

# Rapport de stage

---

Dans le cadre de mon master à l'université Paris Dauphine, j'ai effectué un stage sous la direction de Guillaume Daudin du 1<sup>er</sup> avril au 29 août 2015. Ma mission s'intègre dans le cadre du projet TOFLIT18 qui vise à mieux comprendre "les transformations de l'économie française par le prisme du commerce international, 1716-1821". Au-delà de l'assistance que je devais fournir aux chercheurs travaillant sur ce projet, j'ai également eu pour tâche de constituer un mémoire sur l'évolution de la consommation de produits médicaux sur cette période. Ce thème a constitué le sujet principal de mon mémoire.

Dans ce présent rapport, je décris les organismes auxquels j'ai été rattaché, les missions que j'ai pu mener, et ce que ce stage m'a apporté personnellement.

## Plan du rapport

I.	Le DIAL et le projet TOFLIT18 .....	2
a.	Le DIAL.....	2
b.	Le projet TOFLIT18.....	2
II.	Les missions du stagiaire .....	4
a.	Le nettoyage et la normalisation des bases de données .....	4
b.	L'exploitation de la base de données (le mémoire) .....	6
c.	Les tâches annexes.....	7
	Conclusion .....	8
	Annexes .....	9
	Les règles temporaires de normalisation .....	9

## **I. Le DIAL et le projet TOFLIT18**

### **a. Le DIAL**

Mon stage s'est déroulé au sein de l'Université Paris Dauphine où j'étais rattaché au laboratoire de l'UMR DIAL 225 (Développement, Institutions et Mondialisation). Ce dernier concentre des dizaines de chercheurs travaillant sur différents thèmes relatifs à l'économie du développement et à l'économie internationale. Les objets d'études sont aussi variés que les espaces géographiques qu'ils couvrent. La majorité des études conduites portent sur la zone Afrique (Mali, Sénégal, Madagascar, Ghana, Côte d'Ivoire, Burkina Faso, pays du Maghreb), la zone Amérique du Sud (particulièrement le Pérou et le Brésil) et la zone Asie (Vietnam). On dégage deux grands axes d'étude qui sont les « Aspects sociaux et humains du développement », et « Mondialisation et économie politique du développement », ainsi qu'un pôle statistique plus transversal.

Le DIAL étant partenaire du projet TOFLIT18, décrit ci-dessous, j'ai été amené à travailler dans les mêmes locaux que les doctorants de l'université Paris Dauphine affiliés au laboratoire du DIAL. Cette cohabitation était pour moi un plongeon dans la vie des chercheurs et une expérience enrichissante pour quiconque souhaite s'y orienter par la suite. C'était également l'occasion pour moi de découvrir des domaines de recherches, notamment en économie du développement et en économie de la santé, et des méthodes de travail différentes de celles que j'expérimentais au jour le jour pendant ces 4 mois de stage.

### **b. Le projet TOFLIT18**

Le projet TOFLIT18, coordonné par Guillaume Daudin et Loïc Charles et financé par l'Agence nationale de la recherche, est conduit en partenariat par le Centre d'histoire de Sciences Po, le LEDa-UMR DIAL Laboratoire d'Économie de Dauphine, et l'UR11-INED Unité de Recherche Histoire et Population.

Les objectifs de ce projet sont décrits sur le site de l'ANR selon les termes suivants : « Le projet "Les transformations de l'économie française par le prisme du commerce international, 1716-1821" a pour objectif de mieux connaître l'évolution de l'économie française au cours de laquelle la France et l'Europe ont préparé leur entrée dans la révolution industrielle et l'ère moderne. La série des données du commerce extérieur français établies par l'administration est la plus complète et la plus cohérente série dont nous disposons sur l'économie française de cette époque. Ces données ont été

collectées et agrégées au niveau national par le Bureau de la Balance du Commerce à partir de 1716<sup>1</sup>. Au-delà des réorganisations administratives, les techniques de collecte et de présentation des statistiques sur le commerce extérieur français n'ont guère changé jusqu'aux années 1820.

