NBER WORKING PAPER SERIES

CODIFICATION, TECHNOLOGY ABSORPTION, AND THE GLOBALIZATION OF THE INDUSTRIAL REVOLUTION

Réka Juhász
Shogo Sakabe
David Weinstein

Working Paper 32667
http://www.nber.org/papers/w32667

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2024, revised January 2025

We give special thanks to Chris Meissner and John Tang for sharing their trade data for Belgium and Japan. We thank Benjamin Eyal, Isaac Loomis, Zachary Marcone, Ojaswee Rajbhandari, Roshan Setlur, Alex Zhang, and especially Michael Duarte, Hriday Karnani, Verónica C. Pérez, Angela Wu, and Dongcheng Yang for excellent research assistance. We also want to thank Treb Allen, Andrew Bernard, Kirill Borusyak, Serguey Braguinsky, Florian Caro, Davin Chor, John Fernald, Walker Hanlon, Shizuka Inoue, Takatoshi Ito, Debin Ma, Joel Mokyr, Chiaki Moriguchi, Tetsuji Okazaki, Charly Porcher, Jared Rubin, Robert Staiger, Jón Steinsson, John Tang, Dan Trefler and Weiwen Yin for their excellent comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Codification, Technology Absorption, and the Globalization of the Industrial Revolution
Réka Juhász, Shogo Sakabe, and David Weinstein
NBER Working Paper No. 32667
July 2024, revised January 2025
JEL No. F14,F63,N15

ABSTRACT

This paper studies technology absorption worldwide in the late nineteenth century. We construct several novel datasets to test the idea that the codification of technical knowledge in the vernacular was necessary for countries to absorb the technologies of the Industrial Revolution. We find that comparative advantage shifted to industries that could benefit from patents in countries and colonies that had access to codified technical knowledge but not in other regions. Using the rapid and unprecedented codification of technical knowledge in Meiji Japan as a natural experiment, we show that this pattern appeared in Japan only after the Japanese government codified vast amounts of technical knowledge. Our findings shed new light on the frictions associated with technology diffusion and offer a novel take on why Meiji Japan was unique among non-Western countries in successfully industrializing during the first wave of globalization.

Réka Juhász
Vancouver School of Economics
University of British Colunbia
6000 Iona Drive
Vancouver, BC V6T 1L4
and NBER
reka.juhasz@ubc.ca

Shogo Sakabe
ss5122@columbia.edu

David Weinstein
Columbia University
Department of Economics
420 W. 118th Street
MC 3308
New York, NY 10027
and NBER
dew35@columbia.edu

"At present, the learning of China and Japan is not sufficient; it must be supplemented and made complete by inclusion of the learning of the entire world... I would like to see all persons in the realm thoroughly familiar with the enemy's conditions, *something that can best be achieved by allowing them to read barbarian books as they read their own language.* There is no better way to enable them to do this than by publishing [a] dictionary."

Shozan Sakuma, 1858, quoted in Hirakawa (2007, p. 442, emphasis added).

1 Introduction

Although recent econometric evidence finds that modern economic growth started in England around 1600 (Bouscasse et al., 2023), the spread of economic development has been highly uneven. For example, currently, there are only four types of high-income countries in the world: English-speaking countries, countries close to England, resource-abundant countries, and Japan and its former colonies.[1] While economists have made enormous progress in understanding why English-speaking countries, Europe, and Petrostates are rich, data-driven studies of why the Industrial Revolution first spread to Japan and not to any other non-Western country are almost nonexistent. After centuries of resisting economic and social change, Japan transformed from a relatively poor, predominantly agricultural economy specialized in the exports of unprocessed, primary products to an economy specialized in the export of manufactures *in under fifteen years*.[2] Why did Meiji Japan succeed in this structural transformation while so many other countries failed to develop in this period?[3]

We bring several novel datasets to bear on this question and test one of the main theories proposed by Mokyr (2011): namely, that an essential component of the Industrial Revolution was the development of what Stevens (1995) calls "technical literacy," i.e., the *codification* of engineering, commercial, and industrial practices. We call this knowledge "technical knowledge." We hypothesize that codification reduces technology access costs by enabling entrepreneurs to understand technology by reading it in their vernaculars. While there is extensive evidence that what Mokyr (2016) refers to as the culture of Enlightenment created vast amounts of codified technical knowledge in Western European languages, our understanding of codification beyond Europe is much more limited. For example, we have no idea how many books containing technical knowledge a literate person in China could have read in 1870 or the extent to which the number of books containing codified knowledge changed over time. As a result, prior work has been unable to explore how access to technical knowledge in the vernacular contributed to the *spread* of the Industrial Revolution.

An ideal experiment would require both cross-sectional and time-series variation in technical knowledge supply. Cross-sectional variation allows us to examine whether industries that stood the most to gain from the supply of codified knowledge grew faster in countries that supplied this knowledge. Time-series variation lets us explore whether the timing of this faster growth coincides

[1]We define high-income countries as those with a purchasing-power-parity adjusted GDP per capita of 50 percent or more than the US level in 2022 as measured by the World Bank. See Appendix D for more details.

[2]We also see this sudden transformation in the efficiency of modern industry. For example, Clark (1987) finds Japan transitioned from not having any modern textile and weaving mills in 1870 to having modern mills that achieved levels of output per unit of capital that were 96 and 98 percent of those in Britain by 1910. By contrast, the Chinese textile and weaving industries had 79 and 66 percent output efficiency relative to Britain's in 1910.

[3]The Meiji period started with the Meiji Restoration in 1868 and lasted until the emperor died in 1912.

with the provision of this knowledge. The experiences of Meiji Japan and the other codifiers of knowledge in the late nineteenth century provide precisely this empirical setting.

We test the link between codified knowledge and productivity growth in Meiji Japan and the late 19th-century global economy by constructing the first dataset that enables us to quantify the extent of codification by language, the usefulness of this codification by industry, and industry-level export growth in 39 countries and regions in the late nineteenth and early twentieth centuries.[4] We use this novel dataset to construct the first estimates of industry-level productivity growth for many nineteenth-century countries and regions using structural trade theory (Costinot et al., 2012) and the method of moments estimator developed by Amiti and Weinstein (2018). We build this dataset by scraping the catalogs of libraries for every major language, digitizing technical books for every major tradable industry, digitizing the synopses of all British patents issued between 1617 and 1852, digitizing bilateral industry-level trade data for Japan and the U.S., merging these trade data with extant trade datasets to create the first multicountry, bilateral, industry-level, trade dataset for the nineteenth century.

We establish three novel stylized facts about the global spread of the Industrial Revolution and the uniqueness of Japan's nineteenth-century industrialization, "the Meiji Miracle." The first stylized fact is that Meiji industrialization was exceptional in comparison to other regions in the periphery. In the raw data, we find that late 19th century Japan experienced a surge in the share of manufacturing exports that surpassed that of any of the other 38 regions in our sample. Notably, this surge in manufacturing exports happened rapidly and much later than well-known events in nineteenth-century Japanese economic history. In particular, although Japan jumped to free trade in 1858, eighty percent of Japanese exports were still primary products twenty-five years later. Moreover, although many people point to the 1868 Meiji Restoration as a watershed moment in introducing new reforms, fifteen years later, the Japanese manufacturing export share had *fallen* from 29 percent of total exports to 21 percent. In other words, there is no evidence that Japan's natural comparative advantage or state reforms were gradually improving productivity in the manufacturing sector. However, in a brief 13-year period from 1883 to 1896, Japan's manufacturing export share tripled and then remained at around 60 percent of total exports until the Second World War. Consistent with the evidence from the raw data, our structural estimates of productivity growth for Japan also point to an exceptionally strong shift in comparative advantage towards manufacturing sectors in the late nineteenth century. Thus, the puzzle of Japan's development is why Japanese manufacturing exports grew so suddenly, and so much. It was as if Japanese entrepreneurs had suddenly learned how to do modern manufacturing.

The second stylized fact is that in 1870, entrepreneurs in most regions—including Japan—had almost no technical books to read in their vernaculars. We document this by scraping thousands of libraries containing books in 33 major languages and find that in 1870, 84 percent of all *technical* books were written in just four languages: English, French, German, and Italian. People who could not read these four languages were therefore technically illiterate. For example, a person who could only read Arabic would only have been able to read 72 technical books in 1870. Libraries for other major non-European languages, such as Chinese, Hindi, and Turkish, have extensive collections of books but contain similarly small numbers of technical books. By contrast, speakers of major European languages would have had access to thousands of technical books. This puts the achievements of the Enlightenment, with its emphasis on the coding, storing and transmission of technical knowledge" (Berg, 2007, p. 125) in comparative perspective for the first time. In short, outside of the languages of the Enlightenment, literacy in the vernacular was a ticket to reading

[4]Henceforth, we describe our geographical units as "regions" instead of "countries" because many trade flows were reported for colonies and other geographies that were not formally countries.

the humanities and history, not reading science.

The third stylized fact is that the Japanese language is unique in that it started at a low base of codified knowledge in 1870 but experienced explosive growth in the publication of technical books, catching up with the West in the middle of the 1880s. By 1890, there were more technical books in the Japanese National Diet Library (NDL) than in the libraries of Germany or Italy. By the end of our sample period (1910), there were more technical books written in Japanese than in any other language except English and French. This catchup in codification coincides with Japan's sudden industrialization. Japan suddenly began exporting manufactured products shortly after Japanese entrepreneurs could read *in Japanese* how to make these products.

How did Japan achieve such a remarkable growth in the supply of technical books? We show that the Japanese government was instrumental in overcoming a complex public goods problem, which enabled Japanese speakers to achieve technical literacy in the 1880s. We document that Japanese publishers, translators, and entrepreneurs initially could not translate Western scientific works because the Japanese language lacked the words needed to describe the technologies of the Industrial Revolution. The Japanese government solved this coordination problem by creating a large dictionary that contained Japanese jargon for many technical words. Indeed, we find that new word coinage in the Japanese language grew suddenly after a massive government effort to subsidize translations produced technical dictionaries and, subsequently, a large number of translations of technical books.

Beyond producing technical dictionaries, the Meiji government also made substantial investments in codifying knowledge by paying for the large-scale translation of technical knowledge from the West (Montgomery, 2000). Our analysis of the institutional affiliations of these translators reveals that 74 percent of them were government employees, indicating the relative importance of the government in funding this public good.[5] This created two sub-periods in Meiji Japan: a period before the 1880s, in which Japan had completed substantial economic reforms but had codified less than half as much as had been codified in Spanish (and a small fraction of what had been codified in French, English, Italian, and German), and a period afterward in which Japanese people could read Western technical knowledge at a level equal to or exceeding that in the West.

Together, our stylized facts show that the sudden increase in codified knowledge and the sudden increase in manufacturing specialization, which happened shortly thereafter, was unique to Japan in the 19th century. Thus, Japanese manufacturing export growth rates didn't gradually increase as institutions improved; rather, a rapid increase in manufacturing happened only after Japan codified about as much knowledge as Germany had in 1870. Together, we interpret these three stylized facts as suggestive evidence that access to technical knowledge may have been a necessary (although not sufficient) condition for the spread of the Industrial Revolution.

In the second part of the paper, we exploit the natural experiment of Japan's rapid codification of knowledge to test this hypothesis rigorously. This requires both time series variation and cross-sectional variation in technical knowledge; thus, we move our empirical analysis to the industry level. In particular, we develop a method to quantify the supply of useful, codified knowledge generated by the Industrial Revolution for each industry. We use a text-based approach that closely follows how codified technical knowledge was disseminated in this period: through the publication of technical manuals. For example, "The America Cotton Spinner, and Managers' and Carders' Guide," published in 1851, contains a description of every aspect of operating a cotton

[5]The NDL catalog specifies the translator for over 200 technical books translated to Japanese in the 1870s and 80s. We searched for the names of all translators on *JapanKnowledge Lib*, an online database, and made extensive use of Ueda et al. (2003), a biographical dictionary containing entries for more than 75,000 Japanese people and Heibonsha (1974), a biographical dictionary of 30,000 people in Japanese history.

spinning mill from the dimensions of the building, to setting up the gearing which distributes power through the building, as well as the operation, and maintenance of each machine used in production.

For each industry, we calculate the similarity of text from these historical technical manuals (in English) to the text of British patents using cosine similarity, the standard metric in natural language processing (NLP). We call this measure "British Patent Relevance" or BPR. Our BPR measure rises in the similarity of the word use in an industry's technical manuals to that in British patents. Thus, it is a metric for how useful the knowledge codified in British patents is for a particular industry. Reassuringly, industries such as textiles, which benefited the most from the new technologies of the Industrial Revolution, have descriptions of production processes, including flagship technologies such as spinning machinery and steam engines, that also feature prominently in patent texts. As such, we say that the contents of patent texts are relevant for manufacturing textiles. On the other hand, the cosine similarity between word use in manuals and patent descriptions is smaller for industries like charcoal, which suggests that the makers of charcoal benefitted little from Industrial Revolution technologies. Importantly, BPR is a measure of the *supply* of new technical knowledge; it does not use *any* (potentially endogenous) information on what was translated in Japan or elsewhere.

To measure outcomes at the industry level, we use our novel, bilateral, industry-level trade dataset to back out industry-level productivity growth from 1880 to 1910. This methodology is well-suited to data-scarce environments such as ours. Importantly, no data from individual reporting countries is required in the sample.

Armed with these data, we examine the relationship between the supply of technical knowledge and export/productivity growth in Japan and around the world. We show that Japanese export and productivity growth was higher in the industries where the *supply* of technical knowledge produced by the Industrial Revolution was greater. The estimates point to a large effect of access to technical knowledge on growth. Between 1880 and 1910, we estimate that an industry in the 75th percentile of the BPR distribution would have had 15 percentage points per year faster annual export growth and 1.4 percentage points faster productivity growth than an industry in the 25th percentile.

We use several features of our setting to make the case that this relationship is likely to be causal. First, we exploit the fact that Japan was unique among periphery countries in codifying knowledge. If our findings for Japan are causal, this would suggest that other periphery countries, which did *not* have access to codified knowledge, should not have a similar association between industry growth and BPR. Indeed, we find that, on average, other regions do not exhibit a similar effect. Interestingly, other low-income regions and other Asian regions tend to have a negative relationship with BPR, suggestive of divergence, though this negative effect is not always statistically significant. In contrast, for the other major codifying European languages (English, French, German, and Italian), we find a similar positive effect of BPR on growth, though one that is smaller in magnitude. In summary, in the cross-section of countries, we find supporting evidence that only codifying countries have industry growth patterns systematically related to BPR.

Second, we exploit the sharp timing of codification in Japan. In most settings, identifying the effect of codification on growth is difficult, as codification typically proceeds slowly, making it difficult to rule out confounders. However, our third stylized fact showed that Japan can be separated into two relatively well-demarcated periods: one in which Japan looks like other technically illiterate periphery economies and one in which codification in Japanese is comparable to that in the most codified Western European languages. Consistent with our hypothesis, we find a positive and statistically significant effect of BPR on industry growth in Japan *only after* Japan became technically literate. Indeed, until 1890, Japan looked remarkably similar to the rest of the global periphery, and Asia in particular, in which comparative advantage shifted away from industries that heavily used British technology.

Finally, we conduct a series of robustness checks. These include showing that the relationship between the supply of codified knowledge and industry growth in Japan holds *within* manufacturing, and when we control for the steam intensity of sectors. It also seems unlikely that some unobserved factor common across codifying European countries and Japan can account for our results: consistent with our mechanism, once we account for the effect of BPR, there is no systematic relationship between Japanese and European codifying countries' export and productivity growth. Taken together, our results lend support to the idea that low-cost access to technical knowledge, which at the time usually meant access *in the vernacular*, was a necessary condition for the diffusion of Industrial Revolution technologies and modern manufacturing growth more broadly. Moreover, our results suggest that for regions not strongly influenced by the Enlightenment, the codification of technical knowledge was a complex public good that required state provision.

This paper contributes to three strands of the literature. First, our results shed light on why technology diffusion to the global periphery was slow in the 19th century. Economic historians have put forward a number of explanations ranging from the imperialist context of the period (Allen, 2012) to culture (Clark, 1987). Our explanation builds on Mokyr (2011)'s pioneering work on the importance of "technical knowledge" for European industrialization, though with a Gerschenkronian (Gerschenkron, 2015) twist.[6] In particular, our results show that codified technical knowledge was almost non-existent outside of Europe. Thus, for regions that were physically distant from Europe and linguistically distant from major European languages, the provision of technical knowledge required the state's involvement due to its public good-like attributes. This points to a novel arena where the Gerschenkronian argument of the state as a critical agent in late industrialization may apply.

Second, our results inform our understanding of the sources of Japan's unique industrialization. Previous work has examined the introduction of new institutions (Sussman and Yafeh, 2000), modern banking (Tang and Basco, 2023), railroads (Tang, 2014), subsidized firms (Morck and Nakamura, 2007, 2018) and trade (Bernhofen and Brown, 2004, 2005). This careful work has not found large positive impacts of these policies on economic outcomes and sometimes finds the policies were counterproductive. For example, Sussman and Yafeh (2000) conclude that "the great majority of the Meiji reforms—including the establishment of the Bank of Japan and the introduction of 'modern monetary policy, the promulgation of the Meiji Constitution, and the introduction of parliamentary elections—produced no quantitatively significant market response." In the end, they conclude that only land tax reform and Japan's adoption of the gold standard mattered to investors. Our findings thus offer a resolution to the puzzle of what drove the Meiji Miracle.

These results are particularly helpful in placing the "Meiji Miracle" in a comparative perspective. That is, while the more standard modernization efforts of the Meiji government, such as the introduction of banking and railroads, certainly contributed to industrialization, given their fairly widespread adoption in other parts of the global periphery, which were characterized by more modest growth, it is unlikely they can give a full account. In contrast, our paper provides empirical support for the long strand of Japanese economic history that has emphasized the more unique aspects of the Japanese government's efforts to adopt Western technology. In fact, our results

[6] As such, our paper is related to recent studies that examine the role of Enlightenment ideals in Europe and their effect on industrialization (Squicciarini and Voigtländer, 2015; Almelhem et al., 2023). Our focus is on the spread of technology outside of Europe and the context of the Enlightenment.

suggest that the Japanese state may have been uniquely successful in relaxing key constraints to adopting Western technology.

Third, our paper makes a methodological and data contribution to the quantitative study of economic history. In particular, the paper contributes to a long and rich tradition in economic history that uses country-specific data sources to reconstruct GDP and sectoral output (Bolt and van Zanden, 2020). We view our work as providing a complementary approach utilizing more widely available trade data. Our approach may be particularly useful in data-scarce environments, as is the case for many economies in the global periphery during the nineteenth and early twentieth centuries. As the geographic and temporal scope of detailed bilateral trade data widens, we hope this approach may yield new insights about economic growth at this critical juncture in economic history.

The remainder of the paper is structured as follows. Section 2 discusses the data we use and how we measure codification and British Patent Relevance. Section 3 discusses why Japan adopted its technology policy and provides details on the historical context and what the Japanese government did. Section 4 explains how we estimate productivity growth. Section 5 explores how Japanese export and productivity growth in the Meiji period was exceptional by international standards. Section 6 presents our main results, and Section 7 concludes.

2 Data

One of our main contentions is that relative to other regions in the periphery, Japan made a uniquely large investment in codification and technical education that enabled its firms to rapidly assimilate British technology and increase productivity in the sectors that stood the most to gain from Industrial Revolution technologies. Testing this hypothesis requires us to construct several novel datasets. In this section, we describe the main datasets used to conduct the empirical analysis. First, we discuss how we quantify the British supply of technical knowledge by sector. Second, we show how we quantify technology access costs for dozens of languages by measuring the extent of codification in each of them. Third, we create a bilateral trade dataset that enables us to use the structural model of Costinot et al. (2012) to measure industry productivity growth. The appendix contains a complete discussion of all data used, including all data construction steps and sources.

2.1 Constructing the British Patent Relevance measure

A key challenge for this paper is to quantify the supply of technical knowledge by industry available to Japan and other regions. Our approach uses natural language processing to quantify one of the main channels through which codified technical knowledge diffused during this period: the translation and publication of technical manuals. We first describe industrial manuals and show that they contain relevant information on production processes. We then describe how we use natural language processing to quantify the amount of new Industrial Revolution technologies contained in each industry.

2.1.1 Industrial manuals as repositories of codified technical knowledge

We focus on British technology for two reasons. First, Britain is generally perceived as the technological leader during the Industrial Revolution (Broadberry, 1994; Crafts, 1998; Rosenberger et al., 2024).[7] Second, the Japanese government also seems to have considered Britain as the technology leader. As we discuss below, one of the major technology transfer policies was to pay for

[7] More precisely, while recent empirical evidence suggests Britain did not have the technological lead in *all* sectors (Hallmann et al., 2021), the sectors Britain did lead in were the ones more central in the innovation network, such as steam engines (Rosenberger et al., 2024).

ten thousand person-years of foreign instruction to create a modern university system and train bureaucrats and the public more generally in Western technology (Jones, 1980). Since almost two-thirds of the foreign instructors brought to Japan were from Britain and other English-speaking countries, we assume that British technology and technology codified in English were the main sources of codified technical knowledge sought by the government.

Nineteenth-century technical manuals in English give detailed, practical descriptions of the technological and organizational aspects of an industry. Their audience was the practitioner, the entrepreneur setting up a plant, or the manager overseeing production. Their value lay in the fact that they contained precisely the type of technical knowledge entrepreneurs would need to familiarize themselves with for the setting up of modern, factory-based manufacturing, as well as their day-to-day operation.

