

« BASE MONTESQUIEU » LES DONNÉES DU COMMERCE EXTÉRIEUR FRANÇAIS DE 1836 À 1938

Stéphane Becuwe, Bertrand Blancheton, Karine Onfroy¹

Université de Bordeaux, GREThA - UMR CNRS 5113

Cet article présente l'état d'avancement d'une base de données du commerce extérieur français dite « Base Montesquieu » centrée sur les produits. Elle regroupe les données de commerce extérieur contenues dans la série annuelle intitulée « Tableau général du commerce de la France avec ses colonies et les puissances étrangères sur la période 1836-1938 ». L'article présente les données et discute leur robustesse. Il analyse aussi les enjeux associés à la numérisation et la transformation de cette source en base de données, afin de les mettre à disposition des chercheurs.

Mots clés : commerce extérieur, humanités digitales, France.

La recherche en sciences humaines et sociales est, depuis une dizaine d'années, profondément structurée et renouvelée grâce aux humanités numériques. Pour les chercheurs, l'accès à des plateformes d'archives numériques marque le début d'une nouvelle approche des sources historiques et offre des possibilités d'exploitation multipliées fondées sur un travail collaboratif et plus aisément interdisciplinaire.

¹ Les auteurs remercient les trois rapporteurs anonymes de la revue pour leurs suggestions et remarques. Ils remercient Katherine Barbieri, Léo Charles, Loïc Charles, Guillaume Daudin, Béatrice Dedingier, Claude Diebolt, Jean-Pierre Dormois, Giovanni Federico, Alexandre Fernandez, Frédérique Flamerie de Lachapelle, Michel Fouquin, Adrien François, Claire-Lise Gauvain, Michael Huberman, Jules Hugot, Markus Lampe, Sylvia Marzagalli, Samuel Maveyraud, Chris Meissner, Nikolai Nenovski, Kim Osterlink, Antonio Tena-Junguito et Romain Wenz pour leurs appuis et commentaires dans le cadre de ce projet de constitution de base de données.

Cette recherche a été soutenue par la Région Nouvelle Aquitaine AAP 2016, « Ports et développement des hinterlands. Humanités numériques et temps long (1836-2015) » (Maison des Sciences de l'Homme d'Aquitaine, resp : Bertrand Blancheton).

En histoire économique quantitative, la constitution de bases de données désagrégées a permis de nombreuses avancées. Le projet DFIH coordonné, par P.-C. Hautcoeur, collecte et emmagasine, à très large échelle, des données financières (voir Ducros *et al.*, 2018). Plusieurs équipes françaises travaillent actuellement à la constitution de ressources numériques sur le commerce extérieur français en histoire moderne et contemporaine. Le projet « Transformations of the French Economy through the Lens of International Trade, 1716-1821 » (ANR TOFLIT18 <https://toflit18.hypotheses.org>) porté par L. Charles et G. Daudin (2015) offre des chiffres de la balance de commerce, une analyse interactive des flux et des réseaux maritimes nationaux et européens. Le projet navigocorpus de S. Marzagalli, P. Pourchasse et J.-P. Dedieu propose, lui aussi, des données sur le commerce maritime aux XVIII^e et XIX^e à partir des affrètements (capitaines, itinéraires, cargaisons). Dans le prolongement des données de commerce bilatéral rassemblées par K. Barbiéri et son équipe pour 1870-2014 (www.correlatesofwar.org), le projet *RICardo* lancé en 2007 à Sciences Po Paris a abouti, après dix années de travail sous l'impulsion de B. Dédinger, à offrir une base très complète de données commerciales bilatérales à l'échelle mondiale du XIX^e au XXI^e siècle. Une base voisine – nommée TRADEHIST accessible au CEPII (http://www.cepii.fr/CEPII/fr/bdd_modele/presentation.asp?id=32) – a été constituée par M. Fouquin et J. Hugo (2016) en vue d'étudier la dynamique des coûts de commerce et le rôle de la distance dans la régionalisation du commerce mondial en longue période. A. Tena et G. Federico (2016) ont – eux aussi – constitué en concertation avec les équipes françaises une base de données du commerce mondial centrée sur la dimension pays (Federico-Tena World Trade Historical Series). Au plan international, l'équipe de l'université Humboldt à Berlin, termine la constitution d'une base du commerce extérieur allemand à partir de 1880. Les initiatives se multiplient au service d'une recherche coopérative et partagée.

Dans cette perspective, notre contribution vise à présenter le travail de création d'une base de données du commerce extérieur de la France (centrée sur la dimension produit) entrepris par l'équipe d'histoire économique de l'Université de Bordeaux. Cette base est intitulée « Base Montesquieu » en référence aux thèses de l'écrivain philosophe bordelais sur le doux commerce². Par extension, nous analysons les

2 Montesquieu est favorable au développement du commerce international pour des raisons d'abord morales, « l'histoire du commerce est, dit-il, celle de la communication des peuples », avant d'ajouter que « l'effet naturel du commerce est de porter à la paix. » (*De l'esprit des lois*, livre XXI, t. 2, p. 604, Paris, Gallimard, Bibliothèque de la Pléiade, 1949 pour le tome 1, 1951 pour le tome 2).

enjeux associés à l'émergence des humanités numériques en histoire économique (interopérabilité des bases, émergence d'une recherche collaborative, promotion de l'interdisciplinarité avec les sciences exactes, etc.).

Pour la période 1836-1938, le projet « Montesquieu » s'appuie sur la numérisation d'une publication annuelle de l'administration des douanes, le « Tableau général du commerce de la France avec ses colonies et les puissances étrangères » par la suite désigné par « Tableau » (qui devient le « Tableau général du commerce et de la navigation après 1896 »). Des publications annuelles sont disponibles antérieurement (notamment à partir de 1827) mais elles proposent des nomenclatures produits différentes de celles, plus homogènes, de la période 1836-1938 (voir Asselain et Blancheton, 2005). Aussi la « Base Montesquieu » débute-t-elle en 1836. Le commerce extérieur français est ensuite bien connu pour l'après Seconde Guerre mondiale, les données du FMI sur le commerce bilatéral sont disponibles en ligne (dans la base DOT (*Direction Of Trade*)) et les échanges par produit facilement accessibles aux normes SITC jusqu'à un très haut niveau de désagrégation (Digit-5) dans COMTRADE de l'*International Trade Center*.

Les données de commerce extérieur intéressent aujourd'hui non seulement les historiens et les économistes mais aussi les géographes, les politistes, les écologues, les physiciens en quête de séries longues sur les quantités échangées entre pays afin d'évaluer, par exemple, les dynamiques de long terme des émissions et des circulations de CO₂.