Ces documents fournissent des informations d'un intérêt exceptionnel pour les historiens-économistes. Ils peuvent en effet être utilisés pour étudier les effets économiques du commerce international sur des échelles très différentes : au niveau de la France dans son ensemble, au niveau d'une seule région, au niveau d'un port et même au niveau des comportements microéconomiques des agents, par exemple les négociants. Ils peuvent également être employés pour mieux comprendre l'imbrication de, et les interactions entre, ces différentes logiques géographiques.

Le volume et la dispersion des sources primaires rendent la collecte et la retranscription des sources longues et onéreuses. Le groupe inclut donc des chercheurs ayant une grande expérience de la construction et de la gestion de grandes bases (notamment MARPROF, NAVIGOCORPUS, RICardo and SoundToll Registers Online). Leur expertise sera précieuse pour vérifier les données collectées par le projet en les comparant avec d'autres types d'informations sur le commerce (navigation et comptes marchands). La collaboration de chercheurs qui travaillent sur des bases similaires pour des partenaires importants de la France (la Grande-Bretagne et les Pays-Bas Autrichiens) permettra de réaliser des vérifications croisées et des études comparatives.

La base de données qui sera créée par le projet comprendra la valeur des flux de commerce bilatéraux (avec de 20 à 30 partenaires commerciaux) au niveau national entre 1716 et 1821, ainsi qu'une répartition de ces flux par marchandise (600-1000 marchandises différentes) à partir de 1750 au moins. Il inclura aussi les valeurs unitaires et les quantités de biens associés à ces flux entre 1771 et 1792, et des données de commerce au niveau régional (1716-1780). Cette base de données sera rendue accessible à la communauté des chercheurs à la fin du projet.

Les données collectées peuvent partiellement remplacer les séries macroéconomiques françaises inexistantes. Nous les utiliserons donc pour améliorer notre connaissance de l'économie française et des mécanismes économiques à laquelle cette dernière était soumise, au niveau national et au niveau régional. Nous privilégierons deux programmes de recherche et d'exploitation de ces données. D'une part, nous examinerons les évolutions de la spécialisation française, à la fois entre les régions françaises et par rapport au reste de l'Europe. Peut-on lier ces évolutions au développement économique contrasté de la France et de la Grande-Bretagne ? Que nous apprennent-elles sur les

---

<sup>1</sup> Loïc Charles and Guillaume Daudin, "La Collecte Du Chiffre Au XVIIIe Siècle: Le Bureau de La Balance Du Commerce et La Production Des Données Sur Le Commerce Extérieur de La France," *Revue D'histoire Moderne et Contemporaine* 58, no. 1 (2011): 128–55.

déterminants du commerce international ? D'autre part, nous étudierons les effets des chocs politiques sur l'économie française : la France a traversé plusieurs guerres et bouleversements politiques. Elle a connu aussi des changements radicaux de politique commerciale : mercantiliste en début de période, elle a initié une politique d'Exclusif mitigé avec son empire colonial dans les années 1760, signé des traités de commerce dans les années 1780, avant de se refermer après 1793. Quels ont été les effets de ces transformations des législations commerciales sur son économie ? »

Lorsque j'ai commencé à travailler sur ce projet, la base de données n'était pas encore complète. Si les données nationales avaient été intégrées dans leur quasi-totalité, ce n'était pas le cas des données locales. Pourtant, la base de données comportait déjà un nombre très important de flux de marchandises et le traitement de ces données était rendu impossible par sa taille. Il nous fallait donc ajouter de nouvelles données et en simplifier la lecture. Ce travail devait se faire en étroite coopération avec les développeurs informatiques en charge de la création d'une plateforme informatique d'échange des données. Mettre la base de données à disposition du plus grand nombre, professionnels comme amateurs, requiert de créer une interface commode d'utilisation sans en altérer le contenu via des simplifications abusives. Ces deux axes sont devenus les leitmotivs de mon stage : accroître le volume de données disponibles tout en les simplifiant.