2.1.2 Quantifying the supply of technical knowledge by industry

We use a text-based approach to quantify how much new technical knowledge was created by the Industrial Revolution in each industry. This consists of four main steps (we provide a complete description in Appendix F). First, we use the text of technical manuals as a measure of production methods at the technology frontier. Second, we take patent text as a measure of technical knowledge created during the Industrial Revolution. Technical manuals and patent text thus constitute the data sources we rely on. The third step entails using standard natural language processing techniques to represent the text as data. Fourth, we compute the similarity of text in industry technical manuals and patent text for each industry. This similarity score captures how relevant Industrial Revolution technologies were for an industry's production processes. We now describe each step in turn.

The previous section established that technical manuals contained information about state-of-the-art production and organizational methods by industry. We use the full text of these historical technical manuals directly. Specifically, for each three-digit, standard industrial trade classification (SITC) revision 2 sector in our data, we create a curated list of technical manuals from the *HathiTrust Digital Library*. Table 1 shows a random sample of the 460 books we selected from the *HathiTrust Digital Library*. We take the full text of these technical manuals to represent frontier knowledge of codifiable production techniques. Figures 1 and 2 illustrate the type of information we collect, using word clouds for unigrams and bigrams in two industries: textile yarn and fuel wood and charcoal. Reassuringly, high-frequency unigrams and bigrams contain words associated with the production processes. The most common unigrams in books explaining textile yarn production include words like "spindle," "shaft," and "card," and common bigrams are "front roller," "driven pulley." In contrast, the unigrams and bigrams used in technical manuals about fuel wood and charcoal production are words and phrases like "billet," "hearth," "coal process," and "smoke vent".

We measure the potential supply of British Industrial Revolution technology to all industries by digitizing the synopses of all British patents issued between 1617 and 1851 from Bennet Woodcroft's "Subject Matter Index of Patent of Invention." While patents capture only a subset of innovation (Moser, 2005), the part they do may be more relevant for measuring innovation that could diffuse at a (potentially very large) distance. The main alternative form of protecting intellectual property is secrecy, which inherently limits the technology's ability to diffuse. Likewise, knowledge that cannot be codified is harder to diffuse and absorb and will be absent from technical manuals. For these reasons, the patent text should proxy new technologies of the Industrial Revolution that could most easily diffuse. Figure 3 plots the unigrams and bigrams for British patent synopses. High-frequency unigrams include "engine," "spin," "weave," "steam," "loom," and "boiler;" high-frequency bigrams include "steam engine," "fibrous substance," and "motive power". Thus, these

Table 1: Random Sample of Book Titles from the *HathiTrust Digital Library*

SITC	Industry Description	Book Title
232	Natural rubber latex; rubber...	India rubber and gutta...
786	Trailers, and other vehicles,...	A complete guide for coach...
112	Alcoholic beverages	Hops; their cultivation,...
023	Butter	Butter, its analysis and...
764	Telecommunication equipment,...	The speaking telephone,...
882	Photographic and...	On the production of positive...
263	Cotton	Cotton in the middle states :...
274	Sulphur and unroasted iron...	A theoretical and practical...
271	Fertilizers, crude	American manures; and...
897	Gold, silver ware, jewelry...	Diamonds and precious stones,...
098	Edible products and...	Peterson's preserving,...
898	Musical instruments, parts...	Musical instruments ...
553	Perfumery, cosmetics, toilet...	A practical guide for the...
212	Furskins, raw	The trapper's guide: a manual...
046	Meal and flour of wheat and...	The American miller, and...
844	Under garments of textile...	Garment making a treatise,...
641	Paper and paperboard	Paper & paper making ancient...
664	Glass	The art of glass-blowing, or,...
268	Wool and other animal hair...	Sheep husbandry; with an...
061	Sugar and honey	The Chinese sugar-cane; its...

Note: This table provides a sample of books we used describing the technology in each industry. We randomly picked 20 industries, and for each industry, we randomly picked one of the books assigned to it.

Figure 1: Word Clouds for Textile Yarn

(a) Unigrams (b) Bigrams

Figure 2: Stemmed Word Clouds for Firewood and Charcoal

(a) Unigrams (b) Bigrams

Figure 3: Stemmed Word Clouds of Patent Synopses

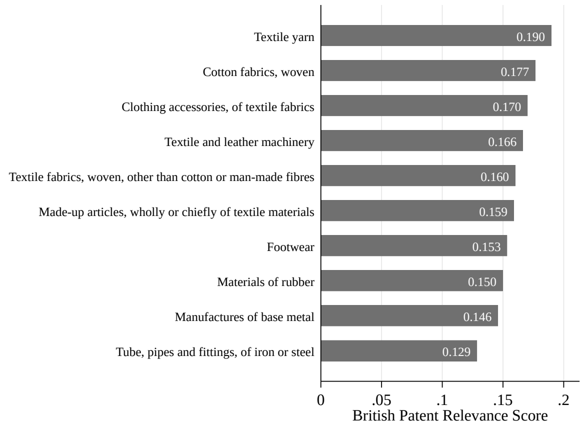
(a) Unigrams (b) Bigrams

unigrams and bigrams capture the technologies and concepts used in the Industrial Revolution. Moreover, unigrams like "spin," "bobbin," and "silk" are high-frequency words in both patent synopses and textile-yarn manuals, consistent with our prior that British patents were relevant for textile-yarn manufacturing.

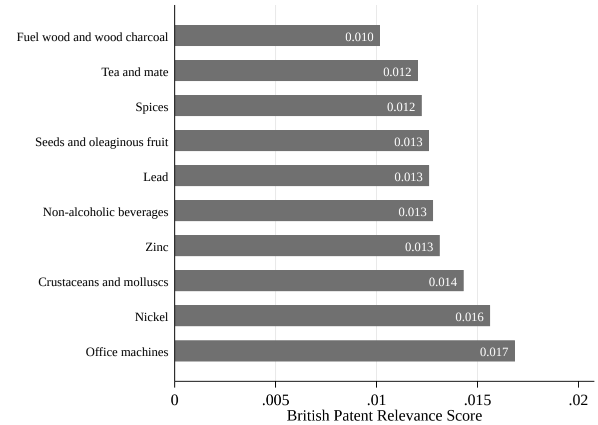
With these two data sources, we have measures of the state-of-the-art production methods by industry (the technical manuals) and new technologies of the Industrial Revolution (British patent text). The third step is to represent this textual information as data, which we do by taking a vector representation of the text. Each set of manuals associated with industry i or the set of patents is represented by a vector of length n, where n is the vocabulary size across the entire corpus. The vocabulary includes unigrams and bigrams and employs the term frequency-inverse document frequency (TF-IDF) weighting to account for the fact that some words appear more frequently across all documents.

The fourth and final step is to quantify the technological relevance of British patents for the

Figure 4: Industries Ranked by British Patent Relevance



(a) Top 10



(b) Bottom 10

processes described in the technical manual for an industry. We assume that if an industry's manuals use words and phrases that are similar to those in patent synopses, the patents are likely relevant for the industry. The standard metric for measuring the similarity of two texts (documents) in natural language processing is cosine similarity, which motivates our adoption of this measure. In our setup, it equals to the cosine of the angle between the vector representation of the word frequencies in manuals and the frequencies in patent synopses. Formally, the cosine similarity between the vectorized Bennett Woodcroft patent text $\boldsymbol{BW}$ and the vectorized technical manuals $\boldsymbol{TM_i}$ for industry i is

$$BPR_i \equiv \frac{\boldsymbol{BW} \cdot \boldsymbol{TM_i}}{\|\boldsymbol{BW}\|\|\boldsymbol{TM_i}\|} = \frac{\sum_{j=1}^{n} BW_j TM_{ij}}{\sqrt{\sum_{j=1}^{n} BW_j^2}\sqrt{\sum_{j=1}^{n} TM_{ij}^2}}, \tag{1}$$

which we call *British Patent Relevance (BPR)* because it measures the relevance of knowledge contained in British patents for each industry. Figure 4 plots the bar chart for the industries with the ten highest and lowest cosine-similarity scores. Reassuringly, high BPR industries include textile, footwear, machinery, and manufactured intermediate-input sectors, whereas low BPR industries contain mostly unprocessed raw materials, which were largely unaffected by Industrial Revolution technologies.

Our measure has several advantages for our setting. Most importantly, focusing on knowledge codified in technical manuals captures one of the key channels through which nineteenth-century Japan acquired Western knowledge: the translation of these documents. Moreover, this measure naturally accounts for how a given technology benefited different industries through input-output linkages. For example, since industries that make use of steam engines will likely have technical manuals that use the bigram "steam engine," our cosine-similarity measure will naturally quantify which industries benefited more from steam engines—a distinct advantage relative to using the industry classification of patents as a measure of relevance which only match the final output sector with patent about making that final output.

2.2 Measuring the codification of knowledge around the world

We collected data on codified technical knowledge available in local vernaculars every year for 33 languages, encompassing the twenty languages with the most speakers. We define the set of books containing technical knowledge as those with a subject that can be classified as applied sciences, industry, technology, commerce, and agriculture. We exclude books on theoretical technical knowledge, such as books in the hard sciences or subjects that do not directly benefit firms (e.g., medicine). After defining a common set of subject codes, we scraped the catalogs of national or other major libraries for books in the vernacular published in each year and report cumulative totals for each language (See Appendix H for details).

For many major European and Asian languages (e.g., English, French, German, Chinese, and Japanese), we scraped the national libraries of countries where the language is the native tongue of a substantial fraction of the population. For many other languages (such as Arabic and Russian), we could not find a scrapable national library. Instead, we scraped WorldCat, an online catalog of over 15,000 libraries worldwide covering dozens of languages. We also supplemented data from the National Diet Library (NDL) for Japan by scraping an additional 81 major libraries to obtain close to the universe of books anywhere in the country. We, therefore, present two samples of books for Japan. The "Japanese: NDL" sample is based on the holdings of just the NDL and is thus methodologically comparable to the methods used for other major codifying languages; and the "Japanese: All" sample contains all books contained in any major Japanese library and is a more comprehensive measure for Japan that we use when international comparisons are not needed. Using the publication year of each book in our sample, we construct the time series of codified knowledge by spoken language. This yields what is, to the best of our knowledge, the first systematic dataset on codified technical knowledge available in the vernacular for major languages.

2.3 Cross-region, bilateral, industry level trade flows

We construct the first cross-region dataset of harmonized, bilateral, industry-level trade flows quinquennially for 1880-1910 using detailed historical trade records for Japan, the United States, Belgium, and Italy ("reporting countries," henceforth).[8]

We combine existing, region-specific data sources and add newly digitized trade data from various sources. Specifically, we digitized data on US trade flows (exports and imports), Japanese exports for 1875, and quinquennial Japanese imports between 1875 and 1910. We use existing data for Belgian exports and imports in manufactures from Huberman et al. (2017); for Japanese exports from Meissner and Tang (2018); and for Italian exports and imports (for major trading partners only) from Federico et al. (2011).[9] An observation in this dataset, x_{ijk}, refers to an export flow in sector k from origin i to destination j. Using the fact that an export flow from i to j is equivalent, in theory, to imports from j to i, we can use import flows from reporting countries for unobserved regions' export flows.

[8]Recent years have seen a proliferation of high quality, cross-country, bilateral trade datasets (see, e.g., Fouquin and Hugot 2016; Pascali 2017; Xu 2022). Yet because these data are not disaggregated at the industry level, they cannot be used for our purposes.

[9]We do not include Germany's digitized trade data in the combined dataset because Germany's historical trade statistics before 1906 present several distinct methodological challenges that make comparison over time and across countries difficult (Hungerland and Wolf, 2022). First, until 1888, some parts of the German Empire were not part of the German customs union, and kept their own records, making the construction of a single dataset accounting for all German trade challenging. Second, during our sample period, the classification scheme for products was revised multiple times: at different points in time, between 400-1,200 distinct products were listed, making it difficult to construct a consistent classification over time (see

Table 2: Summary Statistics

Variable	N	Mean	SD	p25	p50	p75
Double-Relative Productivity Growth of Industry k in the Country i ($\tilde{y}_{ik}$)	1246	0.00	0.04	-0.01	0.00	0.02
Double-Relative Productivity Growth of Industry k in Japan	56	0.00	0.05	-0.01	0.01	0.02
Exporter's Industry Growth Rate	1397	-0.10	0.38	-0.05	0.03	0.09
Exporter's Industry Growth Rate in Japan	71	-0.05	0.37	-0.02	0.04	0.15
Britsh Patent Relevance	125	0.07	0.09	0.03	0.05	0.08

Japanese trade data does not include its colonies, so Japanese territorial expansion over this period does not affect our results. We define the set of non-Japanese Asian Regions (ASIA) as French East Indies, Hong Kong, China, Korea, Portuguese East Indies, Siam, Straights Settlements, and India. We used the Maddison data to divide the set of *non-Japanese* exporters into three terciles—High (H), Medium (M), and low (L)—according to estimated GDP per capita in 1870. For regions that do not correspond to modern countries, we use the average GDP per capita of the countries in that region.

We harmonized product lines in a manner consistent with the other pre-existing data sources used in the dataset. We conducted extensive validation exercises to ensure that similar product lines were consistently concorded to the same three-digit SITC category across all datasets. Region names (and boundaries) were harmonized within and across datasets. All trade values were converted to yen (at current exchange rates) using historical exchange rates from Fouquin and Hugot (2016).

Our dataset consists of export values for 39 regions in 91 industries. Table 2 contains the summary statistics.

3 Japanese Industrialization and Technology Policy: Historical Context

Nineteenth-century Japan presents an interesting study of late industrialization. In a very broad sense, Japan in the 1870s was similar to other poor, predominantly agricultural areas of the global periphery which had missed out on the first wave of industrialization. The most recent estimates of GDP per capita and wages for Japan confirm that it was a low-growth, low-wage society diverging from northwestern Europe until at least the 1870s (Bassino et al., 2019; Kumon, 2022).[10] In this section, we introduce and provide novel evidence for the main stylized facts that motivate our empirical analysis. We first examine Japan's unique industrialization amongst economies in the periphery. Second, we discuss Japan's technology policies and show how they contributed to Japan's rapid and unprecedented codification of technical books in the periphery.

Hungerland and Wolf (2022, Figure 6A)).

[10]There is some debate about whether proto-industrialization was experienced in the late Tokugawa period. Building on Saito and Takashima (2016), Bassino et al. (2019) estimate annual GDP per capita growth of 0.26 percent for the period 1721-1874, driven by growth in the secondary and tertiary sectors. Based on these findings, the authors argue that Japan may have improved its relative position within Asia even before the Meiji Restoration in 1868. However, Kumon (2022) casts doubt on these findings and argues that there is no direct evidence for growth in the secondary and tertiary sectors for this period.

Figure 5: Manufacturing Share of Exports (Japan)

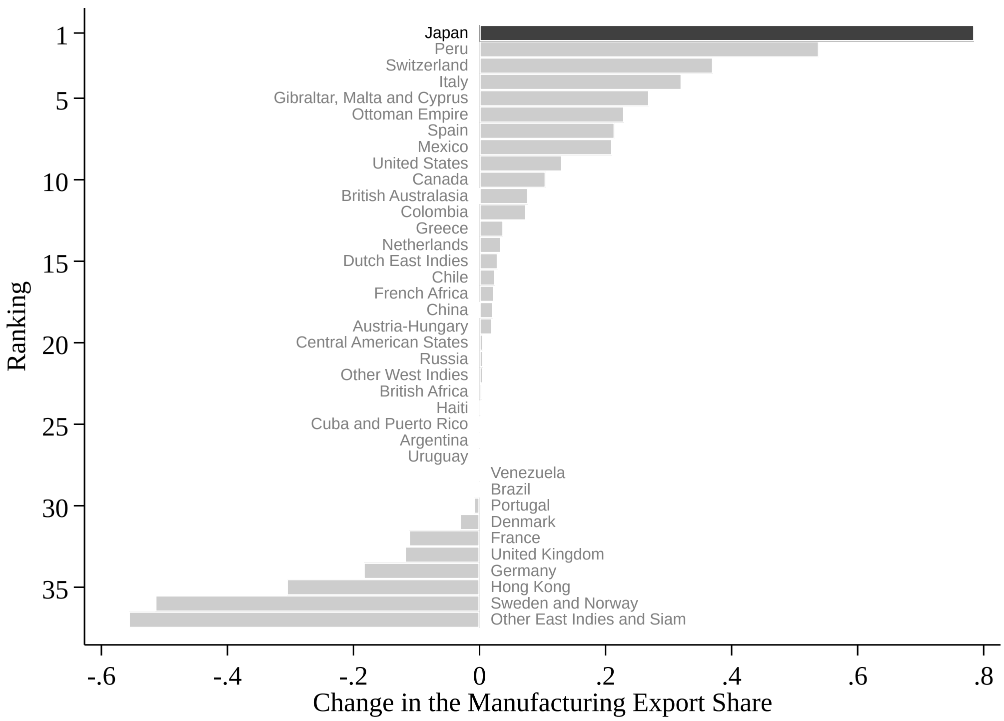
Note: Data sourced from Oriental Economist (1935) *Foreign Trade of Japan: A Statistical Survey.* Tokyo: Toyo Keizai Shinposha.

3.1 Japan's shifting export composition

Starting in the closing decades of the nineteenth century, Japan began to industrialize (e.g., Yamamura 1997). To understand the timing, speed, and scale of this change, Figure 5 plots the manufacturing share of exports in Japan during this period. Consistent with other evidence of a stagnating economy cited above, the trade data show an economy highly specialized in the exports of primary products until the early 1880s. Indeed, between 1868 and 1883, manufacturing exports as a share of total exports fell by almost a third. This downward trend suddenly reversed in the 1880s. Between 1883 and 1896, the share of Japanese manufacturing exports tripled and remained at this high level for over forty years. Since this shift happened twenty-five years after Japan opened to trade, it is hard to explain the shift in terms of Japan simply having a comparative advantage in manufacturing.[11] As a result, by the turn of the twentieth century, Japan was specialized in the export of manufactured products. Moreover, Figure 6 shows that no other economy displayed a similar change in its export composition, which raises the question of what happened in Japan and why it didn't happen elsewhere.

[11]Similarly, Japan's Meiji reforms began in 1868 and many of the most important ones, like tax reform, foreign missions, peak hiring of foreigners, educational reform, postal reform, telegraph construction, banking reform, military reform, judicial reform, etc., were implemented by 1875, so it is not obvious why these reforms should have caused the manufacturing export share to continue to fall only to sharply reverse in the 1880s.

Figure 6: Change in the Manufacturing Share of Exports by Country (1880-1910)



Note: Percentage point change in the share of manufacturing exports relative to total exports of the country. For non-reporting countries, we use imports to reporting countries to estimate exports. We include the following SITC categories in manufacturing: code 6 (Manufactured goods classified chiefly by material), 7 (Machinery and transport equipment), 8 (Miscellaneous manufactured articles), 95 (Armoured fighting vehicles, war firearms, ammunition, parts, n.e.s.), 96 (Coin (other than gold coin), not being legal tender).

3.2 Meiji technology policy

We hypothesize that a state-led technology absorption effort, unique in scale, provided widespread access to the technical knowledge needed to adopt the technologies of the Industrial Revolution. After centuries of self-imposed isolation, the U.S. forcibly opened Japan to foreigners in 1854 and to trade with Western countries in 1858. The Tokugawa shogunate, which had ruled Japan since the 1600s, was overthrown in the 1868 "Meiji Restoration" and replaced by an oligarchy of rival nobility ruling in the name of Emperor Meiji. From its inception on April 5, 1868, the Meiji Government stated that the assimilation of Western knowledge would be a central policy tenet. The Charter Oath, Emperor Meiji's five-sentence statement of the objectives of the fledgling government, declared that "knowledge shall be sought throughout the world so as to strengthen the foundations of imperial rule." (Hirakawa, 2007, p. 338) Thus, all members of the new government were required to support strengthening Japan by absorbing Western ideas.

However, Japanese historians argue that many members of the Tokugawa shogunate had already realized in the aftermath of China's ignominious defeat in the First Opium War (1839-1842) that Japan needed a strategy to absorb Western science (Bolitho, 2007, p. 157). Early Japanese reformers, most prominently Shozan Sakuma, began developing plans for how Japan could co-exist with the West. Sakuma developed a strategy for modernizing Japan, which he summarized with the slogan "Eastern morality, Western technology." While there was little concrete action until U.S. warships entered Edo harbor in 1853, the arrival of the Americans caused the shogunate to spring into action. Almost immediately after the Americans arrived, the Japanese government established the Institute of Barbarian Books (*Bansho Torishirabesho*), which was tasked with developing English-Japanese dictionaries to facilitate technical translations. This project was the first step in what would become a massive government effort to codify and absorb Western science.

This section discusses the three components of Japan's technology absorption effort. First, we describe the effort to codify Western technical knowledge. Second, we show that by investing in elementary and university education, the government ensured the population had the necessary skills to absorb and use the technical knowledge they supplied. Third, we discuss how the government raised enough tax revenue to finance these costly policies.

3.2.1 The effort to codify Western technical knowledge

Linguists and lexicographers have written extensively on the difficulty of scientific translation between dissimilar languages (c.f. Clark 2009; Kokawa et al. 1994; Lippert 2001; Montgomery 2000). Technical translation is relatively easy in languages in which jargon is based on the same Greek and Latin roots. Thus, a speaker of French or German can easily guess that an English technical word like "telegraph" should be translated as "télégraphe" or "telegraph," respectively, and it would be easy for readers of all three languages to remember from their knowledge of Greek root words that telegraphs involve the transmission of words across distance.