Notre démarche se déroule en quatre temps. Dans une première section, nous indiquons la nature des données disponibles au sein de la source. La section 2 analyse les évolutions touchant les séries disponibles, la motivation de certains choix des autorités françaises et les biais associés à l'interprétation des données. La section 3 présente deux bases partielles, l'une constituée à l'aide des résumés rétrospectifs et l'autre consacrée au trafic portuaire. La section 4 présente l'aboutissement du projet *via* la constitution d'une base du commerce bilatéral pays/produits à un niveau de désagrégation très fin et les enjeux associés à ce processus de numérisation (interopérabilité des bases, thesaurus produits, ...).

1. Les dimensions structurelles des Tableaux du commerce

Les Tableaux numérisés pour la période 1836-1938 offrent six entrées principales pour rechercher les statistiques du commerce extérieur de la France (voir graphique 1).

1.1. Dimension géographique

Les valeurs des flux en direction et en provenance des partenaires commerciaux sont disponibles à plusieurs niveaux d'agrégation. Ici d'emblée la question de la nomenclature évolutive des partenaires (entités) se pose (processus d'intégration politique de certaines zones, changement de noms de pays, ...). Les partenaires se rapportent à un pays, une région ou un groupe de pays. Ils sont identifiés tout au long de la période en fonction de cette typologie mais aussi selon le continent auxquels ils appartiennent. Ces entités ont été définies d'après la liste des « Entités RIC » du projet RICardo <http://ricardo.medialab.sciences-po.fr> qui s'est appuyée sur la liste des pays de *Correlates of War* (correlatesofwar.org). De nouveaux noms de partenaires ont aussi été créés. À titre d'illustration, le nombre de partenaires commerciaux recensés est de 53 en 1840, 55 en 1858, 54 en 1893, 79 en 1925 et 101 en 1938.

1.2. Dimension temporelle

Les données sont annuelles, disponibles même durant les périodes au cours desquelles la France est en guerre (1870-1871, 1914-1918) y compris pour le niveau de désagrégation-produit le plus fin.

1.3. Dimension produit

Les produits échangés sont proposés à plusieurs niveaux de désagrégation et de nomenclatures plus ou moins fines. Des synthèses rétrospectives regroupent les principaux produits en 107 catégories pour les importations et en 135 catégories pour les exportations. Ce niveau de désagrégation se situe entre les 67 modalités du niveau 2 de la SITC rev 3 de 1988 et les 261 modalités du niveau 3.

Ensuite au niveau bilatéral, des flux de quantités sont disponibles pour les « sous-produits ». Une valeur totale par sous-produit apparaît (elle est issue d'une multiplication par les contemporains d'une valeur unitaire moyenne par la somme des quantités en provenance ou à destination de l'ensemble des pays). Au niveau des importations, le nombre de catégories est égal à celui du nombre de sections dans le

tarif général. Une correspondance peut être établie ici avec le niveau 4 ou 5 de la SITC rev 3 (voir Annexe 4 pour un tableau de correspondance dans le fichier excel en ligne attaché à l'article). En 1850, par exemple, on recense 875 produits en importations et 1132 en exportations. Ainsi le poste « tissus de soie » figurant dans la synthèse rétrospective, se décline en 30 sous-produits en 1850, 27 en 1913 et 68 en 1938 (Becuwe et Blancheton, 2019).

1.4. Nature et mesure des flux

Les flux des produits précédemment cités sont ventilés soit en exportations, soit en importations en termes de valeurs et de quantités physiques sauf lorsqu'il s'agit de valeurs déclarées. Les différentes unités de mesure utilisées dans les tableaux sont : tête ; ruche ; franc, mille francs ; mille ; mille mètres ; kilogramme ; gramme ; quintal métrique ; stère ; pièce ; mètre ; litre ; hectolitre ; tonne ; hectogramme ; mètre carré ; douzaine ; paire ; douzaine de paires ; tonnage de jauge ; tonnage de déplacement.

1.5. Les droits de douane

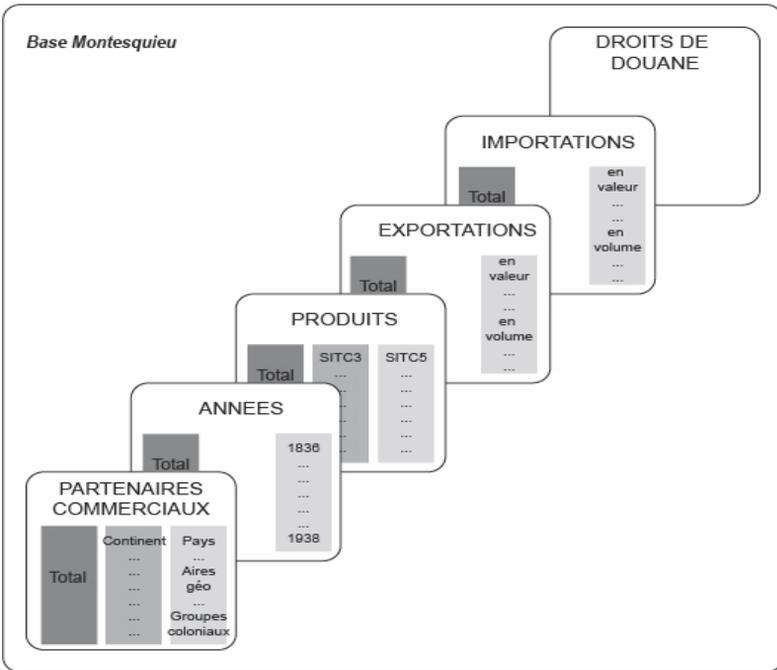
Pour les importations par pays de provenance les droits de douane perçus sont disponibles aux niveaux 2 ou 3 de la STCI (résumés rétrospectifs). Les droits perçus sur les importations de niveau 4 ou 5 sont accessibles mais non ventilés par pays de provenance (droits totaux tous pays confondus pour un sous-produit). Sur cette base, même si sur la période les tableaux entrées-sorties de comptabilité nationale ne sont pas disponibles, des études en termes de protection effective sont possibles grâce à ces données. Il s'agirait d'étudier l'évolution des droits pour les intrants aux différents stades du processus de production. Jusqu'en 1865 (inclus), des droits de sortie sont appliqués sur les exportations et sont disponibles à tous les niveaux d'agrégation.

1.6. Commerce général et commerce spécial

La confrontation aux statistiques historiques de commerce extérieur pose des questions traditionnelles sur l'identification de la nature des flux et l'évolution des pratiques quant à leurs valorisations (voir Lampe, 2008, pour une mise en perspective de ces enjeux à l'échelle de l'Europe).