C'est également à partir de ce projet et des données qui étaient ainsi mis à ma disposition, que je devais mener à bien mon étude sur la consommation de produits médicaux entre 1718 et 1839. Mon travail sur la base de données et mon travail de recherche, matérialisé par le mémoire, allaient être menés parallèlement pendant toute la durée de mon stage : les nouvelles données arrivant continuellement m'empêchaient toujours de finaliser définitivement mon travail de recherche.

## **II. Les missions du stagiaire**

### **a. Le nettoyage et la normalisation des bases de données**

Une partie importante de ma contribution au projet TOFLIT a été d'aider à la mise en place d'une base de données utilisable par le plus grand nombre. Le projet TOFLIT regroupe des sources très variées (locales, nationales) et dont la nomenclature peut varier au cours du temps. L'assemblage et la normalisation ne sont donc pas aisés. En particulier lorsqu'il s'agit de trouver le juste milieu entre la normalisation des données, permettant la simplification des données et la correction des erreurs typographiques ou orthographiques, et la conservation des données au plus proche des sources.

Pour cela il a donc fallu trouver un système de retranscription convaincant qui permet de conserver les données originales, mais également de permettre la simplification et l'agrégation des données provenant de différentes sources. En effet les sources peuvent contenir des erreurs de typographie mais aussi des variations selon les connaissances des commis chargés de la retranscription. Ainsi selon le savoir des uns en matière de toile, une même pièce pouvait être rapportée sous le nom de « Toile », « Toile diverse », « Toile de Bretagne »... Le système que nous avons retenu nous permet de conserver chaque étape du processus. Nous avons également utilisé GitHub pour garder une trace des modifications effectuées et pour simplifier le travail d'équipe.

La retranscription : Elle se fait à partir des originaux. La retranscription a été menée à la fois par les équipes du projet et par un prestataire. Il s'agit uniquement de retranscrire toutes les informations issues des livres du Bureau de la balance du commerce (ou des autres offices produisant des statistiques) en format Excel exploitable informatiquement. Cette retranscription ne doit souffrir d'aucune modification. Cependant certaines erreurs de lecture ou de recopie sont inévitables du fait des sources parfois difficilement lisibles.

La normalisation : A partir de ces sources, nous devons obtenir des données permettant une analyse rapide en fonction des sujets de recherche. Nous avons donc entrepris une normalisation des sources pour obtenir un format unique. Nous avons également entrepris de corriger les dénominations des biens, pays et unité de mesure pour essayer de regrouper les dénominations identiques. Nous avons donc établis plusieurs étapes. La première concerne les corrections orthographiques. La question est de savoir à partir de quand est on en mesure d'affirmer que la différence entre deux termes provient d'une erreur orthographique ou non. Pour cela il a été nécessaire d'édicter des règles compréhensibles et standardisées. Parmi celles-ci, le traitement des pluriels, les corrections des termes latins, etc... Plus de détails peuvent être trouvés en annexes où figurent les règles temporaires de normalisation. Puis nous pouvons procéder à une simplification des dénominations pour obtenir des groupes de marchandises regroupées en fonction de leurs dénominateurs communs. Chacun doit pouvoir naviguer entre les différentes classifications. Il est donc toujours possible de partir d'une forme complexe pour arriver à une forme simplifiée et vice et versa.

Au cours de mon stage j'ai eu l'occasion de participer à la fois à la retranscription et à la normalisation. Néanmoins j'ai été beaucoup plus actif au niveau de la normalisation puisque, avec Pierre Gervais, historien et membre du projet TOFLIT, nous avons édicté les règles de normalisation. De plus cette étape était primordiale pour la réalisation de mon mémoire, puisque comme on le verra par la suite, j'ai eu besoin d'identifier tous les produits utilisés dans la pharmacopée française

de l'époque. Il m'était donc nécessaire procéder à une première normalisation pour pouvoir mener mon projet à bien.