Translation of English jargon into languages with root words not based on Greek and Latin is much harder. For example, a typical speaker of Arabic would be hard-pressed to guess the meaning of the Arabic word "tiligraaf," from its spelling any more than a typical English speaker could guess the meaning of "algebra" (which comes from the Arabic word "al-jabr") from its spelling. People who did not understand languages closely related to English needed to translate vast amounts of English jargon to understand modern production techniques. Consistent with this, in Appendix Table A.1, we provide suggestive evidence that linguistic distance was a barrier to technology diffusion and, ultimately, economic growth in this time period. Specifically, we show that GDP per capita in 1870 and 1913 tended to be lower in countries and regions speaking a plurality language that was more linguistically distant from English, *conditional on physical distance*.[12] While we do not interpret this negative relationship causally, we take it as suggestive empirical evidence consistent with scientific translations increasing technology access costs and inhibiting technology diffusion.

The linguistic problem Japan faced in translating Western science was two-fold. First, no words existed in Japanese for canonical Industrial Revolution products such as the railroad, steam engine, or telegraph, and using phonetic representations of all untranslatable jargon in a technical book resulted in transliteration of the text, not translation. Second, translations needed to be standardized so that all translators would translate a given foreign word into the same Japanese one, a classic example of a coordination problem.

Creation of new words and dictionaries to facilitate translation. Solving these two problems became one of the Institute of Barbarian Books' main objectives. After carefully studying how to solve these problems, Japanese translators decided to base the translation of English jargon on Chinese glyphs whose meaning was known to all literate Japanese (c.f. Kokawa et al. 1994; Lippert 2001; Clark 2009). For example, in order to translate the word "telegraph," Japanese translators created a word using glyphs that combined the Chinese characters for "electric" (*den*) and "message" (*shin*). While a Japanese reader coming across the term "electric message" might not recognize that it means telegraph on the first reading, it is easy to remember it once one learns the definition. Lippert (2001) argues that Japanese government translators' decision to create Japanese jargon based on Chinese root words is an important factor in making foreign scientific texts much easier to understand in Japanese than in alphabetic languages not based on Latin and Greek, where jargon is just transliterated. In other words, the Japanese government's investment in carefully rebuilding the Japanese language to accommodate new words may have been a public good that lowered technology access costs for Japanese people.[13]

The importance of this strategy for codification was not lost on Japanese reformers. For example, Sakuma wrote of the first English-Japanese dictionary, the ETSJ (*Eiwa-Taiyaku-Shuchin-Jisho* or "A Pocket Dictionary of the English and Japanese Language"), which was published by the Institute in 1862, "I would like to see all persons in the realm thoroughly familiar with the enemy's conditions, *something that can best be achieved by allowing them to read barbarian books as they read their own language.* There is no better way to enable them to do this than by publishing this dictionary" (Hirakawa, 2007, p. 442, emphasis added). A much larger dictionary supplanted this small dictionary, the FSEJ (*Fuon-Sozu-Eiwa-Jii* or "An English and Japanese Dictionary") in 1871, which contained two to three times as many words and a significant amount of English jargon.[14]

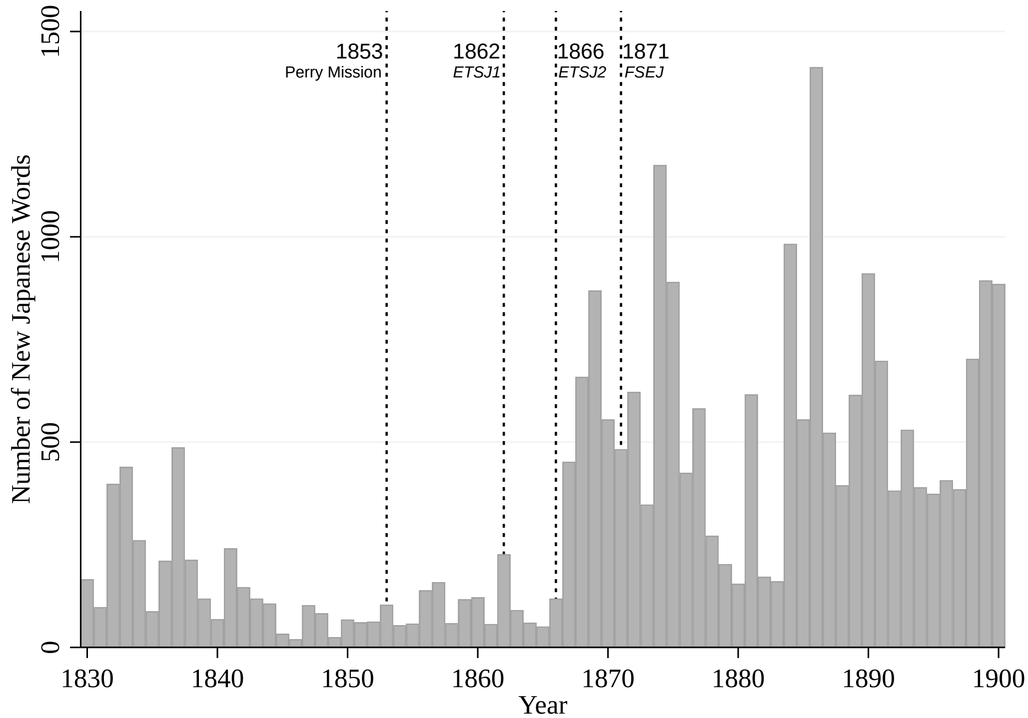
In Figure 7, we present suggestive evidence that solving the coordination problem associated with translating jargon facilitated the emergence of new words in Japanese. We obtained the first recorded use of Japanese words based on the revised edition of the *Nihon Kokugo Daijiten* (Large Japanese Dictionary), published by Shogakukan, encompassing 300,000 Japanese words. Word creation in Japan before the 1860s was surprisingly low—typically, only around 100 new Japanese words were created each year. Even in the first decade after Japan opened to the West following the Perry Mission, the rate of new-word creation in Japan remained essentially unchanged. This result is quite surprising given that in 1854, the Americans brought many pieces of new technology to

[12] Appendix Figures A.1 and A.2 show the scatterplots for this relationship.

[13] Since people living in Japanese colonies (e.g., Taiwan and Korea) were required to learn Japanese, literate people in these regions would have also benefited from these easy-to-remember translations.

[14] Kokawa et al. (1994, pp. 80-119) is the source for our information on dictionaries. Publication and release dates are difficult to pinpoint exactly in this period. The Pocket Dictionary was first released in 1862 with a print run of only 200 copies but was reprinted and distributed much more widely in 1866. Similarly, the FSEJ was printed on a linotype machine in 1871 but officially published in 1873.

Figure 7: Word Creation in Japan



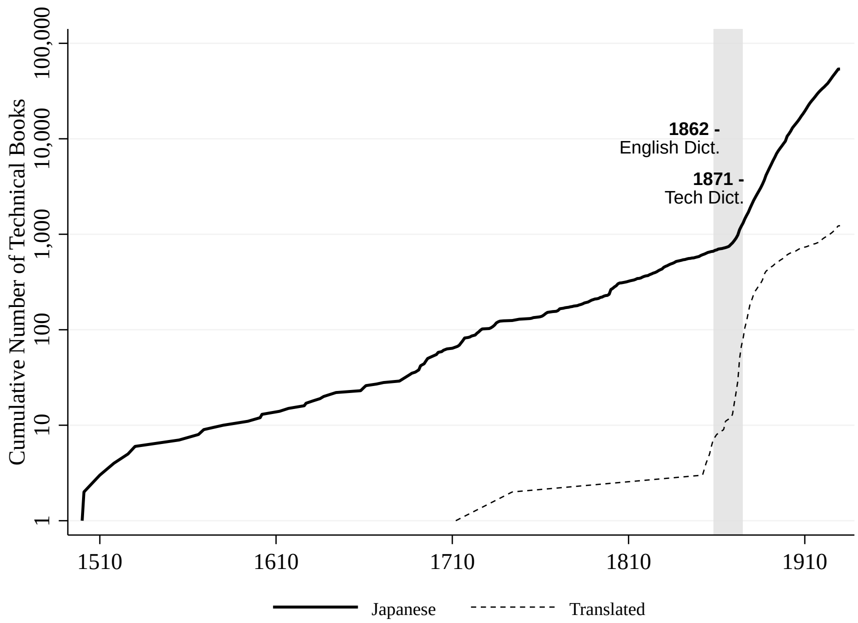
Note: Number of new words created in Japanese from *Nihon Kokugo Daijiten*. The dictionary contains information on the first known time a word was used in a document, which we use to construct this graph. Dashed lines refer to the Perry Mission and publication dates of English-Japanese dictionaries.

show to the Japanese, such as a working locomotive, a telegraph machine, cameras, etc. Exposing the Japanese to Western technology did not, in and of itself, lead to the emergence of new words. However, starting around the creation of the first English-Japanese dictionary (ETSJ1) in 1862 and accelerating with the large print run of this dictionary in 1866 (ETSJ2), the number of new words in Japanese rose to around 500 per year. Word creation accelerated to over 1000 words per year following the release of the extensive English-Japanese dictionary, the FSEJ. Thus, to the extent that new word creation tracks new ways of codifying and conceptualizing the world, this evidence suggests that the government-led dictionary creation efforts in the 1860s helped solve the coordination problem inherent to introducing new ideas.

State-led codification of technical knowledge. Alongside the public provision of dictionaries, the public sector played an outsized role in *translating* technical books. A search through the biographies of every person who translated a technical book between 1870 and 1885 contained in a major Japanese biographical dictionary, Heibonsha (1974) revealed that 74 percent of the translators were government employees. This number is likely a lower bound because Heibonsha (1974) does not necessarily list every job a person held.

What were the effects of this effort to codify technical knowledge? In Figure 8, we examine the evolution of technical knowledge in the Japanese language. We do so by scraping the catalogue of the National Diet Library and 81 additional Japanese libraries to construct a time series of all technical books from 1500 to 1930. Between 1600 and 1860, the number of technical books in

Figure 8: Codified Technical Knowledge in Japan



Note: Codified technical knowledge for each year refers to all technical books written in Japanese in the NDL catalog or any other Japanese library linked to the NDL. A book is considered "Translated" if the NDL flags the book as a translation.

Japanese grew by 1.6 percent per year. The rate almost sextupled to 8.8 percent per year between 1870 and 1900, starting just as staff at the Institute for Barbarian Books produced the 1862 and 1871 English-Japanese dictionaries. After centuries in which the number of technical books written in Japanese doubled every 44 years, the number suddenly began to double every eight years. In other words, Japan's emergence from its Malthusian equilibrium is associated with a massive increase in the growth rate of codified technical knowledge. We see an even sharper increase in translated technical books. Japanese translators had only succeeded in translating 8 Western technical books between 1500 and 1860; by 1900, they had translated 608 books. As the figure shows, the growth rate of new technology entering Japan changed suddenly and sharply after the government produced English-Japanese dictionaries and subsidized technology absorption.

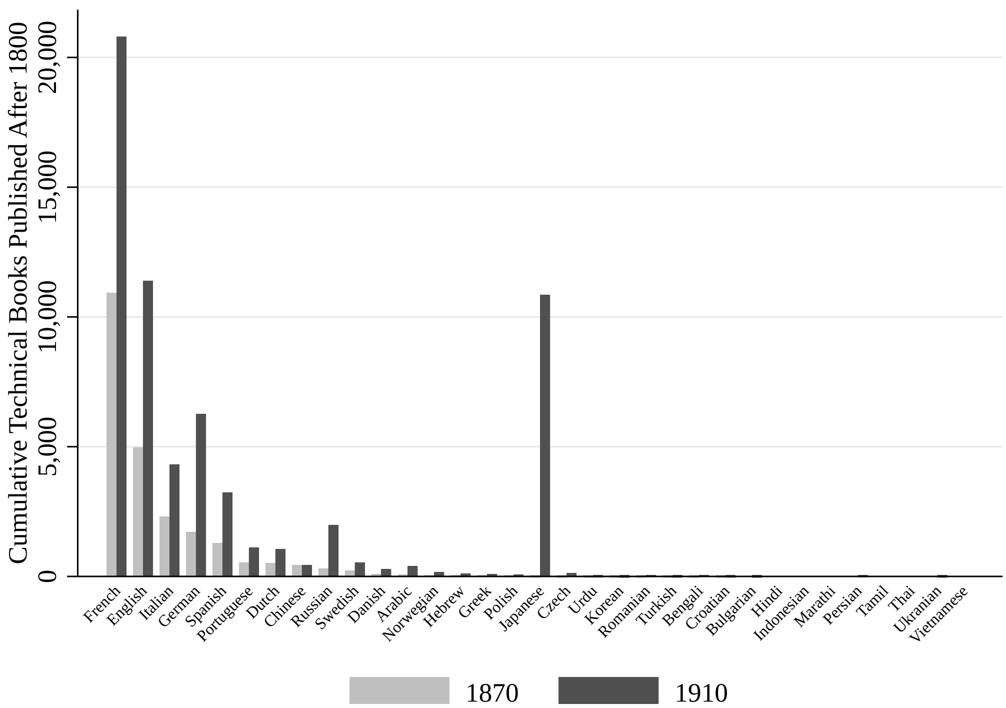
Despite Japan's rapid progress in codification at the end of the nineteenth century, the level of Japanese codification in 1870 was quite ordinary. Figure 9 presents the extent of codification of technical knowledge in 1870 and 1910. Two features of the data stand out. First, in 1870, 84 percent of all technical books were written in four languages: English, French, German, and Italian.[15] This puts into comparative perspective the achievements of the Enlightenment. There was little to no codification in any non-European language, meaning that people who could not speak European languages were technically illiterate. Second, this figure also puts into comparative perspective the achievements of Japan's public provision of technical knowledge *after* 1870. Starting from a level in 1870 that was comparable to other languages spoken in the global periphery, by 1910, within the

[15]We probably underestimate codification in English because we could not scrape the British Library, which suffered a cyber attack in 2023, resulting in security measures that prevented us from scraping it. Our data for English is from the Library of Congress, which provides a measure of technical books written in English and available outside of England.

space of a few decades, Japan had amassed a body of codified technical knowledge comparable to that of the languages of the Enlightenment. Moreover, nowhere else in the global periphery do we see a similar increase in codified technical knowledge. This suggests that outside of the set of countries contributing to the Enlightenment, state intervention was necessary to create codified technical knowledge.

These findings motivate our empirical strategy. Understanding the effects of codifying technical knowledge on economic outcomes such as productivity growth is typically difficult—codification in most settings happened gradually over time, making it hard to rule out other explanations. The swift creation of codified technical knowledge in Japanese however creates two relatively well-demarcated periods; one in which Japan had no access to technical knowledge, and one in which it did.

Figure 9: Codified Technical Knowledge in Major World Languages



Note: Technical knowledge is measured as the number of books in the following subjects: agriculture, applied sciences, commerce, industry, and technology. The languages are ordered based on the total number of technical books published up to and including 1870.

How important was the supply of codified knowledge for absorbing Western technology in Japan? Historical evidence suggests that the translation of Western technical books played a central role in developing one of Japan's most important nineteenth-century industries: cotton textiles. Consider the story of Masatatsu Ishikawa, who established the first cotton textile mill in Japan. Horie (1960) reports that "while employing [Ishikawa as an advisor], the lord [Nariakira Shimazu] showed him a book. Because this was in English, he sent it to Nagasaki for translation into Dutch, and it turned out to be a book on the cotton spinning industry. The attention of the lord, who had been previously interested in machine spinning, was abruptly caught by the book, and the plan for building a cotton spinning mill was made... [Thus began] the Kagoshima Cotton

Spinning Mill, the forerunner of the modern spinning industry in Japan, which began operation in 1867." Braguinsky (2015) reveals that translating the book from English, a language Ishikawa never learned, into Dutch, a language Ishikawa understood, took a whole year. As one can tell from the passage, without English-Japanese dictionaries, technical books often could not even be translated directly into Japanese. This roundabout means of learning technology meant that it took eleven years from the time that Ishikawa was hired by Shimazu as a technical advisor before he could establish the mill. It is a clear case of language differences raising technology access costs and translation lowering them. One can find many other examples of Japanese entrepreneurs using books to guide their investment and production decisions. For example, Tamagawa (2002) writes, "Many large mills such as Osaka, Kanegahuchi and Kurashiki and others founded their own training schools for male workers, teaching fundamental spinning technology. The textbooks and the study aid books on cotton spinning were mainly translated versions of the Platt Bros.' catalogues and instructions."

3.3 Education Policies

Beyond spending on technology policies directly, the Meiji government also deployed education policies. This is important, as supplying technical knowledge in the vernacular would make little sense unless entrepreneurs and managers had the human capital to absorb it. Compulsory elementary school education began in 1872, although most Japanese parents refused to send their children to government schools because, in the words of an 1877 Ministry of Education report, the "people do not yet see education as useful and parents are complaining" (Rubinger, 2000, p. 170). Government pressure quickly overcame the anti-education attitude of non-elite Japanese. The fraction of boys and girls attending school rose from 39.9 percent of eligible boys and 15.1 percent of eligible girls in 1874 to 58.2 percent of boys and 22.6 percent of girls by 1879. By 1890, 90.6 percent of boys and 71.7 percent of girls were enrolled in elementary school (National Institute for Educational Policy Research, 2011). Since child labor was common at this time period, many of these elementary school graduates would have been in the labor force by the time they were teenagers.

These schools offered high-quality education by the international standards of the day. Rubinger (2000) argues that data from mandatory intake examinations for Japanese army conscripts provides us with a representative sample of young Japanese males that we can use to assess education levels. If one defines literacy as being able to write a formal letter in Japanese as judged by the Imperial Japanese Army, new conscripts in all but one of Japan's forty-seven prefectures in 1909 had literacy rates above 90 percent. Mathematics education was equally impressive. Conscripts that had completed six years of education were expected to answer word problems that required them to know algebra in order to solve, and those with eight years of education were expected to be able to compute bond yields. In other words, by the 1880s most Japanese young men could have read technical books.

The Japanese government faced a more complex problem in building a university system because there were almost no Japanese with advanced knowledge of STEM fields. To alleviate this problem, the Japanese government hired 2,400 foreigners to come to Japan as instructors or advisors (Jones, 1980), with many of them employed by newly founded public universities in the 1870s. Only four countries supplied a hundred people or more—Britain, the US, France, and Germany—with almost all other Europeans only sending ten or fewer people. The foreigners hired by the government provided Japan with 9,506 person-years of technical training, of which over half was deployed either in educational institutions or in ministries that oversaw the building of Japan's transportation, telegraph, and postal networks, as well as public works projects. Japan chose

Britain as the most popular source country for instructors in Industrial Revolution technologies, accounting for 46 percent of the total person-years of training. Adding in person-years from the U.S. and Canada reveals that 59 percent of the training was by people whose native language was English, 17 percent was by people from France and Belgium, and 13 percent was by people from Germany. The revealed preferences of the Japanese government in choosing instructors suggest that they saw instructors whose native languages were English and, to a lesser extent, French and German as the key sources of advanced Western technology. As we showed above, these languages contained the most codified technical knowledge. The government not only financed foreigners to come to Japan but also paid for Japanese to study abroad. For example, in 1871, the government sent around one hundred government officials and students abroad for two years on the "Iwakura Mission" to study Western society and technology and send back a wealth of information. In addition to this high-profile study mission, the government subsidized many other Japanese to study abroad. Foreign study trips accounted for up to 0.20 percent of annual government expenditures in the 1870s (Jones, 1980, Table 7). The new universities specialized in practical disciplines like foreign languages, medicine, engineering, mathematics, physics, law, chemistry, and management (Ministry of Education, 1980).

3.4 Paying For The Technology Transfer Policies

One may wonder how the Meiji government was able to raise enough government revenue to pay for these policies. As shown in Figure 10, real government expenditures tripled between 1871 and 1874. Paying for the foreign workers alone required substantial expenditures—about 2 percent of total government expenditures in 1876, one-third of the University of Tokyo budget, one-half of the Ministry of Education budget, and in 1879, two-thirds of the public works budget (Jones, 1980, p. 13).

The key to Japan's newfound ability to pay for these programs was the 1873 Land Tax, which Japanese economic historians have called "the single most important reform of the Meiji Restoration," (Hayami, 1975, p. 47). Interestingly, the idea of instituting a land tax had its origin in the work of government translators in the 1860s. As Yamamura (1986) discusses in detail, Takahira Kanda, a high-ranking Meiji official who had translated a book on economics in the Tokugawa period, realized that Japan could raise enormous amounts of tax revenue with limited efficiency loss by instituting a heavy land tax, as opposed to the earlier output tax, on farmers.[16] Figure 10 shows that the imposition of the land tax enabled the early Meiji government to finance enormous investments in codification and technical absorption. As a result of Japan's impressive ability to raise government revenues, by 1884, Japanese government revenues equaled 83.1 million yen. By contrast, the Chinese government in 1884, still recovering from the chaos of the Opium Wars and Taiping Rebellion, could only raise 114 million yen even though China had ten times Japan's population. This eight-to-one Japanese advantage in per-capita taxation enabled Japan to finance human capital investments and public goods at a rate that Chinese reformers could only dream about.[17] For example, Japanese government expenditures (taken from Ohkawa et al. (1965)) on education alone amounted to 11 percent of the budget in 1880; if China had attempted to implement

[16] Although agricultural taxes before the Meiji period were called "land taxes," they were closer to output taxes in implementation. As Ohno (2018) writes, "A new land tax at the initial rate of 3 percent of the assessed land value replaced the old rice tax that was levied on the annual yield of rice." (p. 39)

[17] Wong (2012) reports that Chinese tax revenue in 1884 was 77 million silver taels. We performed the currency conversion in two ways. The number in the text uses the exchange rate series from (Fouquin and Hugot, 2016) of 1.39. We obtain a similar estimate if we convert silver taels into yen by noting that an 1867 Shanghai silver tael contained 36.0 grams of silver and an 1876 silver yen coin contained 24.3 grams of silver, according to https://en.numista.com. This implies an exchange rate of 1.48 yen per tael.