Deux séries de commerce extérieur sont disponibles : commerce général et commerce spécial. Le « commerce spécial » intègre les flux réellement destinés à la consommation nationale. Le « commerce

Graphique 1. La Base Montesquieu – une structure multidimensionnelle

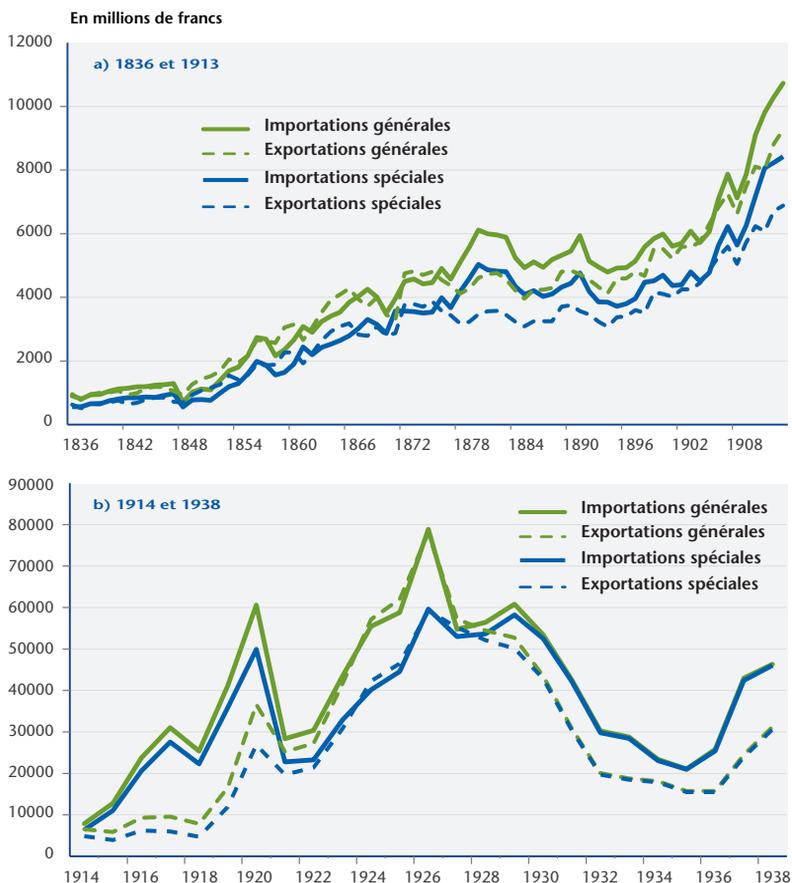


général » inclut, quant à lui, le trafic en transit³. Pour la France, le transit se caractérise par une admission temporaire sur le territoire des marchandises ne dépassant pas en général une durée de trois mois. Le graphique 2 propose une comparaison de l'importance des écarts entre commerce général et commerce spécial pour les importations et les exportations (pour des raisons d'échelle nous dissociions la période 1836-1913 de la période 1914-1938 en proie à l'instabilité monétaire ; cf. Blancheton, 2016). Au niveau macroéconomique, le transit représente en moyenne un quart du commerce international de la France. Des comparaisons internationales systématiques seraient intéressantes. Les travaux de Lampe (2008) suggèrent qu'au milieu du XIX^e la Grande-Bretagne constitue une place forte du transit pour les « commodities ». Certains pays, à l'instar de la Suisse fin XIX^e, manifestent le souhait de développer ce commerce très profitable (Charles, 2016). Une piste de recherche consiste à étudier la

3 Pour plus de détails sur les définitions et méthodes de la statistique douanière, se référer à la thèse de doctorat de Charles-Jacques Schuller, « La statistique douanière et ses méthodes » (1911), éd. par la Faculté de droit de l'Université de Paris : <http://gallica.bnf.fr/ark:/12148/bpt6k74224x>

dynamique et l'importance du transit suivant la position géographique du pays, l'importance de sa flotte maritime, la présence d'entrepôts et les règles de définition des pays d'origine et de provenance des importations (voir Annexe 1).

Graphique 2. Comparaison entre commerce spécial et commerce général pour la France au niveau global entre 1836 et 1913 et entre 1914 et 1938



Du côté des flux exportés de France des biais apparaissent. Avec l'abandon des droits de sortie (en 1866), les douanes n'ont plus intérêt à contrôler la valeur des exportations déclarées par les commerçants. Les marchands peuvent avoir intérêt à sous-estimer le poids et la valeur des produits exportés : la déclaration d'exportation peut servir de base au calcul des droits perçus dans le pays de destination mais aussi de base de calcul du prix du fret par les compagnies de transport.

De même les exportateurs peuvent déclarer à la sortie des flux du commerce spécial alors que les produits sont en réalité en transit. Les formalités pour les flux en transit sont, en effet, plus lourdes. Ces pratiques pourraient être à l'origine d'écarts significatifs entre les flux d'exportations de commerce spécial de la France vers un partenaire et les valeurs des flux importés de France par ce même partenaire toujours pour le commerce spécial. Pour éclairer ces possibles biais, une étude systématique des flux miroirs devrait être entreprise avec plusieurs partenaires commerciaux.

2. L'évolution des données disponibles dans les Tableaux

2.1. La genèse : le besoin de nouveaux chiffres commerciaux au début du XIX^e

Au début du XIX^e siècle, les administrations des douanes fournissaient des données quantitatives sur les poids des marchandises exportées et importées. La valorisation posait des problèmes vraisemblablement trop difficiles à traiter (correction des sous-estimations par les importateurs, voire les exportateurs).

Dans les années 1820, la volonté de constituer systématiquement des données commerciales associant quantités et valeurs, à l'échelle du pays, semble répondre à deux objectifs ambitieux de l'État Français : mieux contrôler les flux entrants et disposer de chiffres pour fonder les dispositifs de politique commerciale envers les partenaires étrangers dans le cadre d'une vision mercantiliste du commerce extérieur.

Déjà la loi du 28 avril 1816, qui entend lutter contre la contrebande, oblige les importateurs à déclarer la provenance, le poids et la valeur des marchandises. En 1823, le gouvernement décide la publication de statistiques annuelles du commerce extérieur. La mission de collecte et de consolidation de l'information est confiée à l'administration des douanes (Schuller, 1911).

