sa  
création*

---

Deux initiatives pilotes :

*les Atlas Linguistiques et Ethnographiques de la France par Région*

L'irruption de l'oral

L'enregistrement conçu comme état préparatoire d'une cartographie

Une illustration : l'usage des bandes magnétiques



*La politique  
(linguistique)  
du CNRS  
depuis sa  
création*

---

L'Institut National de la Langue  
Française

Une vision gaullienne de la  
contribution de la linguistique à  
la politique

Une introduction de la mécano-  
graphie dans l'outillage

Un pilotage par le haut (Imbs)



# *Les évolutions du CNRS*

L'organisation centralisée de la  
recherche

La création des « équipes de  
recherche »

Les transformations de l'enseigne-  
ment supérieur et la mise en place des  
UFR (du statut de l'enseignant-  
chercheur à la mise en place des  
laboratoires)

---

# *Les évolutions du CNRS*

Les transformations sociales et institutionnelles, mais aussi le changement des pratiques scientifiques

L'irruption de l'informatique :

- dans la collecte (numérisation)
  - la gestion documentaire (catalogage)
  - les analyses (TAL)
  - la diffusion (Internet)
  - la valorisation...
-

# *Les évolutions du CNRS*

Un paysage recomposé

en plus de la présence des universités,  
des grands établissements et des  
centres de conservation (BnF,  
Archives Nationales, INA),

en plus d'autres institutions nationales  
de recherche (INSERM, INRIA,  
IN2P3...)

---

# *Les évolutions du CNRS*

l'État a ajouté différentes structures, plus ou moins pérennes, plus ou moins dotées, qui redessinent le périmètre d'exercice du CNRS :

l'ANR (cf. appels corpus)

les PIA (cf. Equipex Ortolang)

l'Alliance Athéna

---

# *Les évolutions du CNRS*

Prendre en compte également :

la « concurrence » internationale

les demandes européennes : la  
création de TGE en SHS en corres-  
pondance avec DARIAH et CLARIN

Le besoin de préserver les données  
selon un certain format : cf. Aix et  
Orléans

---

*Que signifie le  
nom de CNRS  
par rapport  
aux corpus ?*

---

On peut considérer la politique du CNRS par rapport aux corpus de quatre points de vue :

administratif et politique

technique et scientifique






*Que signifie le  
nom de CNRS  
par rapport  
aux corpus ?*

---

Le point de vue administratif

Le CNRS est un organisme dans lequel les SHS sont très présentes avec des spécificités :

- multiplicité des UMR
  - faiblesse des dotations récurrentes
  - modicité des moyens en personnels d'aide à la recherche
  - contraintes politiques (lutte contre la radicalisation, *gender studies*...)
- 


*Que signifie le  
nom de CNRS  
par rapport  
aux corpus ?*

---

Les structures CNRS

HumaNum (et Progedo)

La prise de conscience des conséquences d'une absence de politique de préservation des données de la recherche a abouti à la création de deux structures, ADONIS et IR-Corpus fusionnées dans Huma-Num qui se décline en consortiums



*Que signifie le  
nom de CNRS  
par rapport  
aux corpus ?*

Les fédérations de recherche (ILF et TUL), *De profundis...*

Les MSH comme organisme relais (où affecter les ingénieurs ? où installer les équipements ?)

Les UMR (et USR, UFREMI...)  
comme centres de production et  
d'analyse

---

*Que signifie le  
nom de CNRS  
par rapport  
aux corpus ?*

Scientifiquement, quelles sont  
aujourd'hui les forces et les faiblesses  
du travail sur corpus ?

Où en est-on dans le travail de recueil  
de données, de chaîne de traitement,  
de développement d'outils, de  
diffusion ?

Un exemple pour les jours qui  
viennent...

---

*Bilan et  
perspectives*

Une vision politique ? Le rôle de la  
section 34

Un exercice quinquennal : le rapport  
de conjoncture

La linguistique de corpus, *qu'ès aco* ?

Des propositions à avancer en  
provenance de la communauté, et à  
faire reconnaître (cf. le précédent de  
l'ANR).

---

# *Bilan et perspectives*

Quelques questions (et des réponses à  
contenir en 3.000 signes)

La place des corpus dans la recherche  
linguistique

L'interdisciplinarité : le TAL, les  
mathématiques, les autres SHS, le  
droit (la RGPD)...

Les modalités de l'archivage et de la  
diffusion

---

# *Bilan et perspectives*

Le contrôle de la valorisation

Les évaluations : qu'est-ce qu'un corpus de qualité ? Quels sont les critères à prendre en compte ?

Quelle comparaison internationale ?

Sur la nécessité de ne pas laisser la réponse à des intervenants externes

---