Figure 10: Japanese Government Expenditure

Note: Government expenditure and revenue data are from Toyo Keizai Shimposha (1926) *Meiji Taisho Zaisei Shoran [Meiji and Taisho Financial Details]*, Toyo Keizai Shimposha: Tokyo. [pp. 2 and 640] Before adopting the Gregorian calendar in 1873, Japanese fiscal years varied in duration and did not align perfectly with the Western ones, but the mapping to Western years is approximately correct. These are deflated by the Wholesale Price Index from Ohsato, Katsuma ed. (1966) *Hundred-Year Statistics of the Japanese Economy* Statistics Dept., The Bank of Japan: Tokyo. p. 76.

this one part of the Meiji reform package for its population, it would have had virtually nothing left over for any other government functions.

In summary, the absorption of Western technology was a central aim of the Meiji government. To achieve this goal, the government adopted a multitude of large-scale technology and education policies. The funding of these fiscally intensive policies was made possible by the land tax reform, which itself was a product of Western "technology transfer." Starting in the mid-1880s, the historical record points to a marked shift in the Japanese economy. Pockets of modern, private, factory-based manufacturing began to emerge, predominantly in textiles (Yamamura, 1997). These textile mills used British machinery, inanimate power sources, and a modern industrial labor force. We now examine whether this shift in industrial structure can be linked to the codification and absorption of foreign technology. We note that we view the technology and education policies of the Meiji government as one "package" that made technical knowledge accessible to large swaths of the population. Our interest lies in understanding the effects of these policies.

4 Estimating Productivity Growth

In this section, we show how to use trade data to build a global database to measure productivity growth at the region-industry level. Here, we explain how we estimate productivity growth for our set of regions. In appendix C, we explain how we convert our estimates into annual growth rates.

4.1 Estimating Productivity Growth

Our starting point is the framework of Costinot et al. (2012) who build a multisector Eaton and Kortum (2002) model featuring an economy with multiple countries, multiple industries, and one factor of production, labor. They show that one can write the value of exports from i to j in industry k at time t (x_{ijkt}) as

$$\ln x_{ijkt} = \gamma'_{ijt} + \gamma'_{jkt} + \theta \ln z'_{ikt} + \epsilon'_{ijkt}, \tag{2}$$

where γ'_{ijt} is an importer-exporter fixed effect that captures bilateral trade frictions and exporter-importer aggregate supply and demand forces (e.g., country size and distance) that matter for exports; γ'_{jkt} is an importer-industry fixed effect that captures deviations in importer demand in industry k; θ is the Fréchet scale parameter, z'_{ikt} captures comparative advantage, i.e., factors that shift productivity in a given exporter and industry; and ϵ'_{ijkt} is an error term that captures how trade costs deviate at the industry-exporter-importer level from the exporter-importer average.

Our objective is to estimate $\gamma_{ikt} \equiv \theta \Delta \ln z'_{ikt}$ using trade data. We will estimate it by noting that we can first-difference equation (2) and rewrite it in terms of fixed effects:

$$\Delta \ln x_{ijk} = \gamma_{ij} + \gamma_{jk} + \gamma_{ik} + \epsilon_{ijk}, \tag{3}$$

where we have suppressed the time subscripts and $\gamma_{\ell,m} \equiv \Delta \gamma'_{\ell,m}$ for any index (ℓ, m). Estimating this equation enables us to identify γ_{ik} and therefore $\theta \Delta \ln z_{ik}$ up to the choice of a normalization that pins down the reference exporter productivity, importer demand, and industry productivity.[18] This equation can be rewritten to yield

$$\Delta \ln x_{ijk} = \gamma_{jk} + \gamma_{ik} + \tilde{\epsilon}_{ijk}, \tag{5}$$

where variables without primes correspond to the first differences of variables with primes and $\tilde{\epsilon}_{ijk} \equiv \gamma_{ij} + \epsilon_{ijk}$.

Estimation of equation (5) requires us to drop observations whenever the initial bilateral export flow in a exporter-importer-industry tuple is zero, which is problematic because a large amount of nineteenth-century export growth was due to exporters expanding their set of export destinations over time. This can bias estimates of productivity growth based on a log-difference specification downwards because it cannot account for growth due to the extensive margin. Amiti and Weinstein (2018) [AW] propose an alternative estimation approach that corrects this problem.

Their estimator is closely related to weighted least squares. In particular, if there are no zeros in the export data, the AW estimates will match those obtained using weighted least squares with lagged export weights. A unique property of the AW estimates of γ_{jk} and γ_{ik} is that they aggregate to match the growth rate of total exports in every region-industry in which the industry's *aggregate* growth rate is well defined: i.e., the region initially has positive exports to at least one country in the industry. Similarly, the estimates aggregate to match region-industry import levels as long

[18] One can see this by noting that equation (3) can be rewritten as

$$\Delta \ln x_{ijk} = (\gamma_{ij} + \gamma_i + \gamma_j) + (\gamma_{jk} + \gamma_k - \gamma_j) + (\gamma_{ik} - \gamma_i - \gamma_k) + \epsilon_{ijk}, \tag{4}$$

where γ_i, γ_j, and γ_k are arbitrary normalization constants that define the baseline exporter productivity, importer demand, and industry productivity.

as a region has positive imports from at least one country in the industry in the initial period. Thus, an export-weighted average of the γ_{jk} and γ_{ik} will match *total* export growth in each country and industry.[19] One can formally see that the AW estimator will have this property by writing down the moment conditions used to obtain the estimates. In particular, the estimates will satisfy two types of moment conditions. First, the estimates aggregate to match total exports in every exporter-industry observation i:

$$\frac{\sum_j x_{ijk,t} - \sum_j x_{ijk,t-1}}{\sum_j x_{ijk,t-1}} = \gamma_{ik} + \sum_j \frac{x_{ijk,t-1}}{\sum_\ell x_{i\ell k,t-1}} \gamma_{jk}, \tag{6}$$

where we have added a time subscript, t, to be clear about how time differences are constructed from changes in levels. The left-hand side of the moment condition equals the growth rate of *total exports* in sector k from exporter i, and the right-hand side is the sum of the exporter fixed effect (γ_{ik}) and a bilateral export weighted average of the importer fixed effects (γ_{jk}). This condition, therefore, ensures that an export-weighted average of the parameters aggregates to match total exports. Second, the estimates will aggregate to match total imports in every importer-industry observation j because they impose a second moment condition:

$$\frac{\sum_i x_{ijk,t} - \sum_i x_{ijk,t-1}}{\sum_i x_{ijk,t-1}} = \gamma_{jk} + \sum_i \frac{x_{ijk,t-1}}{\sum_\ell x_{\ell jk,t-1}} \gamma_{ik}. \tag{7}$$

Here, the left-hand side of this moment condition is the growth rate of *total imports* in sector k by importer j, and the right-hand side is the sum of the importer fixed effect (γ_{jk}) and a bilateral export weighted average of the exporter fixed effects (γ_{ik}). Since the estimates satisfy these two moment conditions, the AW estimates aggregate to match every region's export and import growth in each industry.

Once we obtain the estimates of γ_{ik} and γ_{jk}, we run the following regressions to impose normalizations that lead to a meaningful decomposition of global trade patterns:

$$\gamma_{ik} = \gamma_i + \gamma_{1k} + \tilde{\gamma}_{ik}, \tag{8}$$

and

$$\gamma_{jk} = \gamma_j + \gamma_{2k} + \tilde{\gamma}_{jk}, \tag{9}$$

where $\tilde{\gamma}_{ik}$ and $\tilde{\gamma}_{jk}$ are regression residuals. This normalization choice has several useful properties. First, γ_i tells us the growth in exports due to shifts in exporter characteristics (e.g., productivity or size). Second, $\tilde{\gamma}_{ik}$, the "comparative-advantage" component of productivity growth, corresponds to the growth in exports due to shifts in productivity that are orthogonal to changes in exporter factors (i.e., γ_i) and changes in industry factors (γ_{1k}).[20] Finally, recalling that $\gamma_{ik} \equiv \theta \Delta \ln z'_{ikt}$, we

[19]We also considered using the Poisson pseudo-maximum likelihood (PPML) estimator. However, one well-known issue with PPML is that it often fails to converge in datasets with many zeros like ours (Santos Silva and Tenreyro, 2010). While the AW estimator only required us to drop country-industry observations where there were no exports to or imports from *any country* in the initial period, the PPML estimator did not converge unless we used data for countries with at least two export destinations or two import sources in each industry. As a result, while the AW procedure produced 1,358 productivity estimates based on 6,216 observations, the PPML estimator only converged on a subsample that was 36.5 percent as large. The PPML estimator only produced 38 percent as many productivity growth estimates as the AW estimator.

[20]Although we do not use the other normalization constants, we can recover them. $\hat{\gamma}_k \equiv \hat{\gamma}_{1k} + \hat{\gamma}_{2k}$ is the shift in exports that can be attributed to movements in industry k's characteristics (e.g., global productivity growth in k or global demand for k). Similarly, γ_{ij} can be recovered by regressing $(x_{ijk,t}/x_{ijk,t-1} - 1 - \hat{\gamma}_{ik} - \hat{\gamma}_{jk})$ on ij fixed effects.

can define $\Gamma_{ik} \equiv \tilde{\gamma}_{ik}/\theta$ as the change in exporters i's productivity in industry k that cannot be explained by changes in industry factors or general conditions in the exporting country.[21]

In the following sections, we estimate γ_i and Γ_{ik} to understand patterns of productivity growth worldwide. We implement this methodology on annualized trade growth rates for the sample period (1880-1910), so our estimates correspond to averaged annual productivity growth rates. We show how to construct annualized rates in appendix C. All results reported below refer to annualized estimates.

5 The Meiji Miracle in Comparative Perspective

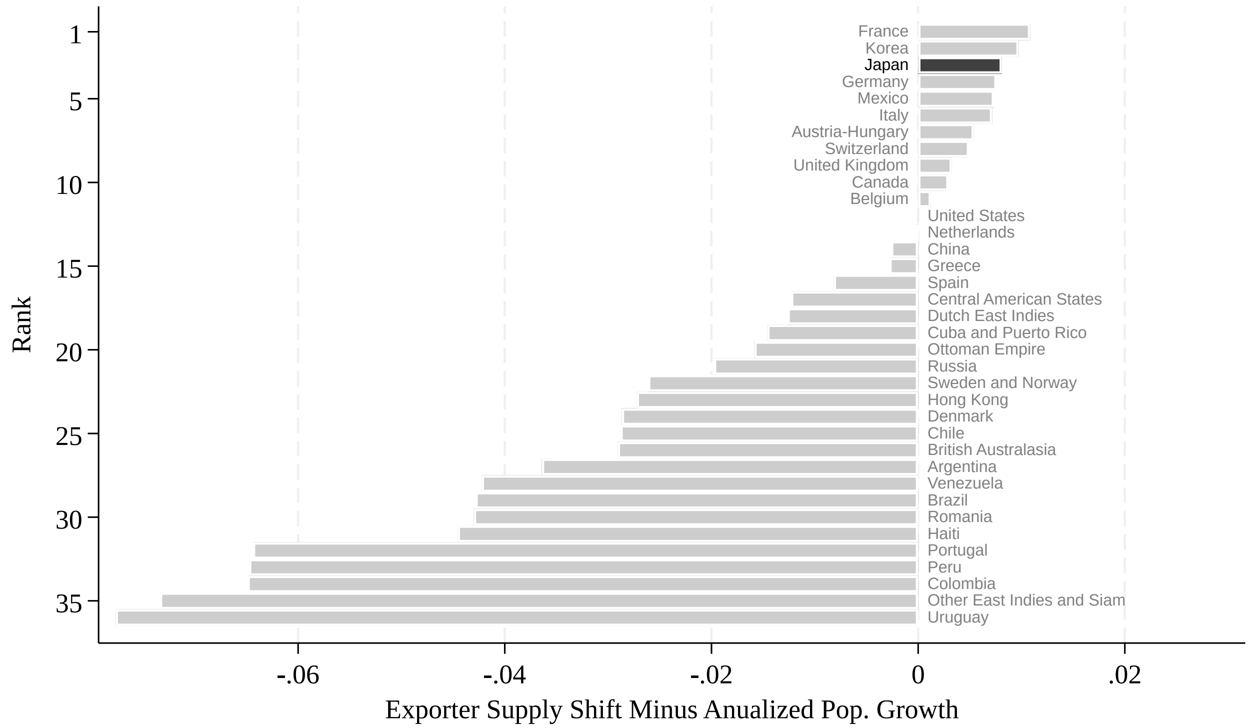
Section 3 examined Japan and other regions' economic performance using using the raw trade data. Here, we utilize the methodology developed in the previous section to provide the first systematic estimates of productivity growth for many regions in the late nineteenth and early twentieth century. Our normalization choice implies that productivity or anything that shifts exporter i's exports conditional on demand conditions will be captured by our estimate of γ_i. We can interpret $\hat{\gamma}_i - \hat{L}$, where $\hat{L}$ is the annual population growth rate, as a measure of exporter productivity, i.e., how much exports in country i grew after controlling for demand conditions and population growth. Figure 11 plots the annualized per capita shift in export supply net of population growth relative to the value for the US, i.e., $\hat{\gamma}_i - \hat{L}_i - (\hat{\gamma}_{US} - \hat{L}_{US})$.[22]

Reassuringly, the ranking of economies broadly aligns with what economic history teaches us about this period. France, Korea, Japan, Germany, Mexico, Italy, Austria-Hungary, Switzerland, the United Kingdom, Canada, Belgium, and the US show robust growth in their export supply shifter. In contrast, economies such as Portugal, Peru, Colombia, and Uruguay show weak performance. Notably, the export-supply shifter for Japan ranks third, confirming that its economy experienced some of the highest export productivity growth globally during this period. Notice that our estimates also suggest that Korea had high productivity growth (alongside Japan), which may be related to the fact that Japan forcibly opened Korea in 1876, and although nominally independent, the Japanese "'reform[ed]' the Korean government and military administration by introducing to the country the kinds of measures that Meiji Japan itself had undertaken" (Iriye, 2007, p. 769)). Our result is consistent with the idea that the Meiji reforms may have also raised productivity in Korea.

[21]We follow Eaton and Kortum (2002) and set $\theta \equiv 8.28$. The choice of θ does not qualitatively affect any of our results; it just raises or lowers all countries' productivity growth proportionally.

[22]Appendix Figure A.3 shows the patterns are similar if we do not account for differences in population growth.

Figure 11: Relative Annualized Per Capita Exporter Supply Shifter by Exporter



Note: Annualized per-capita exporter supply shifts are defined relative to the US, i.e., they are defined as $\hat{\gamma}_i - \hat{L}_i - (\hat{\gamma}_{US} - \hat{L}_{US})$. Annual population growth is computed between {1870,1880} and 1913 using the Maddison data (see Appendix D.5 for details).

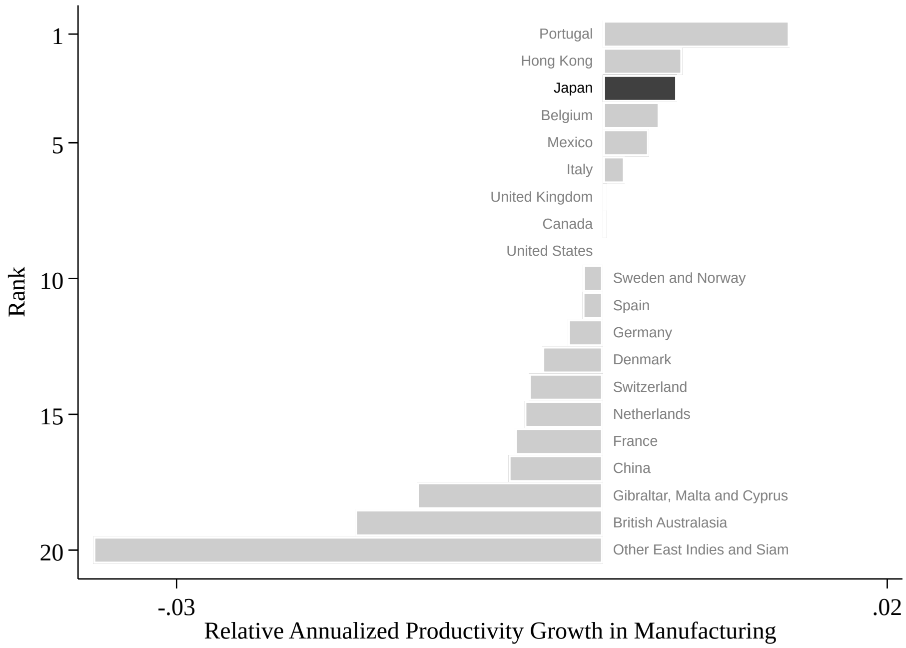
Next, we examine the extent to which productivity growth was biased towards manufacturing. We regress the comparative-advantage component of productivity growth, Γ_{ik}, on broad industry dummies:

$$\Gamma_{ik} = \beta_i^{\text{Agg}} \times I_k^{\text{Agg}} + \beta_i^{\text{Mfg}} \times I_k^{\text{Mfg}} + \beta_i^{\text{Min}} \times I_k^{\text{Min}} + \epsilon_{ik},$$

where I_k^{Agg}, I_k^{Mfg}, and I_k^{Min} are dummies that are one if sector k is in agriculture, manufacturing, or mining, respectively; and β_i^{Agg}, β_i^{Mfg}, and β_i^{Min} are parameters that measure the average growth rate of comparative advantage for exporter i in agriculture, manufacturing, and mining. In words, $(\beta_i^{\text{Mfg}} - \beta_{US}^{\text{Mfg}})$ tells us how fast productivity in manufacturing grew in exporter i relative to the US after controlling for its average growth and the average growth in world manufacturing. Figure 12 reports the results from this exercise for countries in which the manufacturing share of exports in 1880 was not trivial. While Portugal and Hong Kong exhibit strong shifts in comparative advantage towards manufacturing, the results in Figure 11 indicate that these economies had low overall rates of productivity growth, which implies that while they did well in manufacturing, this was relative to their low overall productivity growth. The next seven countries (Japan, Belgium, Mexico, Italy, the UK, the US, and Canada) are all examples of regions that were industrializing over this period by exhibiting rapid productivity growth and having exceptionally high relative productivity growth in manufacturing.

Our structural estimates of industry productivity growth in this period confirm that Meiji Japan's economic performance was exceptional. Average productivity growth was high in inter-

Figure 12: Relative Annualized Productivity Growth in Manufacturing



Note: The plot presents our estimates of productivity growth in manufacturing relative to the US, i.e., $(\beta_i^{\text{Mfg}} - \beta_{US}^{\text{Mfg}})$. β_i^{Mfg} is estimated in equation 5 for regions in which the manufacturing sector's export share in 1880 is at least 0.5% and for regions in which we can estimate productivity growth in at least five non-primary and five primary sectors.

national comparison and shifted strongly towards manufacturing. This result supports the idea that Japan's unparalleled shift towards specialization in manufacturing (Figure 6) was driven by productivity growth biased towards manufacturing—that is, shifting Ricardian comparative advantage. In the next section, we explore whether the growth in technical knowledge in Japanese at this time allowed Japanese entrepreneurs to harness Industrial Revolution technologies, and thus shift comparative advantage towards manufacturing.

6 Codification and Development

The previous sections established that i) Japan experienced strong productivity growth between 1880 and 1910, mainly driven by its manufacturing sectors, and ii) Japan was unique among peripheral economies in providing its citizens with access to codified technical knowledge in their vernacular. This section presents empirical evidence consistent with a causal relationship between these two aspects of Meiji Japan's economy.

Our empirical approach relies on industry-level variation in the extent to which the codification of technical knowledge could increase productivity across sectors. Intuitively, would-be entrepreneurs of textile yarn, which had undergone enormous changes in production methods during the Industrial Revolution, had large productivity benefits to reap from access to technical knowledge. In contrast, producers of raw commodities such as nickel, zinc, or lead – the production of which was barely affected by Industrial Revolution technologies, had far fewer productivity benefits to reap from being able to read technical knowledge. We operationalize how much an industry could benefit from access to technical knowledge with the British Patent Relevance (BPR) measure introduced in Section 2.1.

We test this relationship by estimating regressions of the form

$$g_{ik} = \alpha_i + \beta_J * BPR_k \times I_{iJ} + \beta_r * BPR_k \times I_{ir} + \epsilon_{ik}, \tag{10}$$

where g_{ik} is either annual export growth (raw data) or the growth in comparative advantage ($\tilde{\gamma}_{ik}$) in region i and industry k; α_i is an exporter fixed effect; BPR_k is the British patent relevance measure for sector k; I_{iJ} is a dummy that equals one if i is Japan; I_{ir} is a dummy that equals one if i is part of some other regional grouping r; β_J and β_r are estimated parameters; and ϵ_{ik} is an error term. We partition the regions in our sample into mutually exclusive regions to probe potential confounders.