Charles X considère que l'évaluation des marchandises exportées et importées doit servir de fondement à la législation destinée à régler les relations commerciales de la France avec l'étranger (*Dictionnaire Universel théorique et pratique du commerce et de la navigation*, article Commission permanente des valeurs en douanes, 1861, p.750). Il charge, par l'ordonnance du 11 janvier 1826, le Président du bureau de commerce et des colonies d'ouvrir une enquête spéciale (39 personnes composent cette commission d'enquête). Face au problème

de la valorisation des quantités, l'enquête conclue en mars 1827 que les valeurs moyennes serviront de base à l'évaluation. Elles prennent le nom de valeurs officielles. L'exactitude des valeurs officielles (fondées sur des prix qui n'évoluent pas) est contestée dès les années suivantes mais sans que les pratiques évoluent. Sur ces bases, il convient de prendre les chiffres de la période 1827-1846 avec prudence. Ils sont fondés sur un système de prix qui ne s'est pas adapté à l'environnement productif (gains de productivité différenciés, ...). Par contre, on peut les considérer comme des séries à prix constants (en volume).

2.2. Les changements en matière de valorisation : valeurs officielles et valeurs actuelles

Le besoin de comparer les valeurs officielles avec les valeurs déclarées apparaît de plus en plus criant pour se faire une idée de la dynamique des prix. Léon Foucher à la Chambre des députés lors de la séance du 10 juillet 1847 formule explicitement cette demande. À l'époque pourtant, la France ne semble pas en retard sur la question de la qualité des données commerciales à l'échelle internationale. Seules six autres puissances publient des statistiques annuelles : la Belgique, les États-Unis, l'Autriche, la Russie, l'Association allemande et l'Angleterre. La Belgique⁴ publie depuis 1833 des valeurs officielles (actualisées une fois par une enquête administrative), l'Association allemande publie des données de quantités (non valorisées), l'Autriche et les États-Unis se fondent sur les valeurs déclarées à l'entrée. Pour l'Angleterre les *Statistical Tables* proposent des valeurs actuelles à l'exportation fondées sur les déclarations et des valeurs officielles à l'importation (au niveau global et pas pour toutes les marchandises). Lampe (2008) souligne la grande qualité des statistiques britanniques et attire l'attention sur la distinction proposée dans ces sources entre exportations de produits domestiques et réexportations de produits. Des gains de réexportations très significatifs justifient le maintien du contrôle du commerce international maritime par les Britanniques.

Sur ces bases, le 8 novembre 1847, le directeur des douanes propose au Ministre des Finances, de nouveaux grands principes pour l'évolution des flux commerciaux : le maintien des valeurs officielles pour assurer la continuité mais l'inscription chaque année de prix actuels. Il propose aussi la création d'une commission spéciale chargée

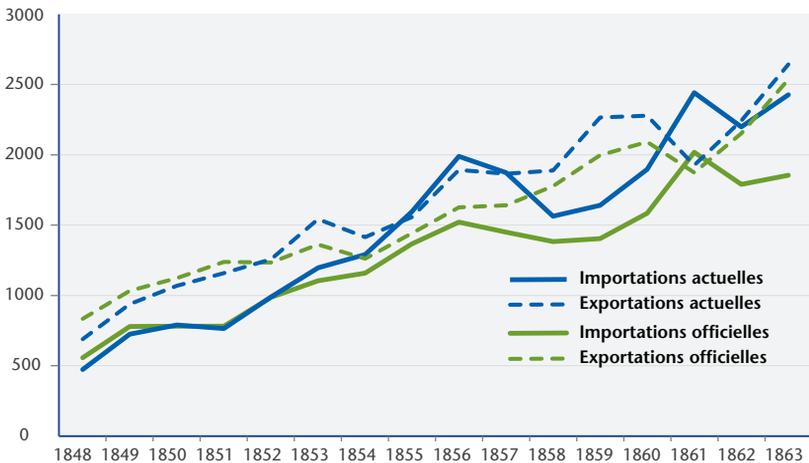
4 Il est intéressant de rappeler que plusieurs tentatives infructueuses de création d'une union douanière entre la Belgique et la France voient le jour entre 1836 et 1843 ; mais elles se heurtent, dans les deux pays, à des forces politiques souverainistes puissantes.

d'actualiser l'établissement des valeurs (elle est composée des membres des deux chambres, des départements des finances et du commerce, de négociants, de fabricants).

Dès lors, dans les tableaux (dès celui de 1847), la commission des valeurs en douane attribue un prix à chaque sous-produit associé à une section tarifaire après une consultation des Chambres de commerce et des chambres consultatives des arts et manufactures. La liste des membres de la commission des valeurs est disponible au début de chaque annuaire (poste par poste : laine, soie, ...). La consultation apparaît très imparfaite, toutes les chambres ne répondent pas...

À partir de 1847, les valeurs actuelles sont fondées sur l'application de prix définis annuellement par la commission. L'année 1847 peut constituer une rupture lorsque l'on considère les séries proposées dans la base Montesquieu. Dès lors, les Tableaux proposent des flux en valeurs actuelles sur la base du travail de la commission et toujours des flux en valeurs officielles. La valeur actuelle représente la moyenne des prix du produit (ou du groupe de produits) pendant l'année entière. Pour les importations il s'agit d'une valeur CAF (Coûts, Assurances, Frets) et pour les exportations d'une valeur FOB (Franco à bord). La France abandonne la publication de valeurs officielles à partir de 1863 (suivant a priori l'exemple anglais de 1854). L'analyse des écarts entre valeurs officielles et actuelles peut être très intéressante sur 1847-1863. Comme le montre le graphique 3 ci-dessous pour les importations,

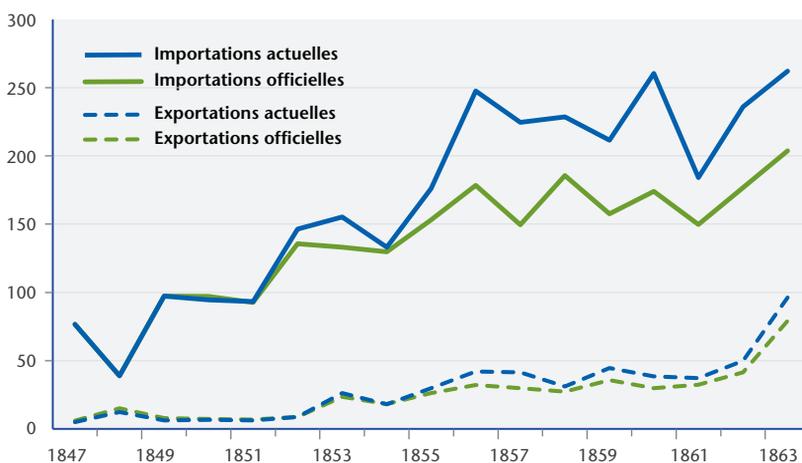
Graphique 3. Comparaison entre les évolutions des valeurs officielles et des valeurs actuelles des exportations et des importations totales de la France entre 1848 et 1863



l'écart entre valeurs officielles et valeurs actuelles se creuse au fil du temps jusqu'à dépasser les 20 %. Pour les exportations les valeurs actuelles se déconnectent beaucoup moins des valeurs officielles.