## **b. L'exploitation de la base de données (le mémoire)**

L'objectif de mon stage, outre de découvrir l'environnement de la recherche, était d'utiliser la base de données TOFLIT18 pour produire une analyse du commerce des produits médicaux entre 1718 et 1839. Cette analyse devait se baser sur le travail antérieur de Patrick Wallis<sup>2</sup> qui a étudié la consommation de drogues médicales en Angleterre entre 1550 et 1800. La méthode consiste à appréhender l'évolution de la consommation de produits médicaux à travers les données de commerce obtenues à partir des registres des principaux ports du pays. Notre exhaustive base de données nous permettait de mener une analyse similaire, en approfondissant encore la méthode puisque nous avons en notre possession les prix et les quantités en plus de la valeur des flux de marchandises échangées.

Pour parvenir à concrétiser notre étude, j'ai tout d'abord lu plusieurs papiers et documents sur l'état de la médecine à cette époque. Il était primordial de bien comprendre le fonctionnement de la médecine pour bien identifier les produits considérés comme médicaux. Comme relaté dans le mémoire, cette tâche s'est avérée plus ardue que prévu du fait de la diversité des pratiques médicales de l'époque. L'art de la médecine n'est pas encore tout à fait institutionnalisé et évolue très vite au cours de cette période. Il coexiste donc des pratiques médicales très traditionnelles basées davantage sur des croyances irrationnelles et infondées, et des pratiques dérivant d'une méthode scientifique et de l'observation. Plus les produits sont nombreux et plus le risque de ne pas prendre en compte l'ensemble des produits médicaux est important. En effet certains produits sont utilisés dans la pharmacopée mais aussi dans d'autres secteurs. Il faut donc sélectionner ceux dont l'évolution du commerce traduit une évolution de la consommation de produits issus de la pharmacopée. Cette étape est primordiale pour la suite de l'étude. La méthode complète est décrite dans le mémoire, notamment les critères de sélection sur les produits les plus controversés. Pour m'assurer de la cohérence de la liste obtenue, j'ai également contacté Christelle Rabier<sup>3</sup>, professeure

---

<sup>2</sup> Patrick Wallis, "Exotic Drugs and English Medicine: England's Drug Trade, C. 1550–c. 1800," *Social History of Medicine* 25, no. 1 (2012): 20–46.

<sup>3</sup> Christelle Rabier, "L'histoire de La Médecine Au Prisme Du Marché: Perspectives Britanniques," *Recherches Britanniques* 1, no. 1 (2011): 17–36;

Christelle Rabier, "La Disparition Du Barbier Chirurgical," 2010;

Christelle Rabier, "Une Révolution Médicale? Dynamiques Des Professions de Santé Entre Révolution et Empire" (*Annales historiques de la Révolution française*, Armand Colin, Société des études robespierristes, 2010), 141–59.

à l'EHESS et spécialiste du sujet, ainsi que Patrick Wallis<sup>4</sup>, dont l'étude s'inspire de son papier, et qui figure également parmi les spécialistes du sujet à la London School of Economics. J'ai pu avec leurs conseils établir une liste plus adaptée au contexte et éviter des simplifications excessives.

Une fois cette étape franchie, un travail de normalisation a été nécessaire pour agglomérer les produits et identifier ceux assimilés à la pharmacopée. Ce travail se rapproche de celui que nous avons décrit dans la section précédente traitant du nettoyage de la base de données. Il permet d'obtenir une liste de produits dont nous étudierons les flux d'import-export sur la période. Nous normalisons également les unités utilisées, les pays, etc... La mise en place et la normalisation des données ont été les deux étapes les plus longues du fait de la quantité de données à gérer. En effet la base de données contient plus de 300.000 flux et près de 35.000 marchandises dont le nom est unique.

Une fois les données prêtes à être utilisées, nous les avons transférées sur STATA pour mener une analyse économétrique. En nous basant sur la méthode des ventes répétées, utilisée initialement pour analyser l'évolution des prix de l'immobilier, nous avons calculé les taux de croissance des importations/exportations en valeurs et en quantités ainsi que l'évolution des prix sur la période.