Since BPR_k is *not* Japan-specific, our measure of British Patent Relevance captures the world *supply* of technical industry-level knowledge. This is important, as our measure is not based on what was written in Japanese, which would be endogenous if the government or entrepreneurs strategically generated knowledge for sectors more likely to succeed.

We hypothesize that $\beta_J > 0$; that is, Japanese industries that benefitted more from the codification of knowledge experienced faster productivity growth in Japan. Column (1) in Tables 3 and 4 show the results from estimating this regression for just the sample of Japanese industries using export growth and productivity growth as outcomes, respectively. Appendix figures A.4 and A.5 plot the corresponding scatterplots. Consistent with our hypothesis, industries with a higher BPR_k experienced faster export and productivity growth during the sample period. The coefficient is both economically meaningful and highly statistically significant. Our estimates imply that a Japanese industry whose British Patent Relevance was in the 75th percentile had export growth that was 15 percentage points per year faster than an industry in the 25th percentile and productivity growth that was 1.4 percentage points per year faster. These large effects help account for the sudden shift of Japanese exports from primary products to manufactures.

A causal interpretation of the parameter of interest, β_J, requires that BPR_k is uncorrelated with the error term ϵ_{Jk}. The main concern in this context is omitted variable bias, namely that unobserved factors correlated with BPR_k drive the pattern of productivity growth in Japan. For example, it is conceivable that BPR_k is correlated with distance to the technology frontier. Or, it could be that some other Japan-specific factors, such as fundamental comparative advantage or institutional reforms implemented during the Meiji Restoration, are correlated with BPR_k. We tackle the various different threats to identification using four strategies.

First, we conduct a placebo exercise in which we examine the relationship between the export and productivity growth of Japan and European codifiers *once we have accounted for the effect of British Patent Relevance*. The idea behind this exercise is that if codification is indeed driving the relationship between BPR_k and industry performance in Japan, there should be no statistically significant relationship between Japanese and European codifiers industry performance beyond that which is contained in BPR_k. To test this hypothesis, we regressed export and productivity growth in the top four codifying European countries (France, Britain, Germany and Italy) on BPR_k. The residuals from these regressions contain components of export and productivity growth that could not be codified. If the key explanatory variable is Western export and productivity growth in general (regardless of codification), we should expect that if we include the residual from this regression as an explanatory variable, it should also have a positive and significant coefficient. We implement this specification in column (2) and find that the residual is insignificant, which implies that non-codified Western technology did not significantly affect Japanese export behavior.

Second, we exploit the fact that we i) know which countries codified and ii) can measure industry performance around the world to examine the relationship between BPR_k and export/productivity growth in a pooled sample including all industry-region pairs for which we

Table 3: Annualized Export Growth and British Patent Relevance

	Export Growth							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
BPR × Japan	3.027***	2.818***	3.027***	3.027***	3.027***	3.027***	3.027***	3.027***
	(0.791)	(0.787)	(0.791)	(0.791)	(0.791)	(0.791)	(0.792)	(0.792)
Residualized Codifier Export Growth		-0.027						
		(0.360)						
BPR × Not Japan			-0.726***	-0.905***	-0.904***	-1.785***		
			(0.211)	(0.240)	(0.254)	(0.390)		
BPR × English-Speaking				1.053**				
				(0.483)				
BPR × French-Speaking					0.844**			
					(0.408)			
BPR × Top-4 Codified						2.039***		
						(0.436)		
BPR × High-Income							-0.320	-0.320
							(0.235)	(0.235)
BPR × Medium-Income							-0.884	-0.783
							(0.543)	(0.556)
BPR × Low-Income							-1.624***	-1.129*
							(0.515)	(0.656)
BPR × Asia								-1.055
								(0.822)
Observations	71	70	1395	1395	1395	1395	1395	1395
R^2	0.123	0.118	0.234	0.235	0.235	0.245	0.237	0.237
Country fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
Sample	Japan	Japan	All	All	All	All	All	All

Note: The dependent variable, "Export Growth," is the annualized export growth rate for a region *i*'s industry *k* between {1880,1885} and {1905,1910}. "British Patent Relevance" is a variable that captures how relevant the synopses of British patents are to the vocabulary of an industry *k*. Japan dummy equals one if the region is Japan and zero otherwise. "Not Japan" is analogously defined. "Residualized Codifier Export Growth" is the average industry residual from a regression of export growth in the top-4 codifiers on BPR. "English-speaking" is an indicator equal to 1 if the region's plurality language is English. "Top-4 Codified" is a dummy for countries that speak one of the 4 most codified languages: French, English, German, and Italian. {High, Medium, Low}-Income are dummies are one when a region is in the top third of the income distribution (High), middle third (Medium), or bottom third (Low) according to the Maddison per capita GDP data; we set these dummies to 0 for Japan. The Asia dummy equals one if the region is in Asia and 0 if it is Japan or not in Asia. Residualized Codifier Export Growth is a measure of non-codified knowledge. It is the residual from a regression between Export Growth and BPR for the Top-4 codifiers. Robust standard errors in parentheses: $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

have data. Column (3) shows that, on average, other countries in our sample did not exhibit the same pattern of export and productivity growth that we found for Japan. Export growth in this period was *negatively* and statistically significantly correlated with BPR_k in countries other than Japan, while there is essentially no correlation with productivity growth. Although the comparative advantage of regions, in general, did not rise faster in sectors that stood more to learn from codified knowledge, the question remains of whether codification was the key or if Japan just did

Table 4: Annualized Productivity Growth Γ_{ik} and British Patent Relevance

	Γ_{ik}							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
BPR × Japan	0.281***	0.282***	0.281***	0.281***	0.281***	0.281***	0.281***	0.281***
	(0.087)	(0.086)	(0.087)	(0.087)	(0.087)	(0.087)	(0.087)	(0.087)
Residualized Codifier Productivity Growth		0.370						
		(0.579)						
BPR × Not Japan			-0.010	-0.022	-0.019	-0.059		
			(0.024)	(0.028)	(0.029)	(0.043)		
BPR × English-Speaking				0.070				
				(0.053)				
BPR × French-Speaking					0.044			
					(0.048)			
BPR × Top-4 Codified						0.097**		
						(0.049)		
BPR × High-Income							-0.004	-0.004
							(0.026)	(0.026)
BPR × Medium-Income							0.047	0.068
							(0.059)	(0.061)
BPR × Low-Income							-0.083	0.015
							(0.062)	(0.075)
BPR × Asia								-0.203**
								(0.096)
Observations	56	56	1244	1244	1244	1244	1244	1244
R^2	0.068	0.076	0.010	0.011	0.010	0.012	0.012	0.016
Country fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
Sample	Japan	Japan	All	All	All	All	All	All

Note: The dependent variable, Γ_{ik}, is the annualized growth rate in comparative advantage for a region i's industry k between {1880,1885} and {1905,1910}. "British Patent Relevance" is a variable that captures how relevant the synopses of British patents are to the vocabulary of an industry k. Japan dummy equals one if the region is Japan and zero otherwise. Not Japan is analogously defined. "Residualized Codifier Productivity Growth" is the average industry residual from a regression of export growth in the top-4 codifiers on BPR. English-speaking is an indicator equal to 1 if the region's plurality language is English. "Top-4 Codified" is a dummy for countries that speak one of the 4 most codified languages: French, English, German, and Italian. {High, Medium, Low}-Income are dummies are one when a region is in the top third of the income distribution (High), middle third (Medium), or bottom third (Low) according to the Maddison per capita GDP data; we set these dummies to 0 for Japan. The Asia dummy equals one if the region is in Asia and 0 if it is Japan or not in Asia. Residualized Codifier Productivity Growth is a measure of non-codified knowledge, it is the residual from a regression between Productivity Growth and BPR for the Top-4 codifiers. Robust standard errors are in parentheses: $^{*}p < 0.10$,** $p < 0.05$,*** $p < 0.01$.

something else that was correlated with codification. If codification was the key difference across countries, we should expect a positive relationship for regions with access to codified knowledge. Indeed, in columns (4) and (5), we see that British Patent Relevance is significantly correlated with faster export growth rates in regions speaking languages with the largest numbers of technical books (English and French) relative to those that did not and positively, but not significantly, associated with productivity growth. In column (6), we try to deal with the imprecision of estimating the impact of codification by pooling across the four languages with the most codification in 1870

(French, English, Italian, and German) and find that Japan and the other top-4 codifiers all experienced faster export and productivity growth in sectors that stood the most to gain from patents relative to regions that did not codify. Taking these results together with the placebo exercise from column (2), we conclude that Japan and European codifiers export and productivity seem to show similar patterns, but only in ways correlated with the codifiable aspects of Industrial Revolution technologies.[23]

We explore the possibility that Japan's experience was a product of its income level or geography in columns (7) and (8). We group countries by income tercile (column 7) and isolate Asia (column 8). No region group displays a similar productivity pattern. On the contrary, the poorest countries, particularly Asia outside of Japan, show a negative correlation, though the patterns are never consistently statistically different from zero. In summary, the pooled specifications suggest that Japan's productivity growth pattern was unusual for a peripheral economy but very much in line with other codifiers. Regional trends or structural factors, such as distance to the technology frontier, are unlikely to explain the relationship.

Our third strategy for addressing confounders is to provide a series of other robustness checks. In Appendix Table A.2 we control for whether the region was a British colony (as opposed to being a plurality English-speaking country), but the coefficient is not significant. An alternative confounder is that BPR may be correlated with an industry's steam-energy intensity, and our regressions are driven by Japan being a late adopter of steam power, which caused it to grow relatively quickly in steam-intensive sectors. We measure industries' steam-power intensity by using French data from the 1860s to measure the amount of steam power used in an industry divided by the wage bill (details of the calculation can be found in Appendix D.9). We find that controlling for steam power does not affect our results.[24]

Appendix Tables A.3 and A.4 drop non-manufacturing sectors to show that the impact of BPR on Japanese export and productivity growth rates did not arise simply because of differences in manufacturing productivity growth relative to non-manufacturing productivity growth rates. In Appendix Tables A.5 and A.6, we show that BPR raised the industry export and productivity growth rates in the European set of top-4 codifying countries. In contrast, countries outside of this set, including Spanish-speaking countries like Spain and Mexico, on average, have a negative relationship between BPR and industry export and productivity growth rates. Although the coefficients are typically imprecisely measured, the pattern is consistent with the pooled estimates of top-4 codifiers reported in Tables 3 and 4. We also demonstrate that our results for Japan are not driven by Japanese exports to any particular geographic region in Appendix Table A.7. Notably, our results hold excluding Asian destination markets which were the primary markets served by Japan (Meissner and Tang, 2018). Similarly, our results are not driven by individual high-growth sectors like textiles or iron and fabricated metal products (Table A.8). Together, these results speak to the fact that broad-based changes were underway in Japan. These patterns hold within the manufacturing sector, are not driven only by Japan's major export destination markets, nor their major export products alone.

Our fourth and final strategy uses the sudden timing of Japan's codification of technical knowledge to examine whether there is a "Japan-specific" confounder. In particular, while Japan was undergoing major economic and political changes in the second half of the 19th century, the previ-

[23]The estimated coefficients for Japan are larger than those for regions in which a plurality speak a codified European language. This makes sense given the late industrialization of Japan implies it had more to learn from British patents.

[24]Steam-engines were used outside of manufacturing too, where British Patent Relevance tends to be low. For example, the mining of coal and other natural resources are steam intensive sectors but do not have a high BPR_k. This likely explains why our results are robust to the inclusion of a measure of steam-intensity.

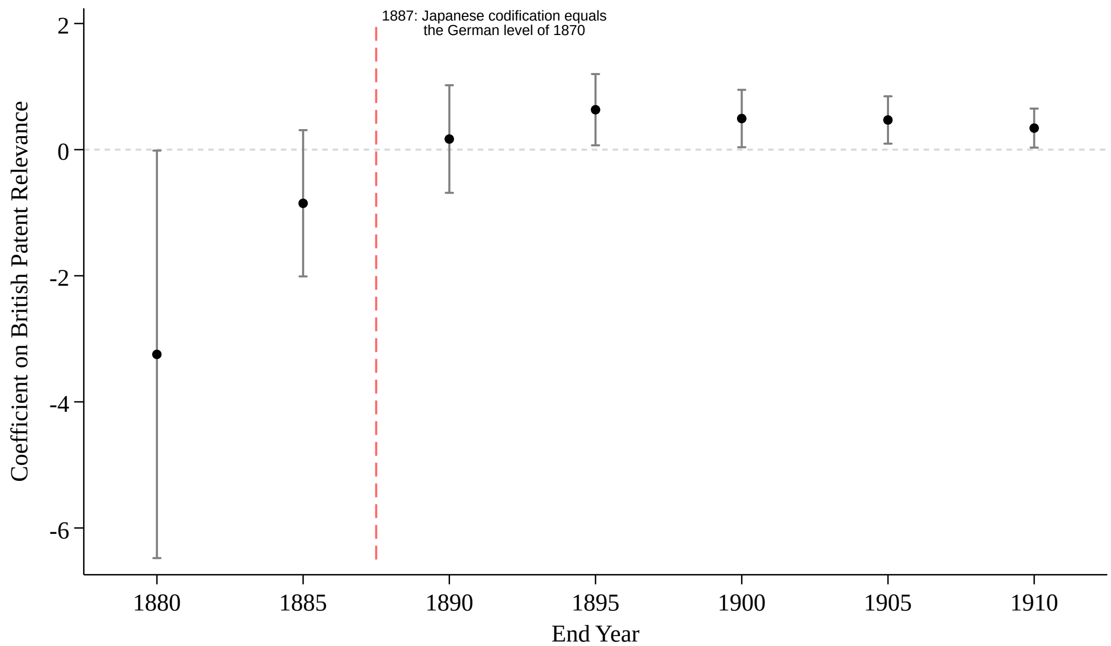
ous sections have established that the change in the composition of exports towards manufactures happened rapidly and immediately after Japanese entrepreneurs had access to technical knowledge in their vernacular. While the timing is suggestive, we now test this formally with industry-level variation. It is not possible to identify a precise year in which Japan became technically literate, but the data suggest that it likely happened in the 1880s. For example, in 1880, there were only half as many technical books in Japanese as had existed in Spanish in 1870. Thus, in 1880, Japan was still a minor codifier compared to the major European codifiers. Between 1880 and 1890, however, the number of Japanese technical books in the National Diet Library grew from 706 to 2,823, surpassing the 1870 number of codified books in every European language except English and French.

If technical literacy mattered for Japan, we should expect to see British patent relevance only matter for Japanese exports after the Japanese could read Western technology. We estimate regressions of annualized Japanese export growth rates by industry between 1875 (a year in which Japan had less than one-sixth as many technical books as Spain had in 1870) and an end year that varies in five-year increments starting in 1880 on British Patent Relevance.[25] Figure 13 plots the estimated coefficients from these regressions, along with the 95% confidence intervals.[26] We interpret the specification for export growth between 1875 and 1880 as a placebo exercise that examines the relationship between export growth and BPR in the years before Japan achieved technical literacy. We obtain a negative and significant coefficient on BPR, indicating that Japanese export growth was slower in industries that benefited the most from the Industrial Revolution. This result is similar to what we saw in Table 3 for countries other than Japan and other Asian economies in particular. In other words, before Japan became technically literate, its export growth patterns looked similar to those of other countries in the global periphery: losing comparative advantage in sectors where the potential to learn from the West was the highest.

[25]Data limitations preclude us from estimating these regressions using productivity growth as the outcome variable. In particular, we have Japanese trade data in 1875 but do not have it for other countries.

[26]Appendix Table A.9 reports the estimated coefficients from the same specifications.

Figure 13: Coefficients from Regressing Japanese Export Growth on BPR by End Year



Note: We plot the estimated coefficient on BPR as well as the 95% confidence interval in a regression of Japanese industry export growth from 1875 to the year displayed in the figure on BPR.

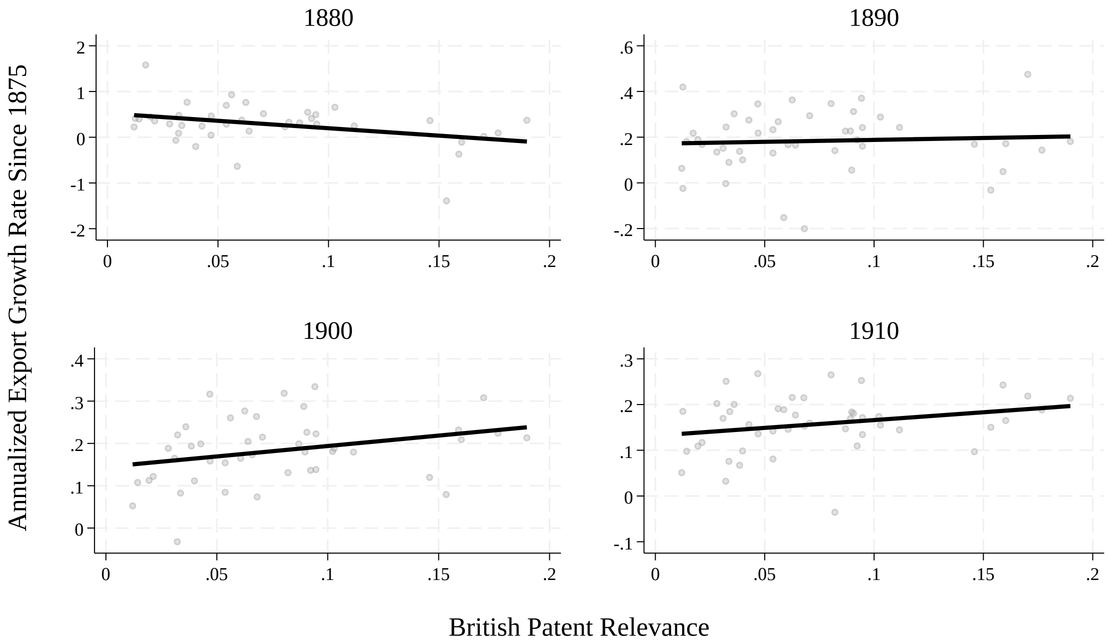
This pattern flips around 1890, a point at which Japan was becoming a major codifier. By 1895, the coefficient of British patent relevance is positive and significant, indicating that Japanese industries that stood the most to learn from British technologies grew faster. This coefficient remains positive throughout the rest of our sample period, suggesting a persistent effect. Figure 14 shows the scatterplots for 1880, 1890, 1900, and 1910 to show that outliers are not driving these results. As one can see from the plots, there is a clear negative relationship in 1880 and a clear positive relationship in later years, and these results are not driven by outliers. The timing of this effect is hard to explain with conventional stories about Japan. We do not detect a significant impact of BPR on exports until 37 years after Japan opened to trade and 27 years after the Meiji Restoration. Neither of these stories can explain this basic pattern of Japanese industrialization—namely, Japanese comparative advantage shifted away from Western sectors that benefited the most from technological progress before Japan could read science and then turned towards these sectors after the Japanese became technically literate.

In summary, the cross-region evidence and the timing of when technical knowledge became predictive of industry growth in Japan provides a strong case for a causal interpretation of codification on Japanese productivity growth. By making technical knowledge widely available in the vernacular, the Meiji government relaxed a critical bottleneck for industrialization. Any alternative explanation of Japanese productivity growth needs to account for both its distinctive pattern in cross-regional comparison and its timing in Japan.

7 Conclusion

This paper shows evidence in support of the argument that the public provision of technical knowledge in the vernacular eliminated an important friction impeding the absorption of Western

Figure 14: Japanese Export Growth and BPR by Decade, 1880-1910



Note: These graphs plot the annualized growth rate between 1875 and year X against British Patent Relevance.

technology in Meiji Japan. Our results show an empirical pattern unique to Japan and other codifiers: industries that had more to benefit from Western technology experienced faster export and productivity growth in relevant sectors. This suggests that regions hoping to emulate European industrialization in the nineteenth-century context, particularly those linguistically or physically distant from Western Europe, needed to provide complex public goods, such as access to technical knowledge, to emulate Britain successfully.

While these public goods were unlikely to be sufficient in and of themselves to foster modern industrial development, our results suggest they may have been necessary. Other ethnically and linguistically distinct countries that received Japanese institutions and were forced to learn Japanese through annexation or colonization, such as the Ryukyu Kingdom (now Okinawa), Ezo (now Hokkaido), Taiwan, and Korea, also have per capita incomes that are now similar to Japan. We leave it to future researchers to examine whether Japanese colonial institutions, like British ones, had any salutary effect on their growth.

The obvious question is why the Japanese government was unique among regions in the periphery in providing these public goods. Our reading of the historical record suggests that it was the severe, existential threat to the Japanese regime caused by the arrival of Western powers which aligned the elite in support of a strategy of aggressive defensive modernization. A central part of this effort, as we have shown, was the absorption of Western science and technology.