Le graphique 4 présente l'évolution des valeurs officielles et actuelles pour la soie (brute). L'envolée de l'écart entre les valeurs actuelles et les valeurs officielles à partir de 1855 montre l'importance des mouvements de prix liés en l'occurrence à un choc (les conséquences de la pébrine qui freine la production).

Graphique 4. Comparaison entre les évolutions des valeurs officielles et des valeurs actuelles des exportations et des importations de soie de la France entre 1848 et 1863



2.3. Contrebande et achats des touristes

L'importance de la contrebande est un enjeu traditionnel des statistiques historiques du commerce extérieur ; elle est source d'imprécisions. Bandes organisées et individus opportunistes cherchent à contourner les restrictions à l'importation principalement pour des produits faciles à transporter et dont l'élasticité-prix de la demande est faible.

L'évaluation de ce phénomène reste assez délicate. Dormois (2009) note que seul le gouvernement britannique publie le nombre des saisies annuelles réalisées par ses douanes. À partir de ce cas, Dormois (2009, p. 134) estime le trafic en contrebande à environ 5 % du commerce total. Cette estimation est conforme aux travaux de Brezis

(1995, p. 60) pour qui la contrebande « n'était pas particulièrement significative » pour la Grande-Bretagne au XIX^e siècle. Pour les autres pays, il n'est pas exclu que le phénomène soit plus significatif surtout lorsque les entraves à l'importation sont importantes, c'est-à-dire à partir des années 1880 lorsque la politique commerciale se durcit en Europe.

Les exportations françaises peuvent être sous-estimées en raison de l'impossibilité d'estimer la valeur des achats de touristes étrangers. Le nombre des touristes européens (anglais, allemands, russes, ...) grandit, sous l'impulsion, du développement du transport ferroviaire à partir du Second Empire dans les zones côtières (Côte d'Azur, Sud-Ouest) et à Paris (voir la thèse de B. Larique, 2006). Leurs achats peuvent être significatifs dans les secteurs du luxe (articles de Paris, textile, ...). À un degré moindre, le même biais peut toucher les importations à travers le rapatriement d'objets par les touristes français.

2.4. Le choix des sections tarifaires en tant qu'instrument de protection

Au niveau le plus fin, le nombre de catégories de produits est égal à celui du nombre de sections dans le tarif général. Pour la France, comme pour beaucoup d'autres pays, le nombre de sections progresse à travers le temps : 1 313 pour le tarif de 1892, 1 469 dans le tarif de 1910. Une correspondance peut être établie entre les sections tarifaires et les niveaux 4 et 5 de la SITC rev 3. Ainsi les postes « tissus de soie », « tissus de coton », « tissus de laine » principales spécialisations de la France durant la première mondialisation se déclinent respectivement en 30, 19 et 11 sous-produits en 1850 et 27, 38 et 28 en 1913. Des flux bilatéraux sont dans ce cas disponibles permettant, par exemple, une analyse de l'intensité du commerce intra-branche bilatéral.

La constitution des nomenclatures douanières recèle des enjeux considérables. Dans le contexte stratégique et discriminatoire de la période qui court de la fin des années 1870 à la Première Guerre mondiale, les nomenclatures tarifaires se complexifient et deviennent de plus en plus désagrégées et précises. Le choix d'une structuration en sections très désagrégées permet, sur le principe, de cibler des produits concurrents plus facilement (exemple la section « cristaux tchécoslovaques ») et de « contourner » la clause de la nation la plus favorisée (Becuwe et Blancheton, 2014).

La nomenclature britannique apparaît relativement simple. Le classement associé au *Customs Consolidation Act* de 1876 propose

514 intitulés à l'importation et 214 à l'exportation. La division en 4 sections de 29 catégories date de 1885 (Dormois, 2009, p. 130). À partir de 1906, la nomenclature allemande comporte un total de 946 entrées. Le tarif italien de 1887 comporte 1202 sections. Une autre piste de recherche consisterait à établir à l'échelle internationale le lien entre politique protectionniste et détail des nomenclatures.

3. Des bases partielles déjà disponibles

3.1. Les résumés rétrospectifs : une « Base produits » agrégée

À partir des résumés rétrospectifs (figurant au début de l'annuaire) une première base a été constituée et exploitée (Becuwe *et al.*, 2018). Sur l'ensemble de la période, les importations distinguent 107 produits et les exportations 135 produits, en provenance et à destination de l'ensemble des partenaires. Les produits sont classés par ordre décroissant de valeur du trafic, entre 80 et 90 % du trafic est identifié et une ligne « autres marchandises » rassemble le solde sans entrer dans le détail. De ce fait, certains produits ne sont plus tracés par la base alors que d'autres apparaissent (automobiles, ...). Ce niveau de désagrégation correspond approximativement au niveau 3 de la classification contemporaine SITC.

Les produits sont classés en trois grandes catégories : produits primaires, agricoles et manufacturés. Cette répartition a été réalisée par l'Administration des douanes. La modalité « autres marchandises » regroupe des produits dont le trafic est faible et le contenu inconnu à ce niveau de désagrégation. L'interopérabilité des bases de données – à l'échelle internationale – impose ici un codage de chaque nomenclature historique en norme SITC ou ISIC. Les noms des produits peuvent évoluer à travers le temps, des produits disparaissent, des nouveaux arrivent sur le marché.

Par rapport à la classification contemporaine ISIC⁵, une différence notable réside dans le classement de produits comme le beurre, le vin, les spiritueux, Dans les nomenclatures anciennes ils sont considérés comme des produits agricoles et en norme ISIC comme des produits manufacturés (section D) en raison de leur intensité capitaliste et d'un degré de transformation important. Pour certains pays exporta-

5 L'International Standard Industrial Classification (rev.3) considère quatre catégories : agriculture, chasse et forêt (A), pêche (B), mines et carrières (C) et produits manufacturés (D).

teurs de vins (Espagne, Algérie) les enjeux sont cruciaux en matière d'interprétation de la structure d'exportation.