Enfin la dernière étape a été la rédaction du mémoire en anglais pour pouvoir le partager avec Patrick Wallis et Christelle Rabier. J'ai également présenté mes résultats, alors très incomplets, lors d'une réunion du projet TOFLIT18. En effet le 5 juin s'est tenue une réunion au cours de laquelle j'ai été amené à présenter pendant une heure mes résultats et les problèmes méthodologiques auxquels on peut se trouver confronter lorsqu'on utilise la base de données TOFLIT.

### **c. Les tâches annexes**

En parallèle aux activités principales de mon stage, j'ai également accompli des tâches annexes de manière récurrente ou occasionnelle.

---

<sup>4</sup> Patrick Wallis, "Consumption, Retailing and Medicine in Early-Modern London," *The Economic History Review*, no. 61 (January 2, 2008): pp 26–53, doi:10.1111/j.1468-0289.2007.00391.x;  
Patrick Wallis, "Book Reviews: The Dying and the Doctors: The Medical Revolution in Seventeenth-Century England," *Medical History* 54, no. 4 (October 2010): 538–40;  
Patrick Wallis, "Exotic Drugs and English Medicine: England's Drug Trade, C. 1550–c. 1800."

- Les réunions TOFLIT : Chaque premier jeudi du mois se sont tenues les réunions TOFLIT qui permettaient à chacun des membres du projet de se tenir informé des avancées du projet. A l'issue de chaque réunion, j'avais la charge d'en rédiger les comptes rendus. Pendant ces réunions, je devais également faire mes retours aux développeurs de mes expériences d'utilisation de la base de données. Ceci devait permettre de mettre au point le système le plus efficient possible pour les utilisateurs de la base.
- Organisation du colloque DIAL : Le 2 et 3 juillet 2015 s'est tenu le troisième colloque international en économie du développement à Paris. Il a permis de créer un espace de discussion des innovations théoriques et empiriques en économie du développement avec un focus particulier sur l'analyse des facteurs qui permettent de dépasser les contraintes au développement. Les stagiaires du DIAL, dont je faisais partie, ont participé à cet événement en participant à l'organisation du colloque : accueil des participants, préparation des salles, etc...

## Conclusion

Avec le recul, je peux affirmer que ce stage a été une formidable opportunité de connaître d'avantage le monde particulier de la recherche. En étant impliqué sur toutes les étapes du projet, de la création de la base de données à la rédaction d'un rapport en passant par le traitement des données et l'analyse économétrique, j'ai pu avoir un aperçu global des problèmes auxquels les chercheurs font faces quotidiennement. J'ai notamment beaucoup appris au niveau organisationnel et méthodique. Le travail sur la base de données avec l'obligation de pouvoir en justifier chacune des étapes, et la concertation avec d'autres chercheurs pour mettre au point les meilleures méthodes d'exploitation des données sont des contraintes qui obligent le chercheur à travailler rigoureusement. L'environnement dans lequel j'ai travaillé était également très stimulant puisqu'entouré par des doctorants, plus ou moins avancés dans leur doctorat, qui partageaient avec moins leurs expériences et les difficultés qu'ils pouvaient rencontrer.

En me laissant libre de mes recherches, M. Daudin m'a poussé à prendre des initiatives et à entrer en contact avec d'autres chercheurs. Ces contacts m'ont été très bénéfiques puisque au-delà de leurs conseils, j'ai pu comprendre certains différents qui opposent les chercheurs en histoire économique. Particulièrement l'opposition entre une approche très historique et une approche plus



quantitative, plus proche de la science économique telle que j'ai pu l'étudier lors de mes années de formation à l'université Paris-Dauphine. Par ailleurs les discussions avec Christelle Rabier ont été assez prometteuses pour qu'elle m'a proposé de participer à un séminaire sur « la Santé au prisme des sciences sociales » qu'elle co-organise. C'est avec beaucoup de plaisir que j'ai vécu cette mince immersion dans les débats méthodologiques que peuvent avoir entre eux les chercheurs.