Historians argue that one important advantage Japan had was that its contact with the West happened late, meaning that it could observe and learn from what happened to China (Platt, 2012).[27] Indeed, the fate of China in the wake of the Opium Wars loomed large in Japanese

[27] Platt writes "the Japanese benefited from the negative example of China. As the Japanese government in the 1850s had avoided its own Opium War by signing foreign treaties without overt hostility, so did

thinking. After the British imposed crushing indemnity payments, the Chinese government was thrust into a perpetual state of near-bankruptcy (Keay, 2010). The indemnity payments, coupled with China's subsequent descent into a brutal civil war, meant that Chinese efforts to modernize through the "Self-Strengthening Movement," similar in spirit, though not in scale, to Japan's efforts to absorb Western science and technology, received little government support and, at least for the reform-minded Taiping rebels, open opposition (Keay, 2010).[28] Senior members of the *shogun's* government in the late Tokugawa period correctly anticipated that Japan would be the next target and efforts were underway even prior to the Meiji Revolution to learn from the West, as we have discussed. Importantly, however, Japan did not need to discover the policy tools themselves. State support of technology absorption, particularly the translation of technical books, was a common strategy for regions hoping to emulate Britain. This has been observed from Bourbon France in the late eighteenth century to the Self-Strengthening Movement in China in the nineteenth century (Juhasz and Steinwender, 2024). Meiji Japan thus took the state-led technology adoption playbook developed elsewhere and deployed it at an unprecedented scale.

influential young samurai in the 1860s look to China at the end of its civil war as a warning of what their country might become without dramatic change. A revolution later that decade gave way to a rapid program of industrialization and social transformation that bore a remarkable similarity in spirit—if not in religion—to what Hong Rengan [one of the Taiping leaders] had envisioned for his own thwarted state. By the 1890s, Japan's modernized navy would decisively overpower the Qing fleet, and Japan would take the island of Taiwan from China as its first major colony. By the early twentieth century, Chinese reformers would be looking to Japan as the model of what their own country must become if it were to have any chance of surviving into the future. But perhaps it didn't have to turn out that way. In an interview with a British reporter in 1909, Japan's elder statesman Ito Hirobumi—four-time prime minister and chief architect of the nineteenth-century reform movement—looked to the violence just beginning to unfold in China in the run-up to the 1911 Revolution and declared it long overdue. In his opinion, the new Chinese revolutionaries were merely finishing the work that the Taiping had started fifty years earlier, and in which he firmly believed they would have been successful if left to their own devices. 'The greatest mistake which you Western people, and more especially you English people, made in all your dealings with China,' he told the reporter, 'was to help the Manchus in putting down the Taiping Rebellion."

[28]For example, Keay (2010) writes, "The indemnity [from the first Opium War] payable (plus interest) in installments, would be a crippling burden on the empire's shattered finances." Keay also argues that the failure of the Chinese government to provide adequate funds for its Self-Strengthening movement came to a head in China's next ignominious defeat in the First Sino-Japanese War in 1894: "While in thirty years of Self-Strengthening, China had yet to find a firearm for every conscript, a field gun for every detachment, or sufficient ships – bought, built, reconditioned – for a couple of fleets, the Japanese had constructed a large modern navy and trained up the only professional army east of India. Defeat followed defeat as the Chinese were quickly driven out of Korea. Within a couple of months, the Japanese were at the Chinese border and not inclined to stop there."

References

Allen, R. (2012). Technology and the great divergence: Global economic development since 1820. *Explorations in Economic History 49*(1), 1–16.

Almelhem, A., M. Iyigun, A. Kennedy, and J. Rubin (2023). Enlightenment ideals and belief in progress in the run-up to the industrial revolution: A textual analysis.

Amiti, M. and D. E. Weinstein (2018). How Much Do Idiosyncratic Bank Shocks Affect Investment? Evidence from Matched Bank-Firm Loan Data. *Journal of Political Economy 126*(2).

Bassino, J.-P., S. Broadberry, K. Fukao, B. Gupta, and M. Takashima (2019). Japan and the great divergence, 730–1874. *Explorations in Economic History 72*, 1–22.

Berg, M. (2007). The genesis of 'useful knowledge'. *History of science 45*(2), 123–133.

Bernhofen, D. M. and J. C. Brown (2004). A direct test of the theory of comparative advantage: the case of Japan. *Journal of Political Economy 112*(1), 48–67.

Bernhofen, D. M. and J. C. Brown (2005). An empirical assessment of the comparative advantage gains from trade: evidence from Japan. *American Economic Review 95*(1), 208–225.

Bible (1885). *The Holy Bible: Revised Version*. Oxford and Cambridge: Oxford University Press and Cambridge University Press.

Bolitho, H. (2007). The Temp? crisis. In J. W. Hall, M. B. Jansen, M. Kanai, and D. Twitchett (Eds.), *The Cambridge History of Japan: The Nineteenth Century*, Volume 5, pp. 116–167. New York: Cambridge University Press.

Bolt, J. and L. van Zanden (2020). The Maddison Project: Maddison style estimates of the evolution of the world economy. A new 2020 update.

Bouscasse, P., E. Nakamura, and J. Steinsson (2023). When Did Growth Begin? New Estimates of Productivity Growth in England from 1250 to 1870. Working Paper 28623, National Bureau of Economic Research.

Braguinsky, S. (2015). *Industrial and Corporate Change 24*(4).

Broadberry, S. N. (1994). Technological leadership and productivity leadership in manufacturing since the Industrial Revolution: Implications for the convergence debate. *The Economic Journal 104*(423), 291–302.

Chanut, J.-M. (2000). *L'industrie française au milieu du 19e siècle: les enquêtes de la statistique générale de la France.*

Clark, G. (1987). Why Isn't the Whole World Developed? Lessons from the Cotton Mills. *The Journal of Economic History 47*(1).

Clark, P. H. (2009). The Kokugo Revolution. *Japan Research Monograph*.

Costinot, A., D. Donaldson, and I. Komunjer (2012). What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas. *Review of Economic Studies 79*(2), 581–608.

Crafts, N. (1998). Forging ahead and falling behind: the rise and relative decline of the first industrial nation. *Journal of Economic Perspectives 12*(2), 193–210.

Department of Finance (1916). *Annual Return of the Foreign Trade of the Empire of Japan. Part I.* Department of Finance.

Eaton, J. and S. Kortum (2002). Technology, Geography, and Trade. *Econometrica 70*(5), 1741–1779.

Federico, G., S. Natoli, G. Tattara, and M. Vasta (2011). *Il commercio estero italiano: 1862-1950*. Laterza.

Fouquin, M. and J. Hugot (2016). Two centuries of bilateral trade and gravity data: 1827-2014.

Gerschenkron, A. (2015). *Economic backwardness in historical perspective (1962)*. Cambridge MA.

Hallmann, C., L. Rosenberger, and E. E. Yavuz (2021). Invention and Technological Leadership During the Industrial Revolution. Technical report, Working Paper.

Hayami, Y. (1975). *A Century of Agricultural Growth in Japan*. University of Minnesota Press and University of Tokyo Press.

Heibonsha (1974). *Nihon Jinbutsu Bunken Mokuroku [Bibliography of Japanese Biographies]*. Tokyo: Heibonsha.

Hirakawa, S. (2007). Japan's turn to the West. In J. W. Hall, M. B. Jansen, M. Kanai, and D. Twitchett (Eds.), *The Cambridge History of Japan: The Nineteenth Century*, Volume 5, pp. 432–498. New York: Cambridge University Press.

Horie, Y. (1960). Business pioneers of modern Japan—Ishikawa Masatatsu and Oshima Takato. *Kyoto University Economic Review 30*(2 (69)), 1–16.

Huberman, M., C. M. Meissner, and K. Oosterlinck (2017). Technology and geography in the second industrial revolution: new evidence from the margins of trade. *The Journal of Economic History 77*(1), 39–89.

Hungerland, W.-F. and N. Wolf (2022). The panopticon of Germany's foreign trade, 1880–1913: New facts on the first globalization. *European Review of Economic History 26*(4), 479–507.

Iriye, A. (2007). Japan's drive to great-power status. In J. W. Hall, M. B. Jansen, M. Kanai, and D. Twitchett (Eds.), *The Cambridge History of Japan: The Nineteenth Century*, Volume 5, pp. 721–782. New York: Cambridge University Press.

Jones, H. J. (1980). *Live machines: hired foreigners and Meiji Japan*. Vancouver: University of British Columbia.

Juhasz, R. and C. Steinwender (2024). Industrial Policy and the Great Divergence. *Annual Review of Economics 16*.

Keay, J. (2010). *China: A History*. HarperCollins Publishers.

Kokawa, T., Y. Shigeru, and Y. Masuda (1994). *Historical Development of English-Japanese Dictionaries in Japan*.

Kumon, Y. (2022). The labor-intensive path: Wages, incomes, and the work year in Japan, 1610–1890. *The Journal of Economic History 82*(2), 368–402.

Lippert, W. (2001). Language in the Modernization Process: The Integration of Western Concept and Terms into Chinese and Japanese in the Nineteenth Century. In M. Lackner, I. Amelung, and J. Kurtz (Eds.), *New Terms for New Ideas. Western Knowledge and Lexical Change in Late Imperial China*. Netherlands: Koninlijke Brill.

Mayer, T. and S. Zignago (2011). *Notes on CEPII's distances measures: the GeoDist Database*. CEPII Working Paper 2011-25.

Meissner, C. M. and J. P. Tang (2018). Upstart industrialization and exports: evidence from Japan, 1880–1910. *The Journal of Economic History 78*(4), 1068–1102.

Ministry of Education (1980). *Japan's modern educational system: a history of the first hundred years*, Chapter Higher Education in the Early Meiji Era.

Mitchell, B. (1975). *European Historical Statistics, 1750-1970* (1 ed.). Palgrave Macmillan London.

Mokyr, J. (2011). *The gifts of Athena: Historical origins of the knowledge economy*. Princeton University Press.

Mokyr, J. (2016). *A culture of growth: The origins of the modern economy*. Princeton University Press.

Montgomery, S. L. (2000). *Science in translation: Movements of knowledge through cultures and time*.

University of Chicago Press.

Morck, R. and M. Nakamura (2007). Business Groups and the Big Push: Meiji Japan's Mass Privatization and Subsequent Growth. *Enterprise and Society 8*(3).

Morck, R. and M. Nakamura (2018). Japan's ultimately unaccursed natural resources-financed industrialization. *Journal of the Japanese and International Economies 47*(C), 32–54.

Moser, P. (2005). How do patent laws influence innovation? Evidence from nineteenth-century world's fairs. *American economic review 95*(4), 1214–1236.

National Institute for Educational Policy Research (2011). Primary Schools in Japan. https://www.nier.go.jp/English/educationjapan/pdf/201109BE.pdf. [Online; accessed 15-June-2024].

Ohkawa, K., M. Shinohara, and M. Umemura (Eds.) (1965). *Estimates of long-term economic statistics of Japan since 1868*. Tokyo: Toyo Keizai Shinposha.

Ohno, K. (2018). *The History of Japanese Economic Development: Origins of Private Dynamism and Policy Competence*. Oxon: Routledge.

Pascali, L. (2017). The wind of change: Maritime technology, trade, and economic development. *American Economic Review 107*(9), 2821–2854.

Platt, S. R. (2012). *Autumn in the Heavenly Kingdom: China, the West, and the Epic Story of the Taiping Civil War*. New York: Alfred A. Knopf.

Rosenberger, L., W. W. Hanlon, and C. Hallmann (2024). Innovation networks in the industrial revolution. Technical report, National Bureau of Economic Research.

Rubinger, R. (2000). Who Can't Read and Write? Illiteracy in Meiji Japan. *Enterprise and Society 55*(2), 163–198.

Saito, O. and M. Takashima (2016). Estimating the shares of secondary-and tertiary-sector outputs in the age of early modern growth: the case of japan, 1600–1874. *European Review of Economic History 20*(3), 368–386.

Santos Silva, J. and S. Tenreyro (2010). On the existence of the maximum likelihood estimates in poisson regression. *Economics Letters 107*(2), 310–312.

Squicciarini, M. P. and N. Voigtländer (2015). Human capital and industrialization: Evidence from the age of enlightenment. *The Quarterly Journal of Economics 130*(4), 1825–1883.

Stevens, E. (1995). *The Grammar of the Machine: Technical Literacy and Early Industrial Expansion in the United States*. Yale University Press.

Sussman, N. and Y. Yafeh (2000). Institutions, Reforms, and Country Risk: Lessons fom Japanese Government Debt in the Meiji Era. *Journal of Economic History 60*(2).

Tamagawa, K. (2002). The role of cotton spinning books in the developments of the cotton spinning industry in Japan. https://www.jsme.or.jp/tsd/ICBTT/conference02/KanjiTAMAGAWA.html (Accessed: 2024-07-10).

Tang, J. P. (2014). Railroad expansion and industrialization: evidence from Meiji Japan. *The Journal of Economic History 74*(3), 863–886.

Tang, J. P. and S. Basco (2023). Banks, credit supply, and the life cycle of firms: Evidence from late nineteenth century Japan. *Journal of Banking Finance 154*.

Treasury Department's Bureau of Statistics (1900). *Foreign Commerce and Navigation of the United States*. U.S. Government Printing Office.

Ueda, M., J. Nishizawa, I. Hirayama, and S. Miura (Eds.) (2003). *Nihon Jinmei Daijiten [The Biographical Dictionary of Japan]*. Tokyo: Kodansha.

Wong, R. B. (2012). Taxation and good governance in China, 1500-1914. pp. 353–377. Cambridge, MA: Cambridge University Press.

Woodcroft, B. (1857). *Subject-matter index (made from titles only) of patents of invention, from March 2, 1617 (14 James I.), to October 1, 1852 (16 Victoriae)*. Great Seal Patent Office.

World Bank (2024). World Development Indicators.

Xu, C. (2022). Reshaping global trade: the immediate and long-run effects of bank failures. *The Quarterly Journal of Economics 137*(4), 2107–2161.

Yamamura, K. (1986). The Meiji Land Tax Reform and its Effects. In M. Jansen and G. Rozman (Eds.), *Japan in Transition: From Tokugawa to Meiji*, pp. 382–399. Princeton: Princeton University Press.

Yamamura, K. (1997). *The economic emergence of modern Japan*. Cambridge University Press.

Online Appendix

A Additional Tables

Table A.1: Linguistic Distance from English and GDP

	Log GDP per Capita					
	(1) 1870	(2) 1913	(3) 2018	(4) 1870	(5) 1913	(6) 2018
Log Physical Distance between Country and the UK	-0.170***	-0.207***	-0.237***	-0.248***	-0.315***	-0.323***
	(0.058)	(0.064)	(0.066)	(0.054)	(0.065)	(0.072)
Number of Weeks Required to Learn the Plurality Language	-0.010***	-0.013***	-0.008*	-0.005**	-0.007***	-0.003
	(0.002)	(0.003)	(0.004)	(0.002)	(0.003)	(0.005)
Observations	61	61	61	55	55	55
R^2	0.395	0.428	0.208	0.369	0.426	0.198
Includes English-speaking Countries	✓	✓	✓			

Standard errors in parentheses
$^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

Note: GDP per capita is from the Maddison Project. The physical distance between the region and the UK is from *CEPII* database using the Great Circle Formula. The number of weeks an English native speaker will take to obtain "Professional Working Proficiency" in the plurality language of a country is estimated by the U.S. Department of State's Foreign Service Institute. See Appendix D for data construction and sources. Robust standard errors are in parentheses. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

Table A.2: Annualized Export/Productivity Growth and British Patent Relevance - British Colonies and Steam Intensity

	Export Growth		Γ_{ik}	
	(1)	(2)	(3)	(4)
BPR × Japan	3.027***	3.175**	0.281***	0.194*
	(0.791)	(1.242)	(0.087)	(0.106)
BPR × Not Japan	-0.871***	-0.155	-0.026	-0.001
	(0.242)	(0.259)	(0.028)	(0.028)
BPR × British Colony	0.694		0.077	
	(0.488)		(0.054)	
Steam Intensity		-0.736**		0.015
		(0.296)		(0.035)
Observations	1395	690	1244	627
R^2	0.234	0.310	0.011	0.067
Country fixed effects	✓	✓	✓	✓
Sample	All	All	All	All

Note: The dependent variable, "Export Growth," is the annualized export growth rate for industry k in region i between {1880,1885} and {1905,1910}. Γ_{ik}, is the annualized productivity growth rate for industry k in region i between {1880,1885} and {1905,1910}. "British Patent Relevance" is a variable that captures how relevant the titles of British patents (1617-1852) are to the vocabulary of an industry k. The Japan dummy equals one if the region is Japan and zero otherwise, "Not Japan" is analogously defined. "British Colony" is a dummy for whether a region was a British colony in the 1880-1910 window. Steam Intensity is constructed as Steam Engine Horsepower/Wage Bill by industry using French manufacturing census data from the 1860s (see Appendix D.9 for details about the data construction). Robust standard errors are in parentheses. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

Table A.3: Annualized Export Growth and British Patent Relevance - Manufacturing Sectors

	Export Growth						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
BPR × Japan	2.775**	2.775**	2.775**	2.775**	2.775**	2.775**	2.775**
	(1.161)	(1.156)	(1.157)	(1.157)	(1.157)	(1.158)	(1.159)
BPR × Not Japan		-0.348	-0.348	-0.544*	-1.004**		
		(0.239)	(0.274)	(0.286)	(0.443)		
BPR × English-Speaking			0.004				
			(0.515)				
BPR × French-Speaking				0.883*			
				(0.487)			
BPR × Top-4 Codified					1.230**		
					(0.501)		
BPR × High-Income						-0.259	-0.259
						(0.261)	(0.261)
BPR × Medium-Income						-0.135	-0.065
						(0.663)	(0.682)
BPR × Low-Income						-0.809	-0.478
						(0.571)	(0.760)
BPR × Asia							-0.731
							(0.872)
Observations	31	661	661	661	661	661	661
R^2	0.160	0.364	0.364	0.366	0.369	0.365	0.365
Country fixed effects	✓	✓	✓	✓	✓	✓	✓
Sample	Japan	All	All	All	All	All	All

Note: The dependent variable, "Export Growth," is the annualized export growth rate for industry k in region i between {1880,1885} and {1905,1910}. "British Patent Relevance" is a variable that captures how relevant the titles of British patents (1617-1852) are to the vocabulary of an industry k. Japan dummy equals one if the region is Japan and zero otherwise, "Not Japan" is analogously defined. "English-speaking" is an indicator equal to 1 if the region's plurality language is English. "Top-4 Codified" is a dummy for countries that speak one of the four most codified languages: French, English, German, and Italian. {High, Medium, Low}Income are indicator variables which use 1870 GDP per capita from the Maddison Project to identify if a region is in the top third of the income distribution (high), middle third (medium), or in the bottom third (bottom); we set these dummies to 0 for Japan. Asia dummy equals 1 if the region is in Asia and 0 if it is Japan or not in Asia. Robust standard errors are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.4: Annualized Productivity Growth and British Patent Relevance - Manufacturing Sectors

	Γ_{ik}						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
BPR × Japan	0.256**	0.256**	0.256**	0.256**	0.256**	0.256**	0.256**
	(0.114)	(0.112)	(0.112)	(0.112)	(0.112)	(0.112)	(0.113)
BPR × Not Japan		0.000	0.001	-0.008	-0.027		
		(0.026)	(0.029)	(0.031)	(0.047)		
BPR × English-Speaking			-0.003				
			(0.057)				
BPR × French-Speaking				0.037			
				(0.052)			
BPR × Top-4 Codified					0.053		
					(0.053)		
BPR × High-Income						-0.015	-0.015
						(0.027)	(0.027)
BPR × Medium-Income						0.099	0.115*
						(0.067)	(0.069)
BPR × Low-Income						-0.059	0.011
						(0.066)	(0.082)
BPR × Asia							-0.157
							(0.098)
Observations	24	587	587	587	587	587	587
R^2	0.080	0.120	0.120	0.120	0.121	0.126	0.130
Country fixed effects	✓	✓	✓	✓	✓	✓	✓
Sample	Japan	All	All	All	All	All	All

Note: The dependent variable, Γ_{ik}, is the annualized productivity growth rate for industry k in region i between {1880,1885} and {1905,1910}. "British Patent Relevance" is a variable that captures how relevant the titles of British patents (1617-1852) are to the vocabulary of an industry k. Japan dummy equals one if the region is Japan and zero otherwise, "Not Japan" is analogously defined. "English-speaking" is an indicator equal to 1 if the region's plurality language is English. "Top-4 Codified" is a dummy for countries that speak one of the four most codified languages: French, English, German, and Italian. Steam Intensity is constructed as Steam Engine Horsepower/Wage Bill at an industry level. {High, Medium, Low}Income are indicator variables which use 1870 GDP per capita from the Maddison Project to identify if a region is in the top third of the income distribution (high), middle third (medium), or in the bottom third (bottom); we set these dummies to 0 for Japan. Asia dummy equals 1 if the region is in Asia and 0 if it is Japan or not in Asia. Robust standard errors are in parentheses. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

Table A.5: Annualized Export Growth and British Patent Relevance: Selected Countries

	Export Growth							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
British Patent Relevance	0.190	0.323	1.105*	0.572	0.340*	0.418	-4.105	-1.273*
	(0.325)	(0.699)	(0.624)	(0.436)	(0.193)	(0.395)	(2.477)	(0.643)
Observations	86	48	88	72	74	86	29	46
R^2	0.002	0.002	0.024	0.016	0.032	0.005	0.146	0.040
Constant	✓	✓	✓	✓	✓	✓	✓	✓
Sample	France	Belgium	UK	US	Italy	Germany	Spain	Mexico

Note: The dependent variable, "Export Growth," is the annualized export growth rate for industry k in region i between {1880,1885} and {1905,1910}. "British Patent Relevance" is a variable that captures how relevant the titles of British patents (1617-1852) are to the vocabulary of an industry k. Robust standard errors are in parentheses. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