3.2. Le trafic portuaire (1857-1938)

À partir de 1857, les Tableaux proposent des données sur le trafic commercial au sein de chaque port français et localités des frontières de terre ou de l'intérieur où les mouvements commerciaux sont les plus importants. Nous avons constitué une base de données du trafic de chaque port sur la période 1857-1938. Le nombre de ports pris en compte évolue de 18 en 1857 à 23 en 1878 et oscille autour d'une vingtaine jusqu'à la Seconde Guerre mondiale. Pour chaque port le trafic produit par produit est disponible à un niveau de désagrégation correspondant au niveau 3 de la classification SITC rev 3. Y figurent le commerce spécial, le commerce général pour les importations et les exportations (en quantité et en valeur) et les droits de douane perçus. Cette base pourrait être raccordée aux données TOFLIT18 très centrées sur le trafic portuaire au XVIII^e siècle (Charles et Daudin, 2011, 2015).

Sur la période en termes de trafic les principaux ports sont dans l'ordre en moyenne (importations et exportations confondues) : Marseille, Le Havre, Paris, Bordeaux, Dunkerque et Boulogne. Cette base devrait permettre d'éclairer les liens entre spécialisations territoriales et développement des arrière-pays sur longue période dans le prolongement des travaux de B. Marnot (2011, 2014).

4. L'aboutissement du projet : une base de données du commerce bilatéral (produits/pays)

4.1. Les enjeux d'une numérisation exhaustive

Afin de disposer d'une base de données qui croise les dimensions pays et produits au niveau le plus désagrégé possible sur plus d'un siècle, l'équipe bordelaise a lancé un projet visant l'exhaustivité de la numérisation en coopération avec la Direction Régionale des Affaires Culturelles d'Aquitaine, la Maison des Sciences de l'Homme d'Aquitaine, la Région Nouvelle Aquitaine qui subventionnent la démarche et le Musée National des Douanes de Bordeaux qui offre un appui logistique à travers la mise à disposition de la série des Tableaux.

Nous avons aujourd'hui numérisé l'ensemble des pages des Tableaux pour les années 1836 à 1938 afin de disposer de statistiques les plus complètes possibles sur les produits échangés par la France

avec ses partenaires commerciaux mais aussi sur les mouvements de marchandises pour chacune des principales douanes de France (18 en 1857, 24 en 1897 et 17 en 1937).

L'aboutissement du projet consistera à proposer une base de données complète et à offrir l'entièreté des ressources disponibles dans les Tableaux. Elles sont de trois types et constituent trois ensembles distincts :

- données au format Image (Tiff, Jpeg et PDF) créées par la numérisation complètes des 103 registres (soit 74 090 pages) ;
- données chiffrées, organisées en bases de données, créées à partir des tableaux de commerce (8 600 pages) ;
- données textuelles sous forme de nomenclatures « Produits échangés » et « Entités géographiques » et d'observations rédigées par l'administration (4 nomenclatures distinctes évoluant dans le temps et 380 pages d'explications).

Aujourd'hui, la collection « Registres des douanes de Bordeaux » est progressivement publiée dans la bibliothèque numérique patrimoniale du réseau documentaire des universités de Bordeaux, Babord_Num. Elle dispose d'un lien URL pérenne moissonné par Gallica et référencé dans les moteurs de recherche, ce qui assure une large visibilité (<http://www.babordnum.fr/collections/show/13>). Les images sont ainsi déjà disponibles sous licence ouverte et librement téléchargeables par les chercheurs.

L'ensemble des fichiers issus de la numérisation et du post-traitement des images ont été archivés de manière pérenne sur le système d'archivage numérique de l'Université de Bordeaux. Pour le format des documents numérisés, le projet a suivi les recommandations du CINES (Centre Informatique National de l'Enseignement Supérieur) et du Référentiel Général d'interopérabilité (version 1.9.7 de mars 2015).

Ces données OCRérisées (OCR Optical Character Recognition) automatiquement lors de l'étape de numérisation (fichiers relationnels XML/ALTO) ont été converties au format tabulaire sous EXCEL avec un logiciel de reconnaissance de caractères (OCR repris manuellement) afin de conserver la mise en forme. Nous avons défini et expérimenté une méthodologie pour vérifier et corriger les données.

À partir de ce corpus OCRisé manuellement, et correspondant à « 333 000 lignes » de marchandises entre la France et ses partenaires commerciaux (soit un peu moins de deux millions de données), nous

avons mis en place une architecture d'identification et de correction par exploration manuelle du corpus d'écarts PDF / fichier texte :

- Pour les données quantitatives (Quantités, Valeurs, Droits perçus pour les exportations mais aussi pour les importations) qui contiennent des erreurs d'OCR que nous souhaitons corriger, une approche par règles a été adoptée à partir de la typologie des erreurs réalisée par nos soins (erreurs de segmentation, d'OCR, de mots, ...). Le résultat de ces tâches automatisées, à l'aide du langage VBA, nous a permis de repérer en moyenne par année 10% de cellules avec une anomalie et d'en corriger 8 %. Les 2 % restant nécessitent encore une intervention manuelle ;
- Pour les données textuelles (noms des produits, des unités de mesure et des entités géographiques), une correction orthographique est d'abord adoptée en regroupant les valeurs proches par nombre d'occurrences. Une méthode de recherche approximative est ensuite appliquée afin de regrouper les diverses écritures d'un même produit. Nous utilisons pour cette étape, le répertoire alphabétique des marchandises figurant à la fin de chaque ouvrage. Le recours à des dictionnaires en ligne (Dictionnaire de l'Académie Française et Grand Dictionnaire Universel du XIX^e siècle par Pierre Larousse) est aussi quelquefois nécessaire.

Nous avons créé deux tables de correspondance au format .csv (une pour la normalisation orthographique et l'autre pour la simplification des noms de produits) afin de ne pas perdre d'information et de privilégier une réutilisation et/ou modification de ce travail réalisé avec les logiciels OPENREFINE et R.

Nous souhaitons continuer le travail et la méthodologie adoptés par le projet TOFLIT18 qui a déposé ses tables de correspondance et classifications sur Github au format .csv. Enfin, nous avons aussi tenu compte des anomalies mentionnées dans la partie « Erratum » de chaque annuaire (32 pages pour toute notre période) pour corriger nos données.

4.2. « Le Trésor produit »

Au sein des « Tableaux », quatre nomenclatures de produits coexistent et couvrent un vocabulaire très large sur les matières premières, les produits manufacturés, les produits agricoles, les objets ordinaires, de luxe, etc. À titre d'exemple, sur les décennies 1850-1870, on comptabilise 1766 intitulés produits. Nous souhaitons structurer un trésor réunissant les concepts figurant dans notre corpus

selon le modèle de données SKOS/RDF. Il s'agit d'un travail de modélisation rendu assez lourd par l'apparition et la disparition du trafic de certains produits et des mouvements d'agrégation ou de désagrégation des nomenclatures.