Pour conclure, au-delà des compétences pratiques que ce stage a pu m'apporter, je suis surtout reconnaissant des personnes que j'ai pu rencontrer : tous les membres du projet et les personnes avec qui j'ai pu travailler ont des profils très variés, desquels j'ai pu beaucoup apprendre.

## Annexes

### Les règles temporaires de normalisation

#### REGLES DE TRAITEMENT

- I. marchandises normalisées orthographiques: traitement orthographique => réduction

**00.** tout items séparés par points virgules (un item = un nom ou un adjectif, "agaric", "brut", "préparé" = 3 items)

*agaric amadouvier => agaric ; amadouvier*

**00.a** articles non séparés (et de, le, etc.) et placés après point virgule

*agoutals de bois => agoutals ; de bois*

**00.b** distributifs non séparés (espèce de, sorte de, etc.)

*akia espèce de salade => akia ; espèce de salade*

**00.c** formes adjectivales peuvent être unifiées (en monnaie, monnayé); formes la plus fréquente, sinon la première occurrence

*argent monnayé => argent ; monnayé*

*argent en monnaie => argent ; monnayé*

**00.d** ne pas séparer les adjectifs d'un substantif utilisé comme qualificatif de leur substantif, ou les substantifs compléments d'un verbe utilisé comme qualificatif de leur verbe

*Ambrain à grande barrique => ambrain ; à grande barrique;*

*apprêtée à relier les livres => apprêtée ; à relier les livres*

**00.e** ôter les parenthèses

**01.** Singulier/pluriel

**01.a** modification de singulier en pluriel ou vice versa uniquement si combinaison d'items apparaît dans la base au singulier et au pluriel, en privilégiant la forme la plus fréquente

*ail => ails [ails plus fréquent]*

NB: Si 50/50, prendre le singulier

NB: Ne pas modifier si combinaison est unique

*ail blanc à casser => ail ; blanc ; à casser*

*aux à fauxhes => aux ; à fauxhes*

**01.b** uniformisation du singulier/pluriel en fonction de l'item au n° d'ordre le plus bas (arrivé en premier) quand conflit nom/adjectif

cas hypothétique : *ail blancs => ail ; blanc, aux blanc => aux blancs*

**02.** tout en minuscules sauf noms propres [=> liste des noms propres]

**03.** Unification des orthographes: unifier uniquement si

**03.a** Différence vient d'une différence singulier/pluriel attestée dans les dictionnaires

*ails = aulx*

**03.b** nombre de syllabe identique (e muet compté comme une syllabe) et phonèmes identiques dans l'ordre sur les deux occurrences

*acier = assié*

**03.c** phonèmes peuvent être considérées comme identiques si les deux conditions

03.c.1 ET 03.c.2 sont remplies

**03.c.1** consonnes de même type d'une des deux façons suivantes

- soit phonétiquement (occlusives vs fricatives.)

*alquifoux = arquifoux*

- soit proches par la forme de la lettre

*alpargalle = alpargatte*

**03.c.2** voyelle identique (hors diphtongue)

*aracq = arak, alpargalle = alpargatte*

**03.d** phonèmes finales considérées comme identiques si terminaison latine / terminaison moderne (e / us)

*antimoine = antimonium*

**03.e** seule différence = placement des espaces

*alqui foux = alquifoux*

**03.c** Pour produits inconnus, première recherche rapide dans les principaux dictionnaires

**04.** Unifier ??? et autres expressions de doute en "illisible"

??? brun => illisible ; brun

**05.** Si recherche dans sources primaires sur nom, noter référence ET définition

**06.** A l'occasion, essayer de rechercher dans le pdf d'origine une erreur de transcription (ad lib)

**07.** Une définition peut servir de support à une modification d'écriture à condition de respecter au moins l'une des conditions en 03

**08.** Moderniser l'orthographe chaque fois que cela est possible

*bierre d'angleterre => bière ; d'Angleterre*