Table A.6: Annualized Productivity Growth and British Patent Relevance: Selected Countries

	Γ_{ik}							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
British Patent Relevance	-0.022	0.068	0.100*	0.019	0.106*	0.032	-0.375	0.013
	(0.034)	(0.064)	(0.052)	(0.083)	(0.060)	(0.046)	(0.252)	(0.082)
Observations	73	43	74	63	62	73	26	42
R^2	0.002	0.008	0.033	0.000	0.024	0.004	0.152	0.001
Constant	✓	✓	✓	✓	✓	✓	✓	✓
Sample	France	Belgium	UK	US	Italy	Germany	Spain	Mexico

Note: The dependent variable, Γ_{ik}, is the annualized productivity growth rate for industry k in region i between {1880,1885} and {1905,1910}. "British Patent Relevance" is a variable that captures how relevant the titles of British patents (1617-1852) are to the vocabulary of an industry k. Robust standard errors are in parentheses. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

Table A.7: Annualized Export Growth and British Patent Relevance: Dropping Regions

	Export Growth, Dropping Exports to						
	(1) English-Speaking	(2) British Colonies	(3) Languages Similar to English	(4) High-Income	(5) Medium-Income	(6) Low-Income	(7) Asian
British Patent Relevance	2.850***	2.850***	2.856***	2.351***	2.869***	2.674***	2.674***
	(0.776)	(0.776)	(0.777)	(0.758)	(0.792)	(0.937)	(0.937)
Observations	71	71	71	70	67	61	61
R^2	0.118	0.118	0.119	0.077	0.121	0.074	0.074
Constant	✓	✓	✓	✓	✓	✓	✓
Sample	Japan	Japan	Japan	Japan	Japan	Japan	Japan

Note: The dependent variable, "Export Growth," is the annualized export growth rate for industry k in region i between {1880,1885} and {1905,1910}. "British Patent Relevance" is a variable that captures how relevant the titles of British patents (1617-1852) are to the vocabulary of an industry k. Each column drops exports to a different subset of countries/regions. (1) Drops English-Speaking countries. (2) Drops British Colonies. (3) Drops countries with a language similar to English, defined as those where it takes six or fewer months for an English speaker to become proficient. (4), (5), and (6) drop High, Medium, and Low-income countries, respectively. (7) Drops exports to Asian countries. Robust standard errors are in parentheses. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

Table A.8: Annualized Export/Productivity Growth and British Patent Relevance: Dropping Sectors

	Export Growth, Dropping			Productivity Growth, Dropping		
	(1)	(2)	(3)	(4)	(5)	(6)
	Cotton-Textiles	All Textiles	Iron and Fabricated Metals	Cotton-Textiles	All Textiles	Iron and Fabricated Metals
British Patent Relevance	3.395***	3.662***	3.135***	0.362***	0.306*	0.289***
	(0.948)	(1.368)	(0.829)	(0.095)	(0.156)	(0.092)
Observations	69	63	69	54	49	55
R^2	0.119	0.091	0.123	0.083	0.037	0.068
Constant	✓	✓	✓	✓	✓	✓
Sample	Japan	Japan	Japan	Japan	Japan	Japan

Note: The dependent variable, "Export Growth," is the annualized export growth rate for industry k in region i between {1880,1885} and {1905,1910}. "British Patent Relevance" (BPR) is a variable that captures how relevant the titles of British patents (1617-1852) are to the vocabulary of an industry k. Each column drops exports to a particular industry or group of industries. (1) and (4) drops cotton textile-related industries. (2) and (5) drops all industries related to textiles. (3) and (6) drops industries related to producing iron. Robust standard errors are in parentheses. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

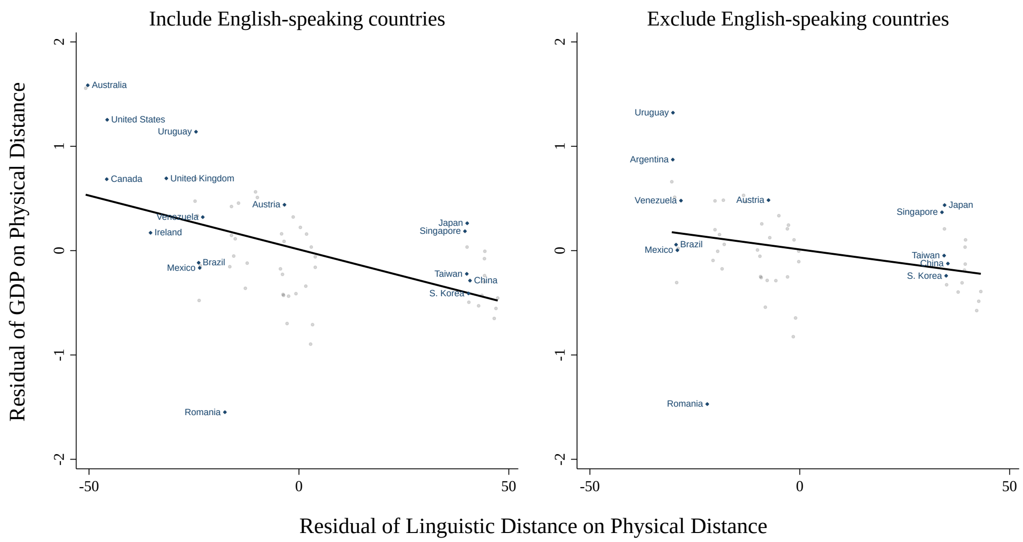
Table A.9: Japanese export growth and British Patent Relevance 1875-1910

	Annualized Export Growth Between 1875 and						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	1880	1885	1890	1895	1900	1905	1910
British Patent Relevance	-3.246**	-0.851	0.168	0.633**	0.493**	0.471**	0.342**
	(1.596)	(0.575)	(0.423)	(0.280)	(0.226)	(0.186)	(0.153)
Observations	40	45	46	47	45	46	47
Constant	✓	✓	✓	✓	✓	✓	✓

Note: The dependent variable is annualized Japanese export growth for the year reported relative to 1875. The number of observations changes across specifications because of the different number of traded sectors in different years. Robust standard errors in parentheses: $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

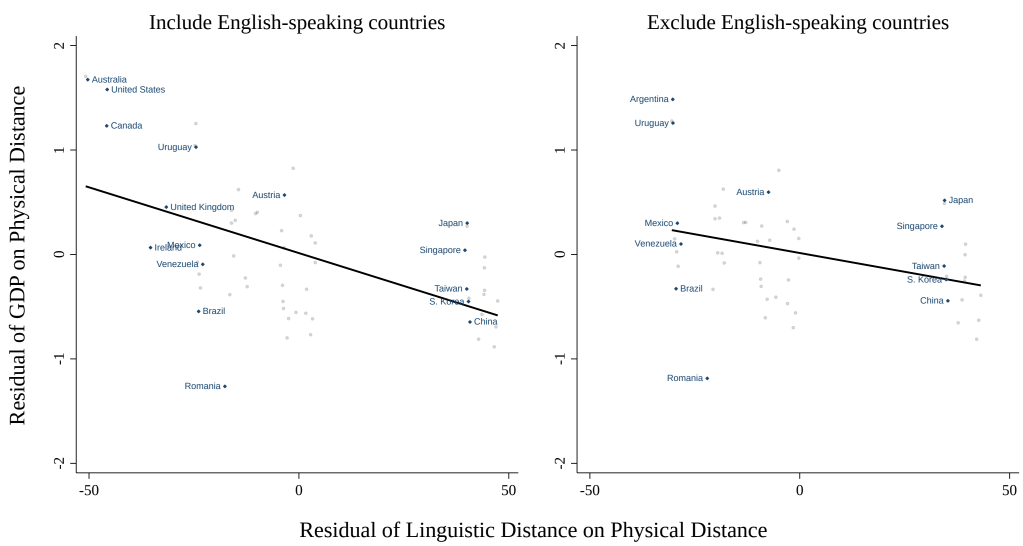
B Additional Figures

Figure A.1: Linguistic Distance Partial Regression Plot for 1870



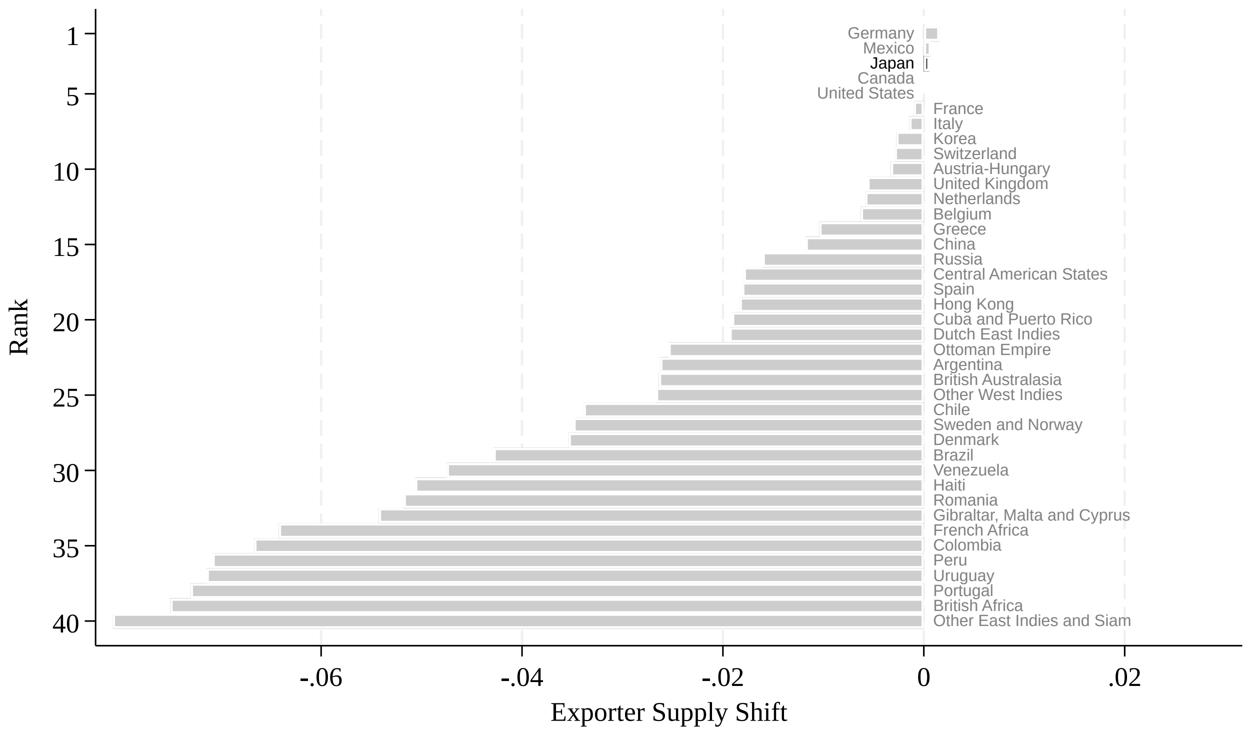
Note: This figure plots the relationship between log GDP per capita in 1870 and linguistic distance after controlling for log physical distance. Data are from the Maddison dataset, the U.S. Department of State's Foreign Service Institute, and *CEPII*, respectively.

Figure A.2: Linguistic Distance Partial Regression Plot for 1913



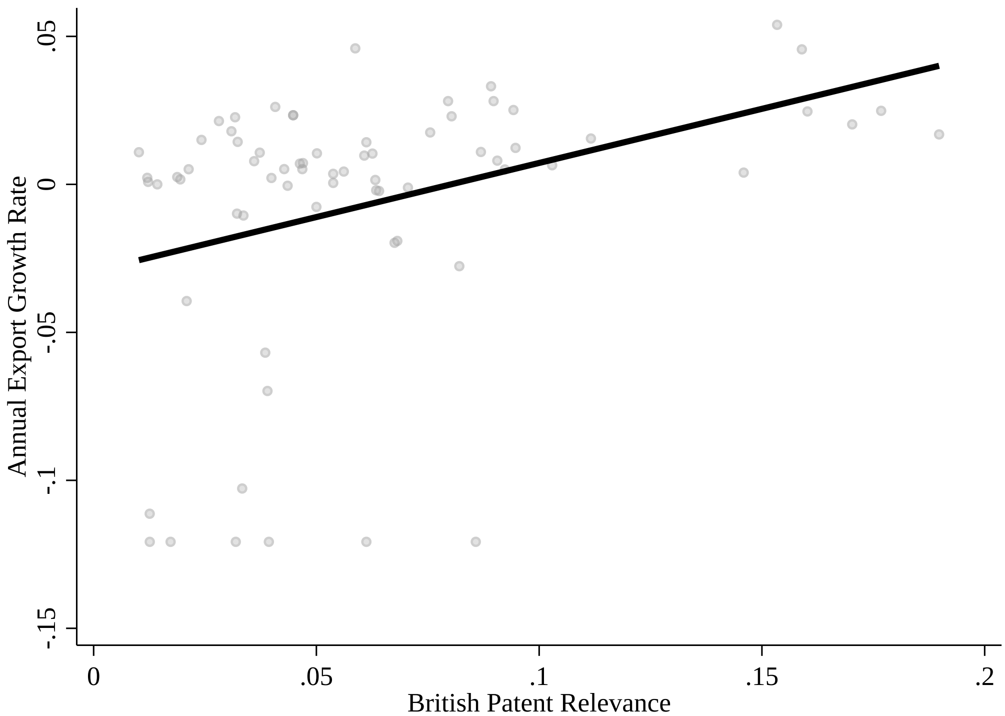
Note: This figure plots the relationship between log GDP per capita in 1913 and linguistic distance after controlling for log physical distance. Data are from the Maddison dataset, the U.S. Department of State's Foreign Service Institute, and *CEPII*, respectively.

Figure A.3: Relative Annualized Exporter Supply Shift by Exporter



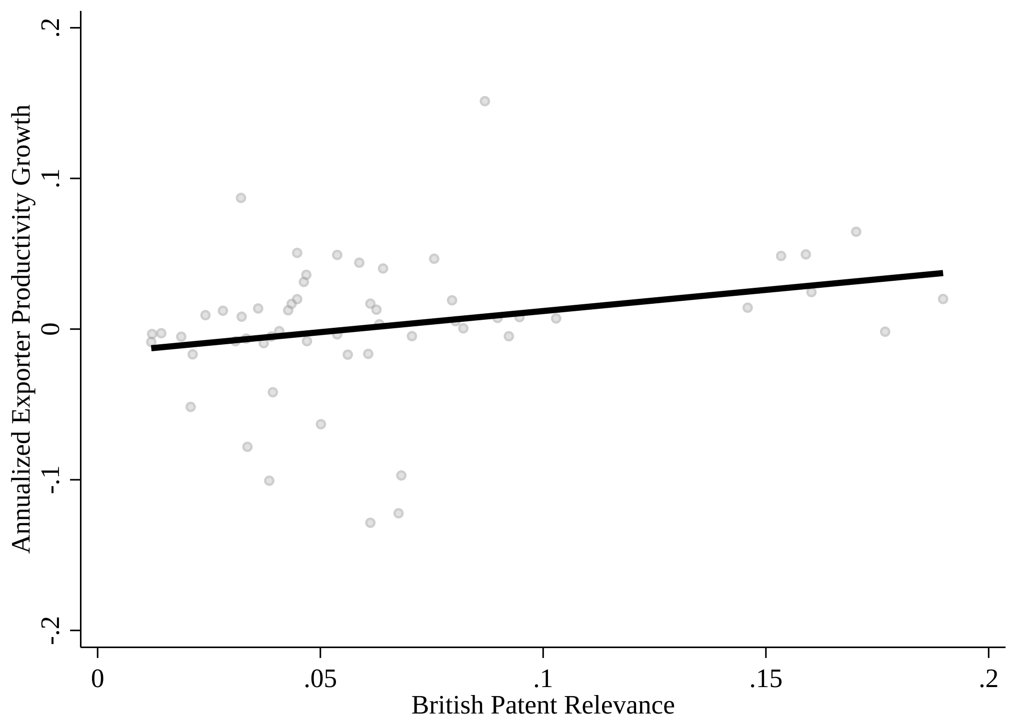
Note: Annualized per-capita exporter supply shifts are expressed as relative to the US, i.e., they are defined as $\hat{\gamma}_i - \hat{\gamma}_{US}$. See text for details on variable construction.

Figure A.4: Annualized Export Growth and British Patent Relevance for Japan



Note: The dependent variable, "Export Growth," is the annualized export growth rate for industry *k* between {1880,1885} and {1905,1910}. British Patent Relevance is a variable that captures how relevant the titles of British patents (1617-1852) are to the vocabulary of an industry *k*. See text for details on variable construction.

Figure A.5: Annualized Prod. Growth Γ and British Patent Relevance for Japan



Note: The dependent variable, Γ_{ik}, is the annualized growth rate in comparative advantage for industry k in region i between {1880,1885} and {1905,1910}. "British Patent Relevance" is a variable that captures how relevant the titles of British patents (1617-1852) are to the vocabulary of an industry k. See text for details on variable construction.

C Constructing Annual Growth Rates

We build the bilateral global trade data by merging bilateral industry export flows from different source countries (Belgium, Japan, Italy, or the U.S.). These data source countries sometimes only report exports in an industry in one of the early years (1880 or 1885) or one of the later years (1905 or 1910). Rather than throw out the industry for all countries when 1880 or 1910 is not reported by one source region, we adopt a procedure to let us be flexible about the start and end dates by computing the average annual export growth rates between any of two potential start years at the beginning of our sample (1880 or 1885) and any of two potential end years at the end of our sample (1905 or 1910).

We set the start year equal to 1880 if the source region reports data in that year or 1885 if data is not available for 1880 but is available for 1885. Similarly, we set the final year equal to 1910 if the source region reports data for that year or 1905 if data is not available for 1910 but is available for 1905. Since this means that the start and final years for bilateral trade growth rates can vary by data source region, we annualize the data so our export and productivity growth rates can be interpreted as average annual growth rates.

We use two procedures to annualize the data. If the reporting region exports the product in 1880 or 1885 (i.e., $\sum_j x_{ijks} > 0$ for $s = 1880$ or 1885), we set s equal to the first year that satisfies $\sum_j x_{ijks} > 0$. We drop the sector if $\sum_j x_{ijks} = 0$ because industry growth rates are undefined if a country does not export anything in the industry in the first period. Similarly, we set f equal to the last year ($f \in \{1905, 1910\}$) that satisfies $\sum_j x_{ijkf} > 0$. We compute the annual growth rate for all bilateral exports satisfying $x_{ijks} > 0$ as

$$g^C_{ijk} \equiv \left(\frac{x_{ijkf}}{x_{ijks}}\right)^{\frac{1}{f-s}} - 1$$

For this sample of exports, we define the implied level of exports in year $s+1$ as $x_{ijk,s+1} \equiv \left(1 + g^C_{ijk}\right) x_{ijk,s}$.

We face a different problem if a country exports the product in year s, i.e., $\sum_j x_{ijks} > 0$, but no bilateral exports are reported between two regions in the industry in the start year, i.e., $x_{ijks} = 0$ for some $\{i, j, k, s\}$. To deal with this problem, we define the average growth rate in exports due to new export destinations as

$$g^N_{ik} \equiv \left(1 + \frac{\sum_{j \in \mathcal{N}_i} x_{ijkf}}{\sum_j x_{ijks}}\right)^{\frac{1}{f-s}} - 1, \tag{A.1}$$

where $\mathcal{N}_i$ is the set of new export destinations, which are defined to be the observations satisfying $x_{ijks} = 0$ and $x_{ijkf} > 0$. In this case, we set the annualized level of exports to new destinations in $s+1$ as $x_{ijk,s+1} \equiv \left(1 + g^N_{ik}\right)^{-(f-s-1)} x_{ijkf}$. In other words, we set the counterfactual amount of exports to new destinations in year $s+1$ equal to the observed amount of exports in year f (x_{ijkf}) deflated by the growth rate in exports due to extensive margin growth between years $s+1$ and f. With these annualized values for exports in hand, we can write the left-hand side of equation 6 as

$$\frac{\sum_j x_{ijkf} - \sum_j x_{ijks}}{\sum_j x_{ijks}} = \frac{\sum_j x_{ijk,s+1} - \sum_j x_{ijks}}{\sum_j x_{ijks}}, \tag{A.2}$$

and the left-hand side of equation 7 as

$$\frac{\sum_i x_{ijkf} - \sum_i x_{ijks}}{\sum_i x_{ijks}} = \frac{\sum_i x_{ijk,s+1} - \sum_i x_{ijks}}{\sum_i x_{ijks}}. \tag{A.3}$$

We then can apply the AW estimation procedure in equations 6 and 7 to estimate the γ_{ik}.

D Variables from External Sources

This section documents the variables we obtained from secondary sources and any changes we made to them. We discuss data from primary sources in the next sections.

D.1 Defining current high-income countries

We make a reference to "high-income" countries in the Introduction. We define a country as high income if its GDP per-capita (PPP adjusted) in 2022 is 50% or more of the US GDP per-capita, based on data from the World Bank (2024). Specifically, we use the variable "GDP per capita, PPP (current international dollars)."

D.2 Identifying the plurality language by country: Ethnologue (2023)

Reference *Ethnologue*, https://www.ethnologue.com/.

We identify the plurality language spoken by each country for the analysis examining the relationship between per capita-income and linguistic distance in Appendix Table A.1 and Appendix Figures A.1 - A.2. To do so, we obtain the modern (2023) plurality language spoken in each country from "Ethnologue".