Nous souhaitons exposer nos terminologies sur la plateforme Loterre (<https://www.loterre.fr>) de l'INIST. L'INIST assurerait la conversion au format XML à partir des données terminologiques fournies au format .csv.

Afin d'augmenter sa visibilité et son adoption par les communautés concernées, ce vocabulaire contrôlé serait soumis pour un référencement dans le portail FAIRsharing (<https://fairsharing.org/>), sous la catégorie Standards > Terminology artifact. Conformément à la charte de Loterre et aux conditions de l'appel à projets, il serait publié sous une licence Creative Commons CC-BY.

La première version de ce thésaurus serait monolingue en français ; des alignements vers des terminologies dans d'autres langues seront étudiés. Ce travail devrait favoriser, le moment venu, l'interopérabilité des bases à l'échelle nationale et à l'échelle internationale. Il permettra aussi de suivre les produits changeant de statut sur une longue période, de tracer la vie, l'apparition ou la disparition des produits mais aussi de comparer les données avec d'autres données thématiques (la base de données INPI des brevets français XIX^e siècle, les traités commerciaux, les recensements de la population, ...). Ce thésaurus produit pourrait être rapproché des 16 000 biens recensés par TOFLIT18 afin d'établir un continuum des produits échangés entre le XVIII^e et le XX^e siècle.

5. Conclusion

Cet article met en perspective une démarche de numérisation et de transformation en base de données de sources annuelles du commerce extérieur de la France pour la période 1836-1938 : le « Tableau général du commerce de la France avec ses colonies et les puissances étrangères » (qui devient le « Tableau général du commerce et de la navigation après 1896 »). L'originalité de cette base est d'être centrée sur la dimension produit et d'offrir au niveau le plus fin des données bilatérales correspondant approximativement au niveau 4 ou 5 de la SITC rev 3. Ce corpus OCRisé manuellement comporte « 333 000 lignes » de marchandises entre la France et ses partenaires commerciaux (soit près de deux millions de données). Il offre un « thésaurus produit » d'une grande richesse. L'article dresse un panorama des

ressources disponibles pour les chercheurs. Il éclaire des questions traditionnelles en statistique historique : valorisation des flux commerciaux, transit, détermination des valeurs unitaires, contrebande, ...

Ces données sont progressivement mises à la disposition du public. La collection « Registres des douanes de Bordeaux » est publiée dans la bibliothèque numérique patrimoniale du réseau documentaire des Universités de Bordeaux, Babord_Num avec un lien URL pérenne moisonné par Gallica. Mais ces matériaux de recherche ont vocation à être structurés et mis en ligne en répondant aux impératifs de l'interopérabilité des métadonnées. En arrière-plan se dessinent les enjeux liés au web de données avec de nouvelles coopérations et partenariats dans et hors de la communauté des historiens économistes.

En histoire économique ces données ont vocation à affiner l'analyse de l'insertion de l'économie française dans les échanges internationaux sur longue période. Comment le commerce extérieur français a-t-il réagi à des chocs comme les guerres et les ruptures de politiques commerciales (Traité de 1860, tarif de 1892, ...). Comment a évolué sa diversification ? Quelles ont été les évolutions du commerce intra-branche ? Comment a évolué la qualité des produits ? Une montée en gamme est-elle perceptible à travers l'évolution des valeurs unitaires à l'exportation ? Quelles ont été les conséquences des politiques commerciales sur les échanges et l'activité économique (vieux débats sur le paradoxe Bairoch (Bairoch, 1992 ; Becuwe et Blancheton, 2013) ?

Mais la construction de cet espace de données plus large offrira, nous le souhaitons, de nouvelles perspectives de recherche en liaison avec les géographes, les politistes sur la question des conflits, des liens entre les politiques commerciales et les votes ou encore avec les physiiciens, les biologistes, les écotoxicologues à propos de la dynamique historique des émissions et de la circulation du CO₂.

Références

- Asselain J.-C. et B. Blancheton, 2005, « Dynamique de l'ouverture internationale. Paradoxes, enjeux, éléments d'interprétation à partir du cas de la France », *Économies et Sociétés-Série AF Histoire Economique Quantitative*, n° 1, pp. 49-179.
- Bairoch P., 1992, « La France dans le contexte international: politique commerciale et commerce extérieur 1890-1990 », dans *Le commerce extérieur français de Méline à nos jours*, CHEFF, Paris, pp. 3-33.

- Barbieri K., O. Keshk O. et B. Pollins, 2009, « Trading Data: Evaluating our Assumptions and Coding Rules », *Conflict Management and Peace Science*, vol. 26, n° 5, pp. 471-491.
- Becuwe S. et B. Blancheton, 2013, « Les controverses autour du paradoxe Bairoch, quel bilan d'étape ? », *Revue d'Economie Politique*, n° 123 (1), pp. 1-27.
- Becuwe S. et B. Blancheton, 2014, « Dispersion of French tariffs (1850-1913) », *Research in Economic History*, n° 30, pp. 163-183.
- Becuwe S. et B. Blancheton, 2019, « French Textile Specialisation in Long Run Perspective (1836-1938): Trade Policy as Industrial Policy », *Business History*,
doi.org/10.1080/00076791.2018.1494732
- Becuwe S., B. Blancheton et C. Meissner, 2018, « Stages of diversification: France, 1836-1938 », *European Review of Economic History*, 2018, vol. 22, n° 4, pp. 430-461, *NBER Working Paper Series*, n° 21777.
- Blancheton B., 2016, « Central bank independence in an historical perspective. Myth, lessons and a new model », *Economic Modelling*, vol. 52, pp. 101-107.
- Brezis E., 1995, « Foreign Capital Flows in the Century of Britain's Industrial Revolution: New Estimates Controlled Conjectures », *Economic History Review*, n° 48 (1), pp. 46-67.
- Charles L. et G. Daudin, 2011, « La collecte du chiffre au XVIII^e 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*, n° 58 (1), pp. 128-155.
- Charles L. et G. Daudin, 2015, « Eighteenth-Century International Trade Statistics. Sources and methods », *Revue de l'OFCE*, n° 140 (4), pp. 7-36.
- Charles L., 2016, *Protection, spécialisation et croissance économique pendant la première mondialisation en France et en Suisse (1850-1913)*, thèse pour le Doctorat en sciences économiques, Université de Bordeaux.
- Dedinger B., 2012, « The Franco-German trade puzzle: an analysis of the economic consequences of the Franco-Prussian war », *Economic History Review*, n° 65, pp. 1029-54.
- Dormois J.-P., 2009, « La défense du travail national ? L'incidence du protectionnisme sur l'industrie en Europe (1870-1913) », PUPS, Paris.
- Ducros J., E. Grandi, P.-C. Hautcoeur, R. Hekimian, E. Pruniaux, A. Riva et S. Ungaro, « Collecting and storing historical financial data: the DFH project », in C. Stuetzler, M. Welker et M. Egger (eds), *Computational Social Science in the Age of Big Data. Concepts, Methodologies, Tools, and Applications*, Neue Schriften zur Online-Forschung, Herbert von Halem Verlag, Cologne, 2018, pp. 355-377.
- Fouquin M. et J. Hugot, 2016, « Two Centuries of Bilateral Trade and Gravity Data: 1827-2014 », *CEPII Working Paper*, n° 2016-14, mai.

- Fouquin M. et J. Hugot, 2016, « Back to the Future: International Trade Costs and the two Globalizations. » *Working Paper CEPII*, n° 2016-14.
- Lampe M., 2008, « Bilateral trade flow in Europe, 1857-1875 », *Research in Economic History*, n° 26, pp. 81-156.
- Larique B., 2006, *L'économie du tourisme en France des années 1890 à la Seconde Guerre mondiale : organisation et développement d'un secteur socio-économique*, thèse pour le doctorat en histoire, Université Bordeaux 3.
- Marnot B., 2011, *Les grands ports de commerce français et la mondialisation au XIX^e siècle*, Paris, Presses de l'Université Paris-Sorbonne.
- Marnot B., 2014, « Les ports français et la mondialisation au XIX^e siècle ou l'émergence de trois modèles de croissance régionaux », *Monde(s). Histoire, espace, relations*, n° 6, novembre, pp. 195-222.
- Marzagalli S., 2016, « Navigocorpus database and eighteenth-century French world maritime networks », in César Ducruet (ed.), *Maritime Networks. Spatial structures and time dynamics*, New York, Routledge, pp. 92-111.
- Pallain G., 1913, *Les douanes françaises : régime général, organisation, fonctionnement*, Paris, Librairie Administrative Paul Dupont.
- Schuller C. J., 1911, *La statistique douanière et ses méthodes*, thèse de doctorat de la Faculté de Droit de l'Université de Paris.
- Tena-Junguito A. et G. Federico, 2016, « World Trade, 1800-1938: a new data-set », *Working Papers in Economic History*, n° WH 16-01, Universidad Juan Carlos III Madrid.

ANNEXE 1. Pays de provenance et de destination

Les définitions de ces deux notions, essentielles pour déterminer les pays au compte desquels les marchandises doivent être prises en charge soit à l'entrée soit à la sortie ont évolué au cours de la période.

Jusqu'en 1869, s'agissant des importations et des exportations par mer, on considérait comme pays de provenance celui d'où le navire avait été expédié en dernier lieu pour la France, et comme pays de destination celui dans lequel se rendait le navire exportateur.

À partir du 1^{er} janvier 1870, toutes les fois qu'il est possible de connaître le point effectif du départ des marchandises ou leur destination réelle, elles sont alors portées au compte du pays d'où elles proviennent réellement ou à la destination duquel elles sont réellement expédiées. De cette façon, par exemple, on fait figurer au compte du Brésil les sucres importés de cet État par des navires qui ont fait escale en Angleterre.

En 1895, les marchandises sont, à l'entrée, quel que soit leur pays d'origine, prises en charge au compte du pays de provenance effective actuelle, c'est-à-dire au compte du pays où a eu lieu le chargement initial à destination de la France. La transaction commerciale s'est, en effet, exclusivement effectuée entre la France et ce dernier pays. Si les marchandises sont originaires des pays expéditeurs, alors le pays de provenance n'est autre que le pays d'origine. Par analogie, la marchandise est prise en charge, à la sortie, au compte du pays de destination effective actuelle, c'est-à-dire du pays sur lequel elle est effectivement dirigée et où elle doit être livrée au destinataire réel. Ce principe de pays de provenance effective a perduré jusqu'en 1933.

À compter de 1934, pour les importations, l'application du système des contingents et des mesures de protection ainsi que la préparation des accords commerciaux ont fait apparaître l'intérêt à adopter désormais l'origine des marchandises, c'est-à-dire le lieu où elles ont été récoltées, extraites ou fabriquées. Pour les exportations, les envois sont imputés à la sortie au compte du pays de la dernière destination connue.

ANNEXE 2.

Répartition des flux entrants et sortants en produits manufacturés, produits de base et produits agricoles

Les tableaux A1. et A2. ([fichier excel en ligne, attaché à l'article](#)) illustrent les données disponibles au sein de la base. La décomposition du commerce extérieur en produits de base, agricoles et manufacturés montre le statut de pays avancé de la France dans les échanges internationaux au milieu du XIX^e siècle et plus largement jusqu'à la Seconde Guerre mondiale. La France exporte essentiellement des articles manufacturés à hauteur de plus de 60 % du total jusqu'à la fin des années 1850. La baisse de la part des articles manufacturés est notable jusqu'à la Première Guerre Mondiale ; après un rebond dans les années 1920, elle diminue de nouveau dans les années 1930 pour atteindre 48 % en 1938. La part des produits agricoles progresse d'environ 20 % dans les années 1830 à 30 % dans les années 1870 avant de décliner assez régulièrement jusque dans l'entre-deux-guerres, révélant le manque de compétitivité d'une agriculture française pourtant très protégée (Asselain et Blancheton, 2005). La part des exportations de produits primaires, négligeable dans les années 1830-1850, progresse à partir des années 1860 pour atteindre 25 % dans les années 1930. Ces tendances révèlent une détérioration à long terme de la qualité de l'insertion de la France dans les échanges mondiaux.

Du côté des importations, les produits primaires sont dominants et atteignent la moitié du total, hormis dans les années 1880-90 où leur part diminue alors que la part des produits agricoles progresse et dépasse parfois les 40 %. Le poids relatif des importations manufacturières reste faible autour de 10 % sauf dans les années 1900 et pendant la Première Guerre mondiale où il dépasse 20 %.

ANNEXE 3. Principaux produits exportés et importés

Chaque année, les produits sont classés par ordre d'intensité du trafic. Les tableaux A3. et A4. ([fichier excel en ligne, attaché à l'article](#)) proposent les huit principales exportations et importations sur l'ensemble de la période. Les principaux postes d'importations concernent des produits primaires destinés à soutenir le développement industriel français (coton, laines, soies, houille). Les principales exportations concernent les produits textiles, le vin, tabletteries, bimbeloteries,

ANNEXE 4. Tableau de correspondance avec les normes contemporaines SITC

([fichier excel en ligne, attaché à l'article](#)).