D.3 Weeks to Learn a Language: Foreign Service Institute (2023)

Reference "Foreign Language Training - United States Department of State," U.S. Department of State, 03-May-2023. [Online]. Available: https://www.state.gov/foreign-language-traning/.

The Foreign Service Institute of the U.S. Department of State estimates the number of weeks required for an English native speaker to reach "General Professional Proficiency" in the language (a score of "Speaking-3/Reading-3" on the Interagency Language Roundtable Scale. We use this measure to proxy linguistic distance for the analysis examining the relationship between per capita-income and linguistic distance in Appendix Table A.1 and Appendix Figures A.1 - A.2.

D.4 Distance to U.K.: GeoDist Database (Mayer and Zignago, 2011)

We control for physical distance in the analysis examining the relationship between per capita-income and linguistic distance in Appendix Table A.1 and Appendix Figures A.1 - A.2. To do so, we use data from *Centre d'Etudes Prospectives et d'Informations Internationales* (CEEPI) which report different measures of bilateral trade distances (in kilometers) for 225 countries. Our measure of the distance between any two countries is the "dist" variable, which is calculated using the great circle formula. They compute internal distances by using the latitudes and longitudes of the most important cities/agglomerations (in terms of population). This means that the distance of a country to itself will never be zero; rather, the distance measure captures how far away major population centers within a country are from each other.

D.5 Historical income and population data: Maddison Project Database

The Maddison Project Database provides information on comparative economic growth and income levels over the very long run. We use the 2020 version of this database (Bolt and van Zanden, 2020), which covers 169 countries up until 2018. We use data on GDP per capita from this source for the analysis examining the relationship between per capita-income and linguistic distance in Appendix Table A.1 and Appendix Figures A.1 - A.2. Further, we also use this source to assign regions into income groups in the main analysis (Section 6).

Classifying regions as high-, medium- and low-income

We classify regions in our dataset by income level using the GDP per capita data from Maddison for 1870. To obtain this variable, we adopt the following steps:

1. The Maddison data uses modern country borders. We first map modern countries to the historic states they were part of in 1880-1914, which will match our trade data (e.g., Hungary and Austria map to Austria-Hungary).

2. The GDP per capita of a historical state that spans two or more modern countries is the simple mean of the GDP per capita of its constituent modern countries.

3. We rank regions by GDP per capita in descending order. Countries in the top third of this distribution are considered high income, countries in the middle third, middle income, and countries in the bottom third, low income.

Finally, we also use the Maddison data to estimate annualized population growth needed for constructing Figure 11.

Estimating annualized population growth

We use the 1870 and 1913 population data to estimate a region's population growth according to the following protocol:

1. Concord the modern countries in the Maddison database with the historic regions we use in this paper.

2. The population of a historic region for a given year is the sum of the population of the modern states that make it up.

3. Compute annualized population growth

$$\text{Annualized Population Growth}_i = \left(\frac{\text{Population}_{i,1913}}{\text{Population}_{i,1870}}\right)^{\frac{1}{1913-1870}} - 1$$

The Maddison Project does not report data for the Russian Empire during this time period; we complement the database by using the Russian population estimates for 1880 and 1910 from Mitchell (1975).

D.6 Historical Italian trade data: Federico et al. (2011)

We obtain Italian trade data for 1880, 1885, 1905, and 1910 from Federico et al. (2011). This dataset harmonizes historical trade records from Italian customs between 1862 and 1950 by concording the different product lines to SITC codes. The source reports bilateral trade at the product level between Italy and its ten biggest trading partners.

D.7 Historical Belgian trade data: Huberman et al. (2017)

We obtain the Belgian bilateral product-level trade data for 1880, 1885, 1905, and 1910 from Huberman et al. (2017). They use the *Tableau générale du commerce extérieur* published by the Belgian government as their primary source and concord product lines to SITC codes. The authors record trade *in manufacturing* at five-year intervals between 1870 and 1910. In 1900, 50% of Belgian exports and 20% of imports were in manufacturing.

D.8 Historical Japanese export data: Meissner and Tang (2018)

We obtained bilateral product level Japanese export data at five-year intervals between 1880 and 1910 from Meissner and Tang (2018). This dataset was constructed from the trade statistics volumes published by the Japanese Ministry of Finance. The authors concorded product lines to SITC codes.

D.9 French Energy Data: Chanut (2000)

We control for the intensity of steam usage of industries in our regressions. We construct this data based on French energy data that comes from Chanut (2000). We manually map French industries to SITC codes. We define the Steam Intensity of an industry as the ratio between the Steam Engine Horsepower of the industry over its Wage Bill. We define the wage bill as:

Wage Bill = (# of Male Workers)*(Avg. Male Hourly Wage) + (# of Female Workers)*(Avg. Female Hourly Wage) + (# of Child Workers)*(Avg. Child Hourly Wage).

D.10 Historical Exchange Rates: Fouquin and Hugot (2016)

Our bilateral-product level trade data converts the value of exports and imports (reported in local currency) into current yen. We use data on annual exchange rates from the *Historical Bilateral Trade and Gravity Dataset (TRADHIST)* from which we obtain the yearly exchange rates for the 1870-1915. Specifically, they provide us the value of one unit of the local currency in pounds.

We calculate the exchange rate from Yen to Belgian francs, Italian lira and US dollars as follows:

$$\frac{£_t / X_t}{£_t / ¥_t} = \frac{¥_t}{X_t}$$

where t refers to year and X to the local currency. The value that we obtain is the value of one unit of the local currency in yen.

E Bilateral Trade Dataset

Our master bilateral, product-level trade dataset is constructed from four main sources:

- American exports and imports in 1880, 1885, 1905 and 1910
- Belgian manufacturing exports and imports in 1880, 1885, 1905 and 1910
- Italian exports to and imports from top trading partners in 1880, 1885, 1905 and 1910
- Japanese exports and imports in 1875, 1880, 1885, 1905 and 1910

As noted in the previous section, the Belgian and Italian trade data, as well as the Japanese export data for most years, has already been digitized and concorded to SITC by others.

We digitized and concorded to SITC the U.S. trade data, Japanese import data, and Japanese export data for 1875. The U.S. data are digitized from yearly volumes of *Foreign Commerce and Navigation, Immigration, and Tonnage of the United States* published by the Treasury Department's Bureau of Statistics (1900). The Japanese trade data was sourced from the yearly volumes of *Annual Return of the Foreign Trade of the Empire of Japan* published by the Department of Finance (1916). From these volumes, we only use the tables from the "Quantity and Value of Commodities Imported/Exported from Various Countries" sections. We use the Meissner and Tang (2018) product-SITC mapping wherever possible for Japan and the U.S. to ensure consistency.

Japan and the U.S. kept detailed records of their trade with foreign countries between 1880 and 1910. Each entry tells us the name of the product, its quantity, units, transaction value, and year, as well as the exporting and importing countries. The construction of these data involves digitizing the records and harmonizing products and country names. To construct the harmonized dataset across different reporting countries, we convert all data to a common currency, harmonize country names, and deal with double reporting issues. The protocols we adopted are described in detail in the subsections below.

E.1 Harmonization of Countries

Country names are not standardized across reporters (Belgium, Italy, Japan, and the U.S.) and years. In order to make comparisons across years and countries, we standardized country names as follows:

1. We made a list of all the country names that appear in all of the trade books from the four reporters.
2. We grouped names that refer to the same country: e.g., Vietnam and French Indo-China both refer to the same political entity at the time.
3. We kept the group if it is used by at least three reporters in the 1880/5 *and* 1905/1910 books for each reporter.

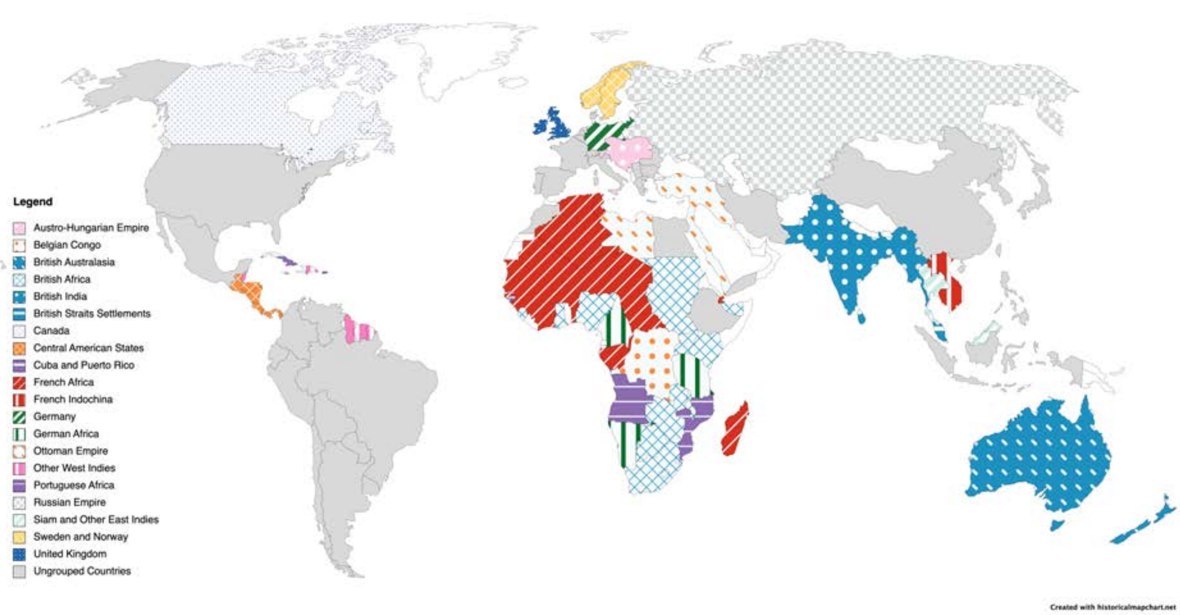
4. If the country group did not meet the previous requirement, then we try to build a regional group that does. For example, Honduras, Nicaragua, and Costa Rica do not have three reporters in the all the required years. If we group all Central American States together, this larger regional group meets our requirements.

5. If a country could not be grouped and did not meet the reporter-year requirement, then we dropped it.

6. If a region was too disaggregated, we dropped it. For example, Singapore and Hong Kong are their own separate categories, each with substantial volumes of trade in our dataset. If one country, in one year, reported "Hong Kong & Singapore," we dropped this observation.

The map illustrates how we grouped countries. Countries in grey were left as they were. We use the map of the world on the eve of World Word I (1914) as a baseline for our country groups.

Figure A.6: Country Groups



Note: Colonies are grouped by imperial power and region (e.g. British Africa, French East Indies). All small, remote islands (e.g. Falklands) were dropped. Countries in white are missing from the dataset, countries in gray are reported as in the raw data. The remainder of the footnote reads from West to East on the map. The West Indies are grouped together with the exception of Cuba and Puerto Rico. British Honduras (although technically in Central America) is considered part of the West Indies due to its political affiliation with other British colonies in the Caribbean. The Ottoman Empire includes Libya, but not Algeria (which fell to the French in 1881). Taiwan is never directly mentioned in any trade statistics and not included in Japanese trade for the time period. Since each book either mentions French India or French Indochina, we conclude that French India refers to French Indochina, not to the French port cities in India. Thailand (then Siam) is grouped with other minor East Indies colonies such as Timor-Leste and British Borneo.

E.2 Double Reporting

Trade between reporting countries appears twice: once as exporters from the origin and secondly as imports by the destination. For all reporting countries except Belgium, we use their export data for their exports to reporting and non-reporting regions. Because Belgium does not report any trade data for non-manufacturing sectors, we use the reporting country's import data from Belgium to fill in these gaps. We use imports by reporting countries from non-reporting countries to construct the exports of non-reporting countries.

F Constructing the British Patent Relevance measure

F.1 Overview

We construct BPR by assuming that the similarity of text in books describing production techniques and the text of British patent data tells us the relevance of patents for that industry. In practice, we start with unigrams (i.e., single words such as "steam") and bigrams (i.e., two-word combinations such as "steam engine" or "steam engines"). We convert these into "terms" by stemming them and converting them into the terms "steam" and "steam engin". We also make use of two types of corpora. The first is the set of books describing production techniques in industry k, and the second is the are British patent synopses. Thus, we have a corpus for each industry and a separate corpus for the patents. To measure the relevance of the British patent corpus to industry k's corpus, we weight each term's frequency in a book by the total number of books divided by the number of books containing the term, i.e., we compute the Term Frequency-Inverse Document Frequency (TF-IDF). For each industry, we build a TF-IDF vector that characterizes its vocabulary (where each element is the TF-IDF of a term); we also build a TF-IDF vector for patents. Finally, we compute British patent relevance of industry k as the cosine similarity between the vector of TF-IDFs for industry k and that for the set of British patent synopses. We explain each of these steps in detail below.

F.2 Building the Terms

To build a term, we start with n-grams, we implement the following steps:

1. We split the raw text into sentences
2. We convert the words in the sentence to lower case, stem the words, replace UK spelling with US spelling
3. We turn each processed sentence in a sentence word list (where the position of a word on a list is the position it has in the sentence)
4. For each sentence word list, we split it into n-grams
5. We count the number of times an n-gram appears in the sentence and sum across sentences
6. Drop n-grams that include at least one stop word, (i.e., "a," "the," etc.)
7. Output a dataset with all the n-grams in the document and their count in the corpus

Example

1. **Text** "A stemmer for English operating on the stem cat should identify such strings as cats, catlike, and catty."
2. **Sentence** "A stemmer for English operating on the stem cat should identify such strings

as cats" "catlike" "and catty"

3. **Processed Word List** "a stemmer for english oper on the stem cat should identifi such string as cat" "catlik" "catti"

4. **Unigrams** "a" "stemmer" "for" "english" "oper" "on" "the" "stem" "cat" "should" "identifi" "such" "string" "as" "cat" "catlik" "catti"

5. **Unigrams without Stopwords** "stemmer" "english" "oper" "stem" "cat" "should" "identifi" "string" "cat" "catlik" "catti"

6. **Final Unigrams with Count** "stemmer" 1 "english" 1 "oper" 1 "stem" 1 "cat" 2 "should" 1 "identifi" 1 "string" 1 "catlik" 1 "catti" 1

F.3 Focusing on Jargon

Many unigrams and bigrams are not technical jargon. In order to focus our analysis on jargon, we drop unigrams and bigrams that are commonly used. We use the Bible to identify commonplace non-technical words that are necessary to write a coherent text but are not helpful in defining an industry's technical vocabulary. We use the 1885 King James Bible because it uses the common, non-technical nineteenth-century words and phrases. We define Biblical words as the 1,000 words occurring with the highest frequency in the Bible. However, if one of these words is used in the description of an SITC keyword, we do not count it as a Biblical word. For example, the stemmed word "brea" is a top 1,000 word in the Bible, but it also happens to be a keyword in the SITC description for cereal products.

F.4 Formally Defining TF-IDF

The term frequency (TF) measure is the count of instances a term appears in a corpus, divided by the number of terms in the corpus. The formula for the TF of term τ in corpus c is

$$\mathrm{TF}(\tau, c) \equiv \frac{F_{\tau,c}}{\sum_{\tau' \in c} F_{\tau',c}} \tag{A.4}$$

where $F_{\tau,c}$ is the raw count of τ in c; and $\sum_{\tau' \in c} F_{\tau',c}$ is number of terms in the corpus. The inverse document frequency (IDF) is a measure of how common or rare a word is across all documents. The rarer the word, the higher the IDF score. We define the IDF for term τ in all corpora C (i.e., the complete collection of books) as

$$\mathrm{IDF}(\tau, C) = \log\left(\frac{N}{N_\tau + 1}\right) \tag{A.5}$$

where N is the total number of books in C; N_τ is number of books in the corpus where the term τ appears.

The TF-IDF is then

$$\text{TF-IDF}(\tau, c, C) = \mathrm{TF}(\tau, c) \cdot \mathrm{IDF}(\tau, c) \tag{A.6}$$

We remove any n-grams that include words in the description of the SITC categories from the sample before estimating the cosine similarities. For example, removing the unigram "cotton" ensures that books describing how to grow cotton are not coded as part of the technology to spin cotton yarn.

Comparing the Vocabulary of Industries and Patents

We define the British Patent Relevance of industry k as the similarity between the TF-IDF vector representation of vocabulary for industry k and patent vocabulary. We use cosine similarity to measure the similarity between the two vectors. If an industry uses the same words at the same frequency as the patent book, then the vectors are the same, and we conclude that British patents are very relevant in the industry. If there is no overlap in words, then the similarity score is low, and we conclude that British Patents are not relevant. See the main text (equation eq:BPR) for the formal definition of cosine similarity.

F.5 Data Sources

All data (unless otherwise specified) was accessed through HathiTrust.

British Patents All patent text comes from the second edition of *Subject-Matter Index of Patent of Invention From March 2, 1617, to October 1, 1851 Parts I (A to M) and II (N to W)*, published by Woodcroft (1857).

Industry For each industry (as defined by SITC-3) we hand-curated a list of books and sections of nineteenth century books relevant in describing the production process of the goods in the industry.

Bible (1885) English Revised Version of the Bible.[1]

G New Japanese Words in the Meiji Period

We utilize the etymology of Japanese words based on the revised edition of *Nihon Kokugo Daijiten* [The Unabridged Dictionary of the Japanese Language], published by Shogakukan (2006). Importantly, it includes the title and year of publication of the Japanese document in which each word is believed to have been first used. We obtained the digitized data for this dictionary from Kotobank.[2] The number of new words by year can be seen on Figure 7.

H Technical Books in the Top World Languages (1800-1910)

H.1 Overview

We report the source libraries for our data on technical books in Table A.10. We tried, where possible, to scrape national libraries. If we could not find a scrapable national library for a language (such as Arabic and Russian), we scraped WorldCat, an online catalog of thousands of libraries worldwide covering dozens of languages. Scraping national library catalogs has an advantage over using WorldCat as the latter source sometimes overstates the number of books because different libraries sometimes report book titles differently (e.g., slight variations in titles or author names).

We minimized the number of possible duplicates by removing spacing and punctuation in book titles and dropping any duplicated book titles published in the same year. In order to minimize the role played by reprints of the same book, we also dropped any duplicates arising from books (possibly published in different years) with the same book ID. Importantly, the number of books

[1]Wikepedia article Revised Version of the Bible: https://en.wikipedia.org/wiki/Revised_Version
[2]Kotobank: https://kotobank.jp/dictionary/nikkokuseisen/

reported for four of our five top codifying languages, French, English, German, and Japanese (but not Italian), were from national libraries, so we can be confident that there is minimal double counting in these book totals.

If we could scrape a national library or WorldCat, we made a judgment call about which source was better. If we saw that for a non-top-4 codifying language, there were more *genuine* technical books than we could find in a national library, we opted for the number from WorldCat. For example, the national libraries of Portugal and Spain have very few technical books in their catalogs relative to the libraries in WorldCat, so we opted to use WorldCat for these languages. Because of the duplication issue in WorldCat and the fact that WorldCat allows us to scrape many libraries for each language, our use of national libraries for English, French, German, and Japanese likely causes us to understate the concentration of technical books in these languages.

We scraped the number of technical books for 33 languages, which include all of the 20 most spoken native languages on earth.[3] We define the set of books comprising technical knowledge as those with a subject classified as applied sciences, industry, technology, commerce, and agriculture. For our purposes, we exclude books on theoretical technical knowledge, such as books in the hard sciences or in medicine.

[3]We assume that if someone speaks Yue or Wu Chinese, they can read Mandarin Chinese, given that these languages all use the same characters.

Table A.10: Catalogs Scraped

Library	Catalog	Languages	Years	Classification System	Tech Topics
Bibliothèque Nationale de France	Link	French	1500-1930	Universal Decimal Classification	Applied Sciences and Technology (6)
Deutsche Nationalbibliothek	Link	German	1500-1930	Dewey Decimal Classification	Technology (600)
National Diet Library	Link	Japanese	1500-1930	Nippon Decimal Classification	Technology (500) Industry (600)
Korean National Library	Link	Korean	0022-1980	Dewey Decimal Classes	Technology and Engineering (600)
Library of Congress	Link	English	1500-1930	Keyword Search	Hand-constructed
National Library of India	Link	Bengali Hindi Marathi Tamil Urdu	1500-1980	Only has three options	Non-Fiction Manually picked tech books.
Shanghai Library	Link not accessible	Chinese	1500-1980	Chinese Library Classification System	Agriculture (S) Industry (T) Transportation (U)
WorldCat	Link	Arabic Bulgarian Croatian Czech Danish Dutch Greek Hebrew Indonesian Italian Norwegian Persian Polish Portuguese Romanian Russian Spanish Swedish Thai Turkish Ukranian Vietnamese	1800-1930	Subject filter in advanced search	Hand-constructed

H.2 Search Filters

$\boxed{1.}$ **Format:** We only search for books. No images, periodicals, articles, or newspapers.

$\boxed{2.}$ **Language:** We always specify the language of the text. For example, when searching the National Diet Library, we only look for books written in Japanese.

$\boxed{3.}$ **Publication Year:** 1500-1930

$\boxed{4.}$ **Subject:** We always search by subject.

- We search by subject code, if possible. Otherwise, we manually picked technical books.
- If subject codes are not available, we use subject keywords. To do this, we first find the underlying subject classification system used by the library (e.g., Dewey Decimal Classification) to get the descriptions of the subject codes we want.